ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Network Models:
Random Networks and Small-World Networks

Session Outline

- **Introduction to (static) Random Networks**
  - Erdös-Rényi Networks
  - Number of links: Binomial/Poisson Distribution
  - Thresholds
  - Giant Components
  - Average Shortest paths
- **Growing (dynamic) Random Networks**
  - Albert Barabasi Model
  - Watts and Strogatz model
- **Small Worlds in real-life networks**
  - Small world phenomenon: Milgram's experiment

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Introduction to (static) Random Networks

Erdös-Rényi Networks
Number of links: Binomial/Poisson Distribution
Thresholds
Giant Components
Average Shortest paths

Introduction

So far we have looked at **how we measure the structure of networks** and at methods for making sense of the network data we get from our measurements.

An obvious next question to ask is, *"If I know a network has some particular property, such as a particular degree distribution, what effect will that have on the wider behavior of the system?"*

In this lecture we consider **models of the structure of networks**, models that mimic the patterns of connections in real networks in an effort to understand the implications of those patterns.

What are network models and why study them?

Informally, a network model is a **process** (radomized or deterministic) for generating a graph

Models of **static** graphs

- **input:** a set of parameters Π, and the size of the graph n
- **output:** a graph G(Π,n)

Models of **evolving** graphs

- **input:** a set of parameters Π, and an initial graph G_0
- **output:** a graph G_t for each time t

Creating models for real-life graphs is important for several reasons

- Create data for **simulations** of processes on networks
- Identify the **underlying mechanisms** that govern the network generation
- **Predict** the evolution of networks

Types of network models

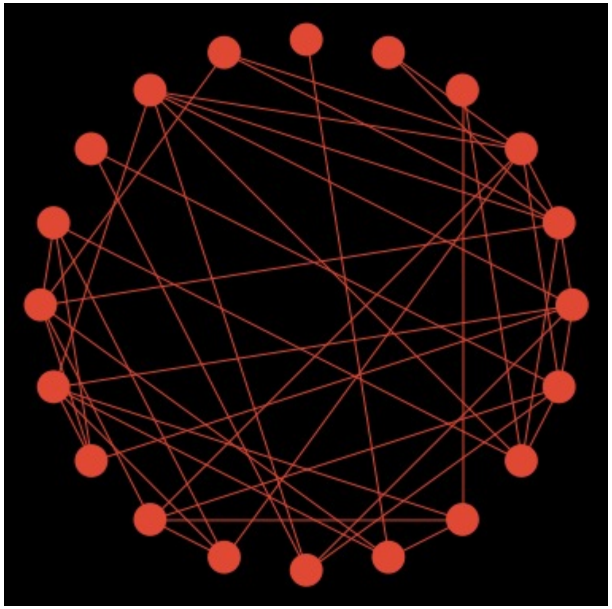
- **Random graph model** (Erdős & Rényi, 1959)
- **Scale-free model** (Barabasi & Alert, 1999)
- **Small-world model** (Watts & Strogatz, 1998)

Erdös-Rényi Networks

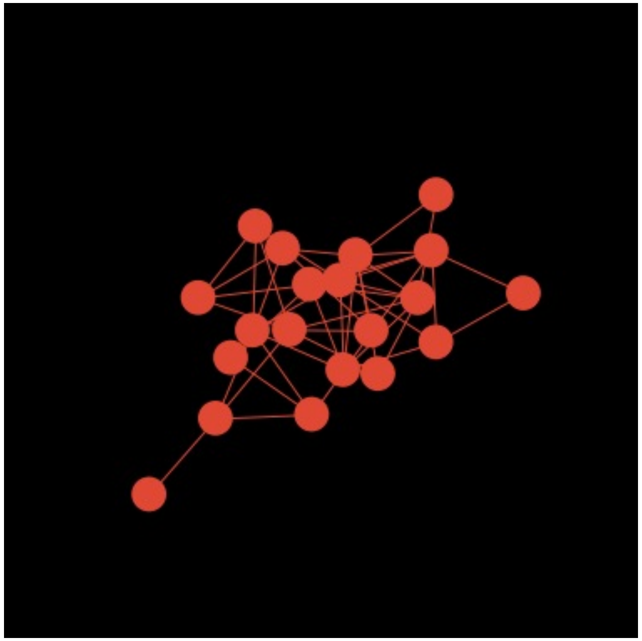
- In graph theory, the **Erdős–Rényi model** is a model for generating random graphs named after Paul Erdős and Alfréd Rényi, who first introduced it in 1959
- The notion behind this network is that we set an edge between **each pair** of nodes with ***equal probability***, independently of the other edges.

Erdös-Renyi: simplest network model

- **Assumptions**
  - nodes connect at random
  - network is undirected
- **Key parameter** (besides number of nodes N) : p or M
  - **p** = probability that any two nodes share and edge
  - **M** = total number of edges in the graph
- **What they look like:**



after spring layout



Degree distribution

- **(N,p)-model:** For each potential edge we flip a biased coin
  - with probability **p** we add the edge
  - with probability **(1-p)** we don't

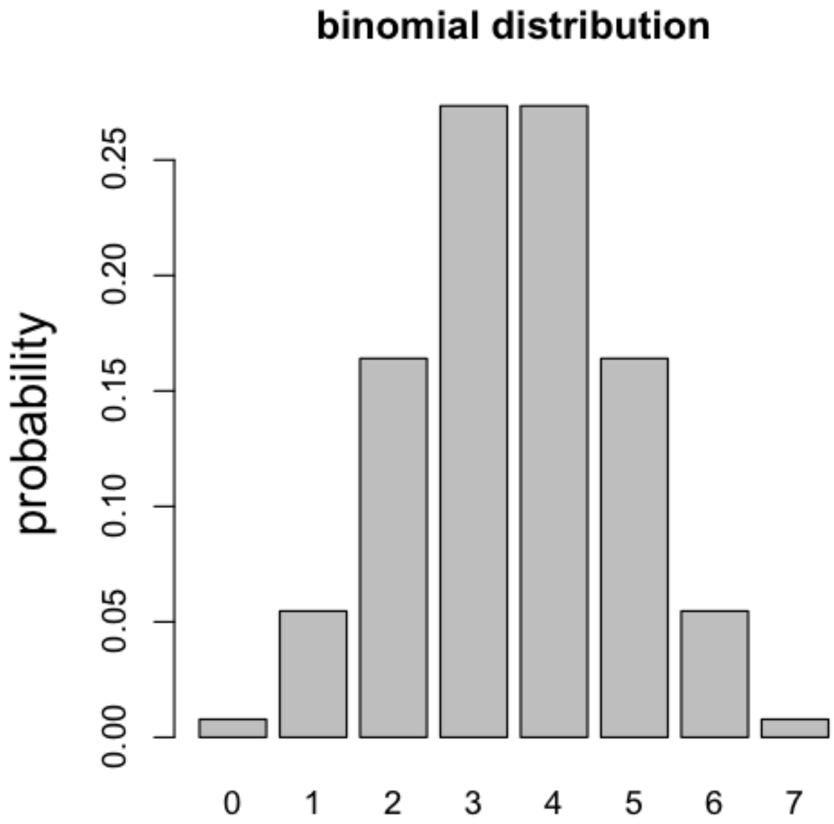
- What is the probability that a node has 0,1,2,3… edges?
- **Probabilities sum to 1**

How many edges per node?

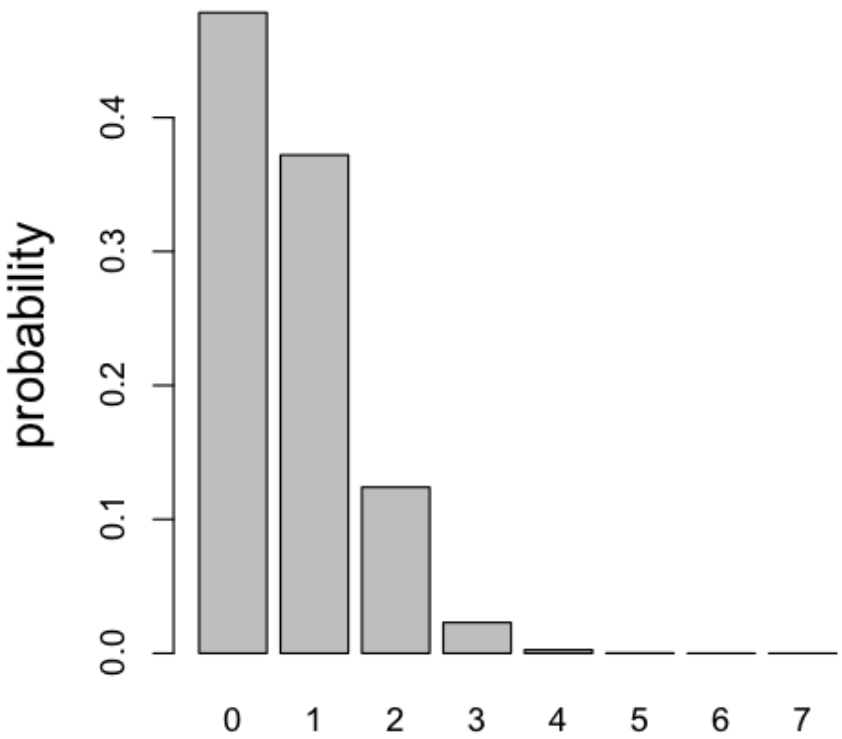
- Each node has **(N – 1)** tries to get edges
- Each try is a success with probability **p**
- The binomial distribution gives us the probability that a node has degree **k:**

$$B(N-1;k;p) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

if p = 0.5


binomial distribution
probability
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7

if p = 0.1



Approximations

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$p_k = \frac{z^k e^{-z}}{k!}$$

