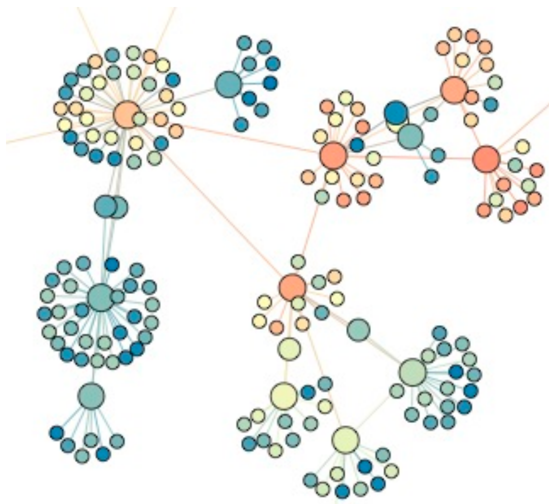
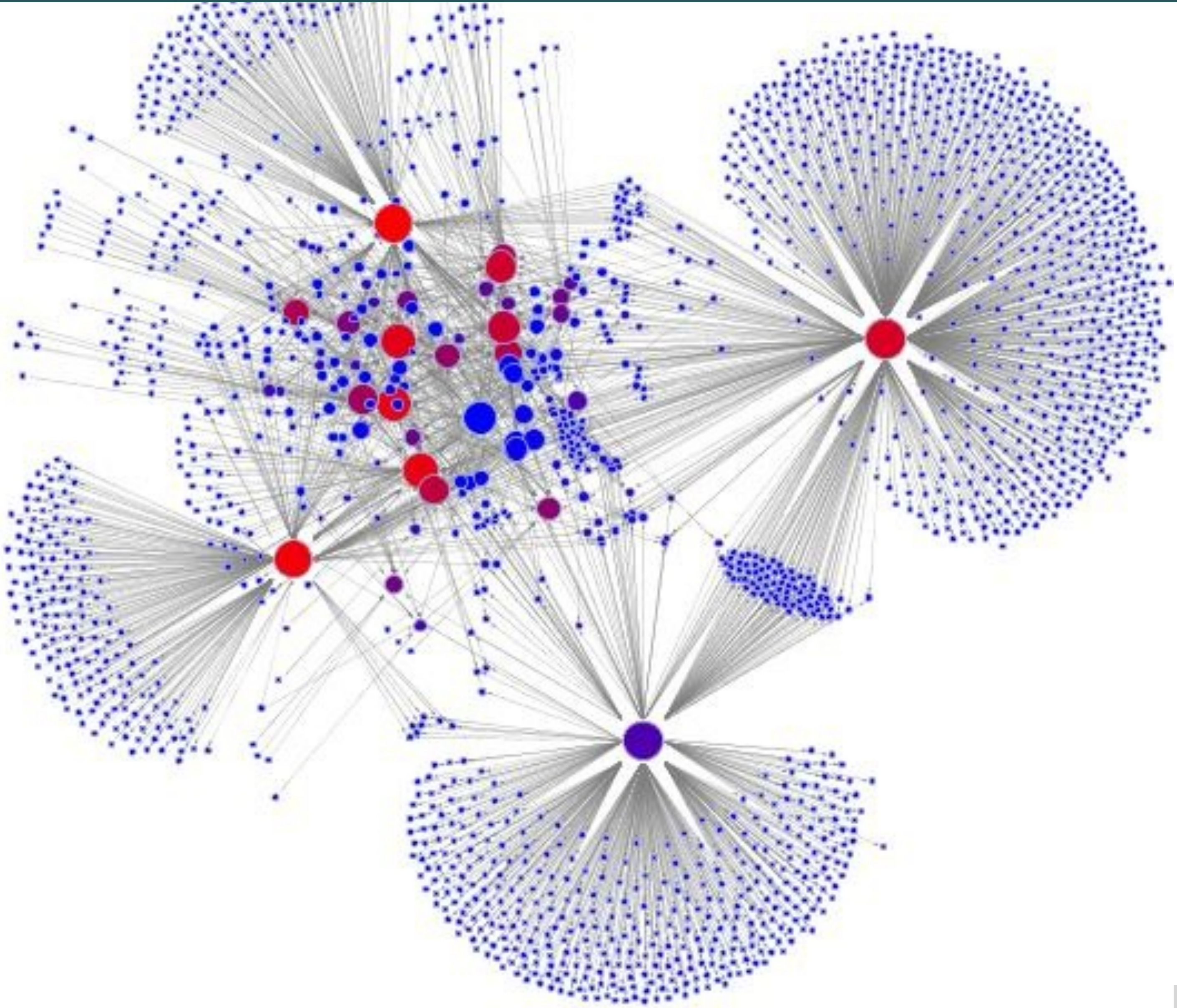


