Static Random Graphs

EE599: Social Network Systems

Keith M. Chugg

Fall 2014

USC University of Southern California

Overview

- Probability Review
- Static Random Models
  - Poisson random graphs (completely random)
  - The configuration model (specify the degree distribution)
    - *Power-law degree distribution*
  - The small world model (Watts-Strogatz)
  - Exponential and Markov models
- Growth models for random graphs

Start with purely random models and see what occurs.

Modify model to better represent real networks

Primary References

- **Static models**
  - Newman, The Structure and Function of Complex Networks, SIAM REVIEW, Vol.45, No. 2, pp.167–256, 2003. (terse and mathematical).
    - Sections IV, V, VI
  - Jackson, Chapter 4. Similar to Newman.
  - Barabasi, Chapters 3 & 4. Less mathematical, buggy.
- **Models for growth of random networks**
  - Newman, The Structure and Function of Complex Networks, SIAM REVIEW, Vol.45, No. 2, pp.167–256, 2003.
    - Sections VII
  - Jackson, Chapter 5.
  - Barabasi, Chapters 5 & 6.

Why Random Network Models?

- Serve as a benchmark
  - If random graph doesn't match real network data, then it points to systematic proerties in real data (e.g., triadic closure)
  - Even "bad" models can lend insight to some real-network properties
  - This can lead to better models
- Given an accurate model
  - Predict information diffusion, epidemic properties
  - Allow us to run "what if" scenarios by changing conditions that were present during the collection of real data
    - e.g., What if friend recommendation algorithm was different? How would a different corporate structure change our profitability?

Probability Review Items

- Some important random variables
  - Bernoulli, Binomial, Poisson, Gaussian
- Bayes Law & Theorem of Total Probability
- Moments and (Moment) Generating Functions
- Linear MMSE estimation
- Statistics
  - Law of Large Numbers
  - Central Limit Theorem
  - Confidence Intervals
  - Linear Regression
- Markov Chains

Reference:

A. Leon-Garcia, Probability, Statistics, and Random Processes for Electrical Engineer- ing, 3rd Edition, Addison Wesley, 2012.

Purely Random Network

- Poisson Random Networks
  - AKA: "Random Networks", Binomial Random Nets, Edos-Renyi Networks
  - Fixed:
    - N = number of nodes
    - p = probability that one of L_max possible edges is present
    - (Degree distribution)
  - Random (varying)
    - L = number of links in graph
    - K_i = degree of node i

Purely Random Network

- Other methods to generate "purely random" networks
  - Fixed and deterministic:
    - N = number of nodes
    - L = number of links in graph
  - Random (varying)
    - K_i = degree of node i
- This and model on previous slide yield largely the same results
  - Model from previous slide is typically adopted as statistical conclusions follow more directly

Poisson Random Network

N nodes

$$L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2} \text{ potential edges}$$

$$p = P(\text{"possible edge (successfully) connected"})$$

$$q = 1 - p = P(\text{"possible edge not (successfully) connected"})$$

All edges connected (or not) independently (Bernoulli trials)

$$K \sim \text{Binomial}(N-1, p) \approx \text{Poisson}(\alpha = (N-1)p)$$

Degree Distribution

$$L \sim \text{Binomial}(L_{\max}, p) \approx \text{Gaussian}(m = np, \sigma^2 = npq)$$