$$p_k = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-z)^2}{2\sigma^2}}$$

Binomial

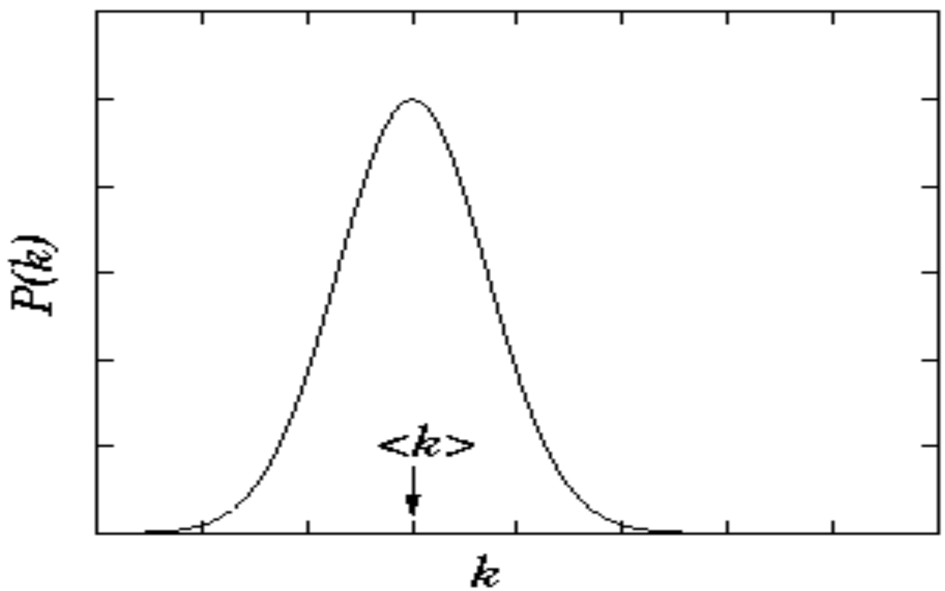
limit *p* small

Poisson

limit large *n*

Normal

Poisson distribution

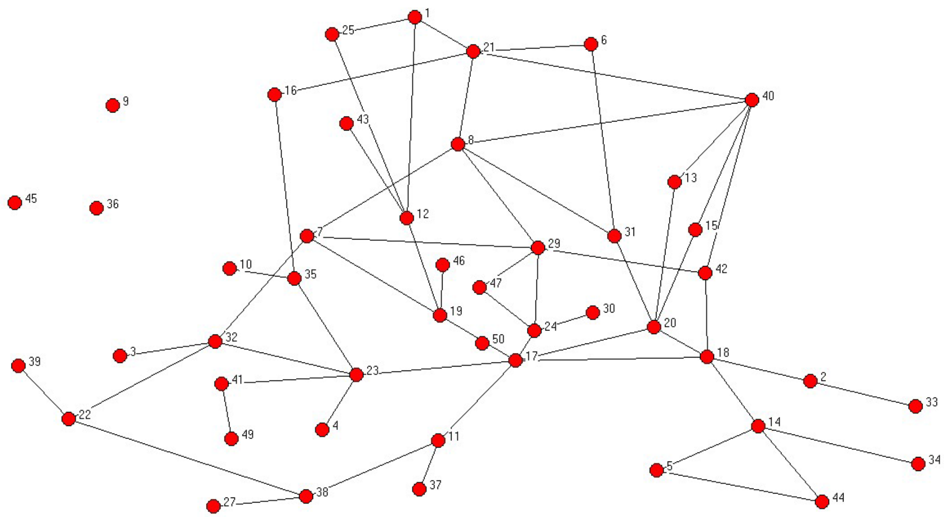


What insights does this yield?
No hubs!

You don't expect large hubs in a random network

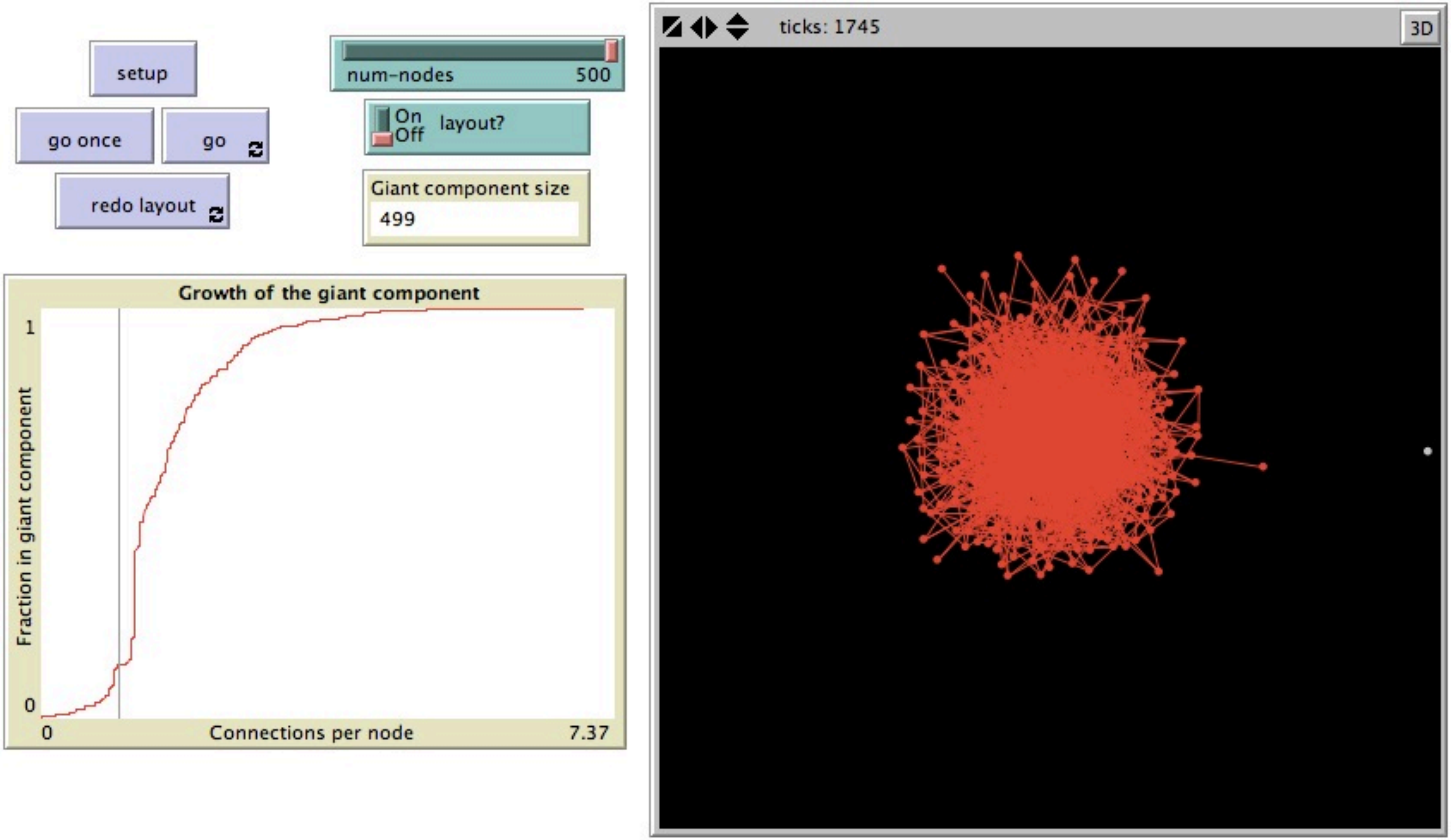
Properties and Thresholds of the Poisson Random Network

- At the threshold of **1/n** we see cycles emerge, and we also see the emergence of a **“giant component”,** which is a unique largest component which contains a nontrivial fraction of all nodes.