SNA 2C: Growth & Preferential Attachment Models

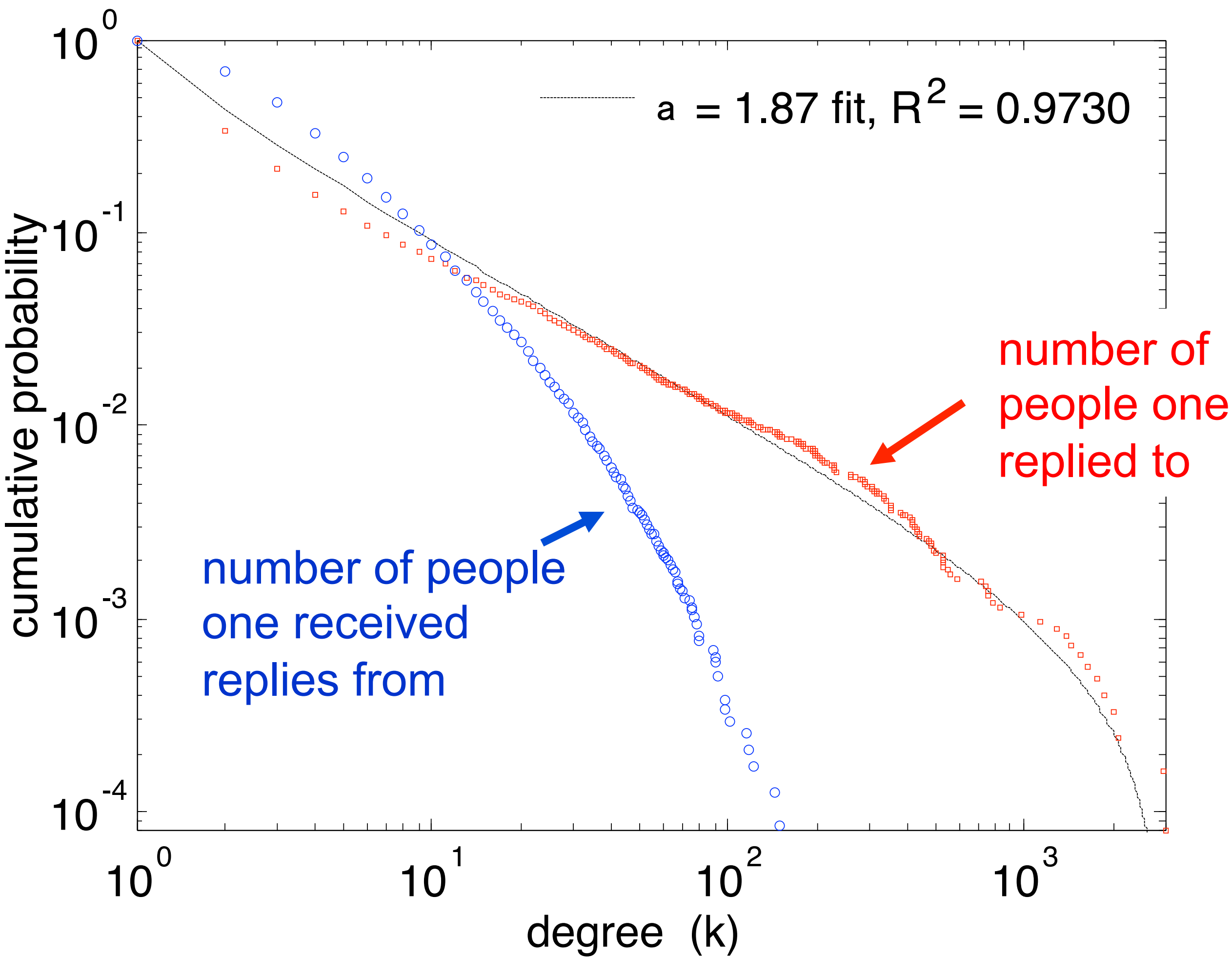
Lada Adamic

school of information

Online Question & Answer Forums



Uneven participation

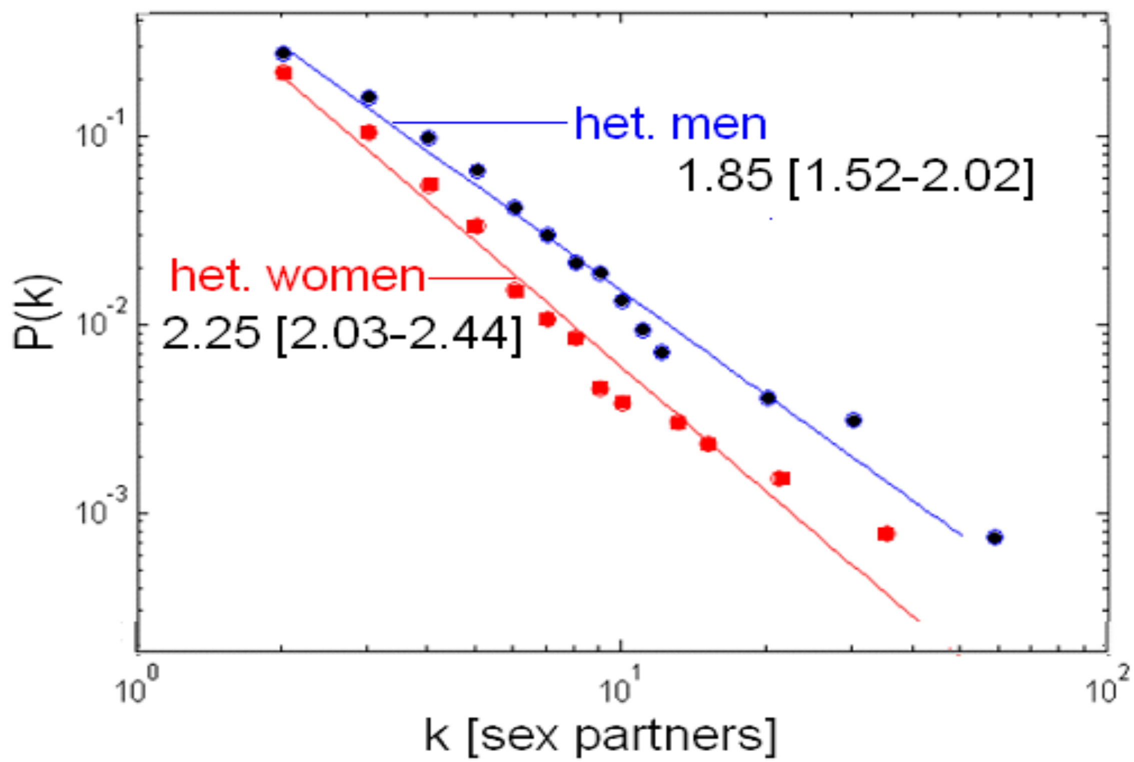


- ‘answer people’ may reply to thousands of others
- ‘question people’ are also uneven in the number of repliers to their posts, but to a lesser extent