Poisson Random Network

N=50

p=1/50=0.02

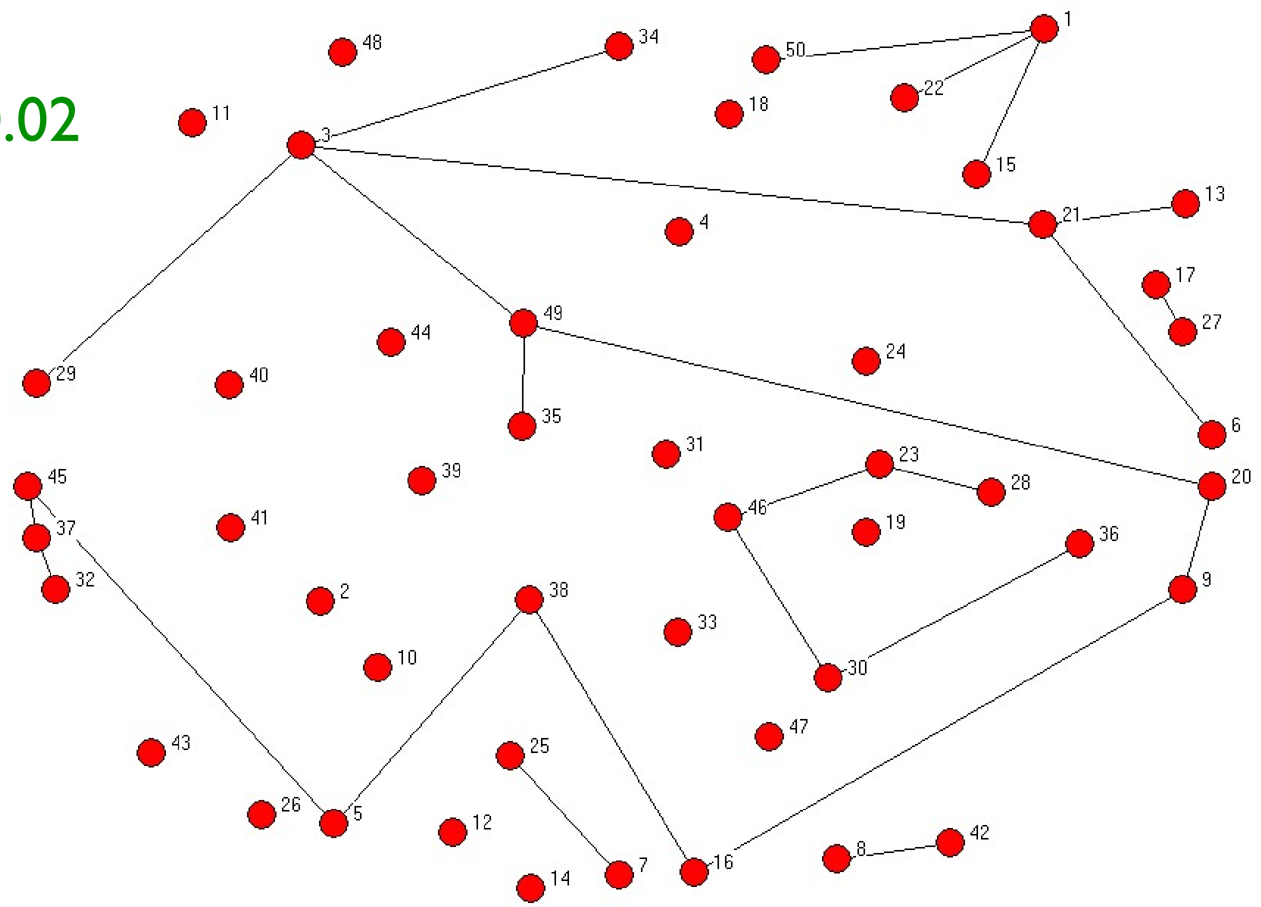


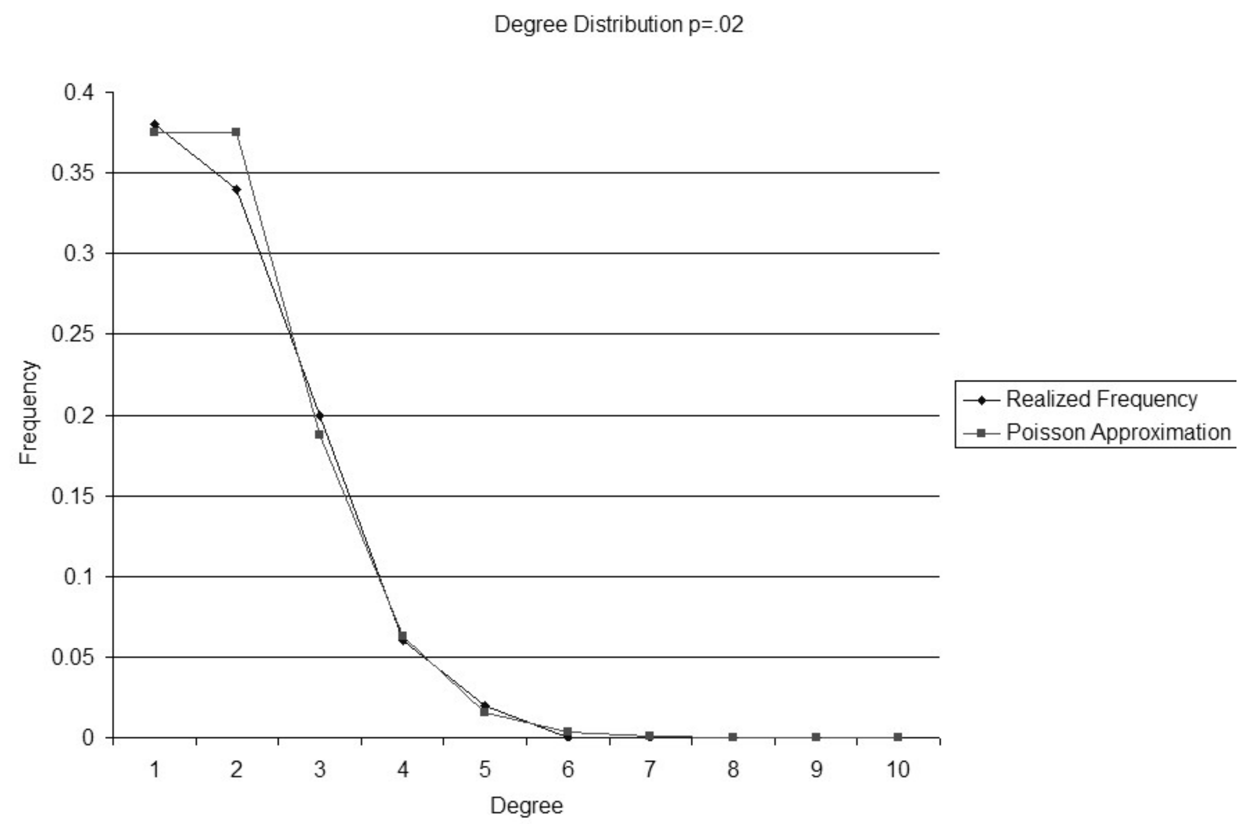
Figure 1.4: A Randomly Generated Network with Probability .02 on each Link

Jackson

- Ave. Degree ~ 1
- No cycles
- A "large" component
- P(K=0) = 1/e = 37.5% = 18.75
- 19 isolated nodes observed in this realization

This is one sample realization

Poisson Random Network



N=50

p=1/50=0.02

Figure 1.5: Frequency Distribution of a Randomly Generated Network and the Poisson Approximation for a Probability of .02 on each Link

Jackson

This is a sample degree distribution based on 1 realization.

If averaged over many realizations (or very large N) there will be excellent agreement