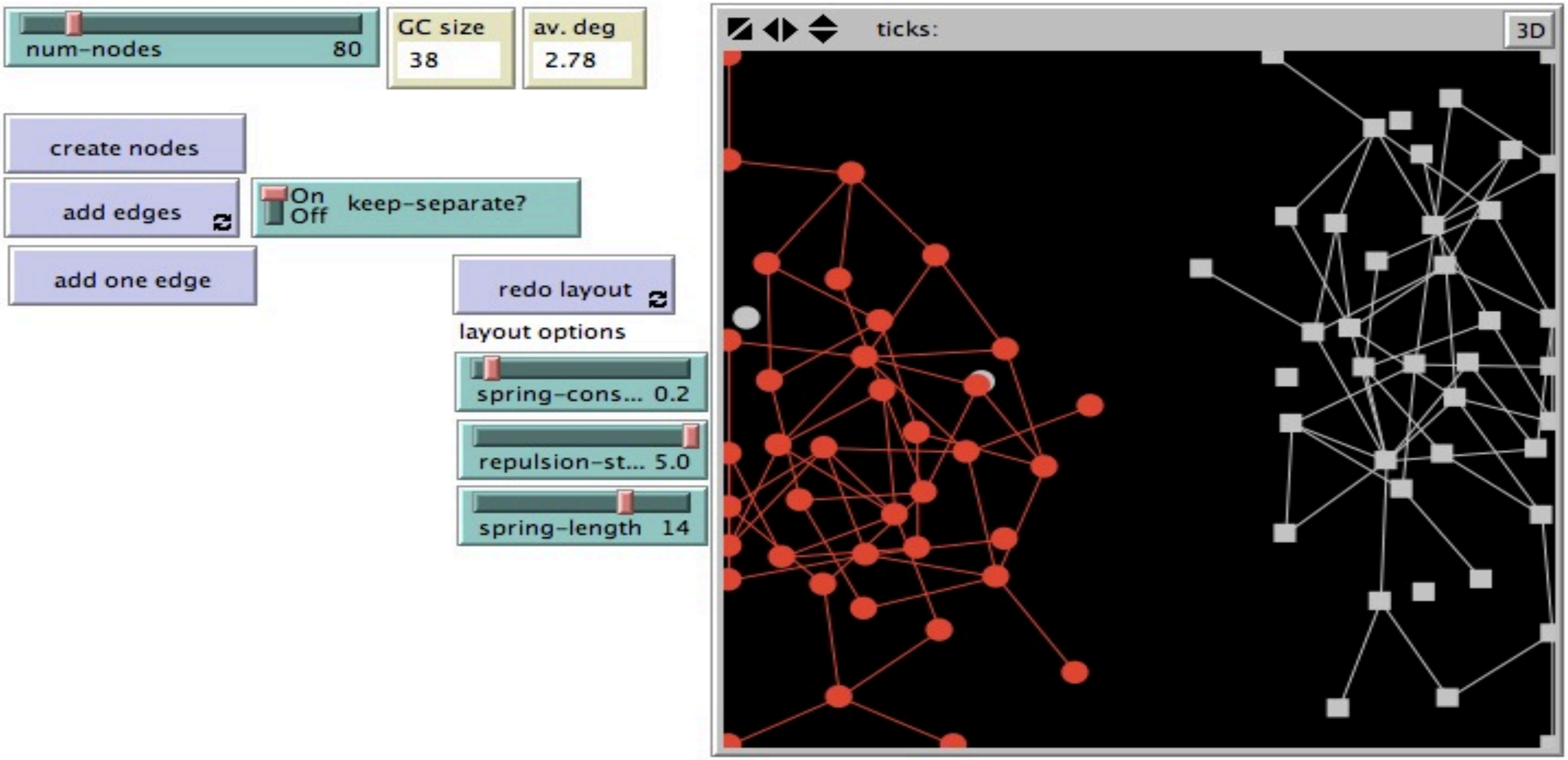
Poisson p=.05, 50 nodes

Emergence of the giant component



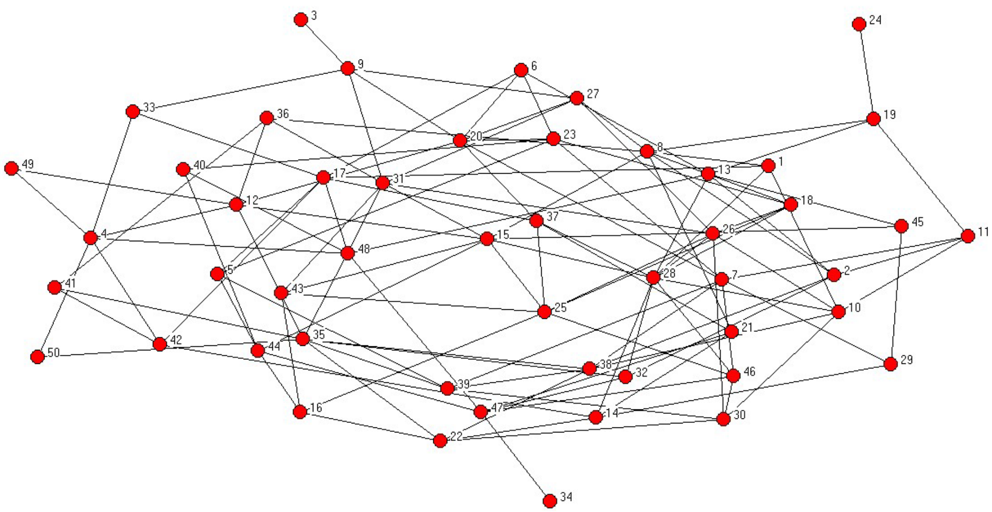
Why just one giant component?

- What if you had 2, how long could they be sustained as the network densifies?



Properties and Thresholds of the Poisson Random Network

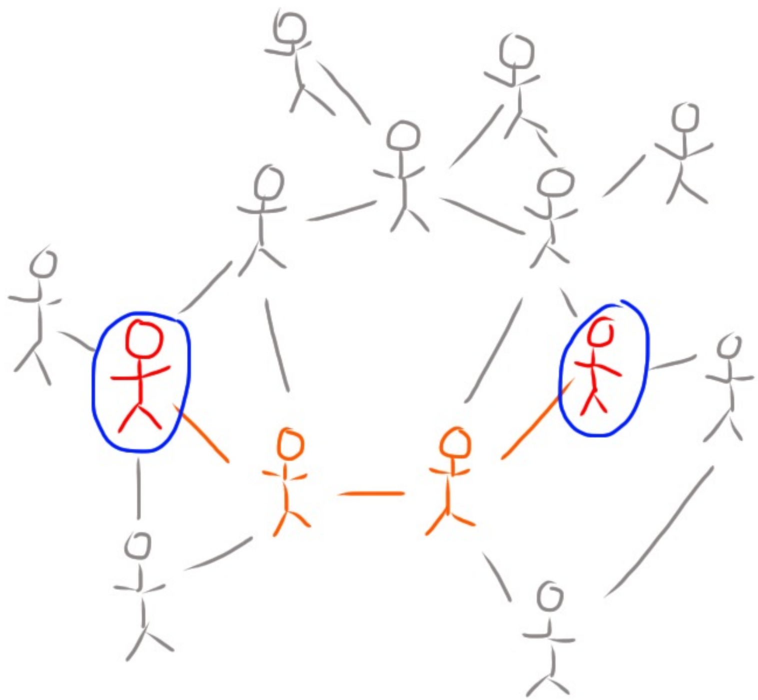
- The giant component grows in size until the threshold of **log(n)/n** , where the network becomes connected.



Poisson p=.10, 50 nodes

Average shortest path

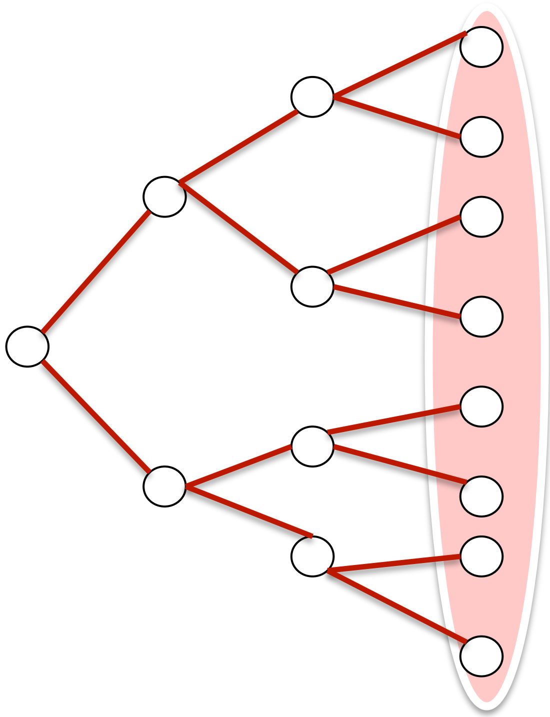
Definition: How many hops on average between each pair of nodes?



Friends at distance m

- Each of your friends has **z = avg. degree** friends besides you
- Ignoring loops, the number of people you have at distance m is $\mathbf{z^m}$

$$N_m = z^m$$

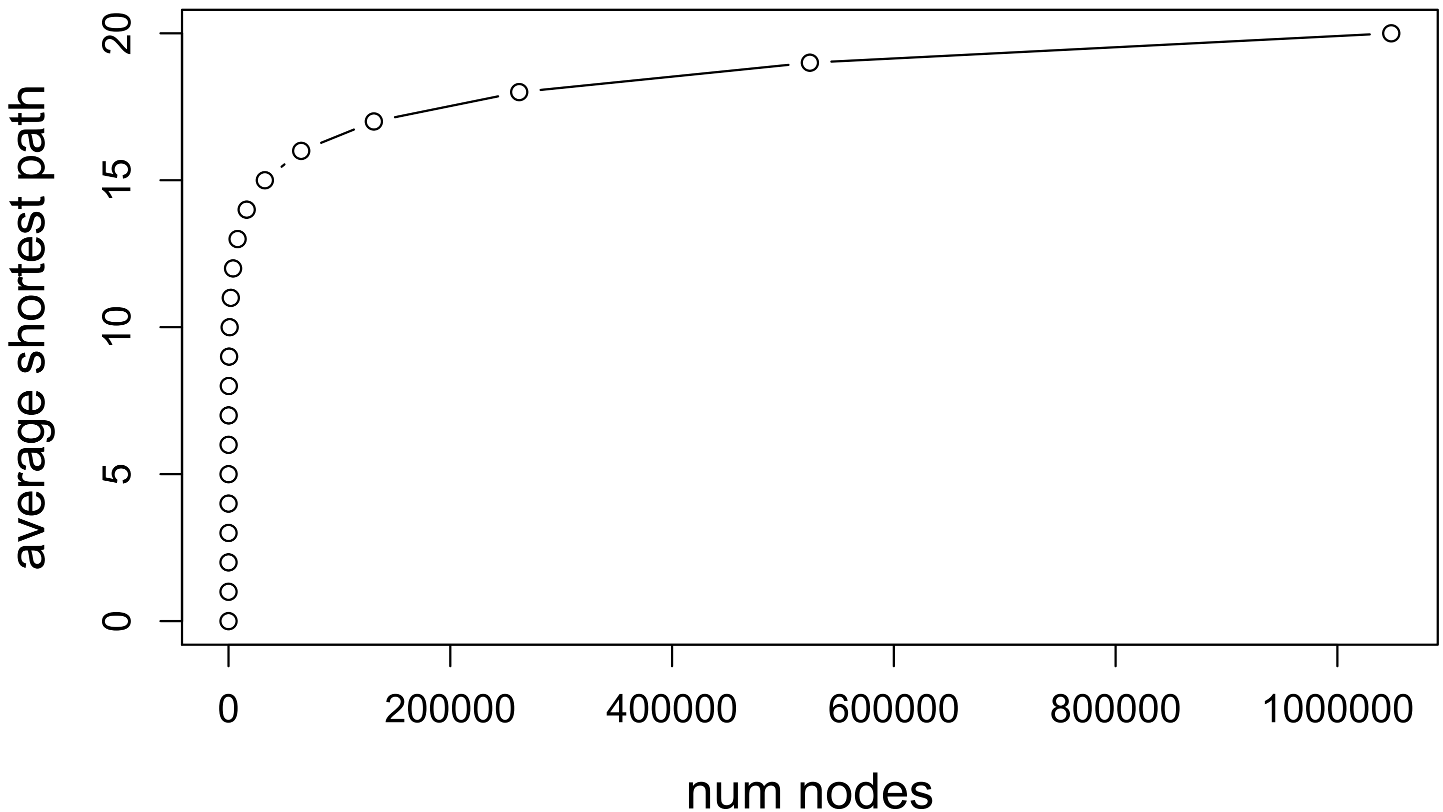


Average shortest path m_{av}

$$m_{av} \sim \frac{\log N}{\log z}$$

What this means in practice

- Erdös-Renyi networks can grow to be very large but nodes will be just a few hops apart