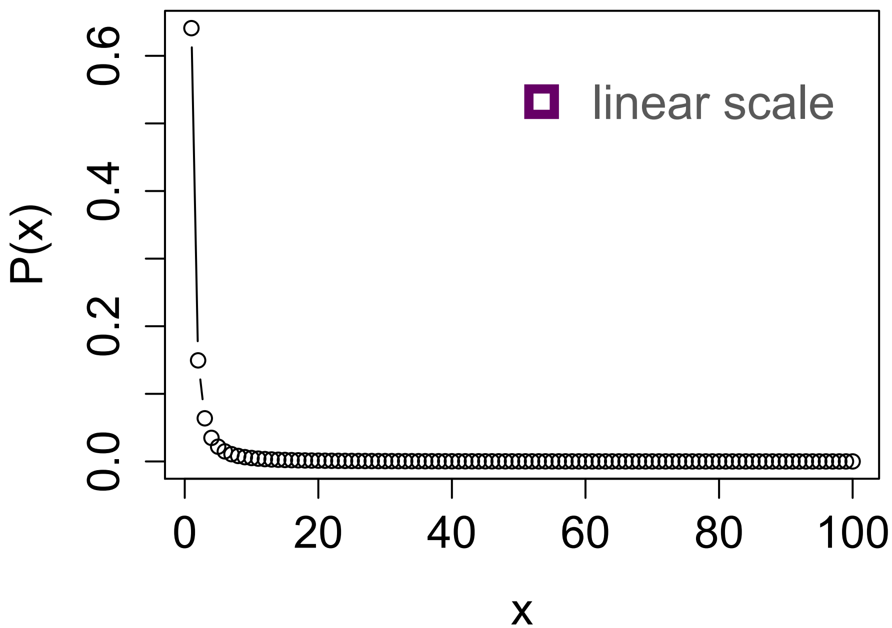
Real-world degree distributions

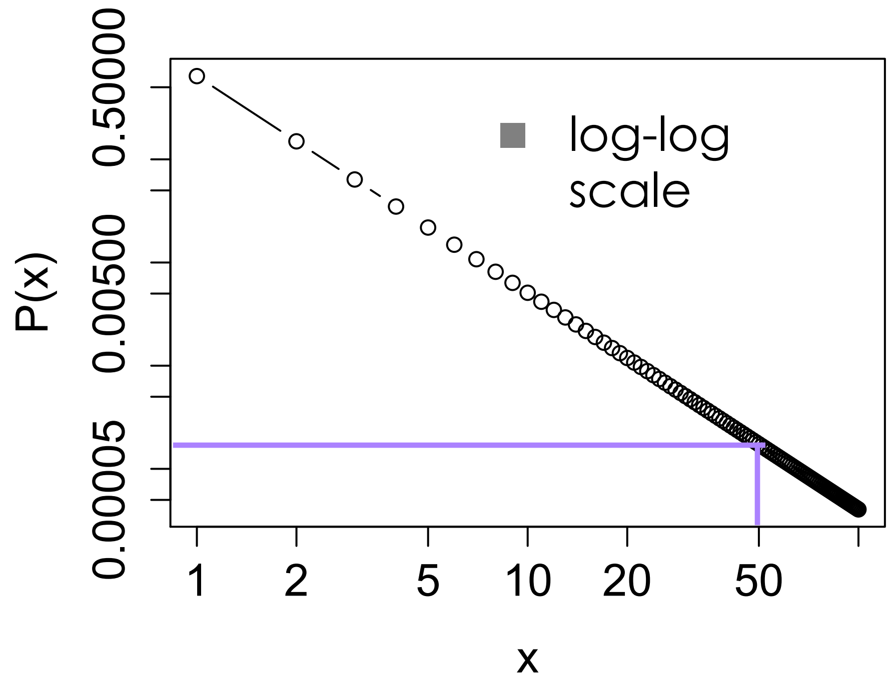
- Sexual networks

- Great variation in contact numbers



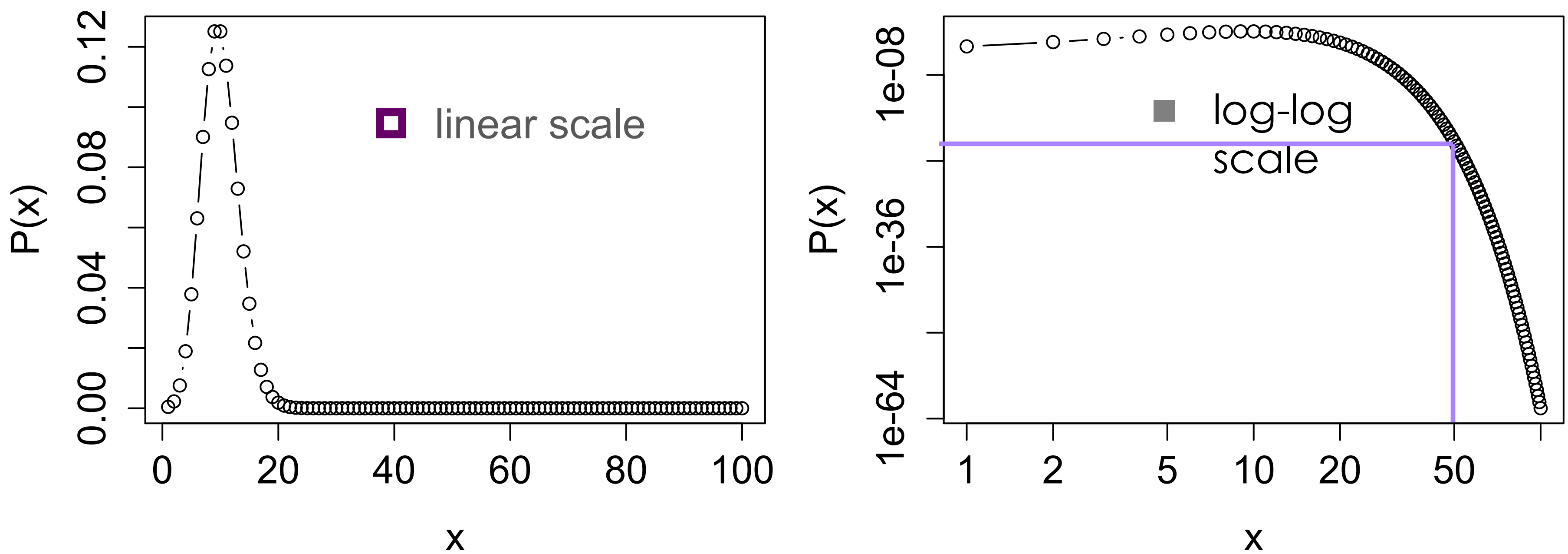
Power-law distribution





- high skew (asymmetry)
- straight line on a log-log plot

Poisson distribution



- little skew (asymmetry)
- curved on a log-log plot

Power law distribution

- Straight line on a log-log plot

$$\ln(p(k)) = c - \alpha \ln(k)$$

- Exponentiate both sides to get that *p(k)*, the probability of observing an node of degree ‘k’ is given by

$$p(k) = Ck^{-\alpha}$$

normalization constant (probabilities over all *k* must sum to 1)

power law exponent a

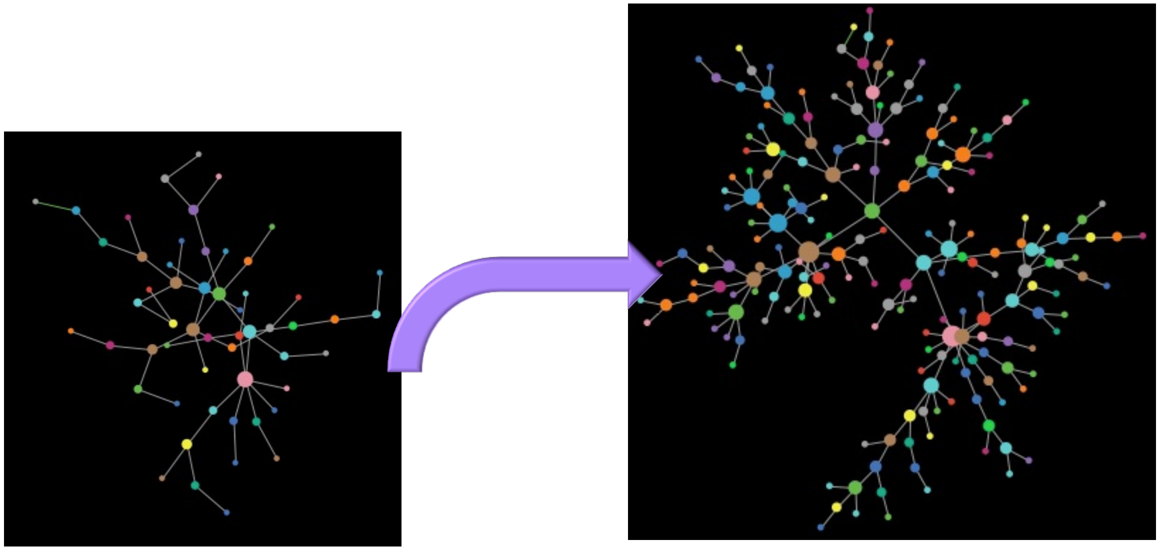
Quiz Q:

- As the exponent α increases, the downward slope of the line on a log-log plot
  - stays the same
  - becomes milder
  - becomes steeper

$$\ln(p(k)) = c - \alpha \ln(k)$$

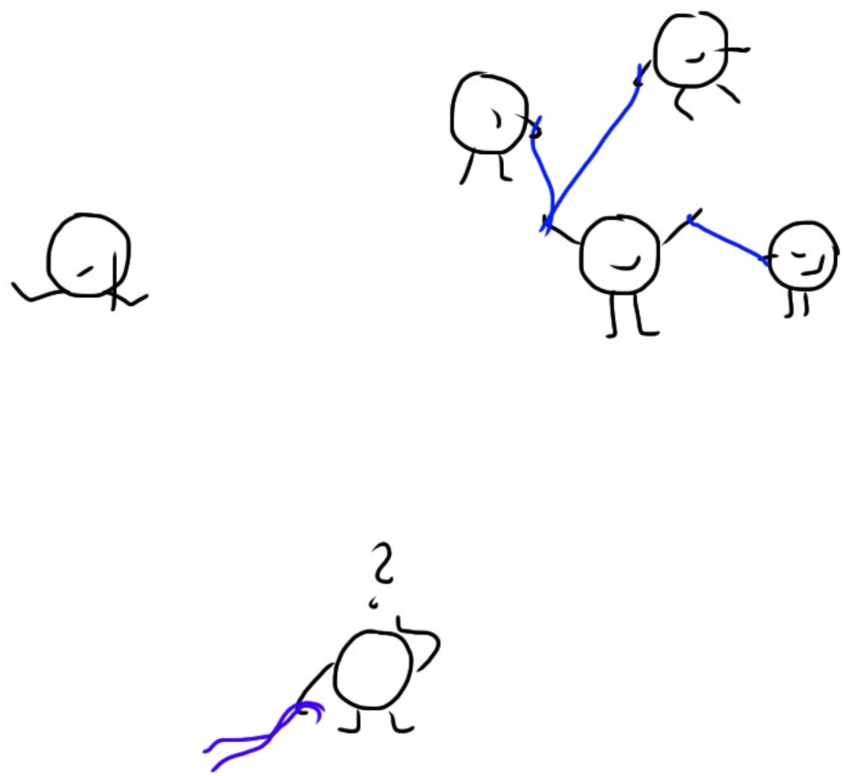
2 ingredients in generating power-law networks

- nodes appear over time (growth)



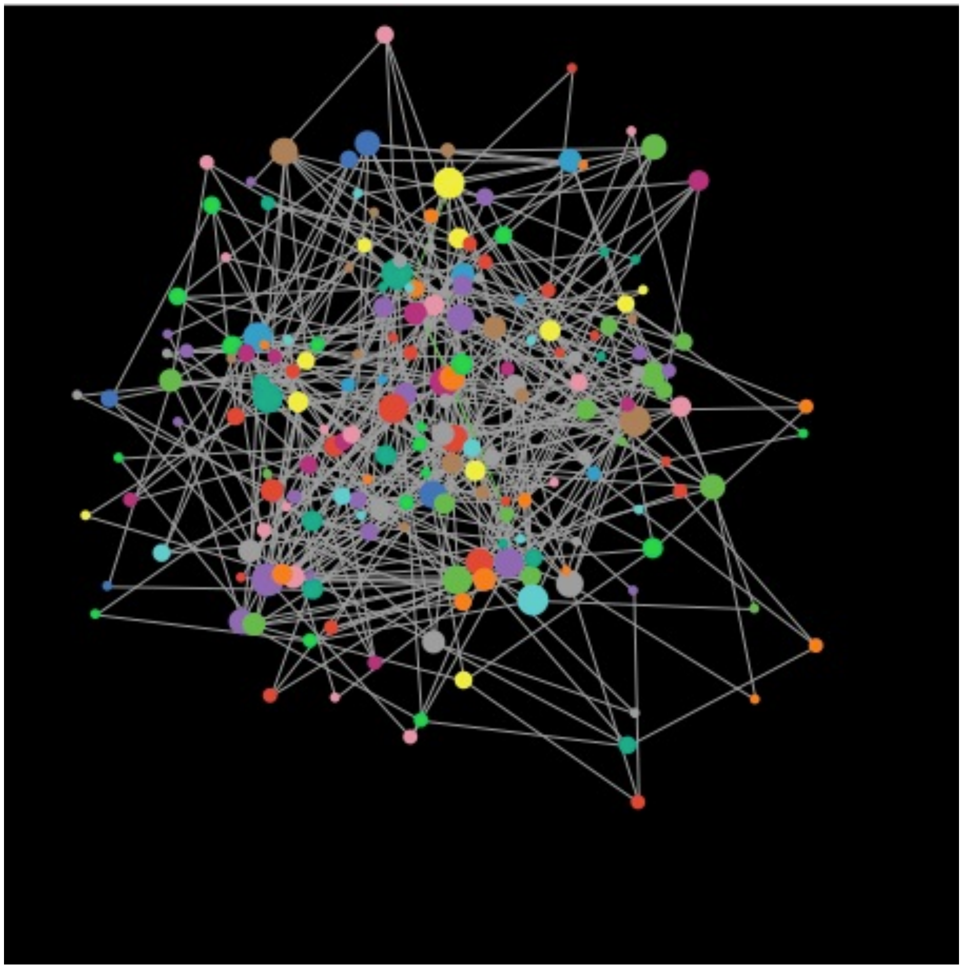
2 ingredients in generating power-law networks

- nodes prefer to attach to nodes with many connections (preferential attachment, cumulative advantage)



Ingredient # 1: growth over time

- nodes appear one by one, each selecting ***m*** other nodes at random to connect to



m = 2

random network growth

- one node is born at each time tick
- at time ***t*** there are ***t*** nodes
- change in degree k_i of node i (born at time i, with $0 < i < t$)

$$\frac{m}{t}$$

there are ***m*** new edges being added per unit time (with 1 new node)