Poisson Random Network

N=50

p=ln(50)/50=0.078

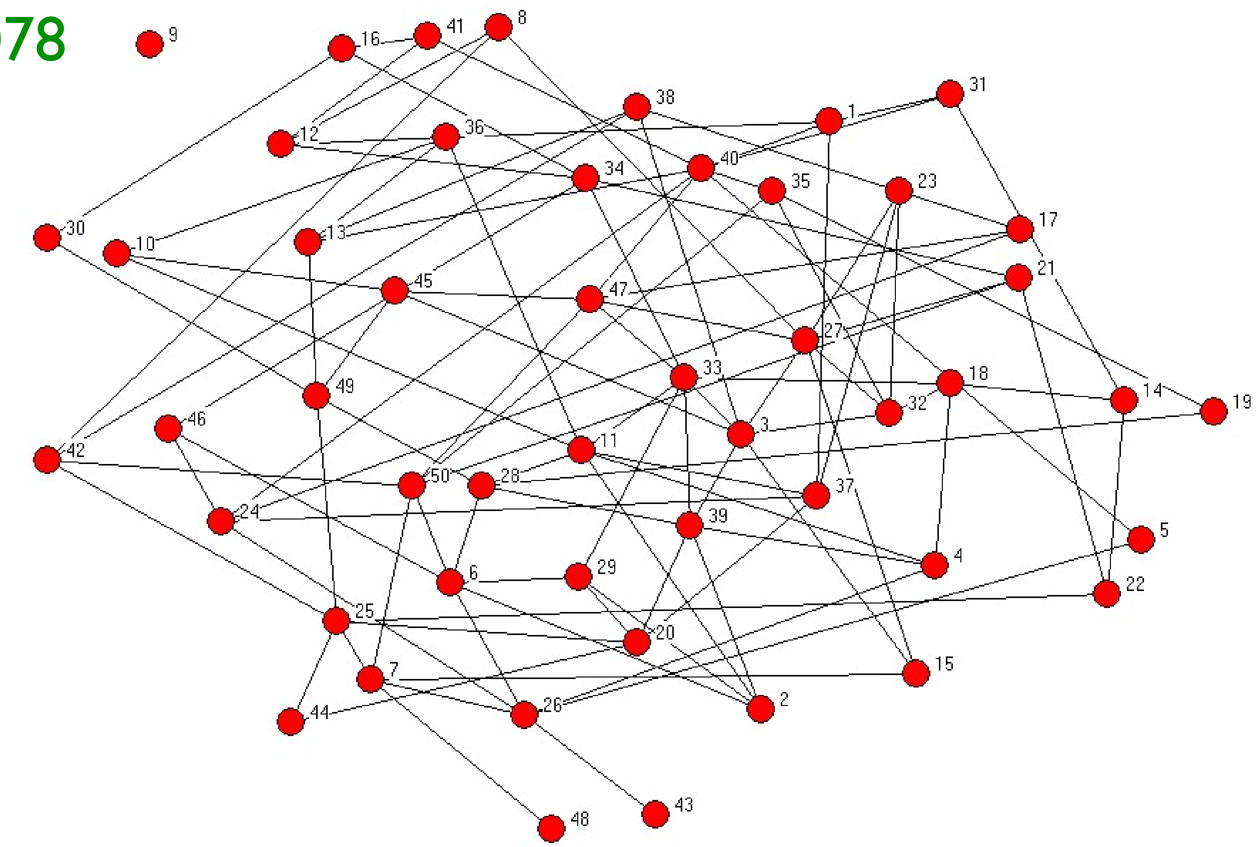


Figure 1.6: A Randomly Generated Network with Probability .08 of each Link
Jackson

- Ave. Degree ~ 3.8
- Many cycles
- A "giant component"
- P(K=0) = exp(-3.8) = 0.022
- 1 isolated node observed in this realization

This is one sample realization

Poisson Random Network

95% confidence region

(check out the sample mean handout)

N=50

p=ln(50)/50=0.078

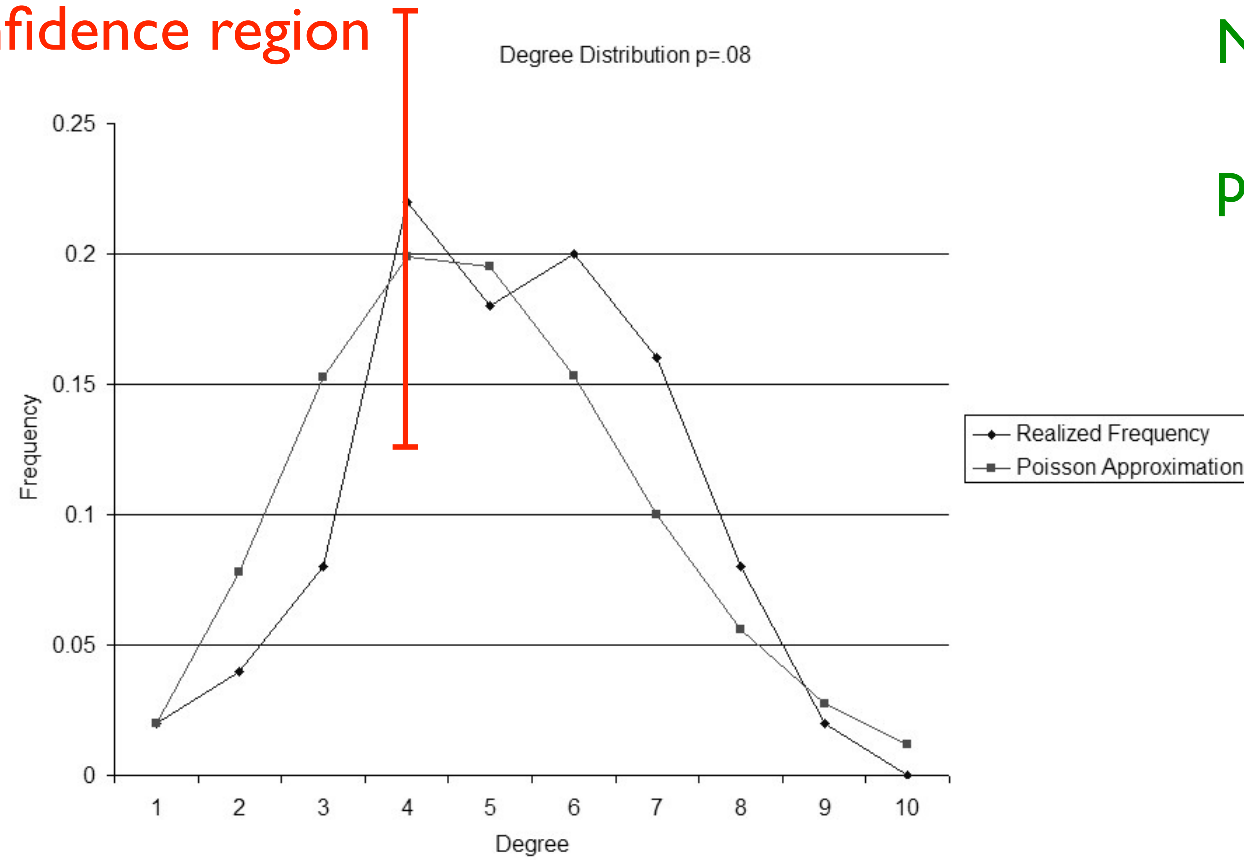


Figure 1.7: Frequency Distribution of a Randomly Generated Network and the Poisson Approximation for a Probability of .08 on each Link

Jackson

This is a sample degree distribution based on 1 realization.

If averaged over many realizations (or very large N) there will be excellent agreement

Aside: confidence for probability estimates

Gaussian approximation to Binomial:

$$p = \hat{p}_n \pm z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

Reasonably accurate p estimates:

$$p = \hat{p}_n \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_n\hat{q}_n}{n}} = \hat{p}_n\left(1 \pm z_{\alpha/2}\sqrt{\frac{\hat{q}_n}{k_n}}\right)$$

Example from last slide:

p=	0.22
q=	0.78
N=	50.00
sqrt(pq/N)=	0.06
MOE=	0.11

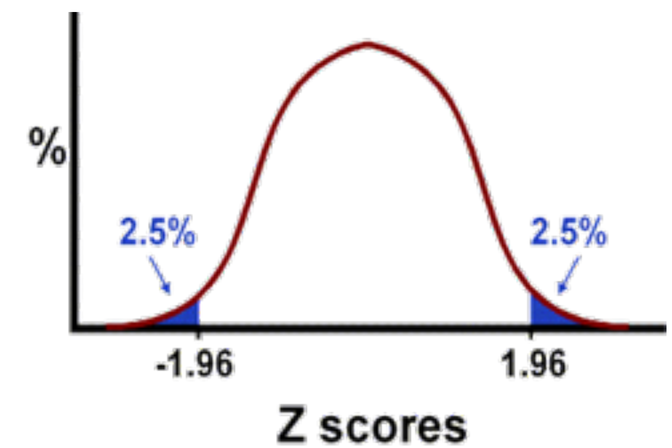


Table of z_(alpha/2), such that P(Z>z_(alpha/2))=alpha/2 when Z is Gaussian, m=0,sigma=1

confidence=1-alpha	0.5	0.8	0.9	0.95	0.98	0.99
alpha/2=	0.25	0.1	0.05	0.025	0.01	0.005
z_(alpha/2)=	0.674	1.282	1.645	1.960	2.326	2.576

Figure 2: Table of values for $t_{\alpha/2}(n-1)$ and $z_{\alpha/2}$ for common values of confidence $1-\alpha$.

