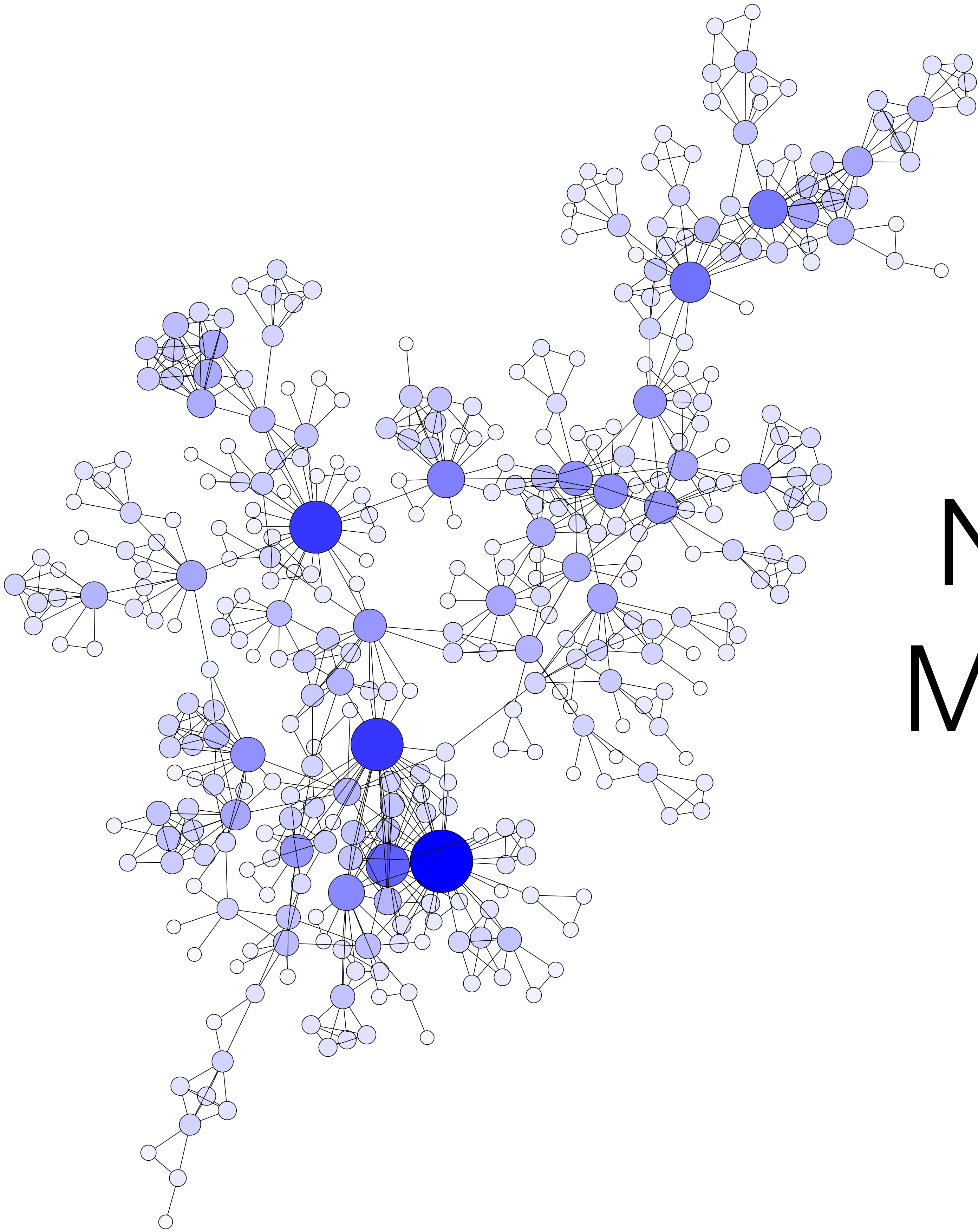
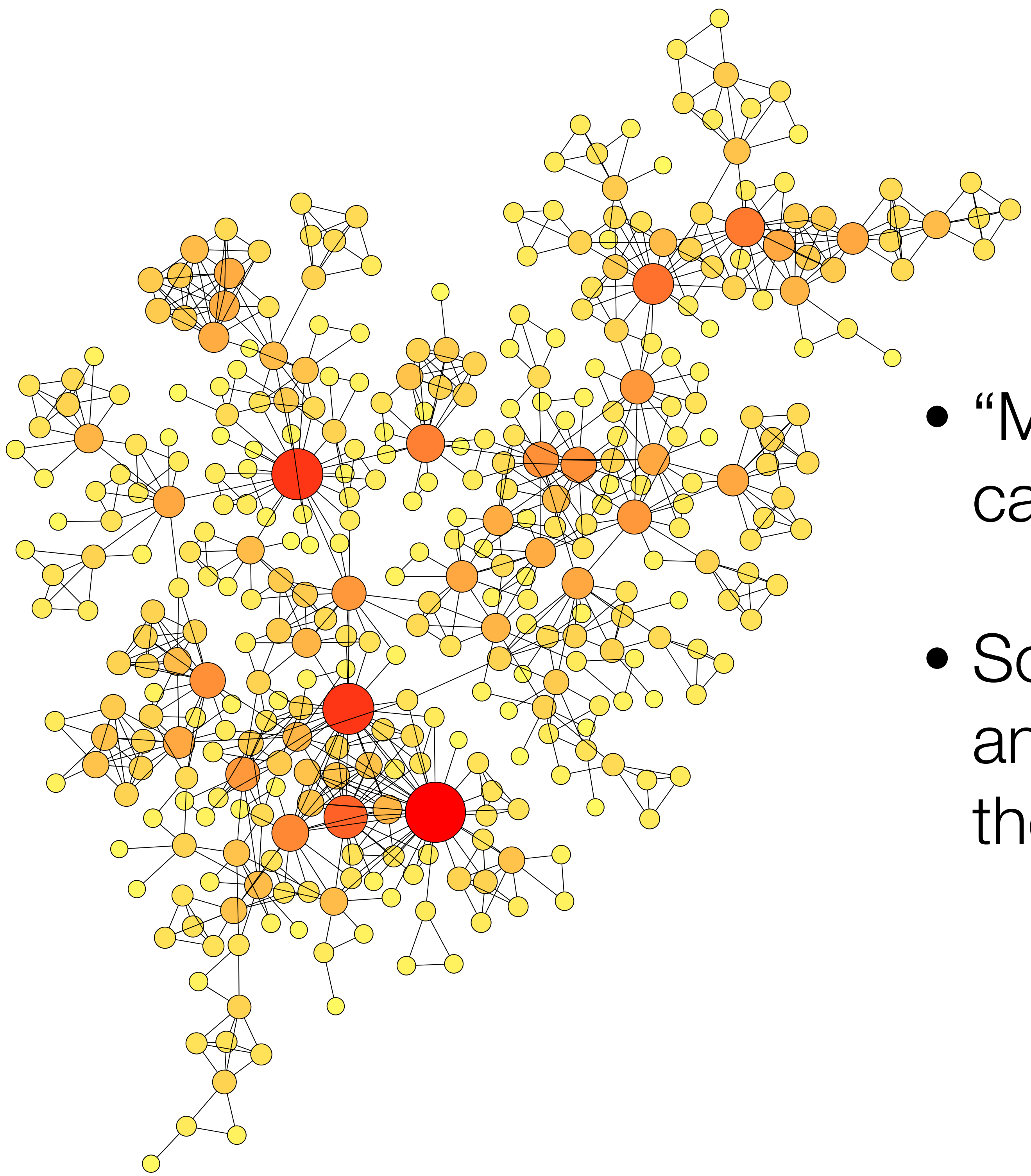


Network Measures

Network Measures: Empirical Observation



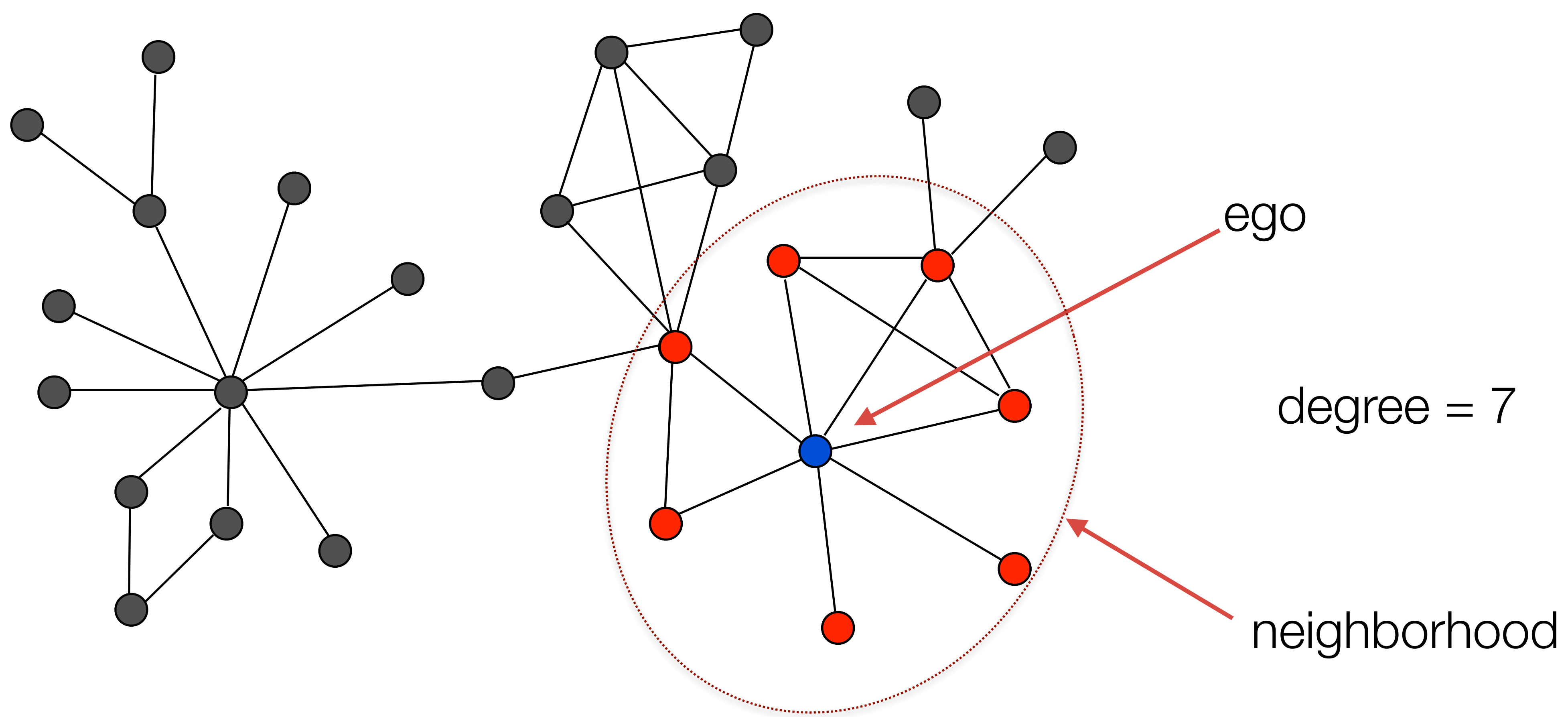
- “Measures” are things that we can quantify
- Some measures are individual, and others are global (related to the network as a whole)

A Node's Neighborhood: Degree

Ego = any single node: i

Neighborhood = the set of nodes ego is connected to: n_i

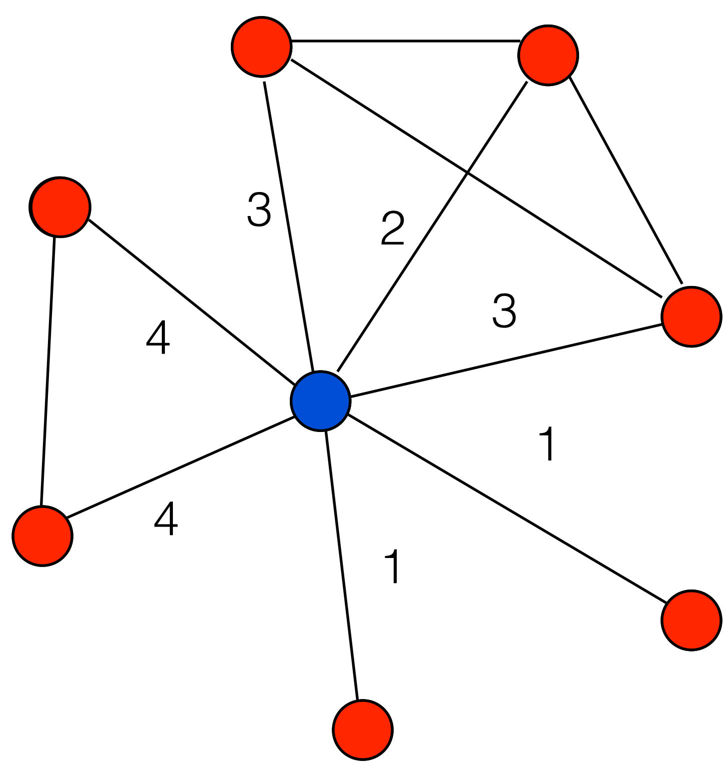
Degree = the number of nodes ego is connected to: $|n_i|$



Degree in a Weighted Network

In a weighted network, there is a second measure of degree: weighted degree:

$$d_i^W = \sum_j w_{ij}$$



weighted degree = 18

Weighted degree tells you something different about nodes than degree does

What does degree mean in an email network? Weighted degree?