Departing from the ER model

We need models that better capture the characteristics of real graphs

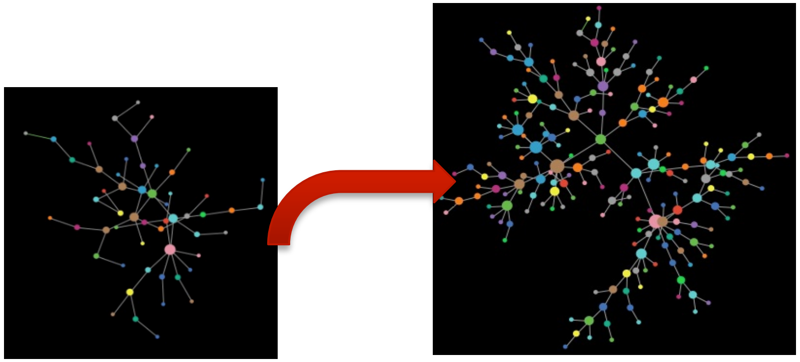
- degree sequences
- clustering coefficient
- short paths

Power Law distribution – Preferential Attachment - Scale-free model

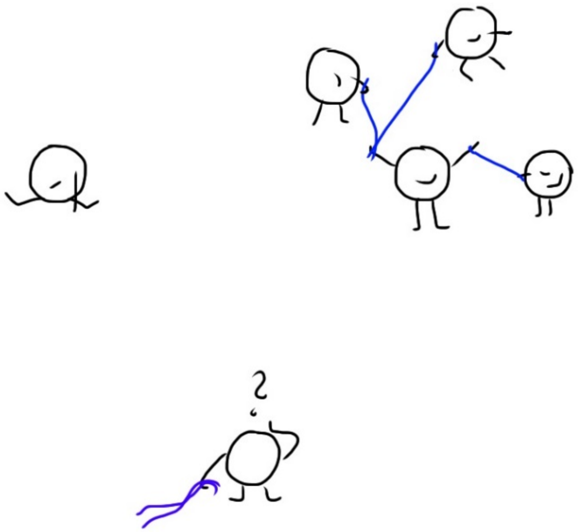
(Barabasi & Alert, 1999)

Two ingredients in generating power-law networks

1. nodes appear over time (growth)



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

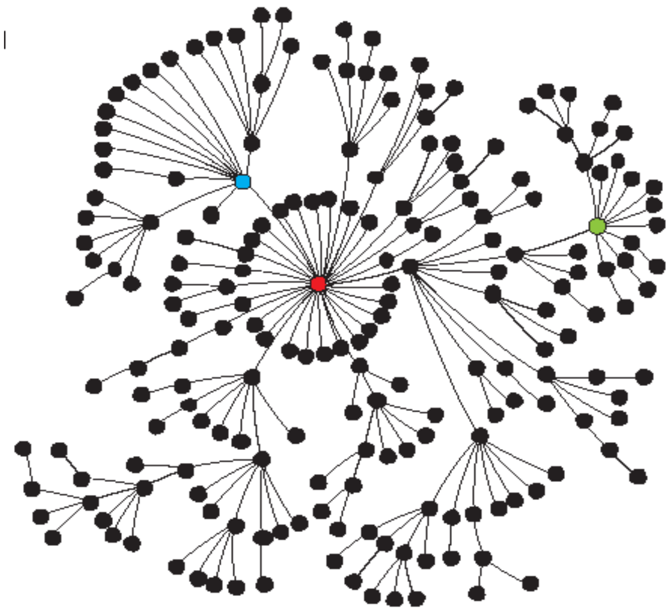


Scale-Free (SF) Networks: Barabási–Albert (BA) Model

- **"Scale free" means there is no single characterizing degree in the network**

- **Growth**:
  - starting with a small number (n_0) of nodes, at every time step, we add a new node with $m(<=n_0)$ links that connect the new node to m different nodes already present in the system

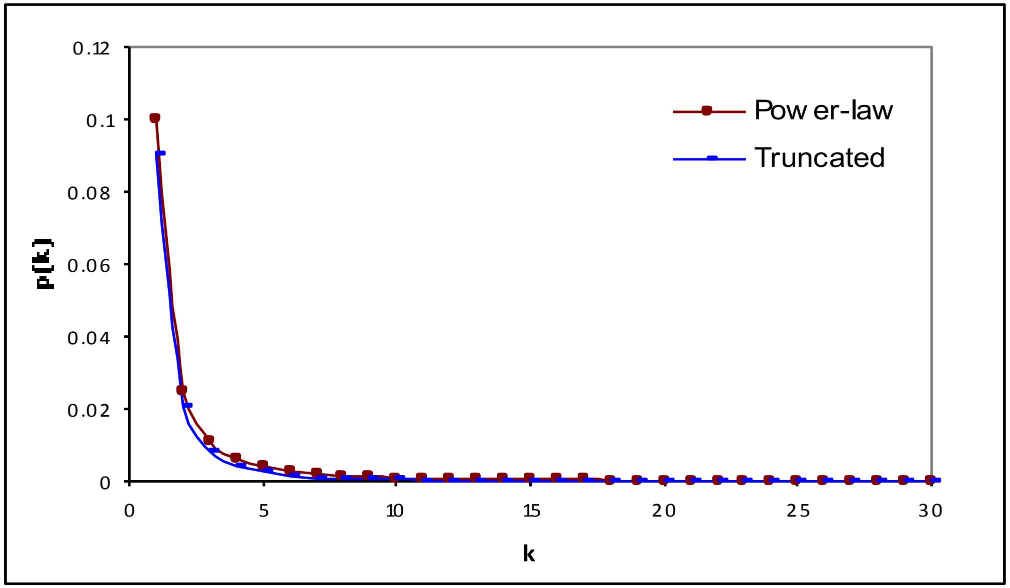
- **Preferential attachment**:
  - When choosing the nodes to which the new node will be connected to node i depends on its degree k_i



Scale-Free (SF) Networks: Barabási–Albert (BA) Model

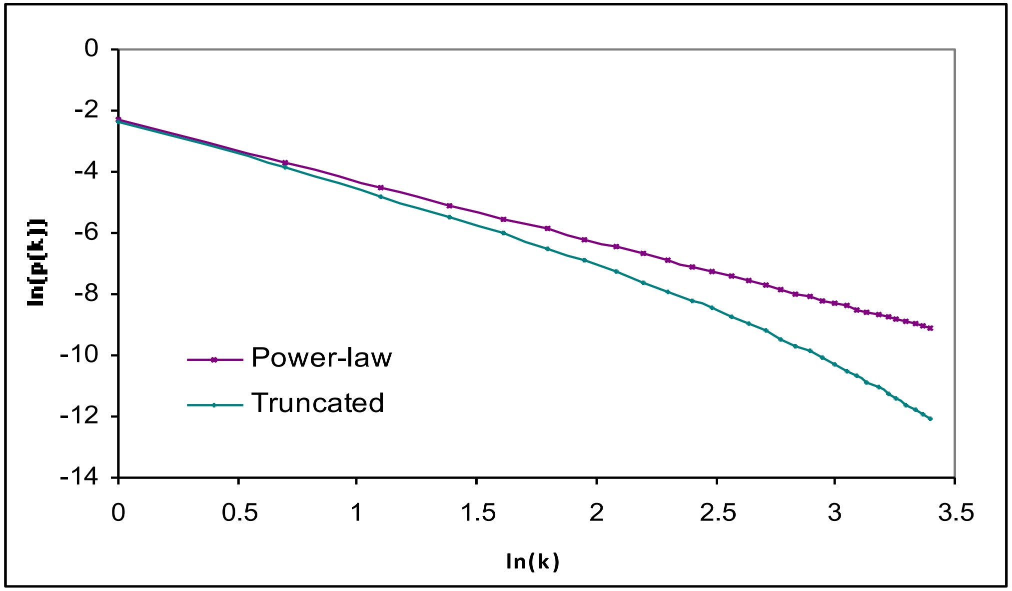
- The degree of scale-free networks follows **power-law distribution** with a flat tail for large *k*

$$p(k) \sim k^{-\gamma}$$



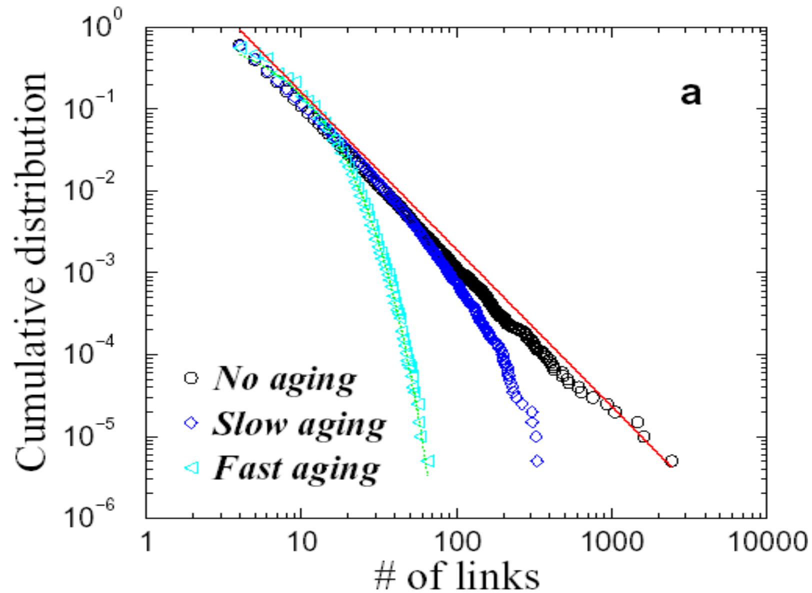
- Truncated power-law distribution deviates at the tail