Chugg, sample mean handout

Asymptotic Thresholds

- Consider a property *A* of a graph and a Poisson graph with link formation probability *p=p(N)*

if:	$\lim_{N\to\infty} \frac{p(N)}{T(N)} = \infty$	**then:**	$\lim_{N\to\infty} P(A\|p(N)) = 1$
	and		
if:	$\lim_{N\to\infty} \frac{p(N)}{T(N)} = 0$	**then:**	$\lim_{N\to\infty} P(A\|p(N)) = 0$

$T(N)$ is a threshold (on p) for A

- Property *A* occurs with probability 1 asymptotically as long as *p(N) grows as T(N)* (and T(N) is the smallest such growth rate)

Asymptotic Thresholds

- Why use this method of analysis?
  - Very difficult to determine specific probabilities for network properties exactly for a fixed value of N.
  - Often interested in very large values of N
  - Predicts "phase transitions"
    - Thresholds mark significant changes in qualitative properties of network

Poisson Asymptotic Thresholds

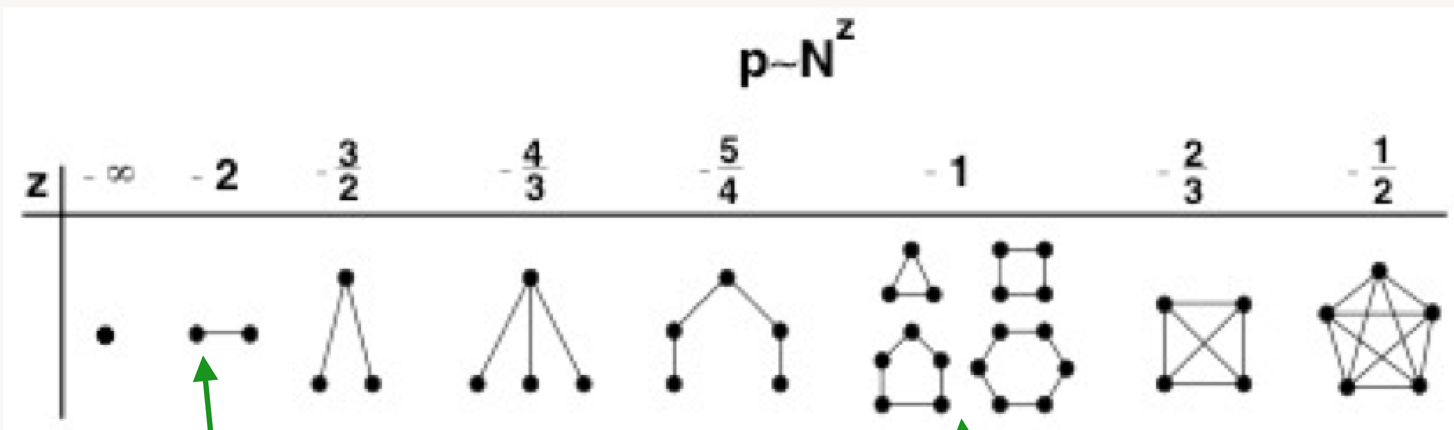


Image 3.7

Evolution of random graph.

Barabasi

links (1/(n^2))

cycles (1/n)

- Thresholds for certain connectivity patterns in Poisson graphs

Poisson Asymptotic Thresholds

N	$\frac{1}{N^2}$	$\frac{1}{N^{3/2}}$	$\frac{1}{N}$	$\frac{\ln(N)}{N}$
50	0.0004	0.0028	0.02	0.078