Degree in a Directed Network

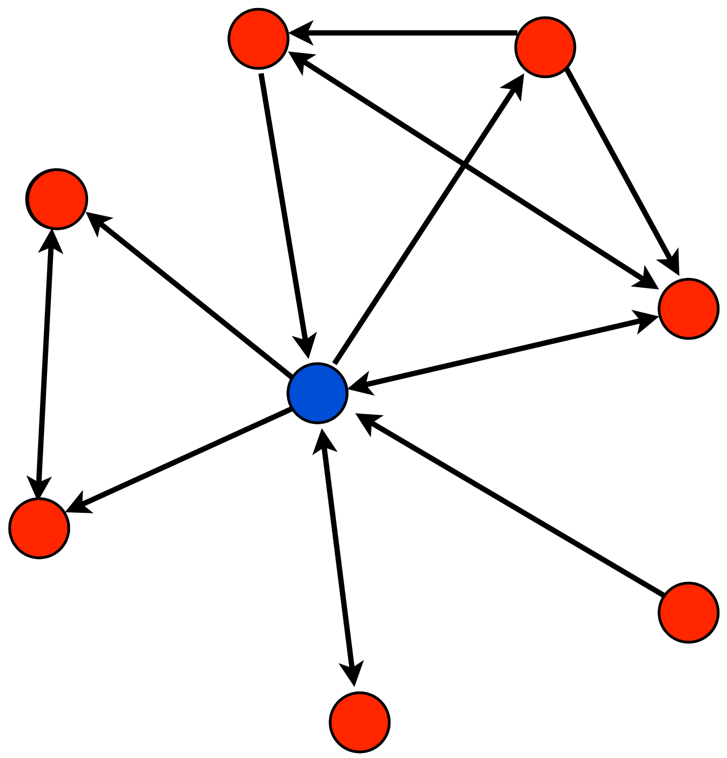
When links are directed, there are two measures of degree:

in-degree = number of nodes who link to ego

$$d_i^I = \sum_j w_{ij}$$

out-degree = number of nodes ego links to

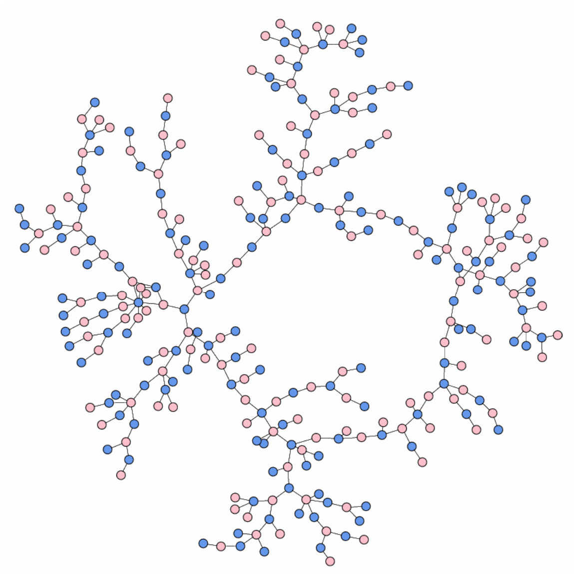
$$d_i^O = \sum_j w_{ji}$$



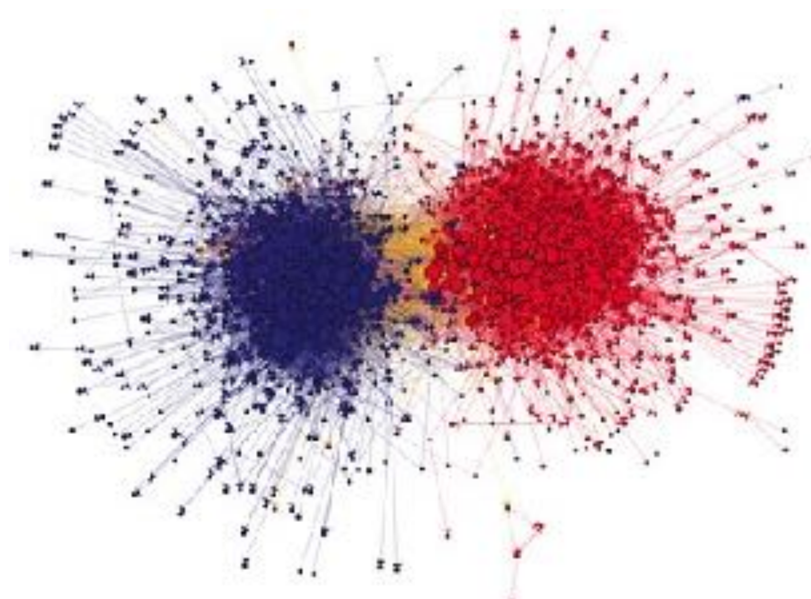
in-degree = 4
out-degree = 5

Interpretation of Degree

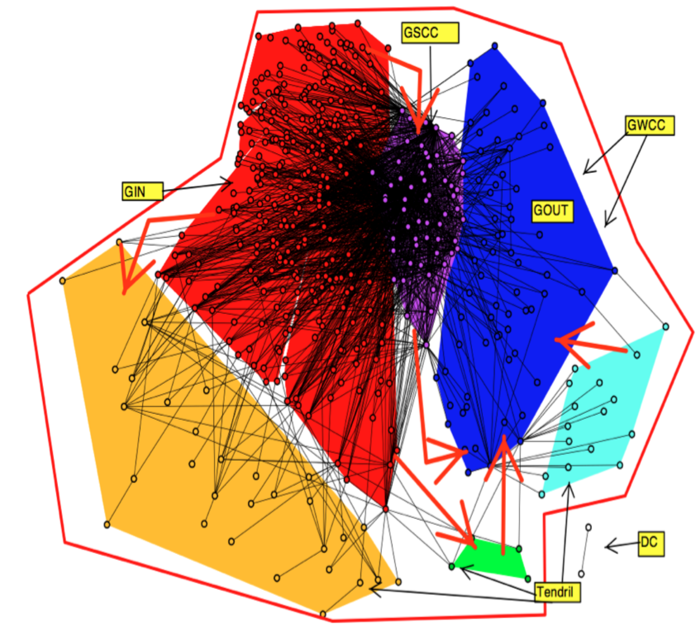
Degree, weighted degree, and in/out degree can have different meaning, depending on what the links mean…



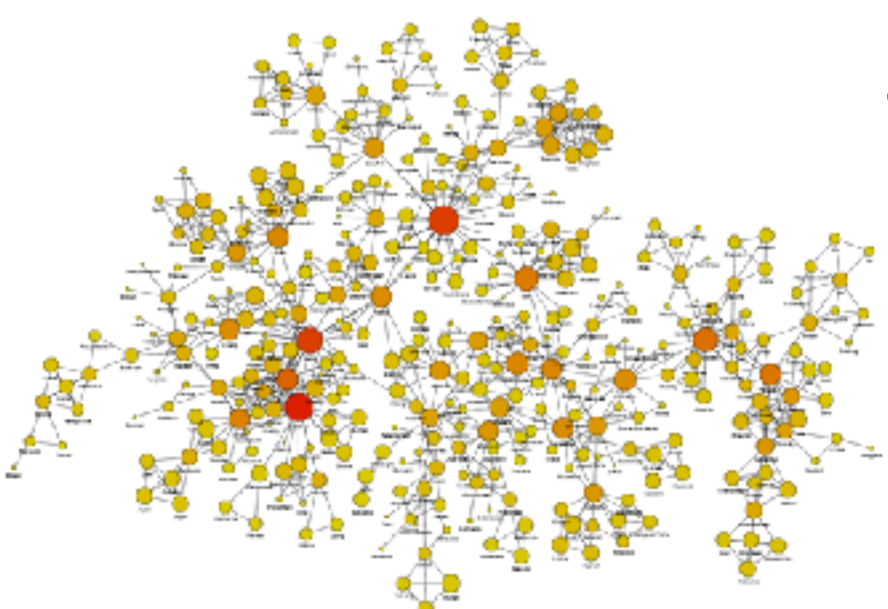
ref: Data by Bearman et al (2004)
Graphic by M.E.J. Newman



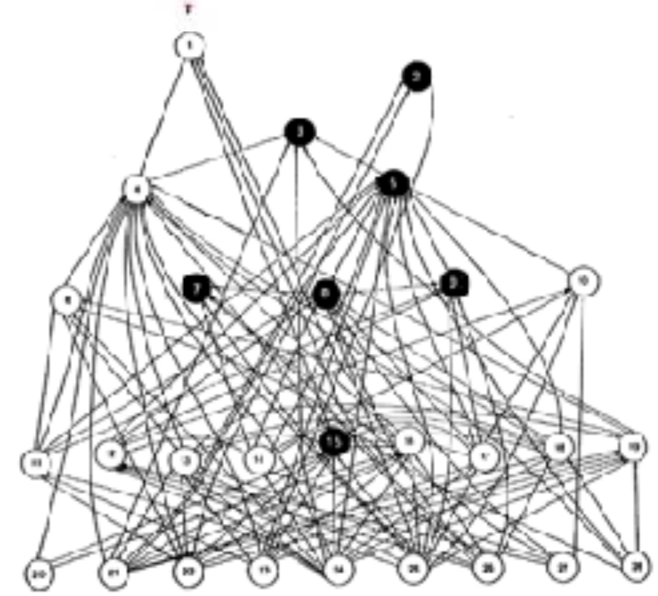
Adamic political blogs



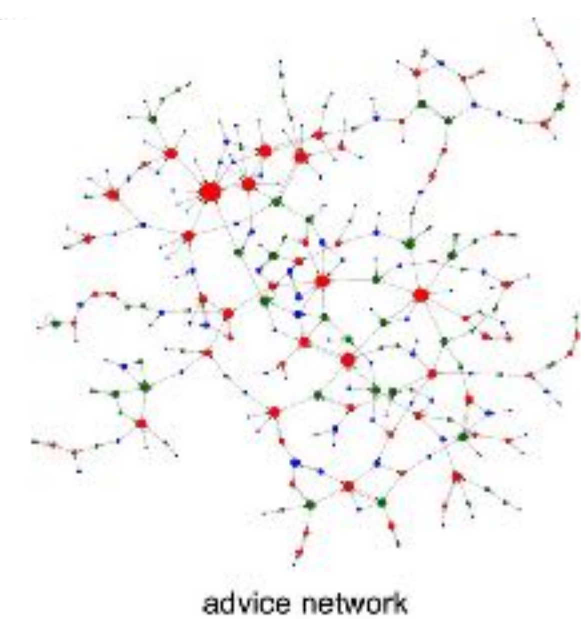
ref: Bech and Atalay 2008



Ref: MEJ Newman (2006)



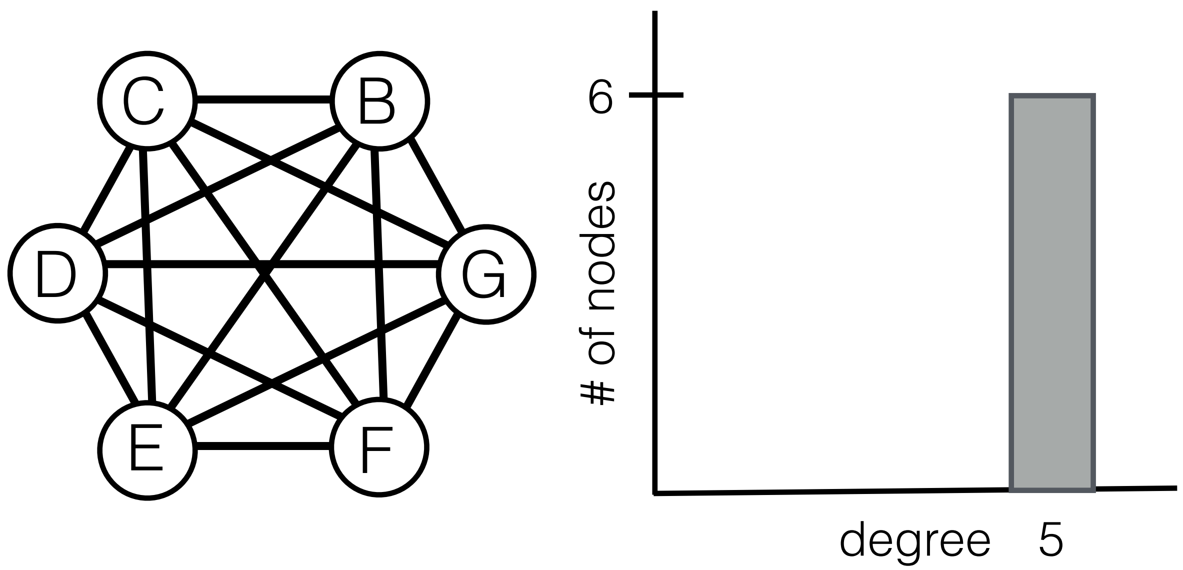
Directed Citation Network: Garfield (1997)