$$p(k) \sim k^{-\gamma} e^{-\frac{k}{\kappa}}$$

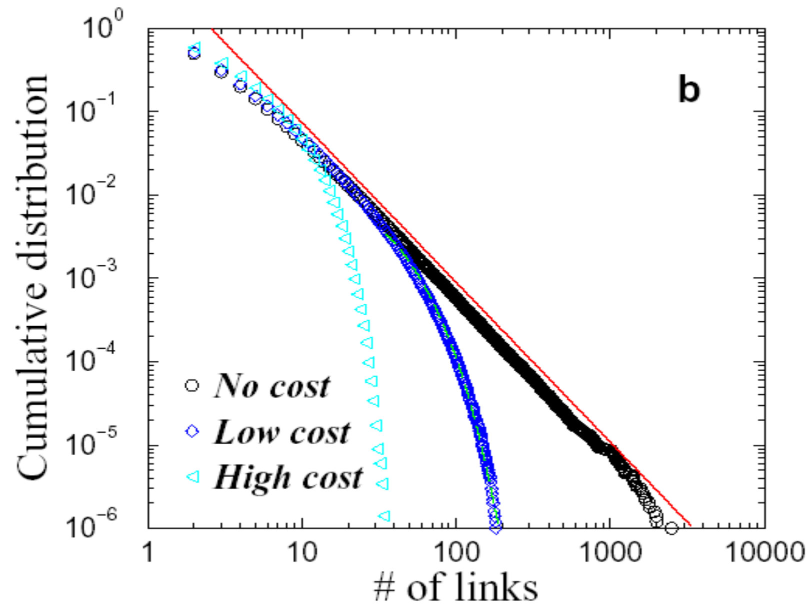


Scale-Free (SF) Networks: Barabási–Albert (BA) Model

- The emergence of scale-free network is due to
  - **Growth effect:** new nodes are added to the network
  - **Preferential attachment effect** (Rich-get-richer effect): new nodes prefer to attach to "popular" nodes



- The emergence of truncated SF network is caused by some constraints on the maximum number of links a node can have such as (Amaral, Scala et al. 2000)
  - **Aging effect:** some old nodes may stop receiving links over time
  - **Cost effect:** as maintaining links induces costs, nodes cannot receive an unlimited number of links



Variations of the BA model

- Many variations have been considered some in order to address the problems with the vanilla BA model
  - edge rewiring, appearance and disappearance
  - fitness parameters
  - variable mean degree
  - non-linear preferential attachment
    - surprisingly, only linear preferential attachment yields power-law graphs

Weaknesses of the BA model

Technical issues:

- It is not directed (not good as a model for the Web) and when directed it gives acyclic graphs
- It focuses mainly on the (in-) degree and does not take into account other parameters (out-degree distribution, components, clustering coefficient)
- It correlates age with degree which is not always the case

Academic issues

- the model rediscovers the wheel
- preferential attachment is not the answer to every power-law
- what does "scale-free" mean exactly?

Yet, it was a breakthrough in the network research, that popularized the area

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Small World Phenomena

Watts & Strogatz, 1998

Introduction

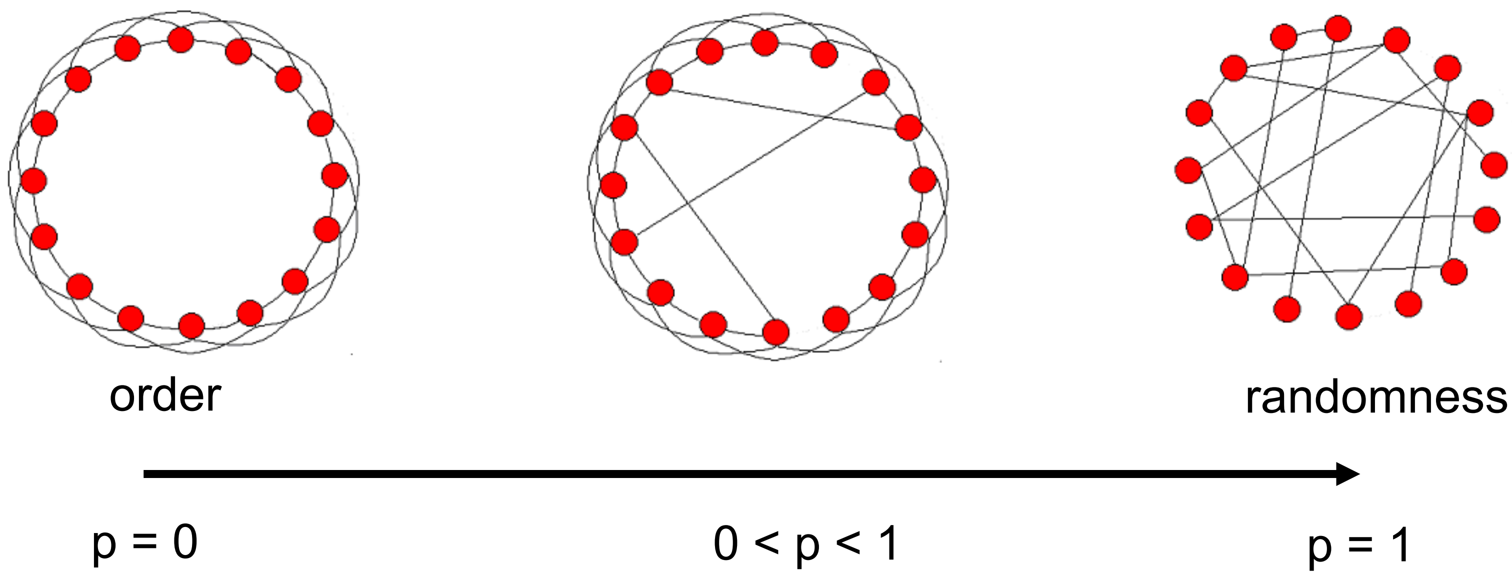
So far we focused on obtaining graphs with power-law distributions on the degrees. What about other properties?

- **Clustering coefficient:** real-life networks tend to have high clustering coefficient
- **Short paths:** real-life networks are *"small worlds"*
  - this property is easy to generate
- Can we combine these two properties?

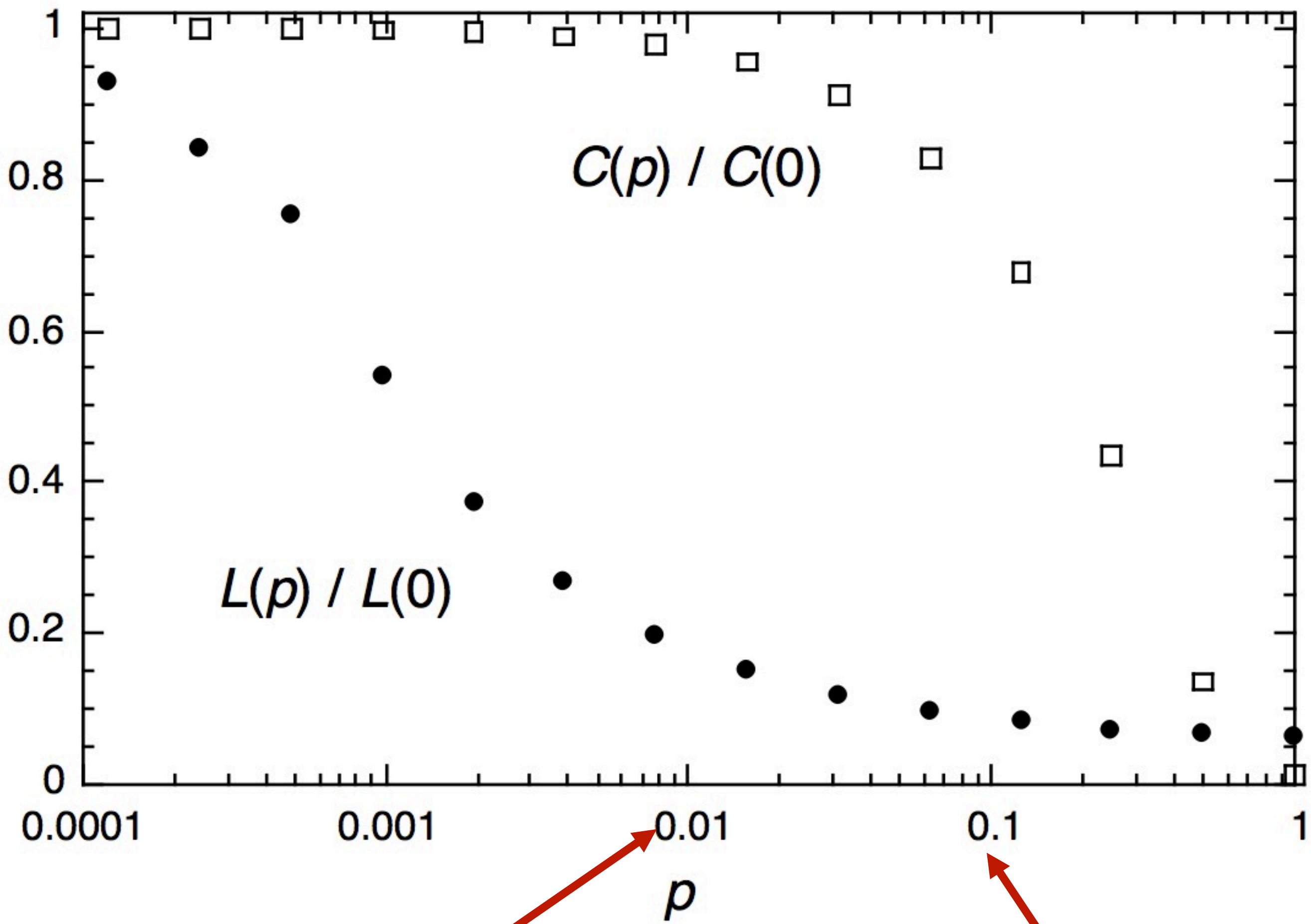
In 1998, Duncan Watts and Steve Strogatz argued that such a model follows naturally from a combination of homophily and weak ties. Homophily creates many triangles, while the weak ties still produce the kind of widely branching structure that reaches many nodes in a few steps.