*

p=0.01

expected degree = 0.5

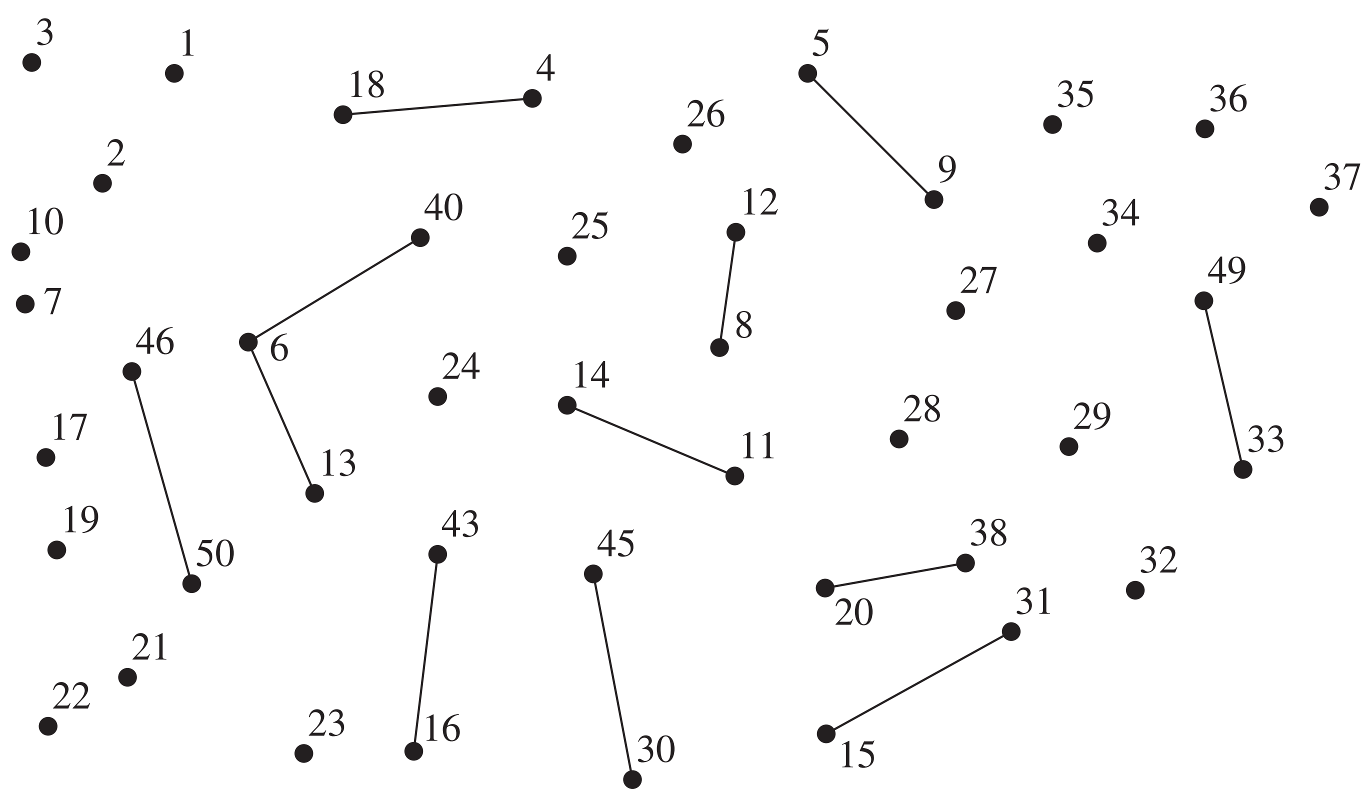


FIGURE 4.4 A first component with more than two nodes: a random network on 50 nodes with $p = .01$.
Jackson

- Links & size 3 components

Poisson Asymptotic Thresholds

N	$\frac{1}{N^2}$	$\frac{1}{N^{3/2}}$	$\frac{1}{N}$	$\frac{\ln(N)}{N}$
50	0.0004	0.0028	0.02	0.078

*

p=0.03

expected degree = 1.5

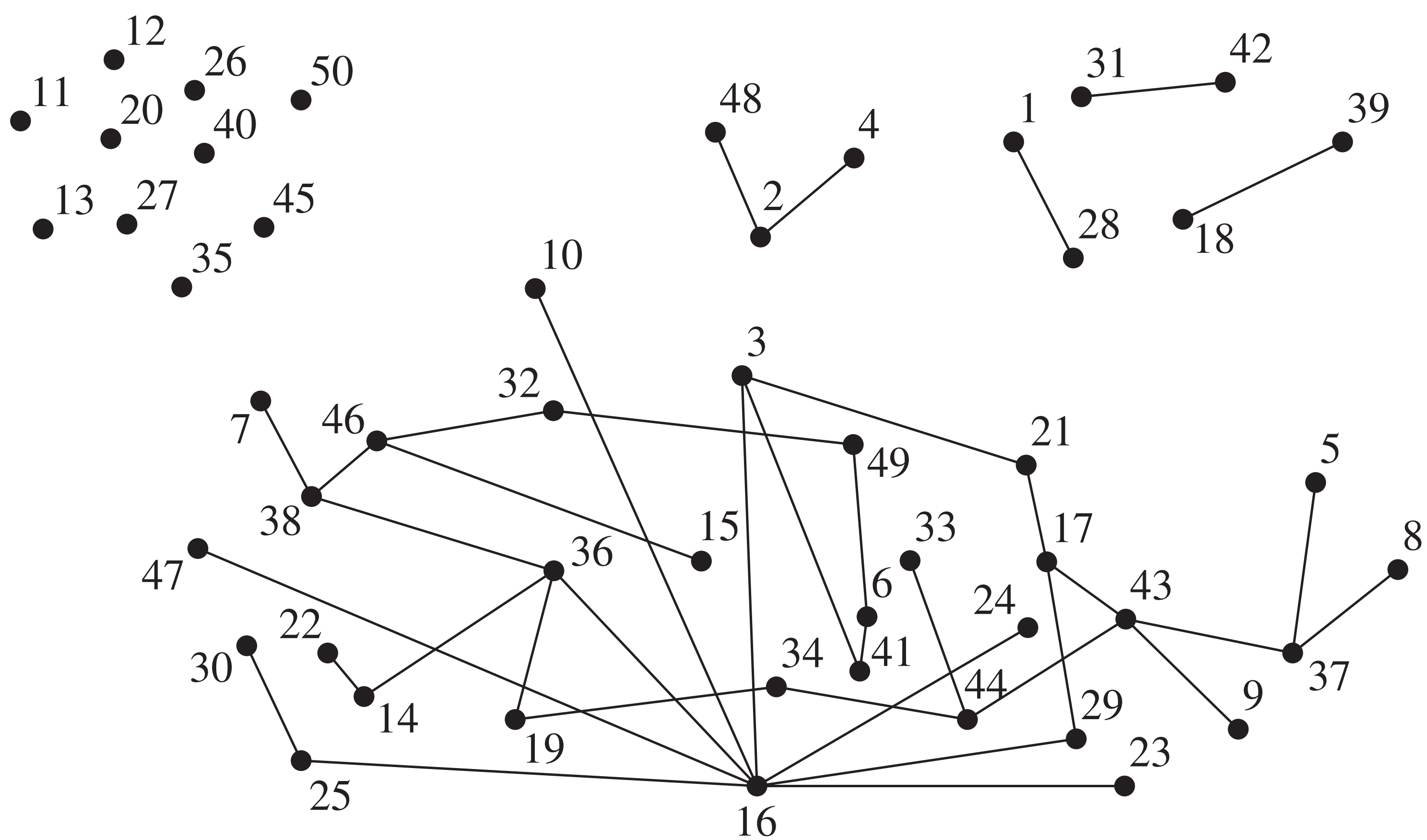


FIGURE 4.5 Emergence of cycles: a random network on 50 nodes with $p = .03$.