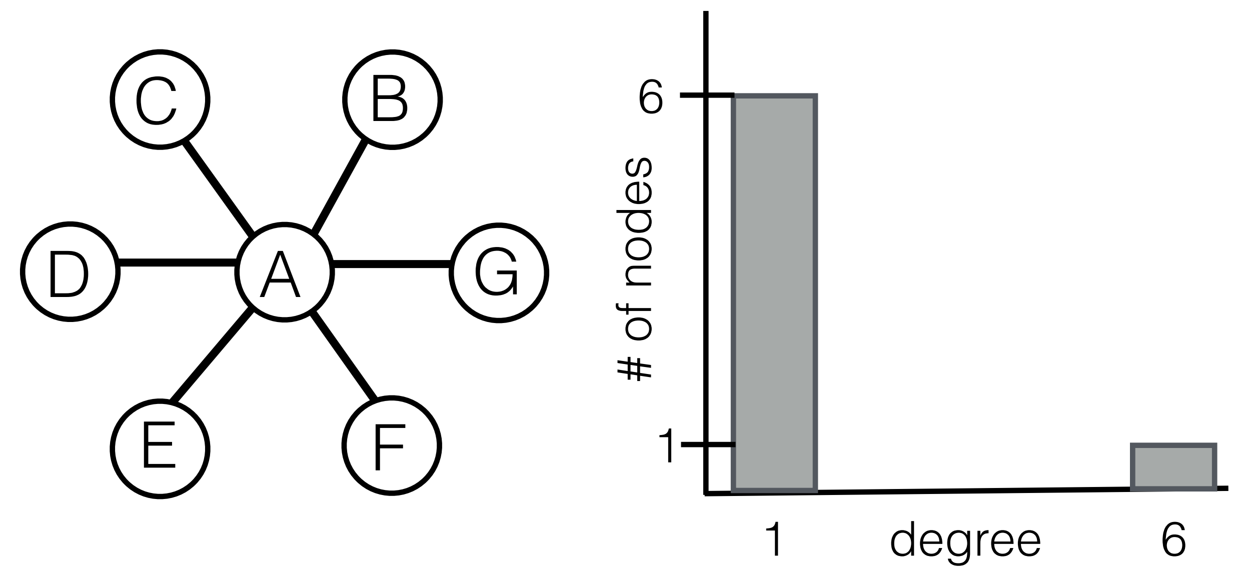


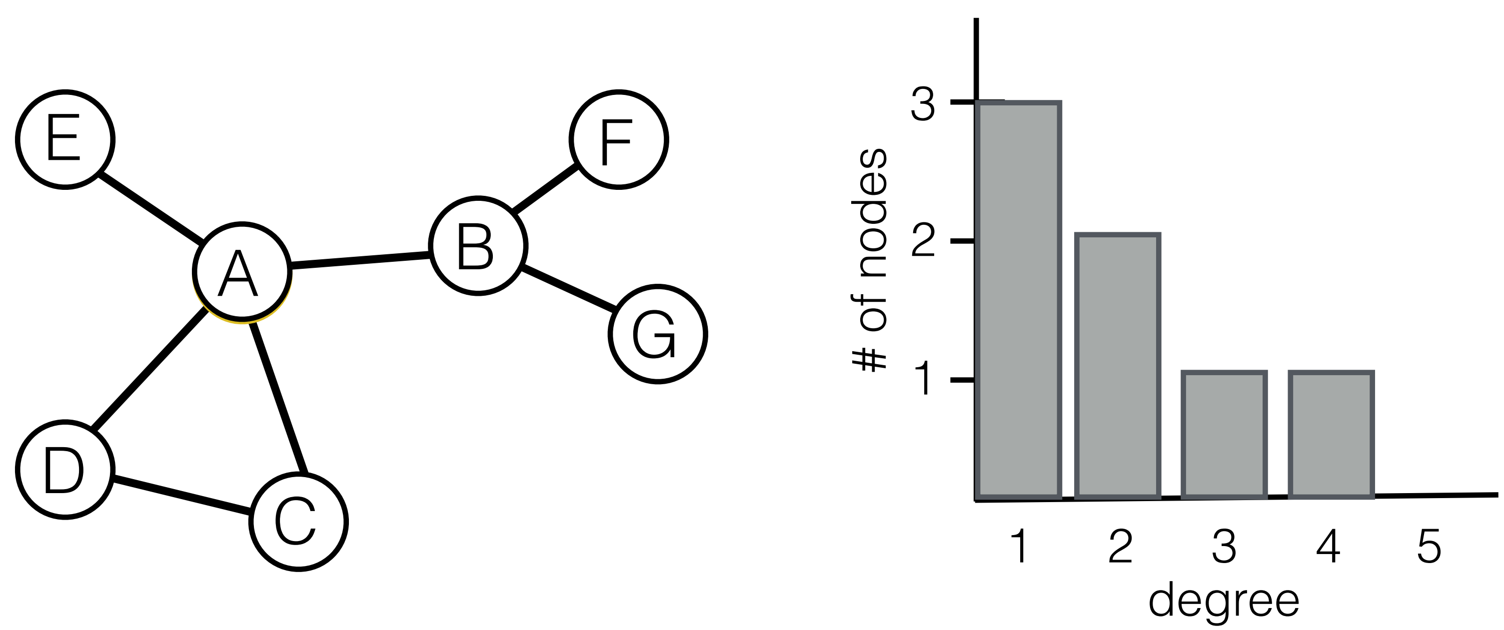
graphic: Wardil and Hauert

Network Measures: Degree Distribution

Degree Distribution = the distribution of degree across the nodes

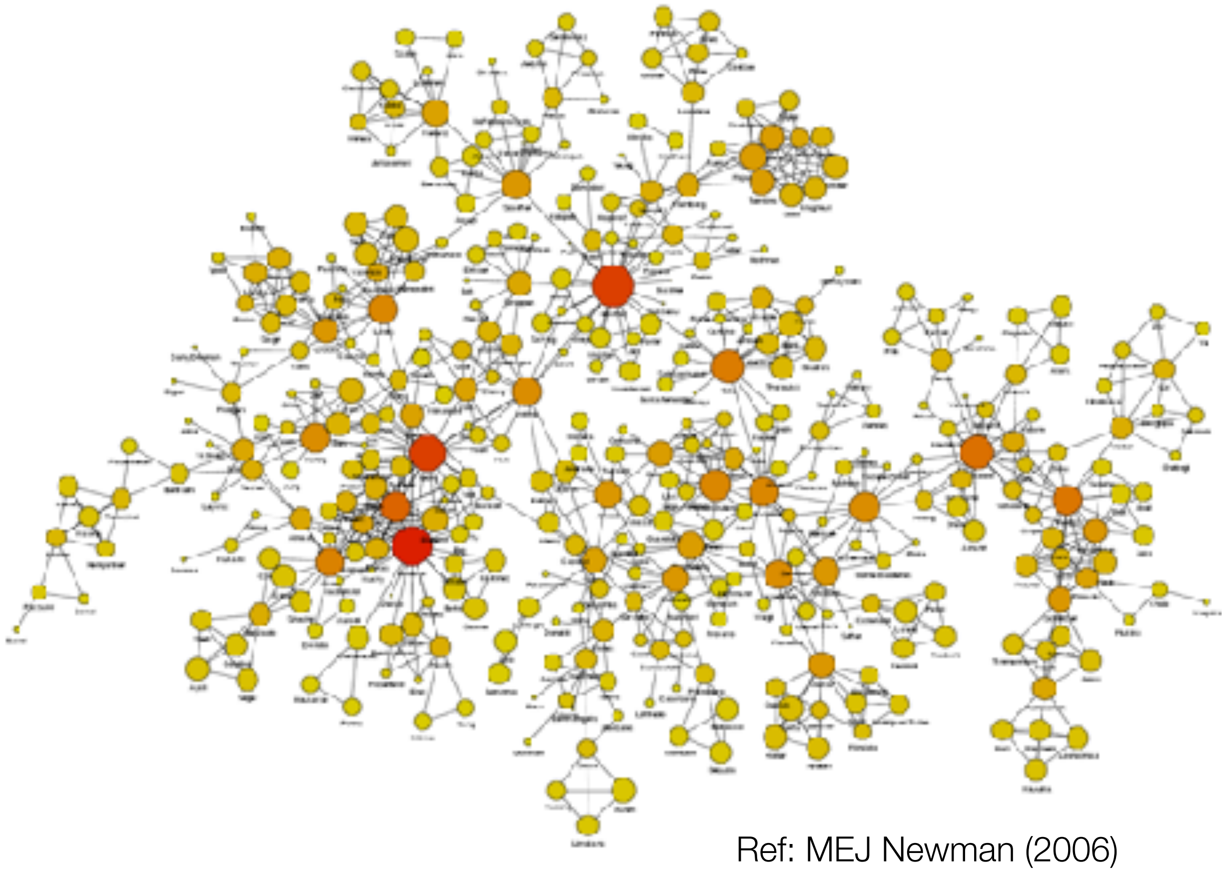




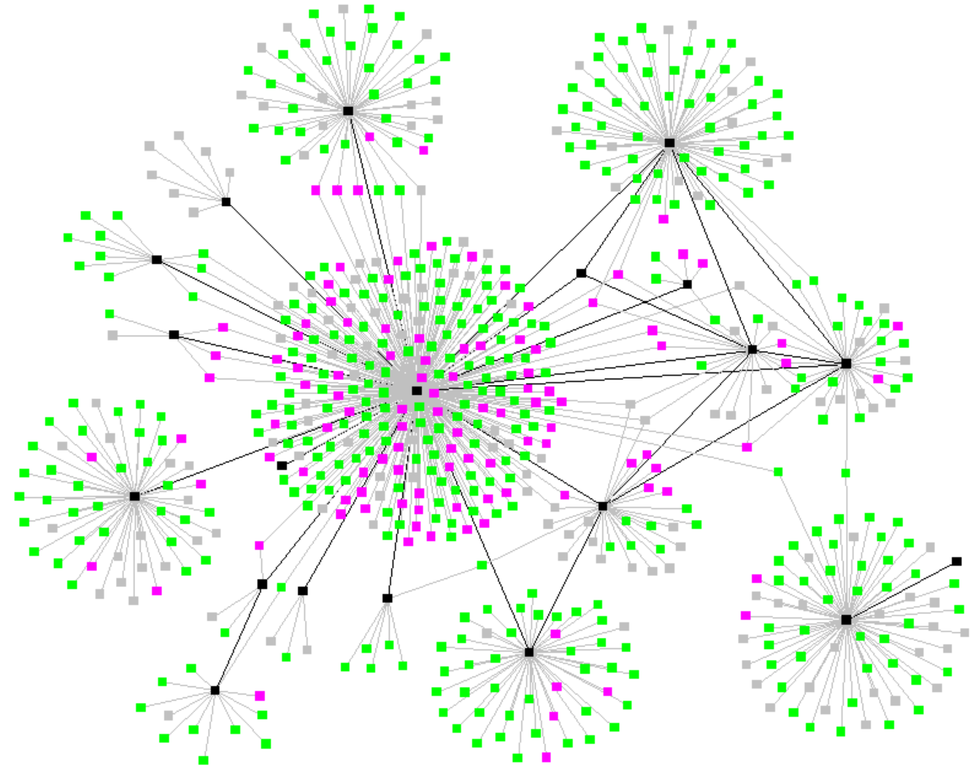


Network Measures: Degree Distribution

Most social networks have a “hub and spoke” structure



Ref: MEJ Newman (2006)