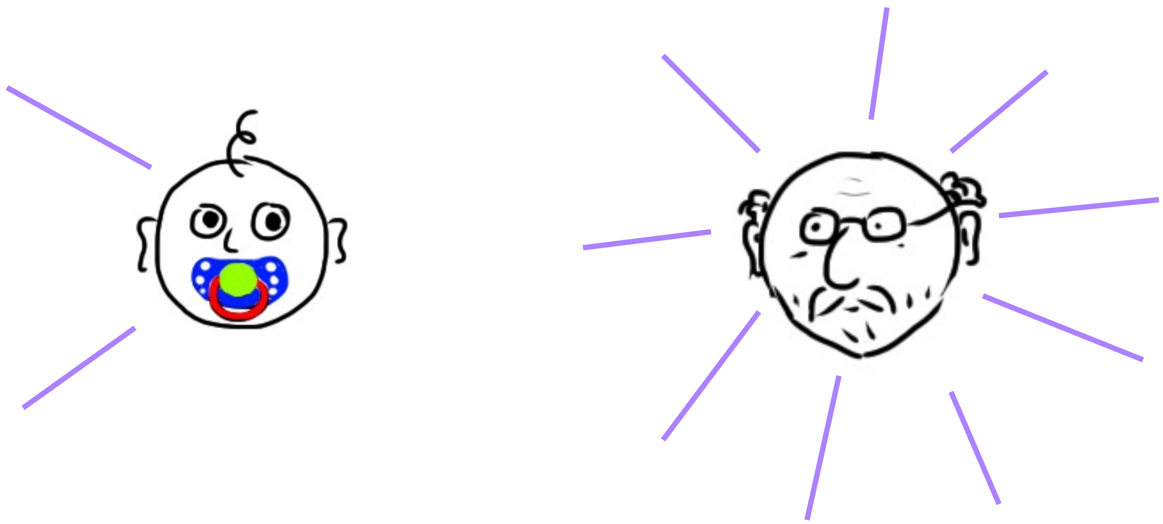
the ***m*** edges are being distributed among ***t*** nodes

age and degree

on average $k_i(t) > k_j(t)$

if $i < j$

i.e. older nodes on average have more edges



Quiz Q:

- How could one make the growth model more realistic for social networks?
  - old nodes die
  - some nodes are more sociable
  - friendships vane over time
  - all of the above

2nd ingredient: preferential attachment

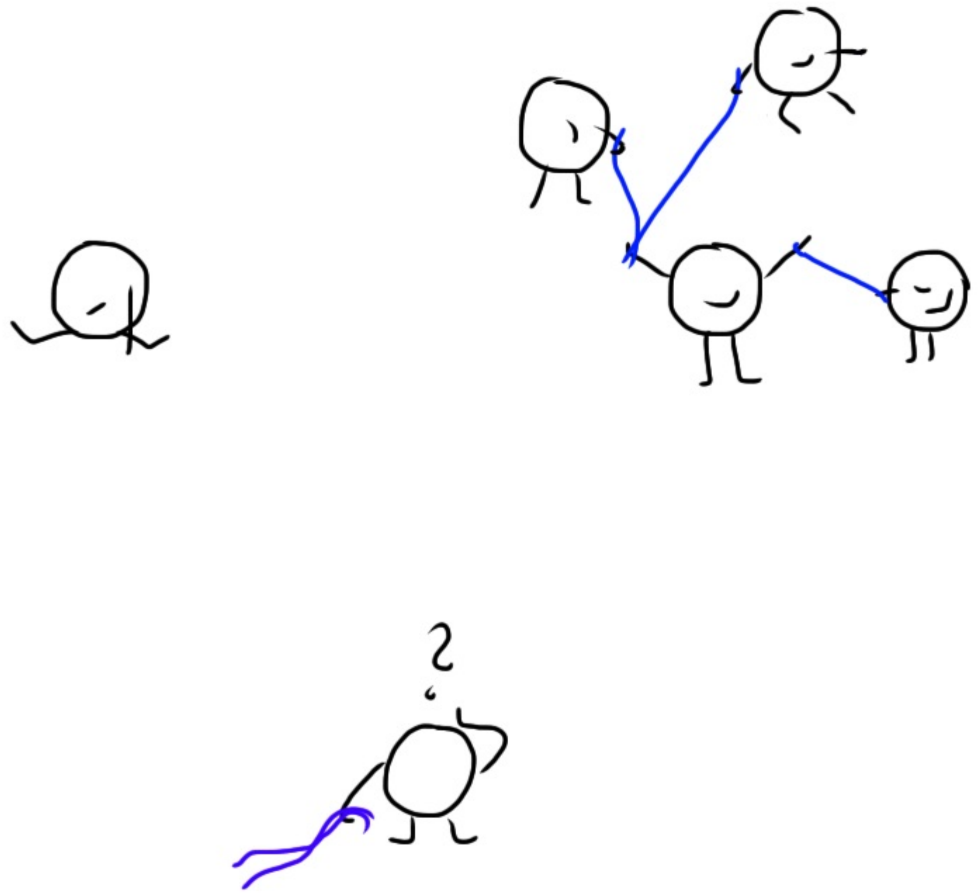
- Preferential attachment:
  - new nodes prefer to attach to well-connected nodes over less-well connected nodes
- Process also known as
  - cumulative advantage
  - rich-get-richer
  - Matthew effect