Jackson

- "Giant component," many isolated nodes, and cycles

Poisson Asymptotic Thresholds

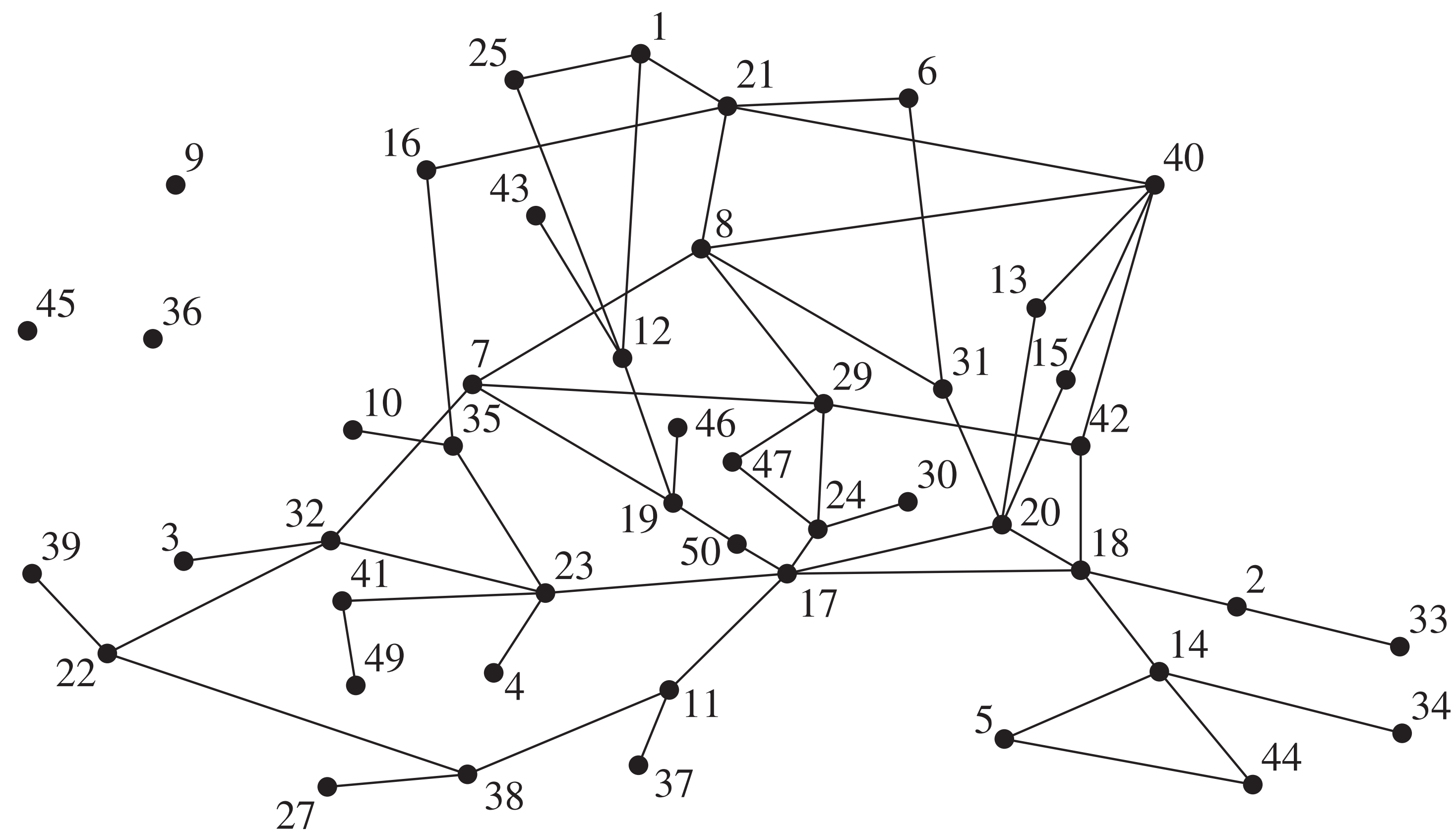


FIGURE 4.6 Emergence of a giant component: a random network on 50 nodes with $p = .05$.

Jackson

N	$\frac{1}{N^2}$	$\frac{1}{N^{3/2}}$	$\frac{1}{N}$	$\frac{\ln(N)}{N}$
50	0.0004	0.0028	0.02	0.078

*

p=0.05

expected degree = 2.5

- Giant component with some isolated nodes

Poisson Asymptotic Thresholds

N	$\frac{1}{N^2}$	$\frac{1}{N^{3/2}}$	$\frac{1}{N}$	$\frac{\ln(N)}{N}$
50	0.0004	0.0028	0.02	0.078

*

p=0.1

expected degree = 5

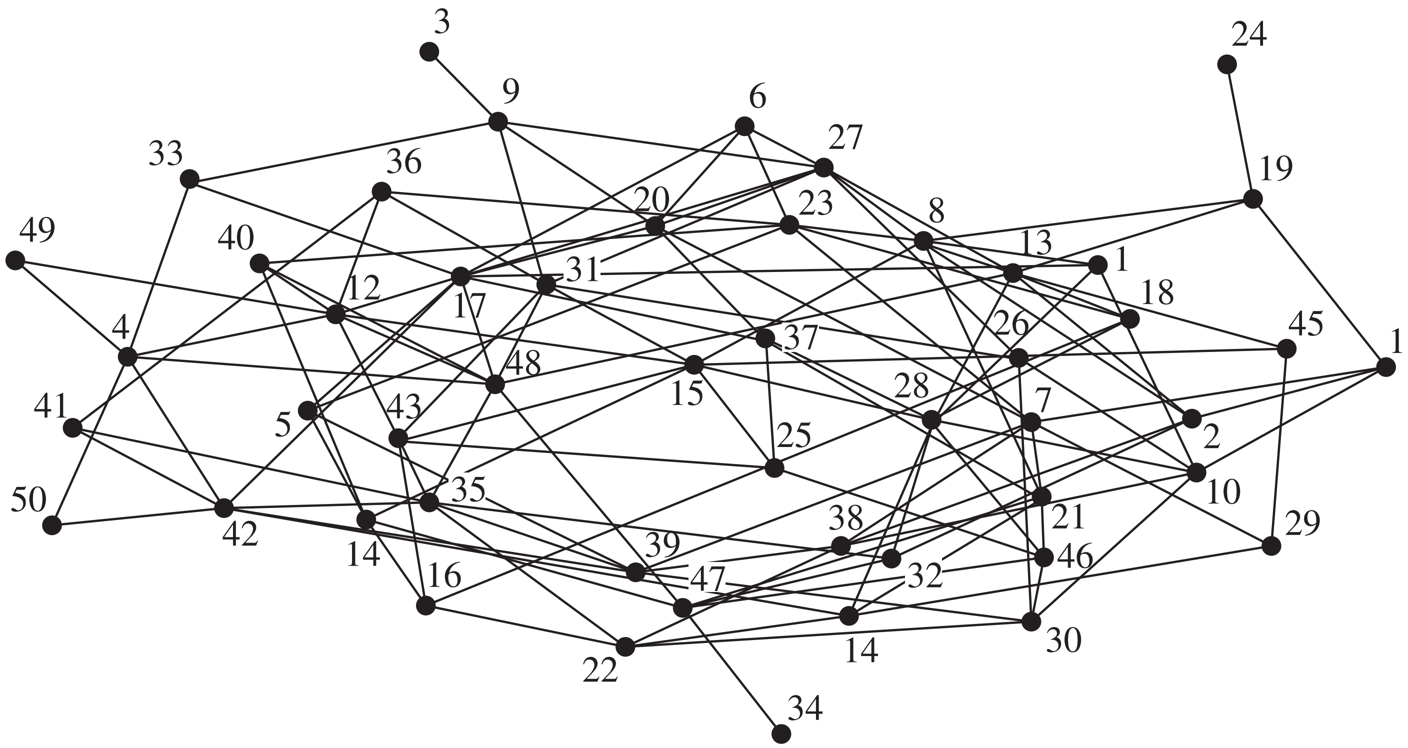


FIGURE 4.7 Emergence of connectedness: a random network on 50 nodes with $p = .10$.