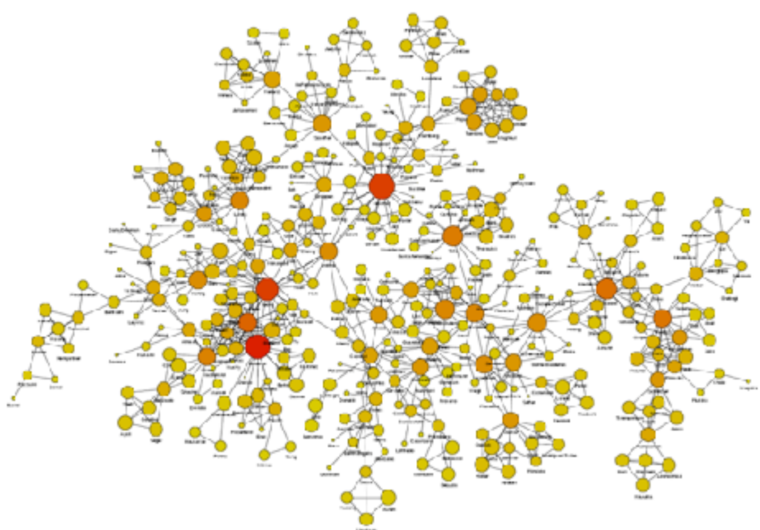


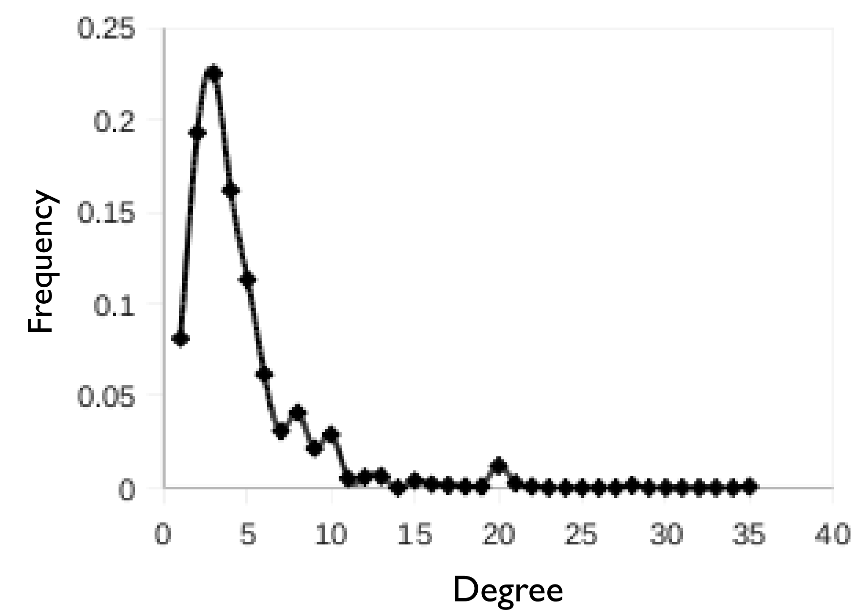
ref: Andre et al

A handful of nodes have lots of connections, but most of the nodes have very few.

Network Measures: Degree Distribution

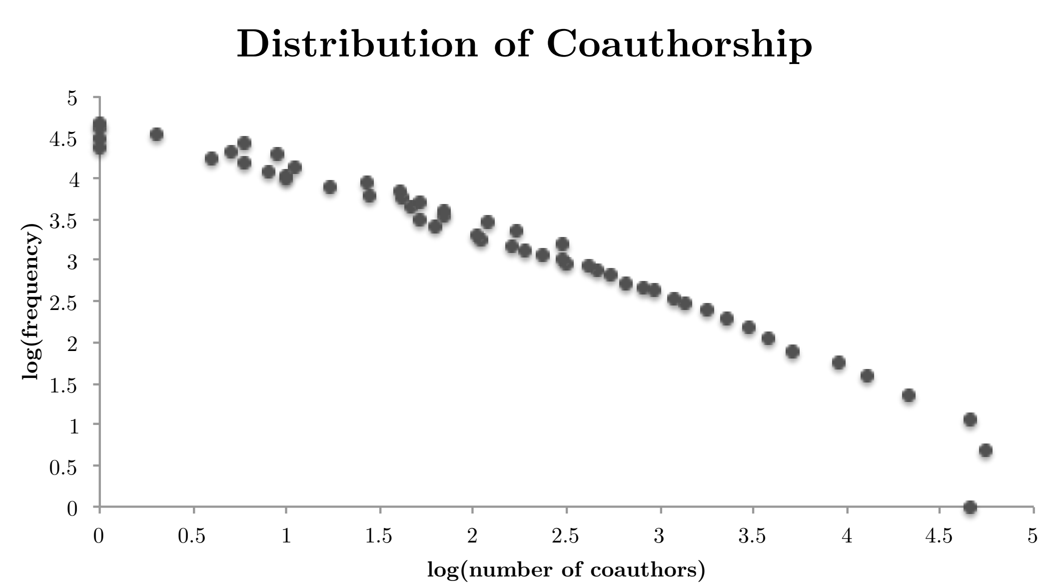
The degree distribution is “long-tailed” (skewed)





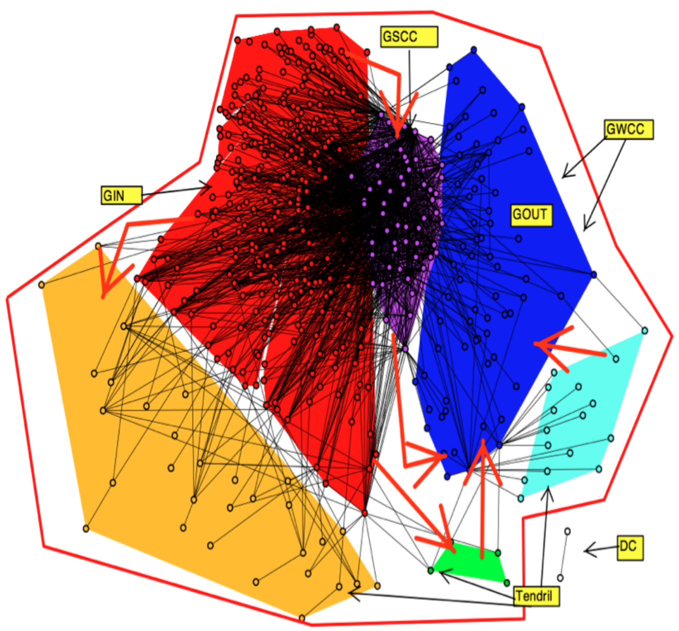
In many cases, it is a *power law* distribution: $P\left(d\right) \sim d^{-\alpha}$

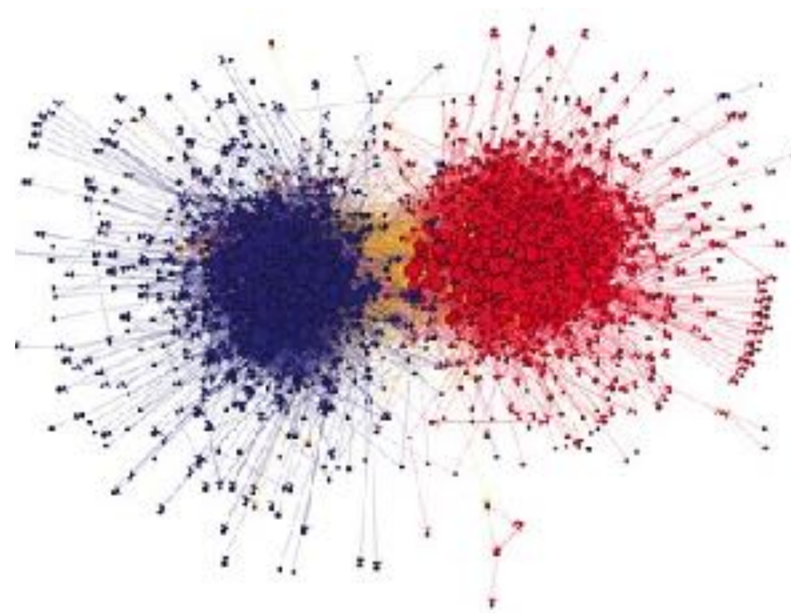
Roughly linear on a log-log plot

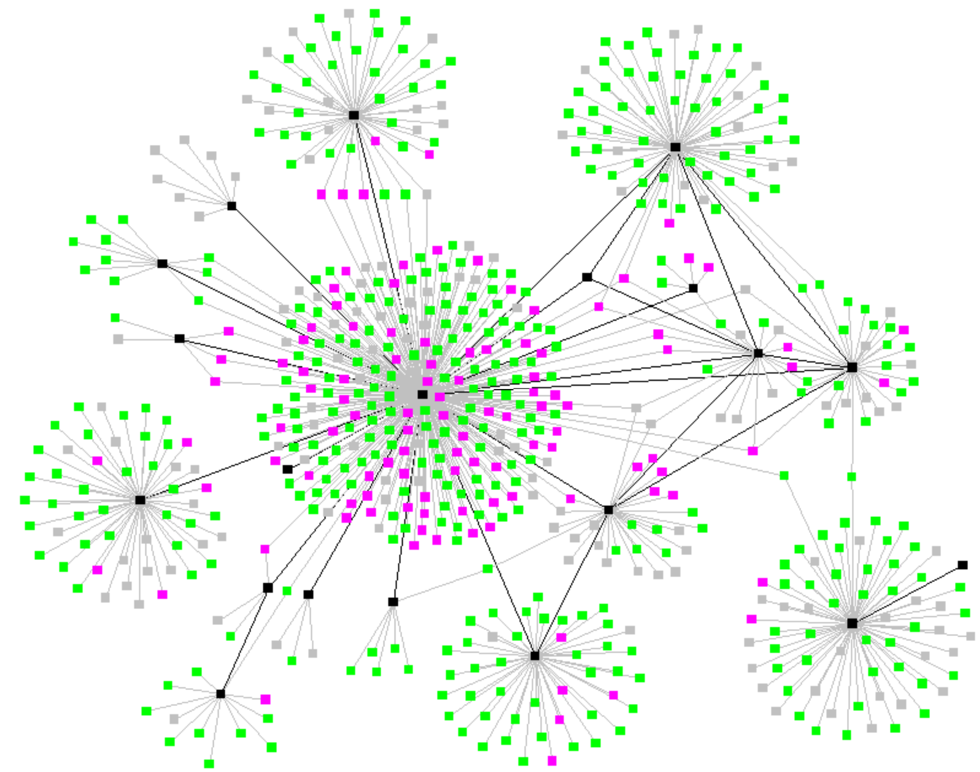


Network Measures: Degree Distribution

The degree distribution is a measure of how evenly-spread connections are across nodes







This has implications for the distribution of influence, the flow of disease, and the fragility of the network

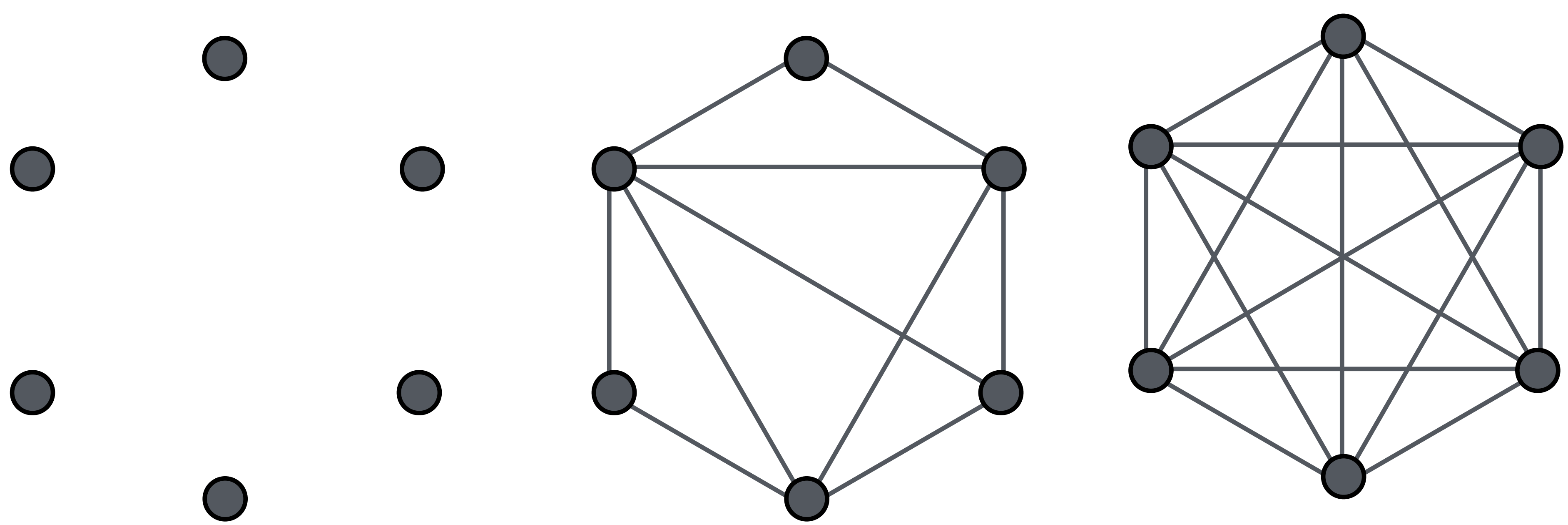
Network Measures: Density

Density is the fraction of all possible links that are made:

$$D = \frac{\frac{\sum_i d_i}{2}}{\binom{N}{2}} \implies D = \frac{\sum_i d_i}{N(N-1)}$$