Price's preferential attachment model for citation networks

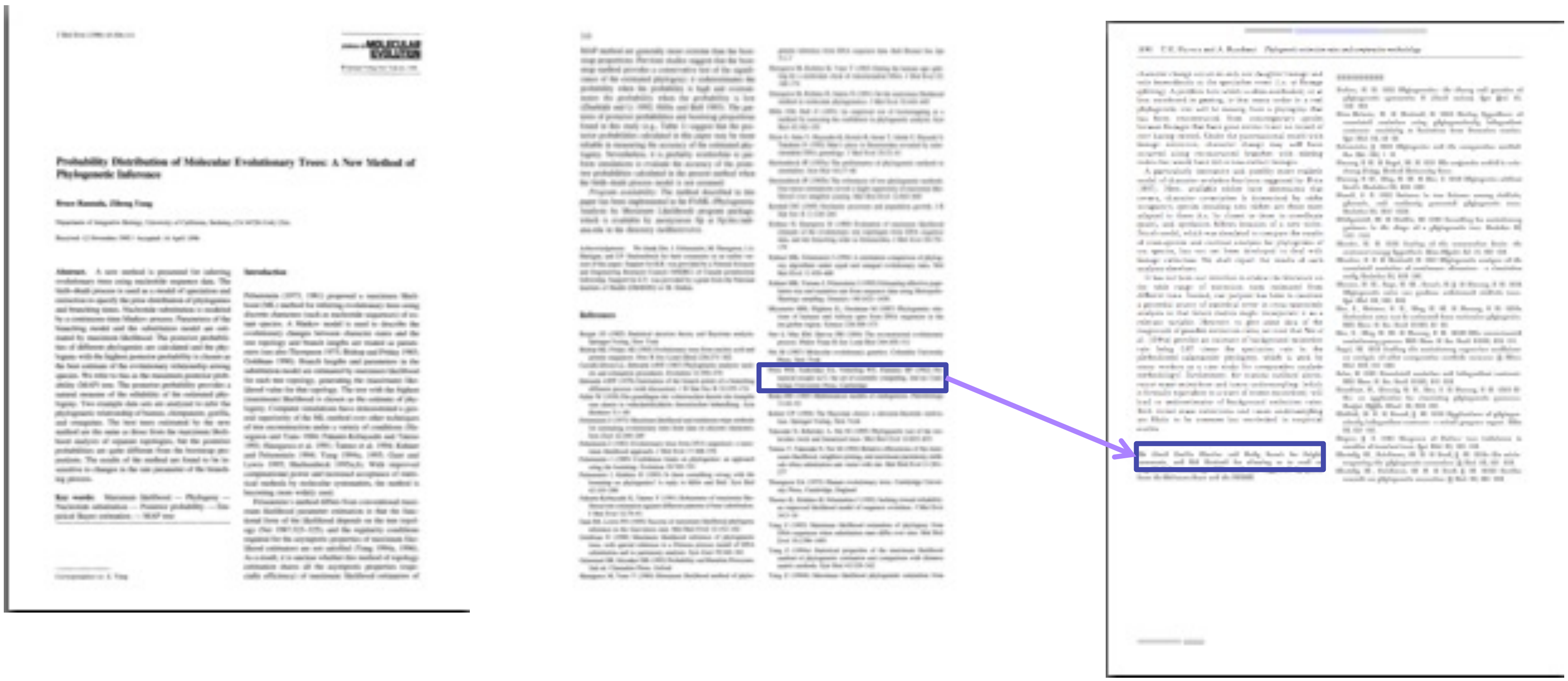
- [Price 65]
  - each new paper is generated with m citations (mean)
  - new papers cite previous papers with probability proportional to their indegree (citations)
  - what about papers without any citations?
    - each paper is considered to have a "default" citation
    - probability of citing a paper with degree k, proportional to k+1

- Power law with exponent $\alpha = 2+1/m$

Preferential attachment



Cumulative advantage: how?

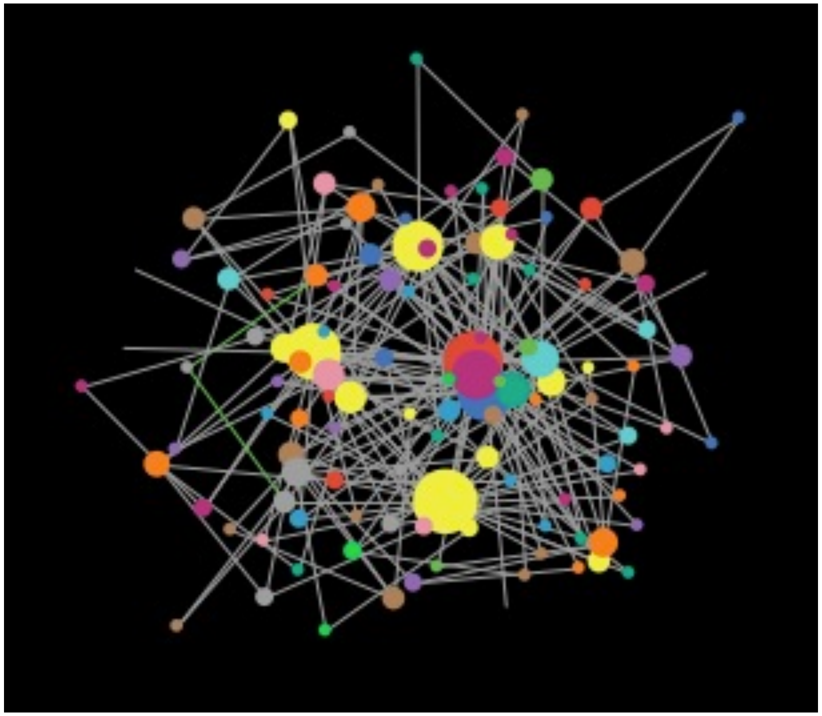


- copying mechanism
- visibility

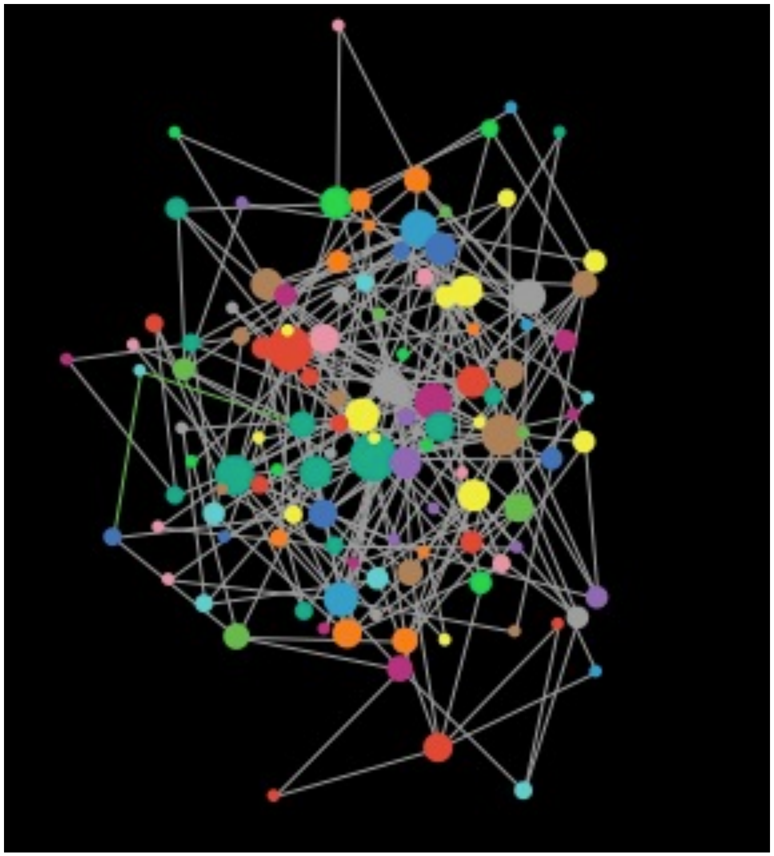
Barabasi-Albert model

- First used to describe skewed degree distribution of the World Wide Web
- Each node connects to other nodes with probability proportional to their degree
  - the process starts with some initial subgraph
  - each new node comes in with *m* edges
  - probability of connecting to node *i*
- Results in power-law with exponent $\alpha = 3$