Jackson

- Dense connectivity with no isolated nodes

Phase Transitions for Poisson Networks

- Heuristic argument ("hand wave")
  - Probability a node is ***not*** in giant component: *u*

$$
\begin{aligned}
u = P(v \notin GC) &= P(\text{All neighbors of } v \notin GC) \\
&= \sum_{k=0}^{\infty} P(\text{All neighbors of } v \notin GC | \text{degree}(v) = k) p_K(k) \\
&= \sum_{k=0}^{\infty} u^k \frac{\alpha^k}{k!} e^{-\alpha} \\
&= \sum_{k=0}^{\infty} \frac{(u\alpha)^k}{k!} e^{-\alpha} \\
&= e^{\alpha(u-1)}
\end{aligned}
$$

$$
1 - u = S = 1 - e^{-\alpha S} \qquad \alpha = \frac{1}{S} \ln\left(\frac{1}{1-S}\right)
$$

Phase Transitions for Poisson Networks

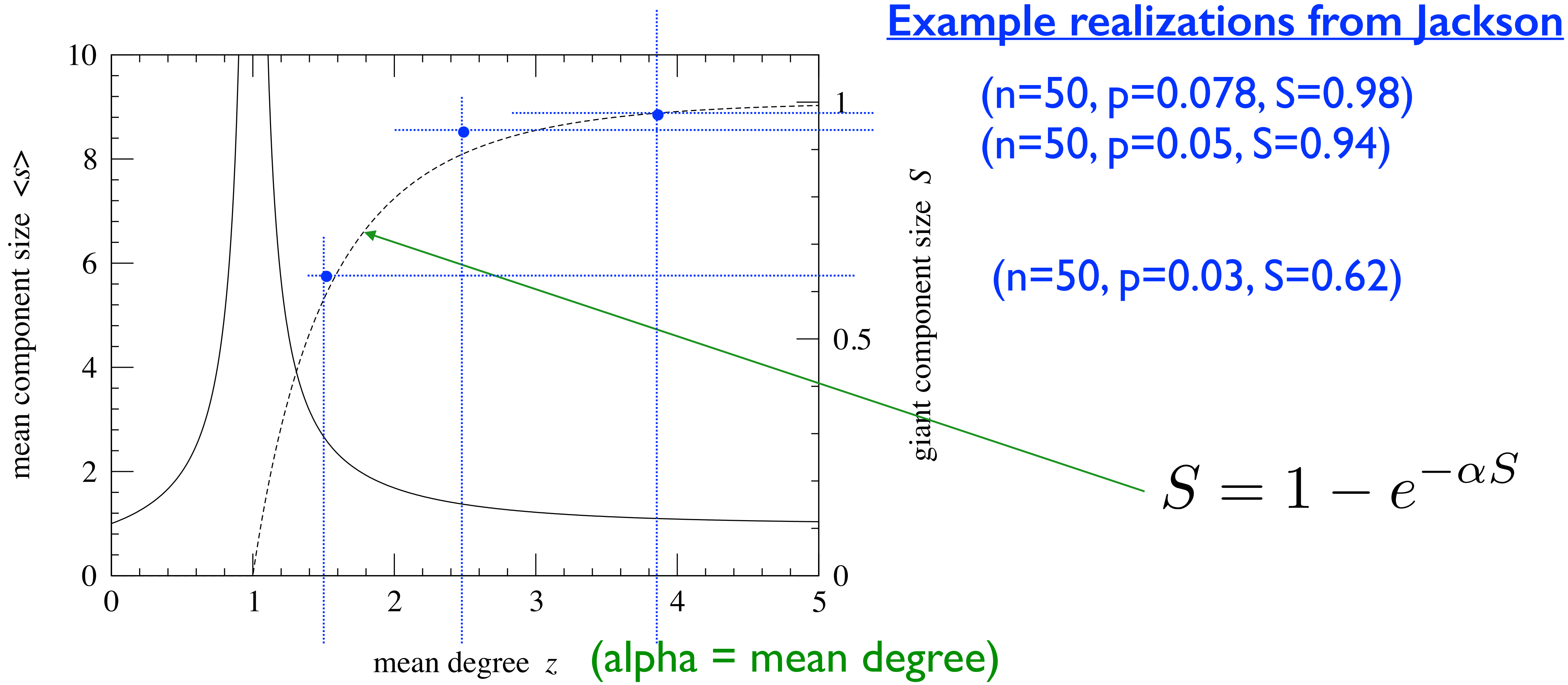


FIG. 10 The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph, Eqs. (20) and (21).

Newman

- Emergence of the Giant Component: *p(N)~1/N (alpha=1)*

Phase Transitions for Poisson Networks

$$P(\text{isolated node}) = e^{-\alpha} \approx \text{relative freq. of isolated nodes}$$