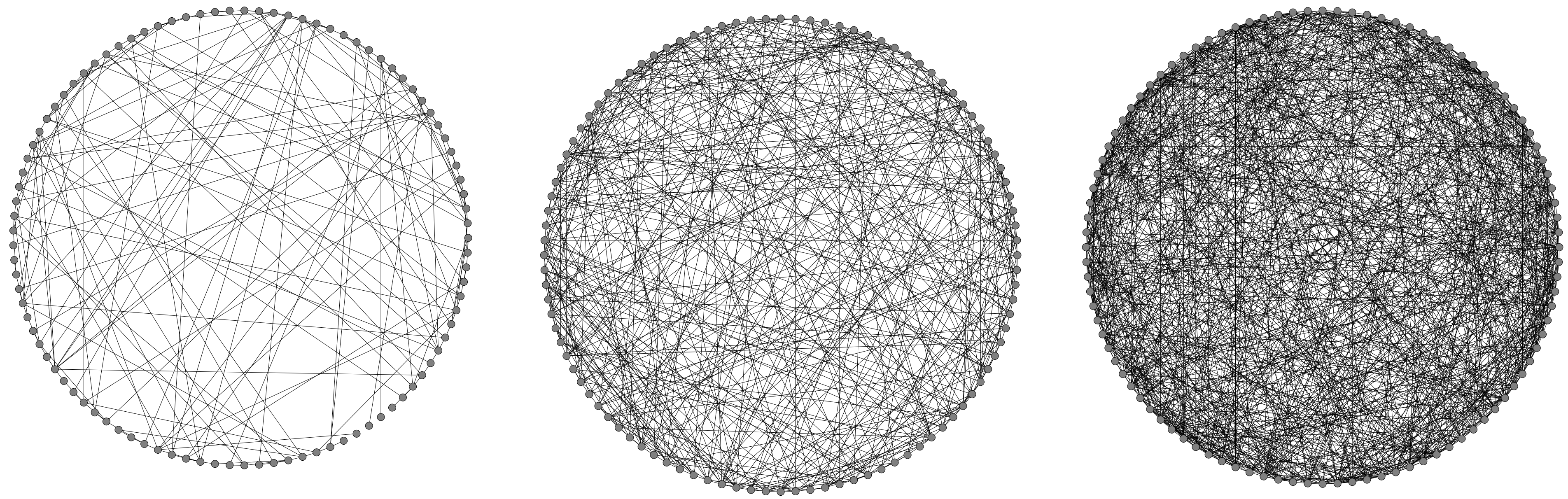
$\frac{\sum_i d_i}{2}$ ← Actual number of connections

$\binom{N}{2}$ ← Max possible connections: $\frac{N!}{(N-1)!2!} = \frac{N(N-1)}{2}$



$D = 0$ $\qquad$ $D = \frac{20}{6*5} = \frac{2}{3}$ $\qquad$ $D = 1$

Network Measures: Density



Density is increasing in average degree

$$D = \frac{\sum_i d_i}{N(N-1)} = \frac{\langle d \rangle}{(N-1)}$$

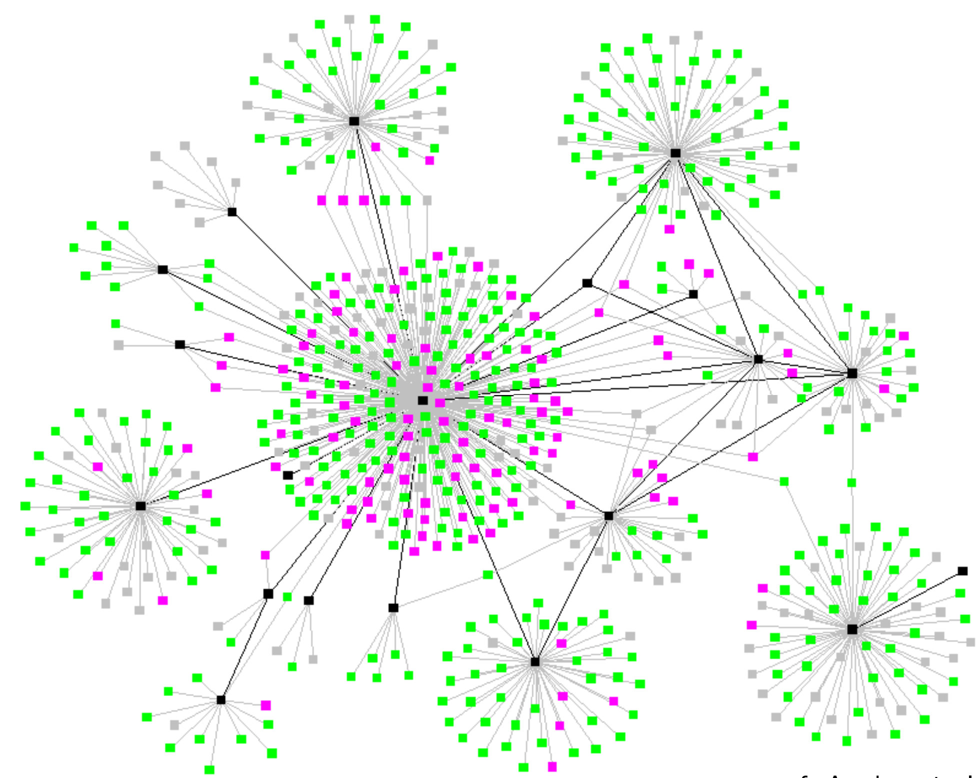
But it is also inversely related to the size of the network

Keep that in mind if you are trying to compare the density of two different networks…

Network Measures: Density

On the whole, social networks tend to relatively “sparse”

ref: Adamic



ref: Andre et al

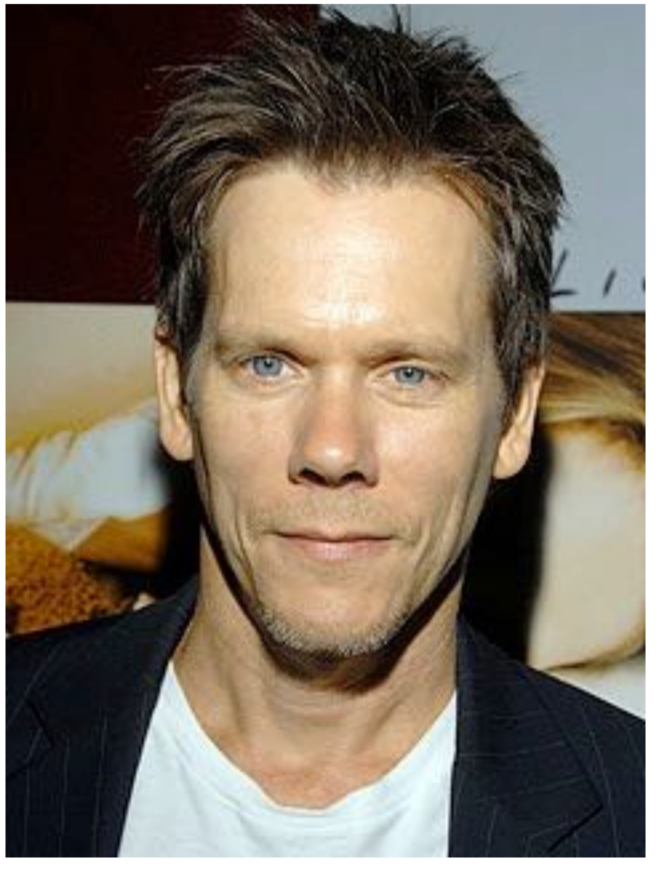
Networks that are more densely connected will aid the diffusion of information…but also disease.

Measures of Network Structure: Paths, Geodesic Paths, Distance

- Path: a sequence of nodes connecting node i to node j
- Geodesic Path: a shortest path between two nodes
- Distance: the length of the geodesic path between two nodes

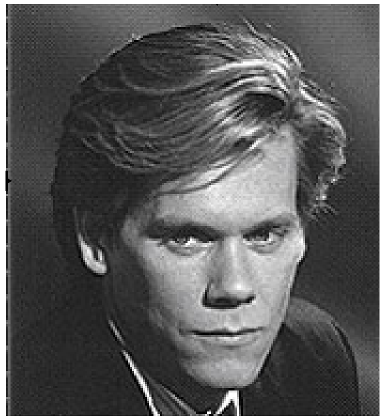
Geodesic distance is a measure of how “close” two nodes in the network are

A demonstration…



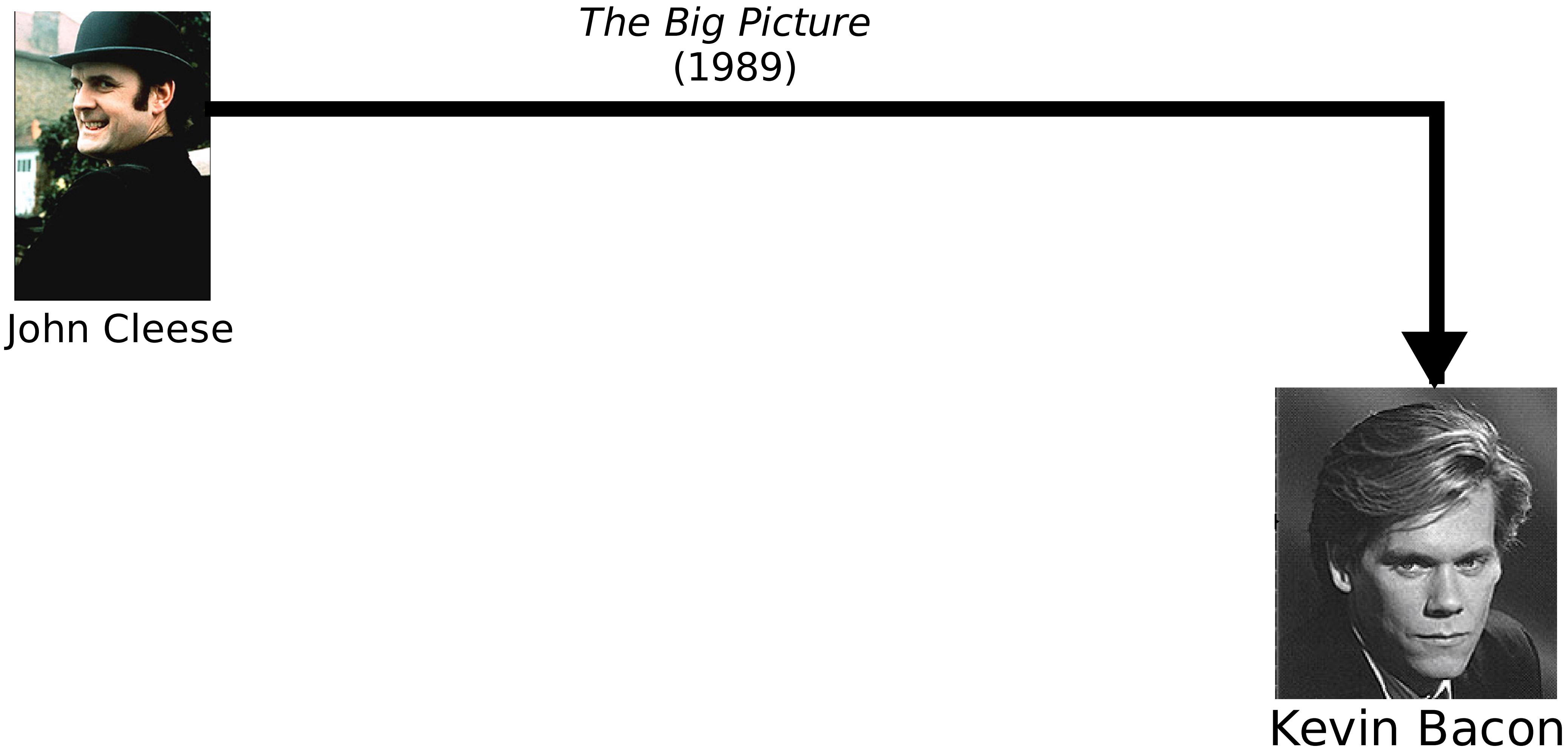
Six Degrees of Kevin Bacon

John Cleese



Kevin Bacon

Six Degrees of Kevin Bacon

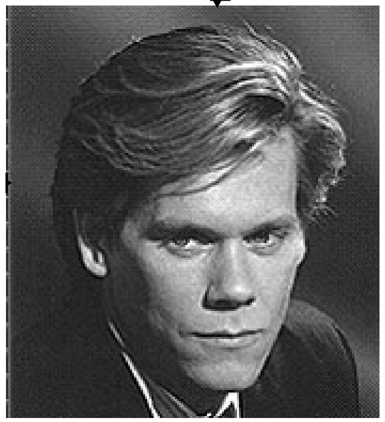


Six Degrees of Kevin Bacon

The Big Picture
(1989)