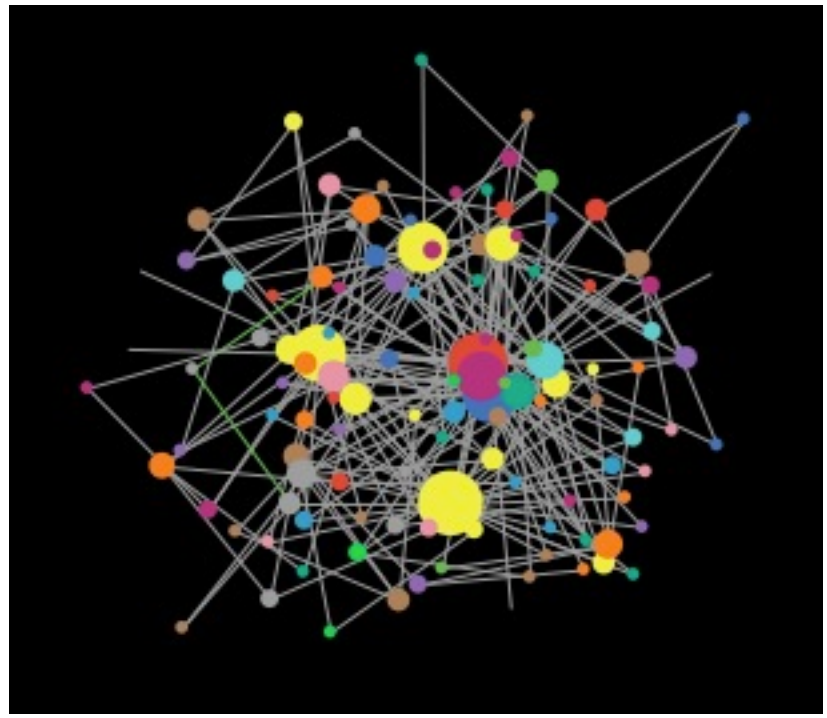
after a while...



contrasting with random (non-preferential) growth



m = 2



random

preferential

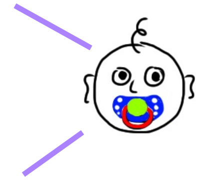
Properties of the BA graph

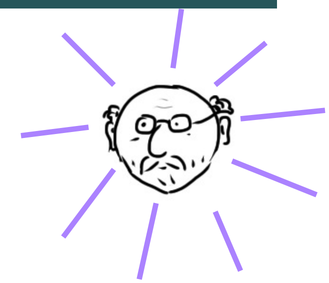
- The distribution is scale free with exponent $\alpha = 3$

  $P(k) = 2\, m^2/k^3$

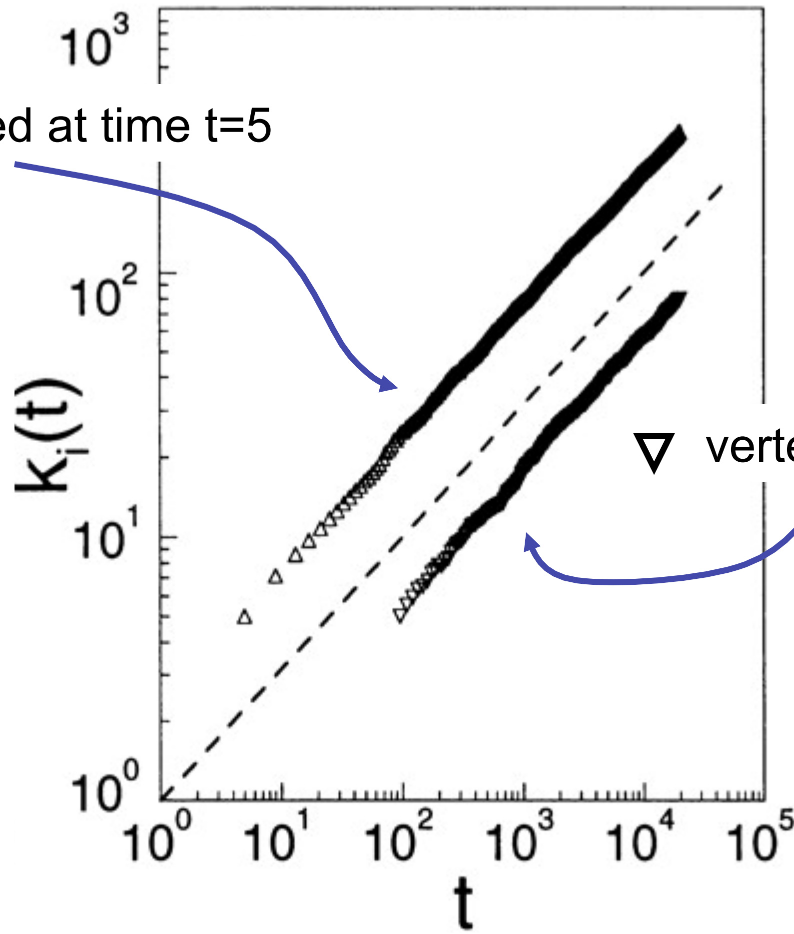
- The graph is connected
  - Every new vertex is born with a link or several links (depending on whether $m = 1$ or $m > 1$)
  - It then connects to an 'older' vertex, which itself connected to another vertex when it was introduced
  - And we started from a connected core
- The older are richer
  - Nodes accumulate links as time goes on, which gives older nodes an advantage since newer nodes are going to attach preferentially – and older nodes have a higher degree to tempt them with than some new kid on the block