Watts and Strogatz model

- Start with a ring, where every node is connected to the next k nodes
- With probability p, rewire every edge (or, add a shortcut) to a uniformly chosen destination.
  - Granovetter, “The strength of weak ties”



Clustering coefficient and ASP as rewiring increases



1% of links rewired

10% of links rewired

Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Comparison with "random graph" used to determine whether real-world network is "small world"

Network	size	av. shortest path	Shortest path in fitted random graph	Clustering (averaged over vertices)	Clustering in random graph
Film actors	225,226	3.65	2.99	0.79	0.00027
MEDLINE co-authorship	1,520,251	4.6	4.91	0.56	1.8×10^{-4}
E.Coli substrate graph	282	2.9	3.04	0.32	0.026
C.Elegans	282	2.65	2.25	0.28	0.05

Now let's look at some real-world examples: distributions differ from theoretical expectations!

Examples of Growing Random Networks

Real World Examples

- Citation networks
- Web
- Scientific networks
- Societies...

Growing and Uniformly Random

- Each date a new node is born
- Forms **m** links to existing nodes
- Each node is chosen with equal likelihood

Example of Expected Degree Distribution

A dynamic variation on the Poisson random network model

- Start with **m nodes** fully connected
- New node forms **m links** to *existing nodes*
- An existing node has a probability **m/t** of getting new link each period
- **No longer binomial**, as probabilities vary with time

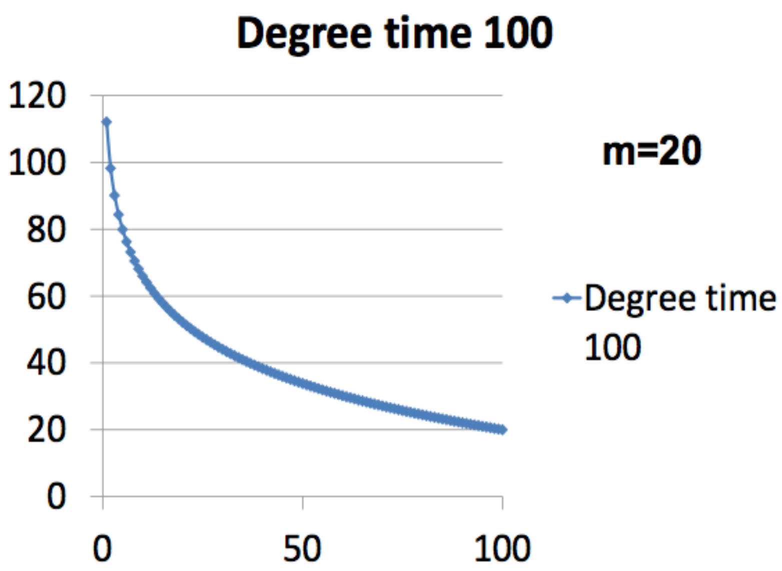
Example of Expected Degree Distribution

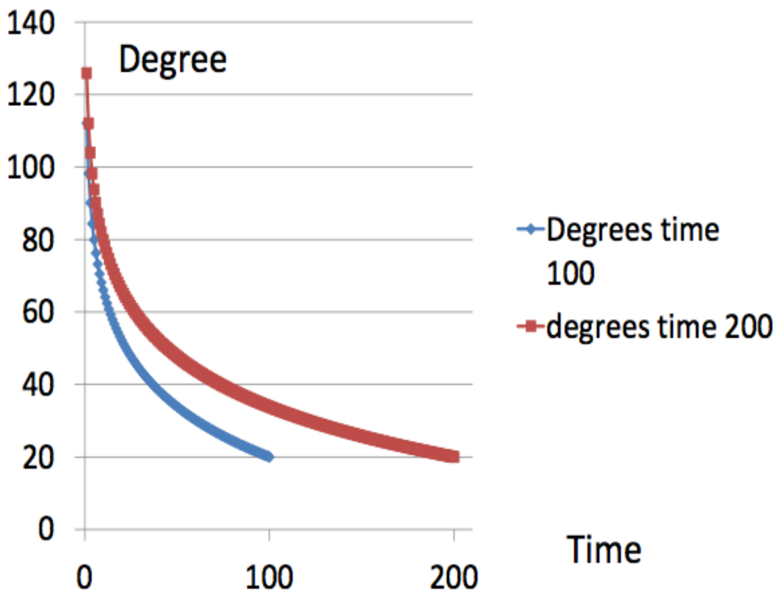
Expected degree for **node i** born at $m<i<t$ is:

$$m + m/(i+1) + m/(i+2) + \ldots + m/t$$

Approx = $m(1+\log(t/i))$ (harmonic numbers)