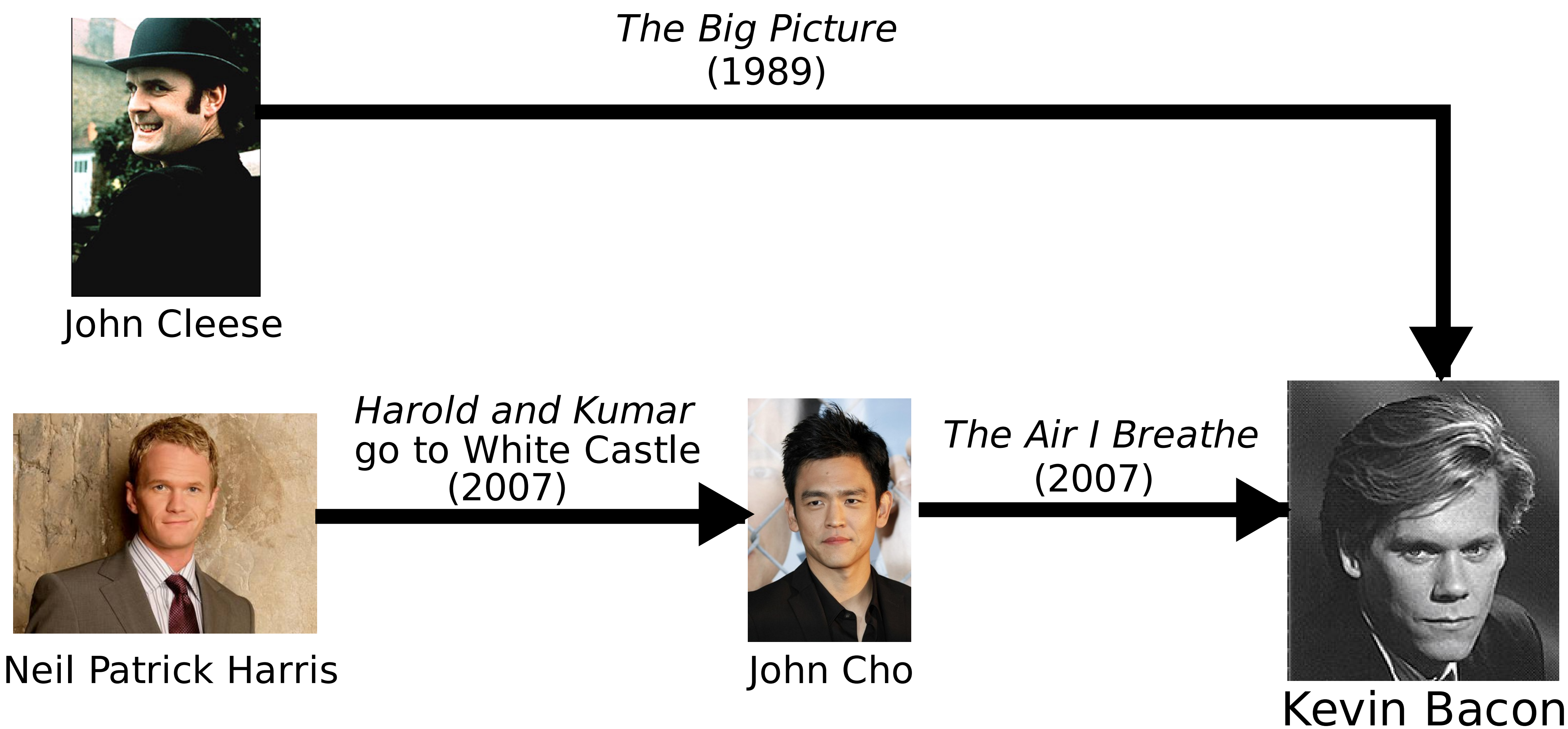
John Cleese

Neil Patrick Harris



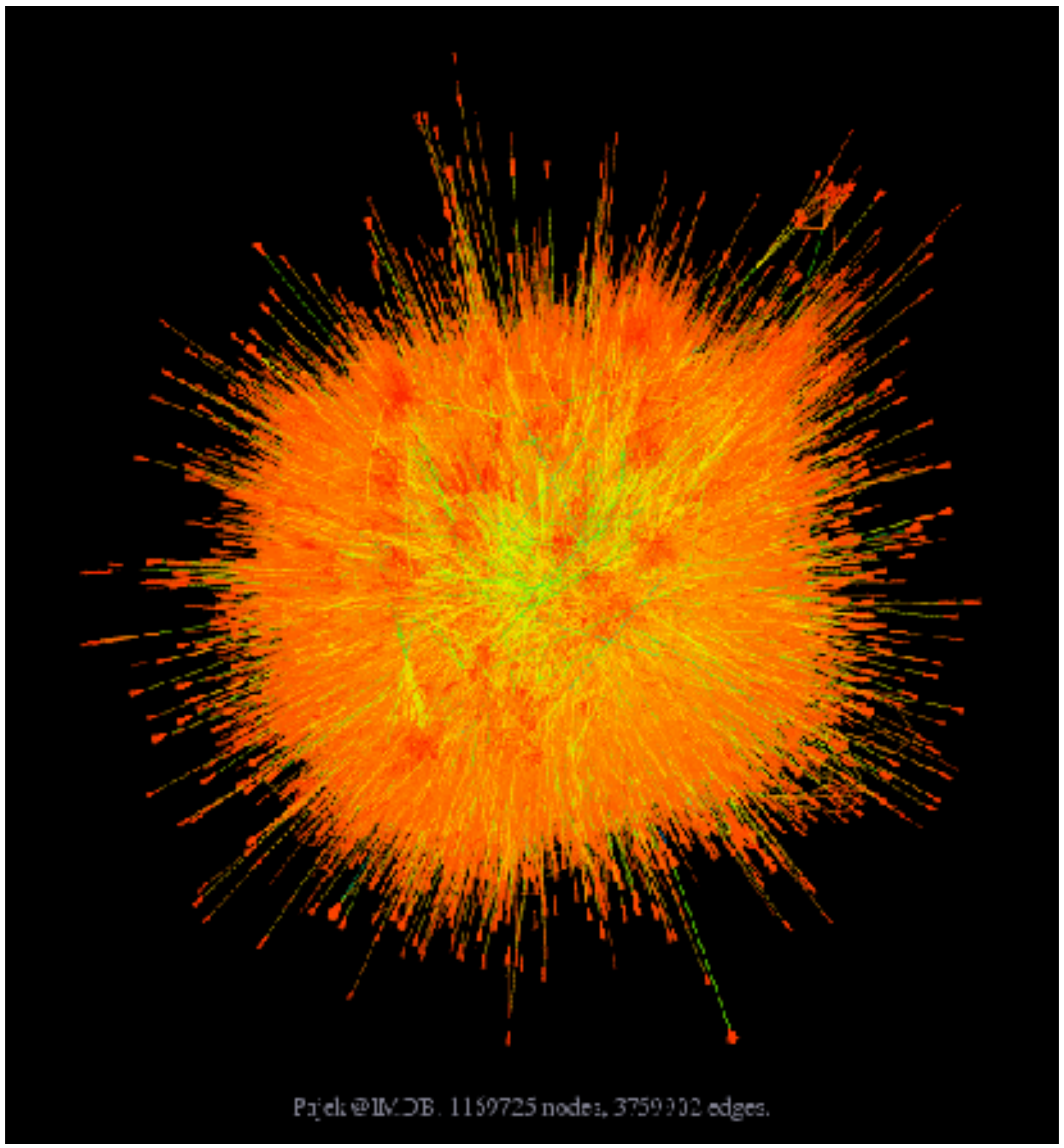
Kevin Bacon

Six Degrees of Kevin Bacon



Diameter / Average Path Length

IMDB network: >1 million actors



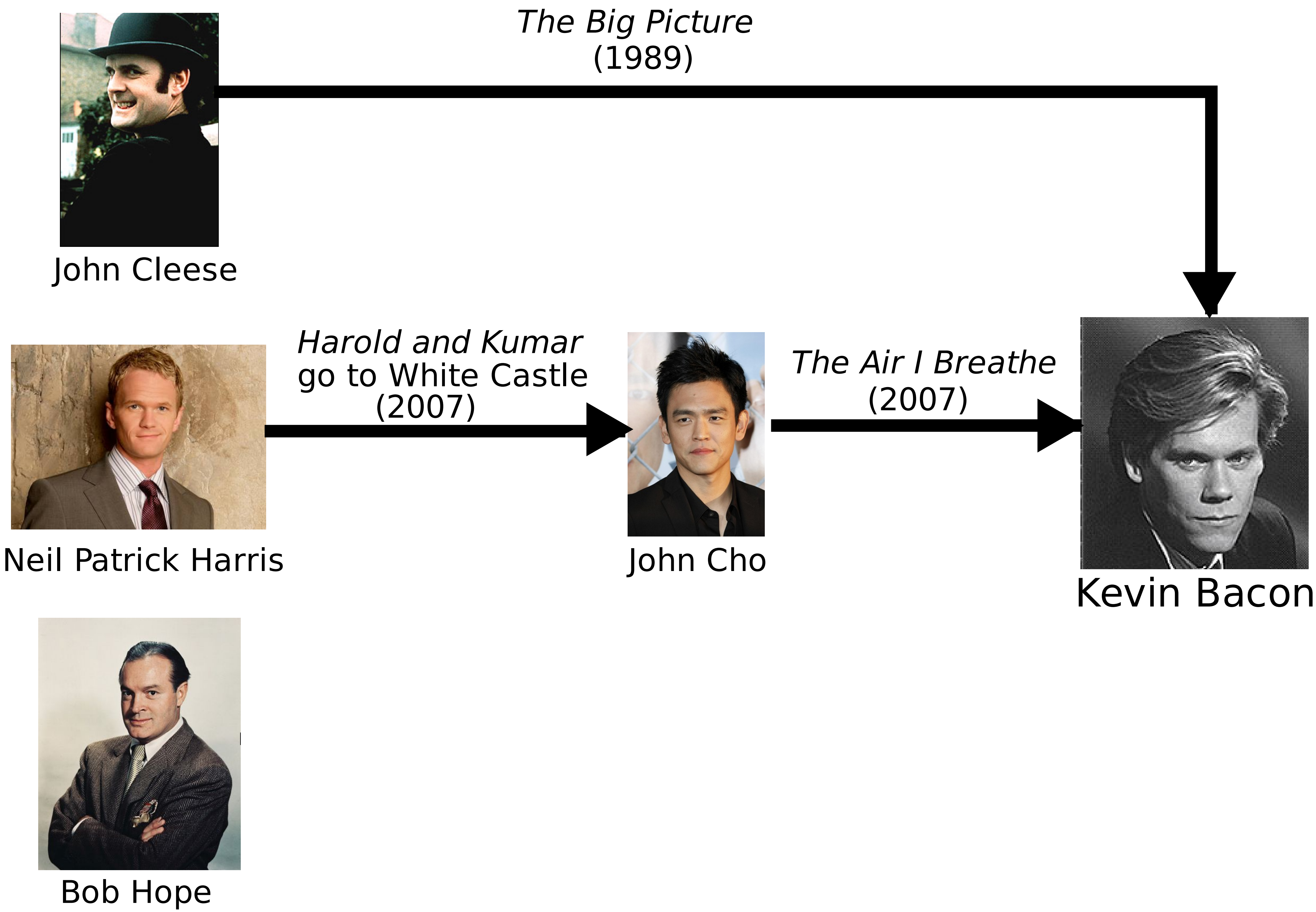
IMDB network (1888-2008)

Average path length = average distance between all pairs of nodes in the network

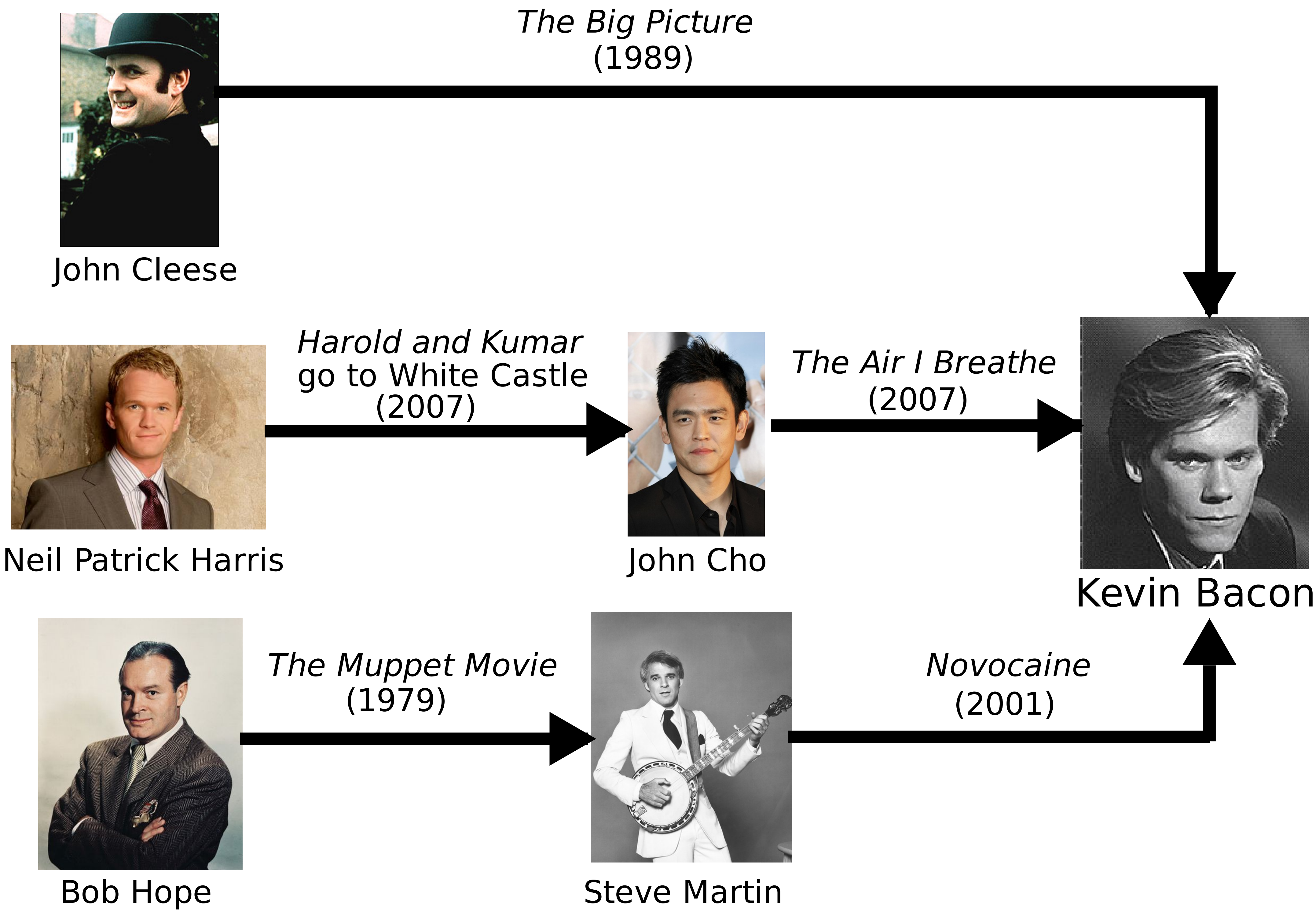
Diameter = length of the longest geodesic in the network

So how far apart are those 1 million people?