Young vs. old in BA model



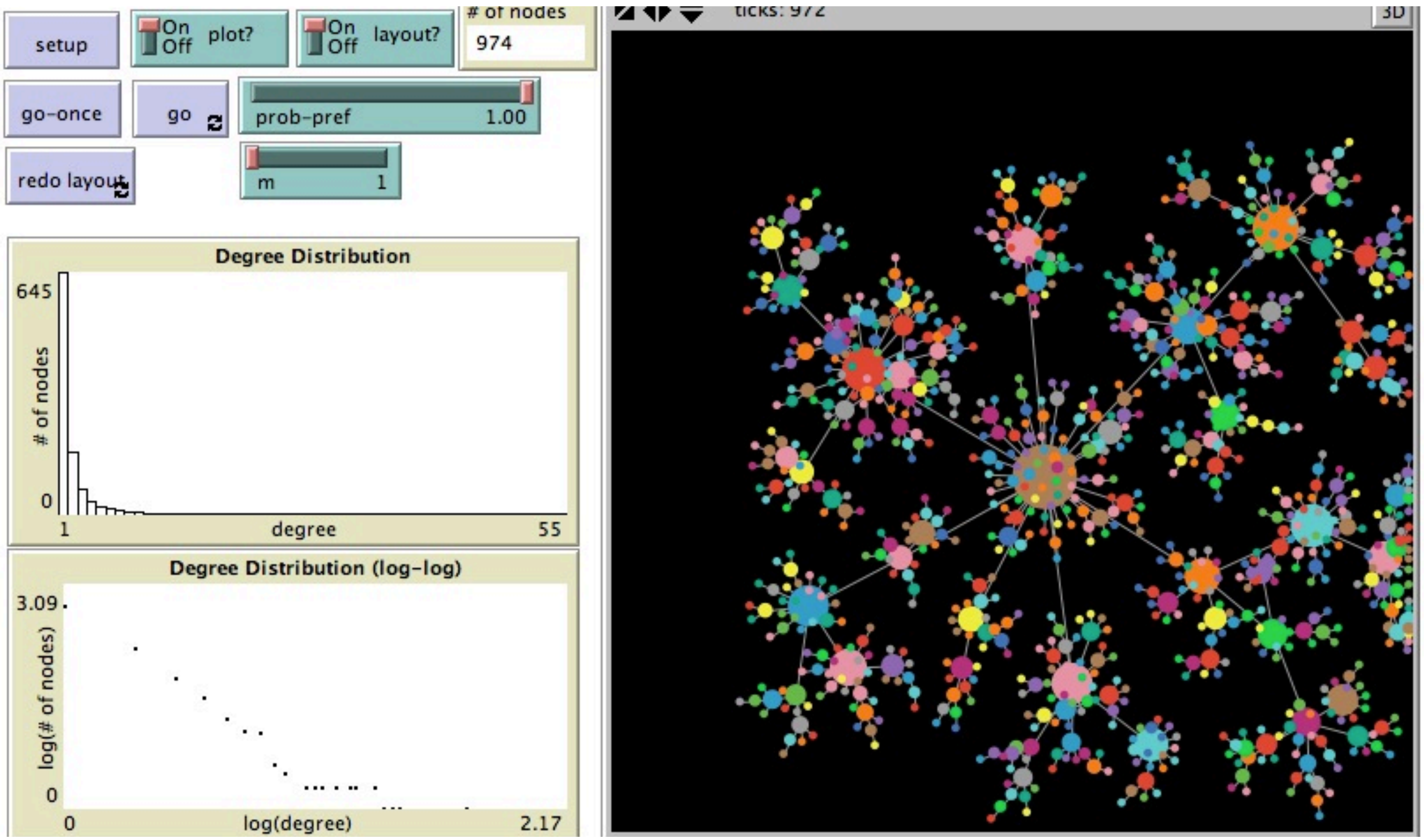


△ vertex introduced at time t=5



▽ vertex introduced at time t=95

try it yourself



http://ladamic.com/netlearn/NetLogo501/RAndPrefAttachment.html

Quiz Q:

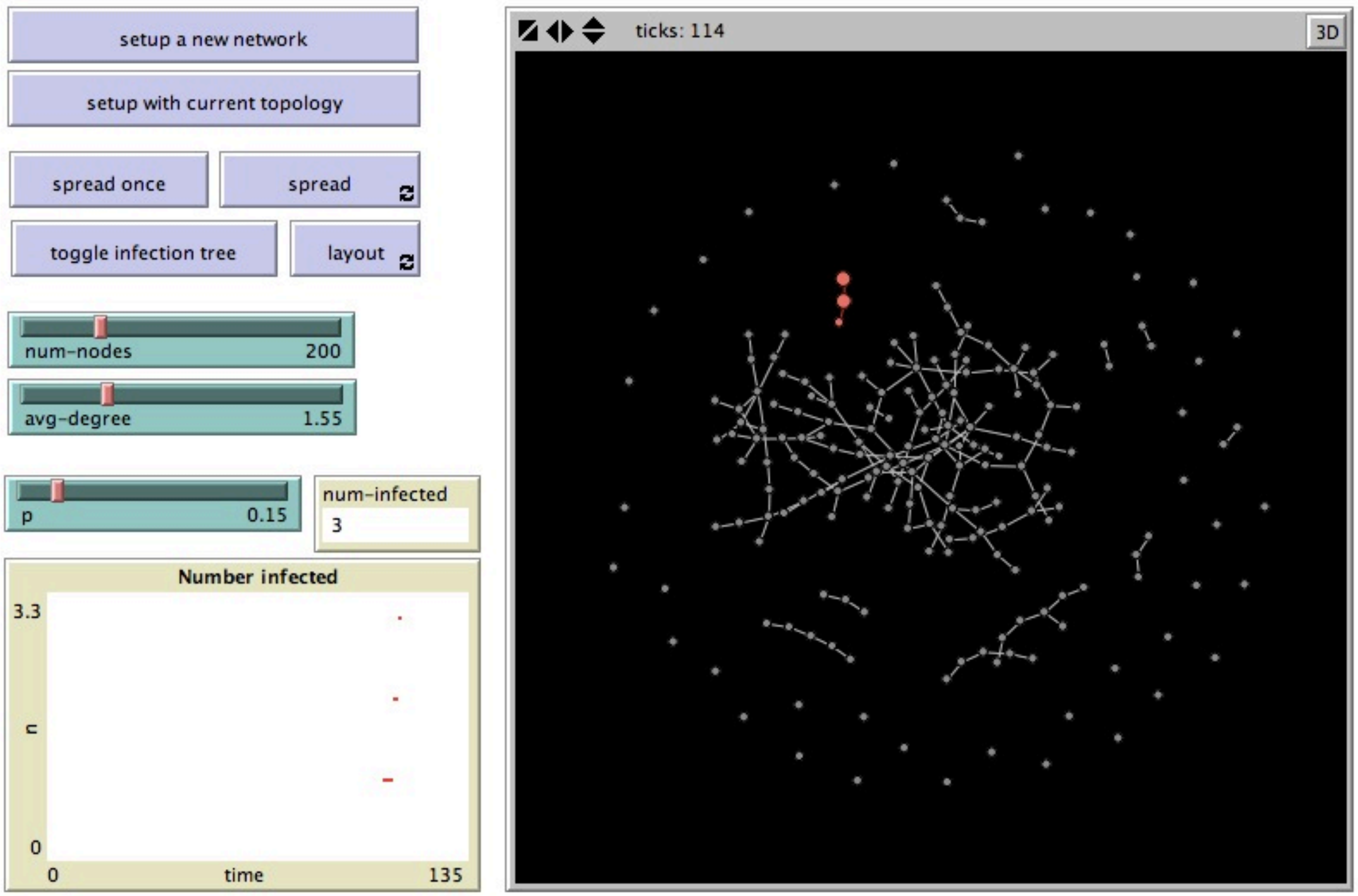
- Relative to the random growth model, the degree distribution in the preferential attachment model
  - resembles a power-law distribution less
  - resembles a power-law distribution more

Summary: growth models

- Most networks aren't 'born', they are made.
- Nodes being added over time means that older nodes can have more time to accumulate edges
- Preference for attaching to 'popular' nodes further skews the degree distribution toward a power-law

Assignment: implications for diffusion

- How does the size of the giant component influence diffusion?



Assignment: implications for diffusion

- How do growth and preferential attachment influence diffusion?