Example of Expected Degree Distribution

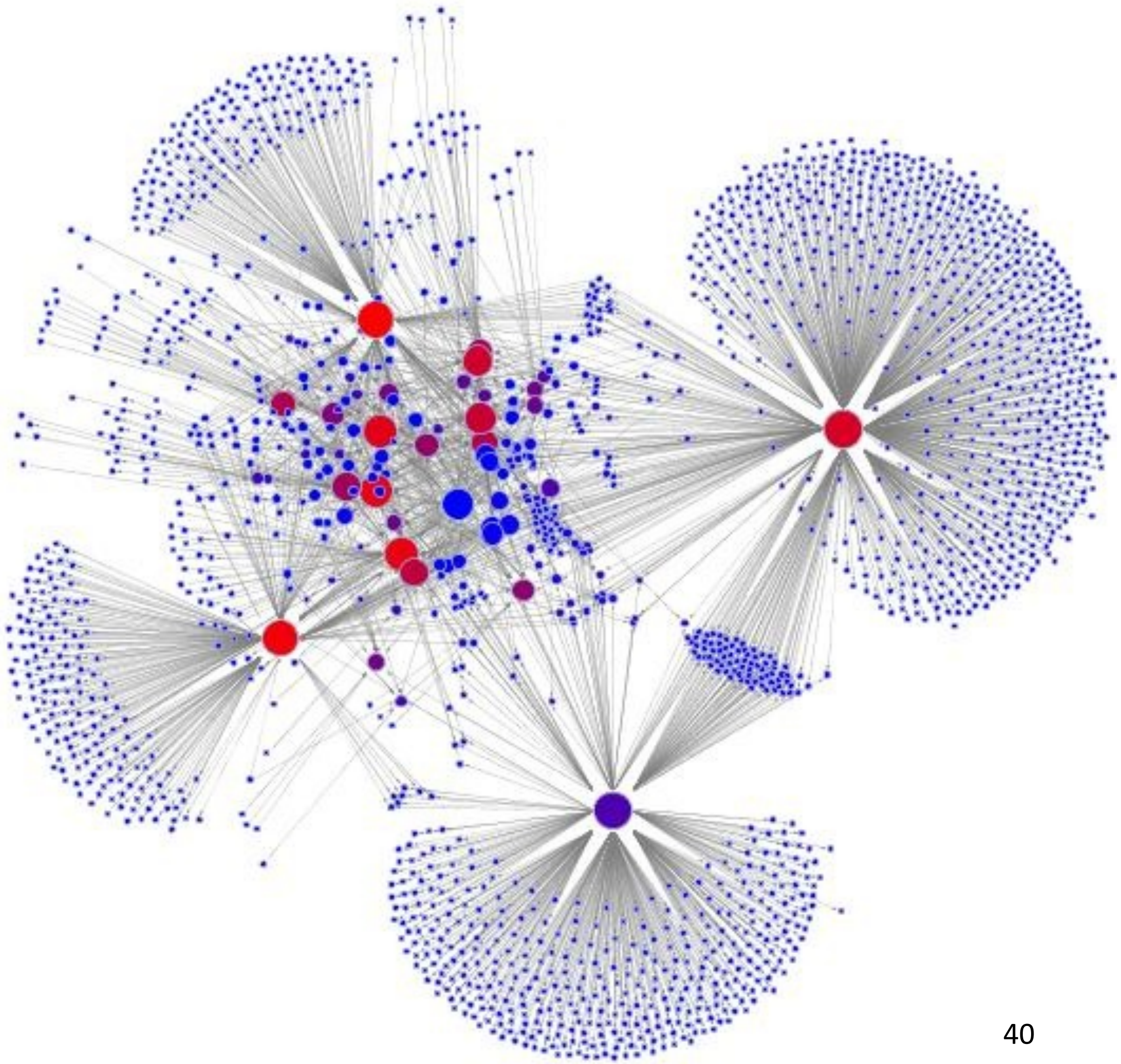




Online Question & Answer Forums

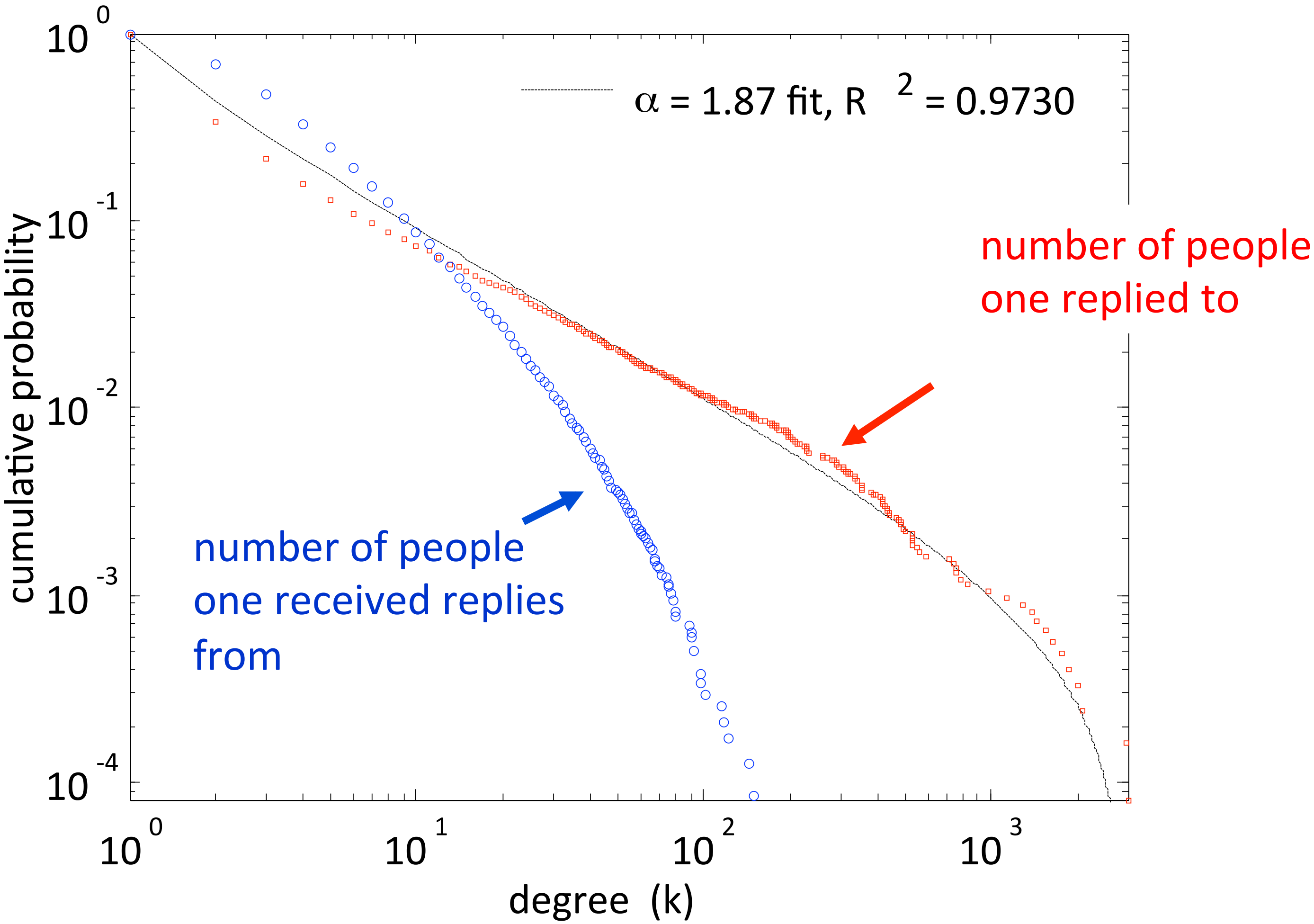
How do you identify the experts ?