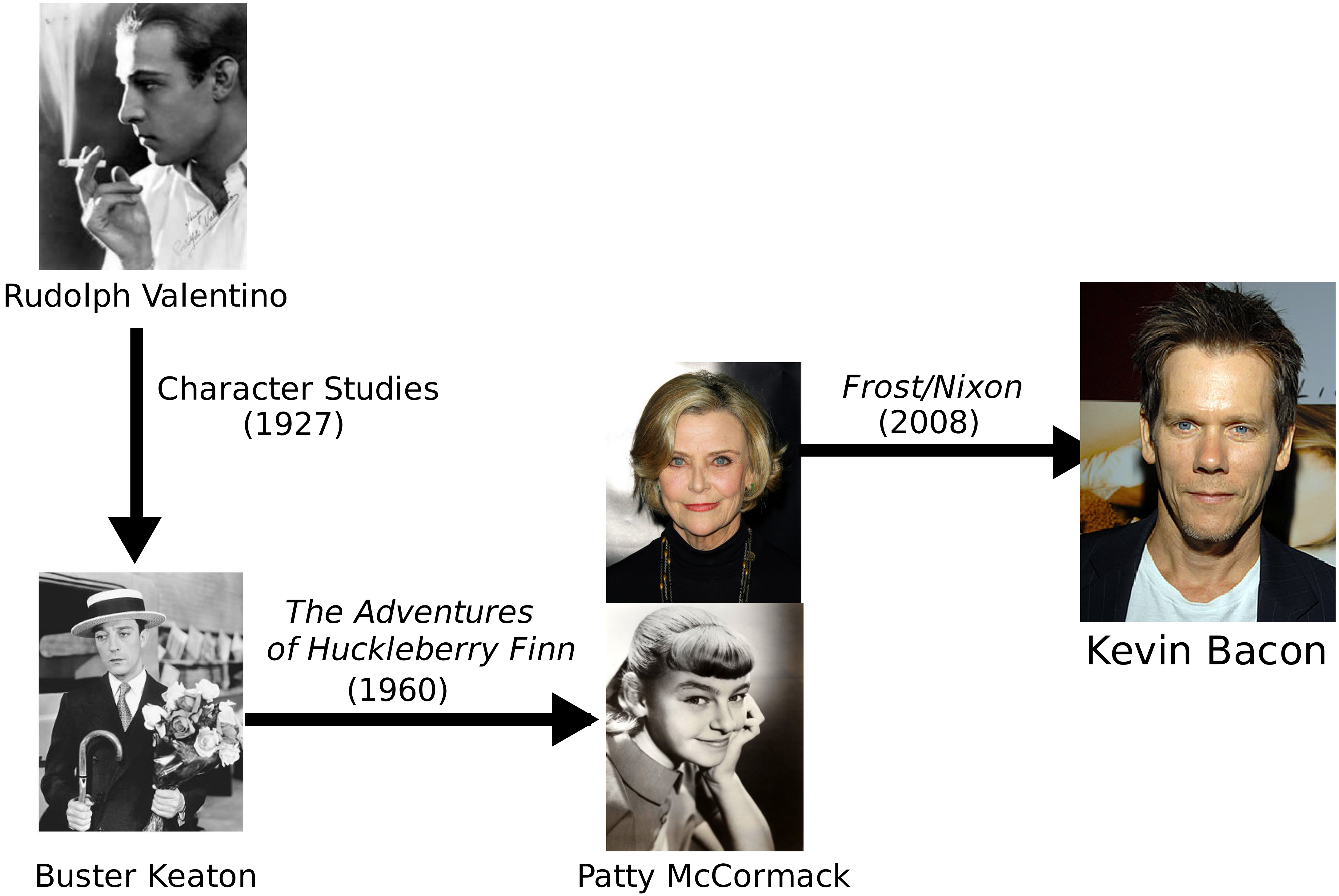
Six Degrees of Kevin Bacon



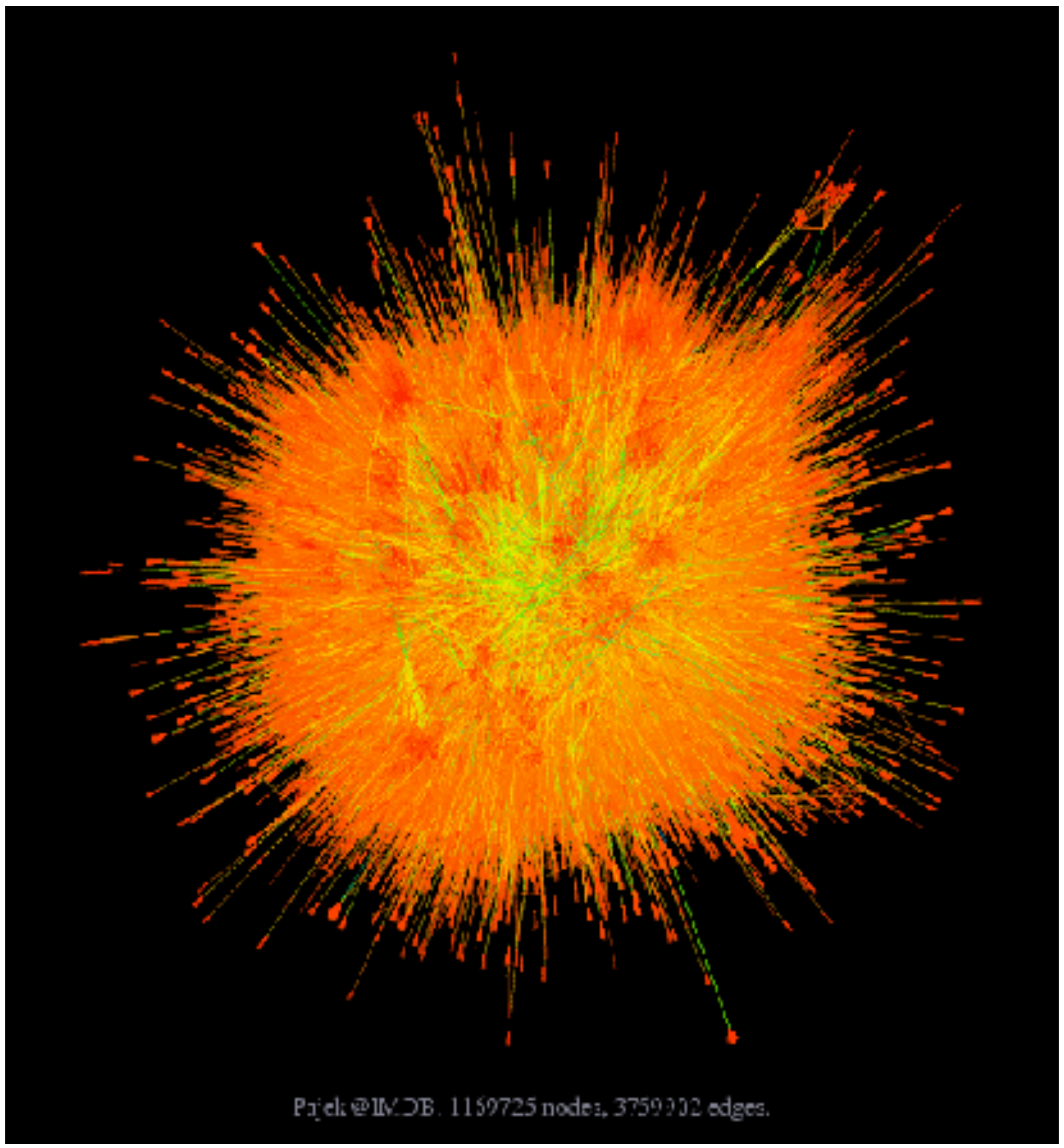
Six Degrees of Kevin Bacon



Six Degrees of Kevin Bacon



Measures of Network Structure: Diameter / Average Path Length



IMDB network (1888-2008)

IMDB network: >1 million actors

Average path length: 2.7 links!

Longest path length: 8 links

So how about human social interaction?

How far apart are *we*?

Measures of Network Structure: “Six Degrees of Separation”

Stanley Milgram’s “Small World” Experiment (1967)

- Theory: the distance between any two people in the world is surprisingly short
    - Based on a conjecture from Guglielmo Marconi
    - Popularized by Hungarian author Frigyes Karinthy
- Question: what is the average path length between two people in the US?

Measures of Network Structure: "Six Degrees of Separation"

Test: See how many hops it takes to get a letter from Wichita, Kansas to a man living in Boston, Mass

- Letters sent to 296 individuals in Kansas
- Target is an individual in Boston
- They can send their letter along to any person they know on a first-name basis
- They sign their name to a roster in the letter and also send a postcard to the researchers

Measures of Network Structure: "Six Degrees of Separation"

Results

- 64 of the 296 letters made it
- Some required only 2 exchanges. Others 10.
- Among those that made it, the average number of hops was 5-6
- Many of the letters got to the target through the same sources (information hubs)
- Geographical distance was easier to cross than social distance

Critiques of this experiment?

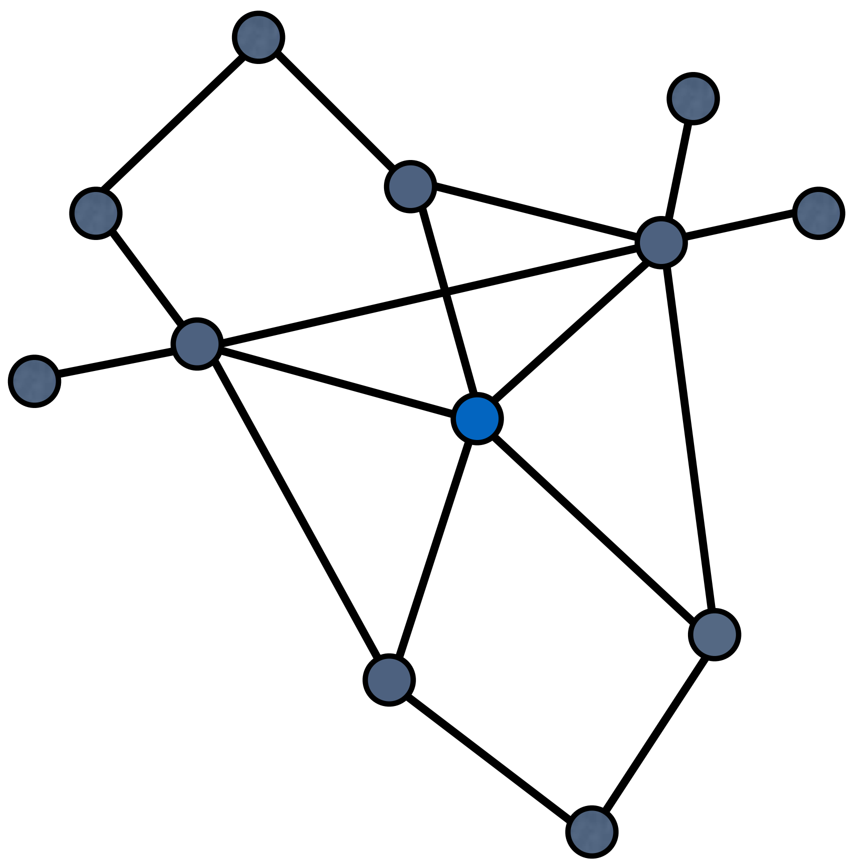
Measures of Network Structure: “Six Degrees of Separation”

Criticisms

- Geographic distance ≠ Social distance
- Initial letter recipients self-selected as “very connected” (not a random sample)
- Long chains more likely to fail (survivor bias)
- Network search ≠ geodesic distance
- Many of the letters got to the target through the same sources (information hubs)

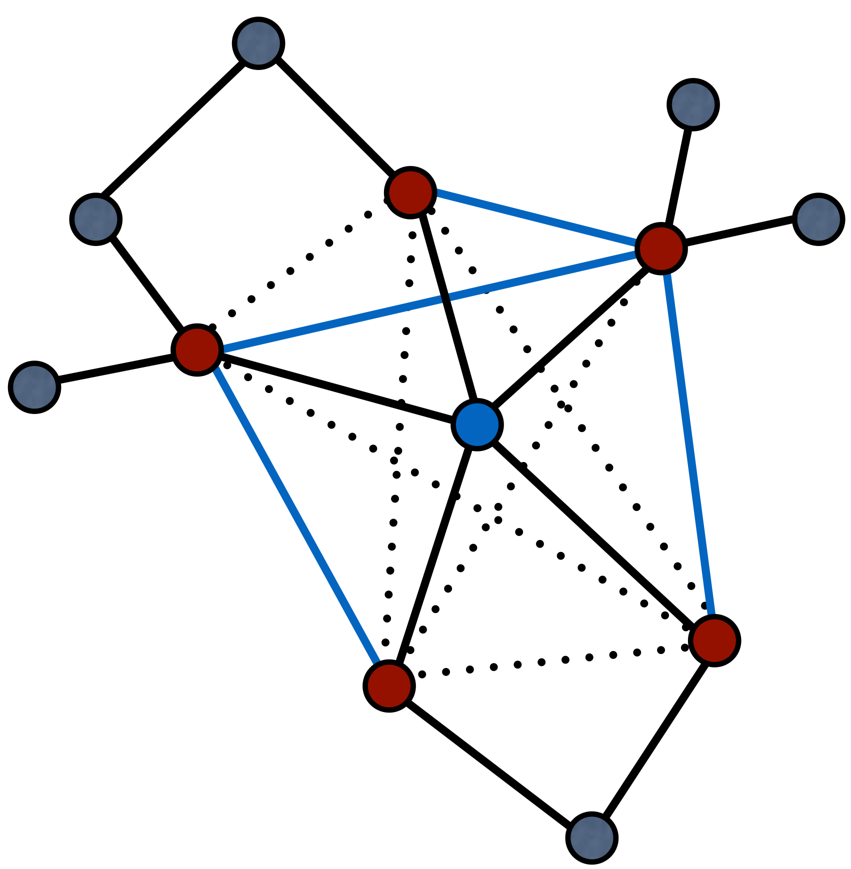
Measures of Network Structure: Clustering

Clustering: the probability that your friends know each other



Measures of Network Structure: Clustering

Clustering: the probability that your friends know each other



The blue node has 5 neighbors (the dark red nodes)

There are 4 edges between those neighbors (blue)

There are a total of 10 possible connections between any 5 nodes

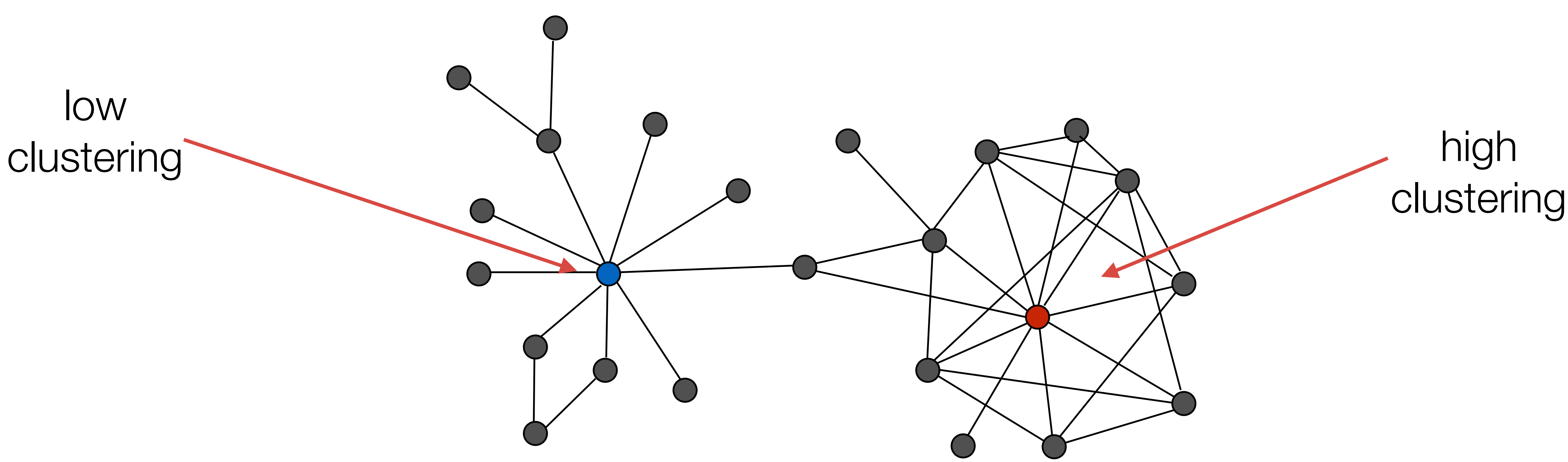
$$10 = \binom{5}{2}$$

So the clustering coefficient for the blue node is 4/10 = .4

Measures of Network Structure: Clustering

More generally, the clustering coefficient for node i is the probability that two of a node's immediate connections are connected to each other

$$c_i = \frac{n_i}{\binom{d_i}{2}}$$ where n_i is the number of a node's neighbors who are connected to each other

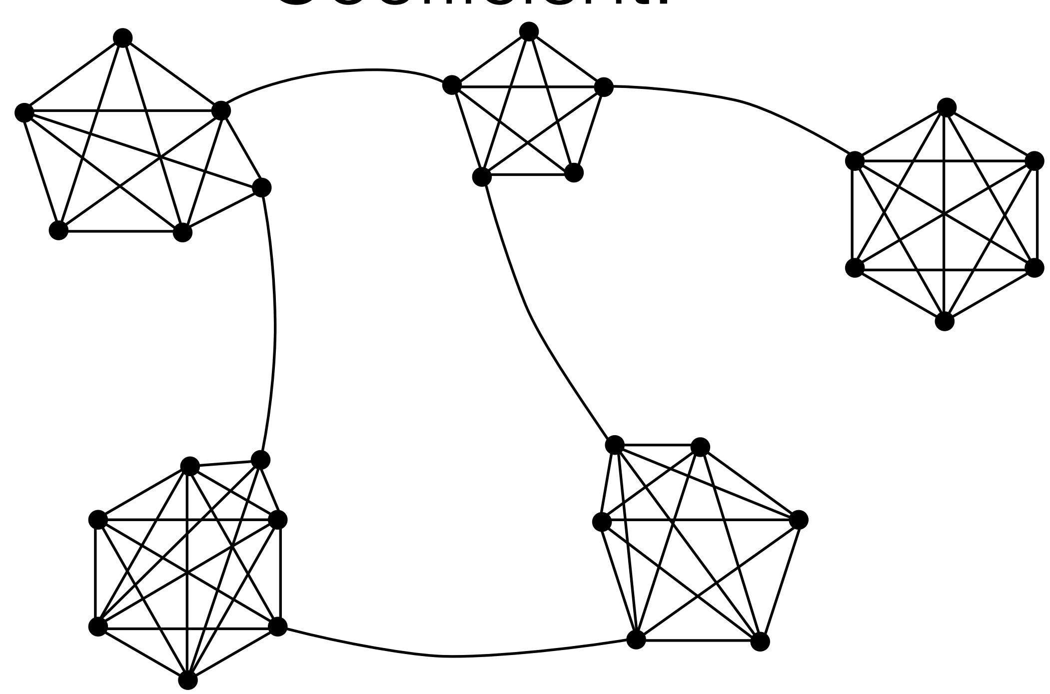


Measures of Network Structure: Clustering

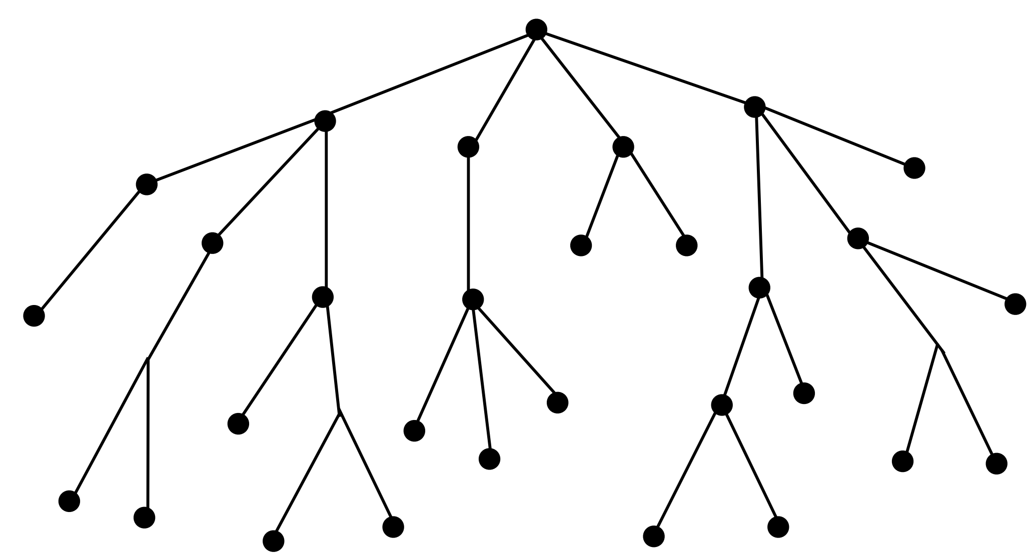
You can also measure the clustering coefficient for an entire network: the average of the clustering coefficients of all of the nodes

$$C = \frac{1}{n} \sum_{i=1}^{n} c_i$$

High Clustering Coefficient:



Low Clustering Coefficient:

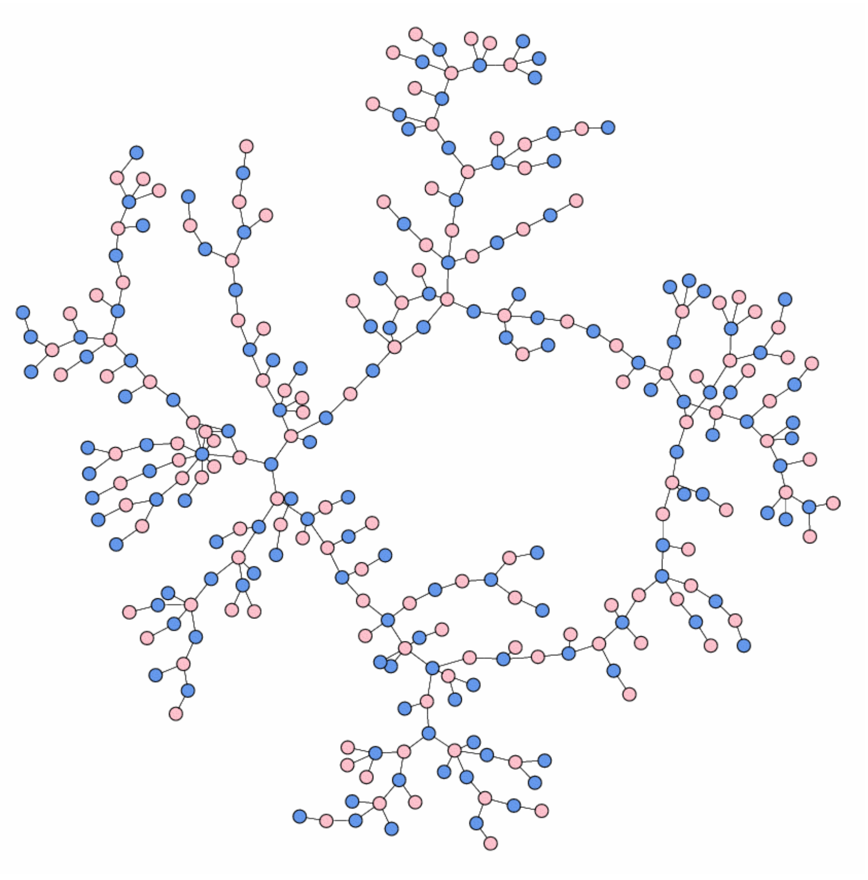


Measures of Network Structure: Clustering

Social Networks tend to have very high clustering (relative to random networks)

→ Your friends tend to know each other

(though there are exceptions)



High School Dating

ref: Data by Bearman et al (2004)
Graphic by M.E.J. Newman

	Empirical
Average Distance	Low
Density	Low
Clustering	High
Degree Distribution