- Replier
- Asker



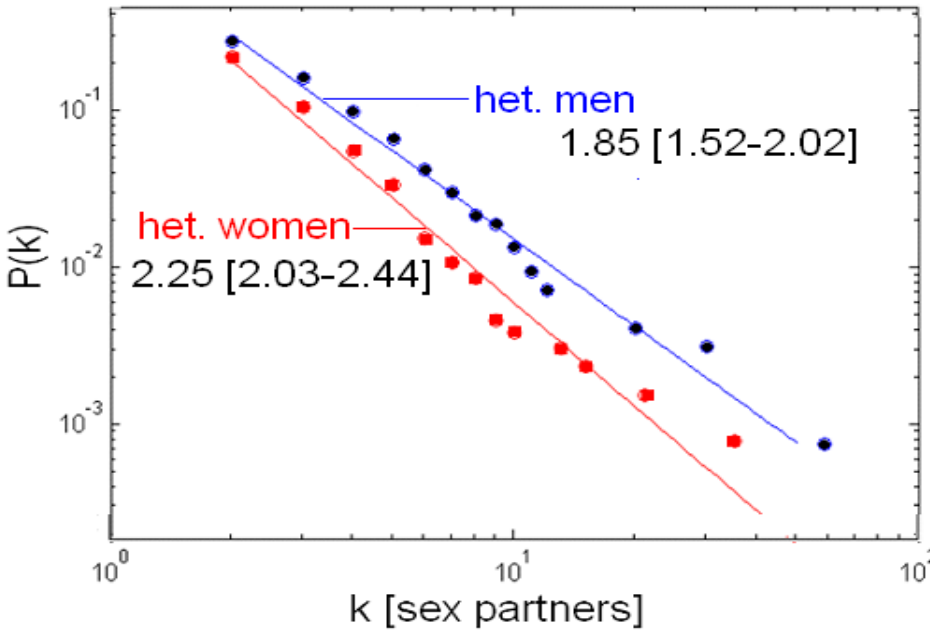
Online Question & Answer Forums

- **'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

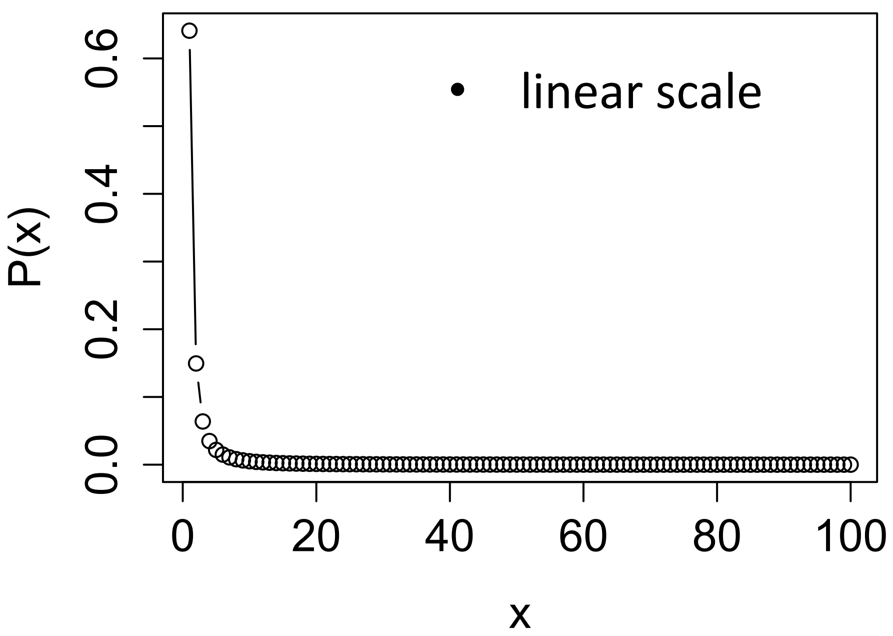


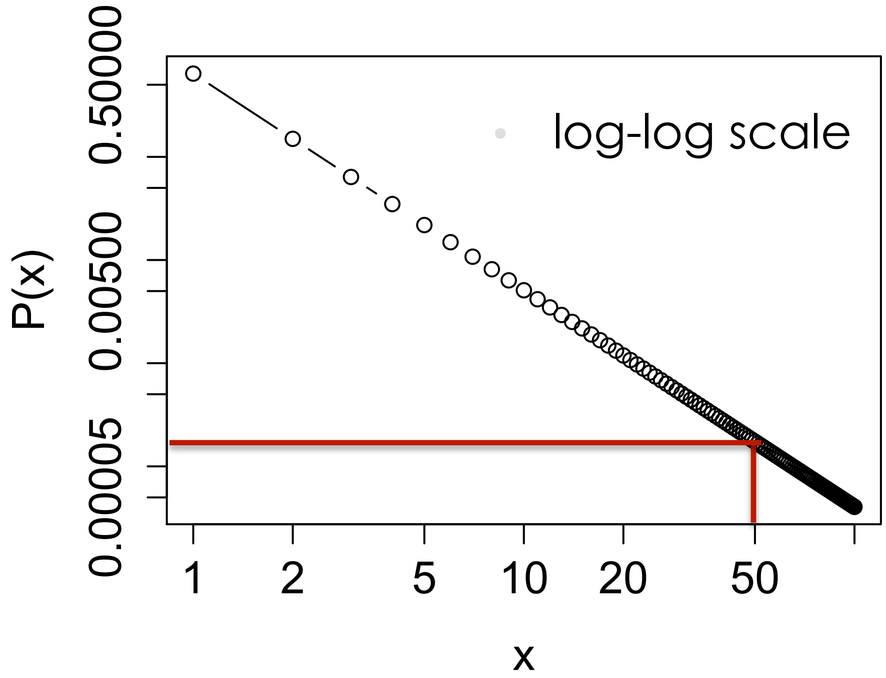
Real-world degree distributions

- Sexual networks
- Great variation in contact numbers



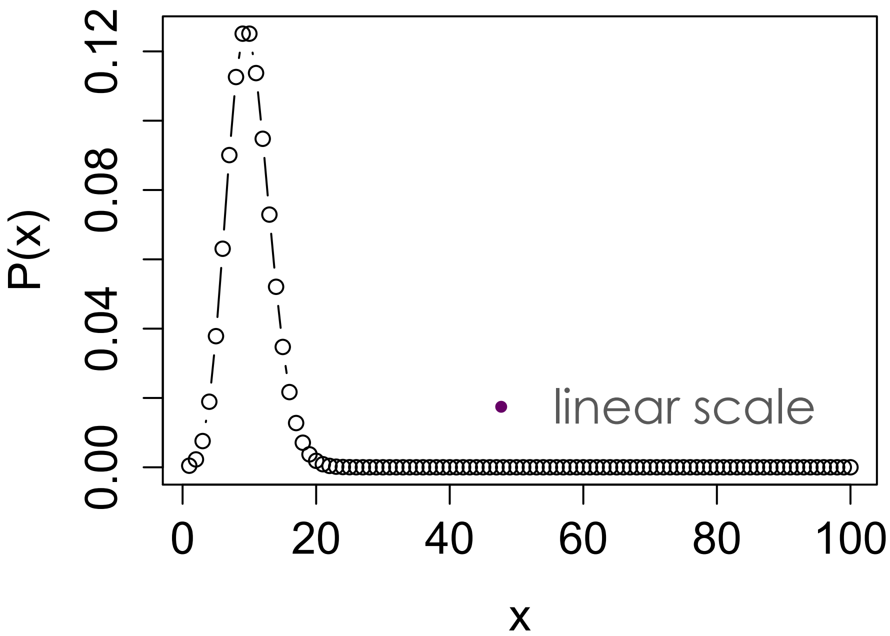
Power-law distribution

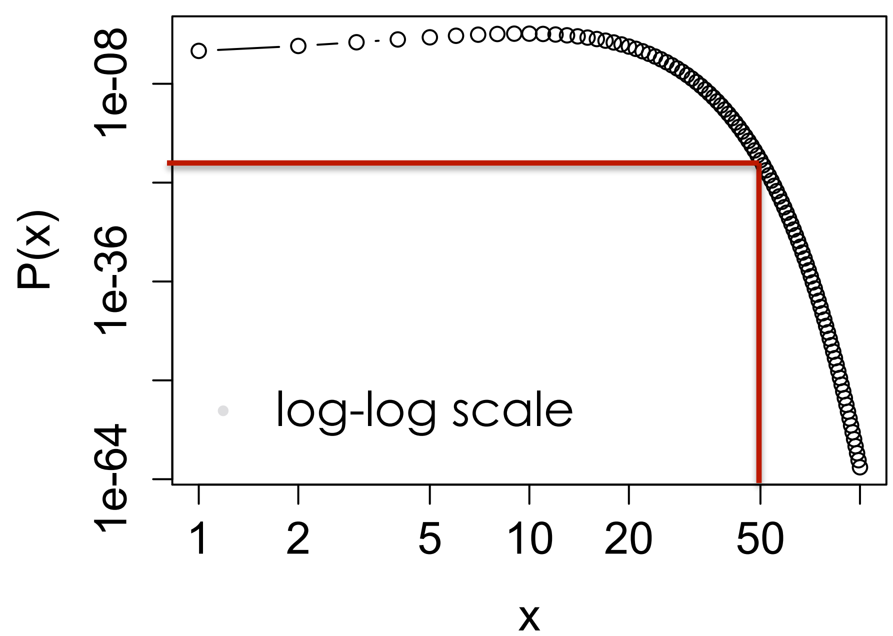




- **high skew (asymmetry)**
- straight line on a log-log plot

Poisson distribution





- **little skew**
- curved on a log-log plot

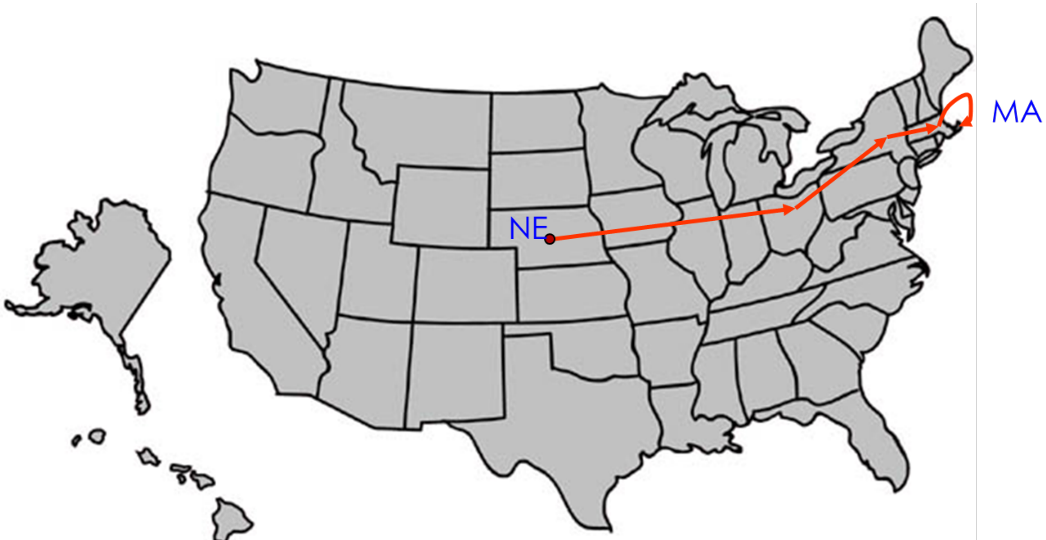
Milgram's experiment

Instructions:

- Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is "closest" to the target.

Outcome:

- 20% of initiated chains reached target
- average chain length = 6.5



"Six degrees of separation"

Milgram’s experiment repeated

E-mail experiment
Dodds, Muhamad, Watts, Science 301 (2003)

- 18 targets
- 13 different countries

- 60,000+ participants
- 24,163 message chains
- 384 reached their targets
- average path length 4.0

Source: NASA, U.S. Government; http://visibleearth.nasa.gov/view_rec.php?id=2429

Interpreting Milgram's experiment

- **Is 6 is a *surprising* number?**
  - In the 1960s? Today? Why?
- **If social networks were random… ?**
  - Pool and Kochen (1978) - ~500-1500 acquaintances/person
  - ~ 500 choices 1st link
  - ~ 500^2 = 250,000 potential 2nd degree neighbors
  - ~ 500^3 = 125,000,000 potential 3rd degree neighbors
- **If networks are completely cliquish?**
  - all my friends' friends are my friends
  - what would happen?
- **Is 6 an accurate number?**
  - What bias is introduced by uncompleted chains?
  - Are longer or shorter chains more likely to be completed?

Three and a half degrees of separation

Each person in the world (at least among the 1.59 billion people active on Facebook) is connected to every other person by an average of three and a half other people.

Some Facebook employees

Mark Zuckerberg
3.17 degrees of separation

Sheryl Sandberg
2.92 degrees of separation

The majority of the people on Facebook have averages between 2.9 and 4.2 degrees of separation. Figure 1 (below) shows the distribution of averages for each person.

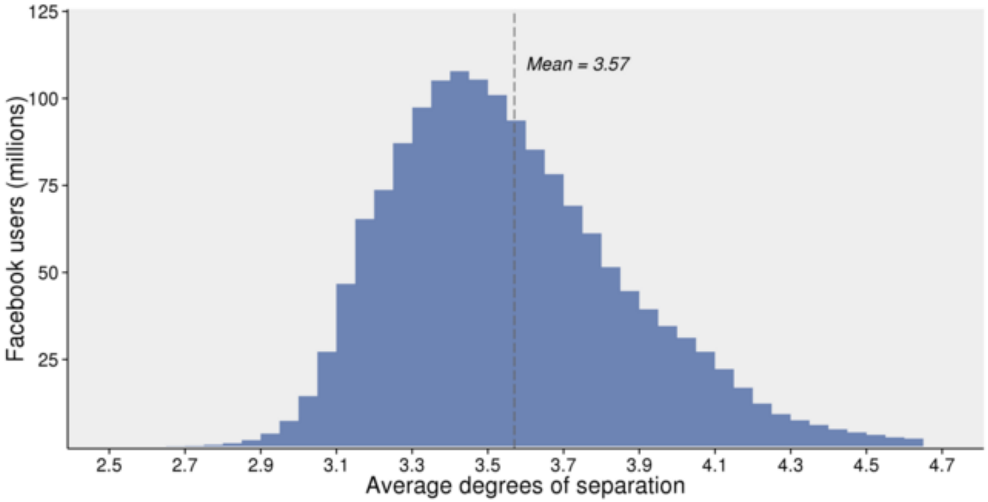
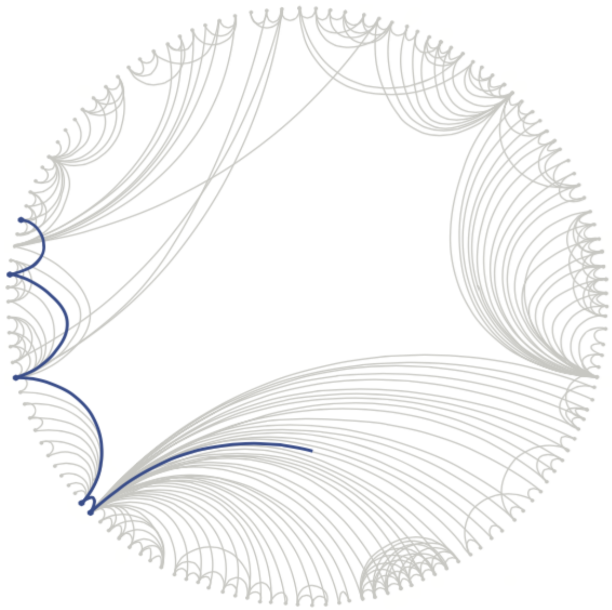


Figure 1. Estimated average degrees of separation between all people on Facebook. The average person is connected to every other person by an average of 3.57 steps. The majority of people have an average between 3 and 4 steps.

Case Study: Facebook Three and a half degrees of separation

Research at Facebook



https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

Key takeaways

- **Radom graphs can help us identify systematic phenomena in real-life networks**

- **Growing networks tend to create hyper-connected nodes**
  - Power-law distributions
  - The phenomenon of Preferential attachment

- **Real-life networks manage to combine large clustering with short distances (small worlds)**