$$e^{-\alpha} = \frac{1}{N} \implies p(N) \sim \frac{\log(N)}{N}$$

- Emergence of the connectedness: *p(N)~ log(N)/N*

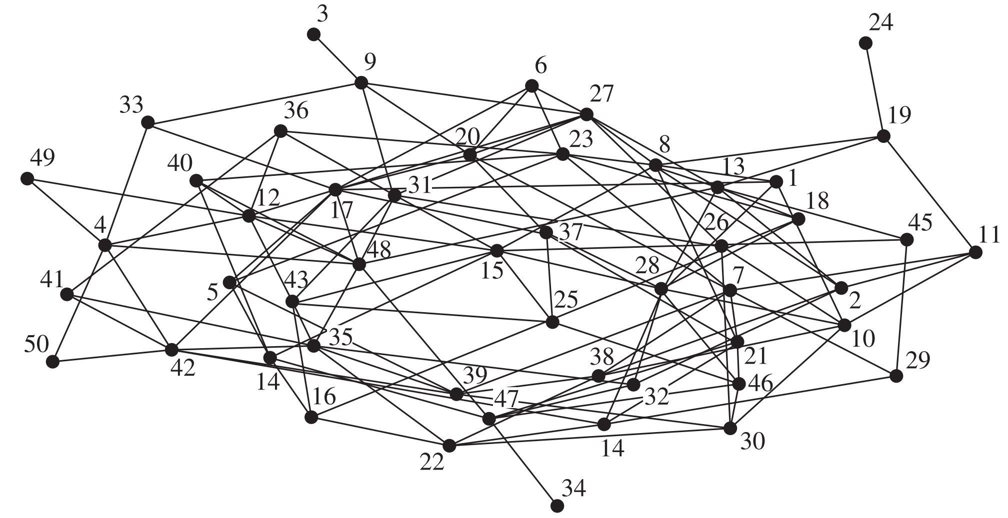
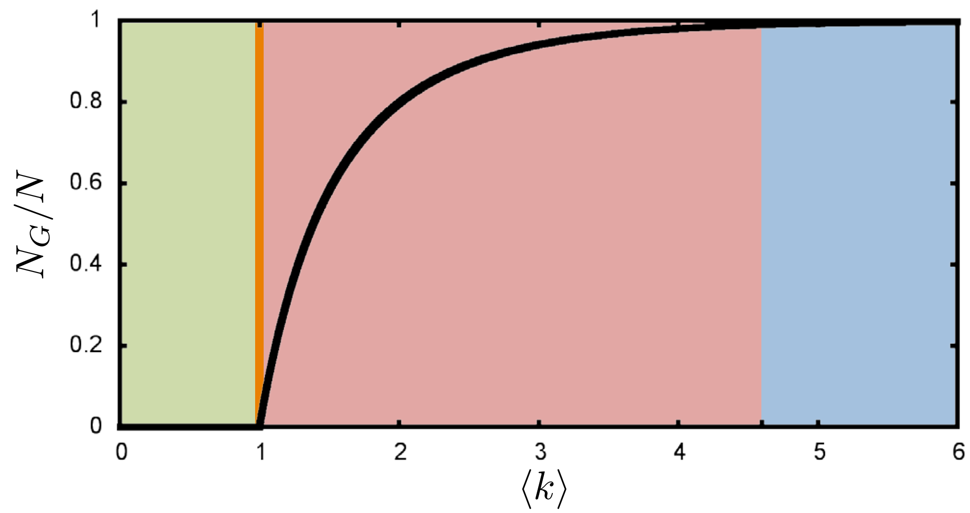


FIGURE 4.7 Emergence of connectedness: a random network on 50 nodes with $p = .10$.

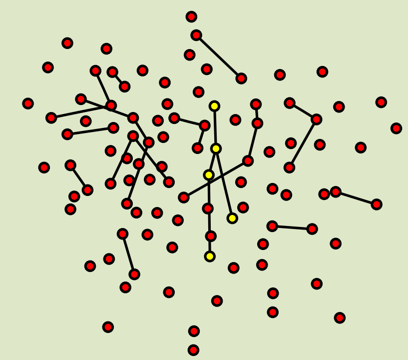
Jackson

Phase Transitions for Poisson Networks

a)



b)

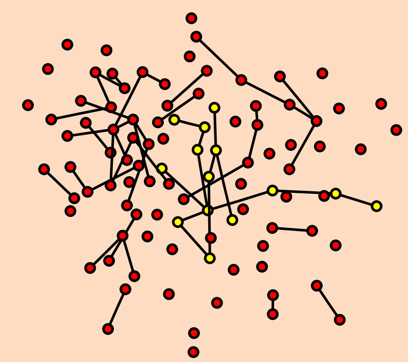


$\langle k \rangle < 1$

Subcritical regime

- No giant component.
- Cluster size distribution:
  $P(s) \sim e^{-\alpha s}$
- The size of the largest cluster:
  $N_G \sim \ln N$
- The clusters are trees.

c)

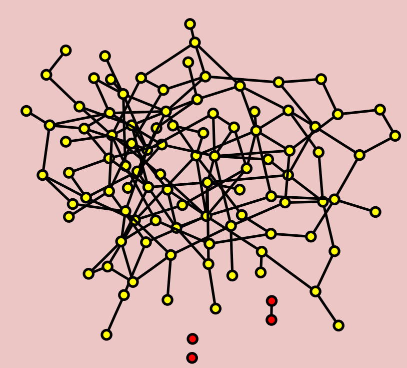


$\langle k \rangle = 1$

Critical point

- No giant component.
- Cluster size distribution:
  $P(s) \sim s^{-3/2}$
- Size of the largest cluster:
  $N_G \sim N^{2/3}$
- The clusters may contain loops.

d)

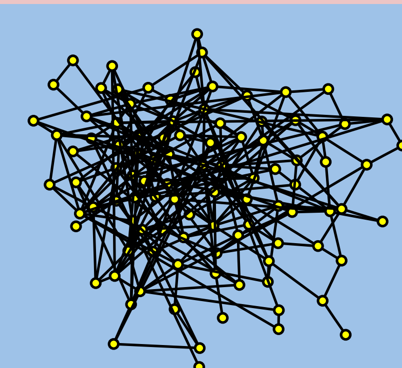


$\langle k \rangle > 1$

Supercritical regime

- Single giant component.
- Cluster size distribution:
  $P(s) \sim e^{-\alpha s}$
- Size of the giant component:
  $N_G \sim (p - p_c)N$
- The small clusters are trees.
- GC has loops.

e)



$\langle k \rangle \geq \ln N$

Fully connected regime

- Single giant component.
- No isolated nodes or clusters.
- Size of the giant component:
  $N_G = N$
- GC has many loops.

Image 3.6

Evolution of a random network.

(a) The relative size of the giant component in function of the average degree ‹k› in the Erdős-Rényi model.
(b)-(e) The main network characteristics in the four regimes that characterize a random network.
Barabasi

Poisson Random Network

$k_{\text{ave}}(N) \sim$	$\frac{1}{N}$	$\frac{1}{\sqrt{N}}$	$\frac{1}{N^{1/3}}$	$\frac{1}{N^{0.25}}$	1	$N^{1/3}$	$\sqrt{N}$	$\ln(N)$
$p(N) \sim$	$\frac{1}{N^2}$	$\frac{1}{N^{1.5}}$	$\frac{1}{N^{4/3}}$	$\frac{1}{N^{1.25}}$	$\frac{1}{N}$	$\frac{1}{N^{2/3}}$	$\frac{1}{\sqrt{N}}$	$\frac{\ln(N)}{N}$
	links (size 2 trees)	size 3 trees	size 4 trees	size 5 trees		size 4 cliques	size 5 cliques	

"subcritical"

Cycles, giant component

"supercritical"

"connected"

Connected Network

Application: Contagion/Diffusion

- Consider modeling the spread of an epidemic with a simplified model
  - People are either completely immune or complete susceptible
  - A fraction π of the population is immune
- *How severe will an outbreak be in this network if a single person is exposed?*
- Generate a Poisson random net with N nodes, prob p
  - Remove a fraction π of the nodes and their links
  - Remaining network is the net of susceptible people
- The size of the outbreak is given by "S curve" of the net of susceptible people

Application: Contagion/Diffusion

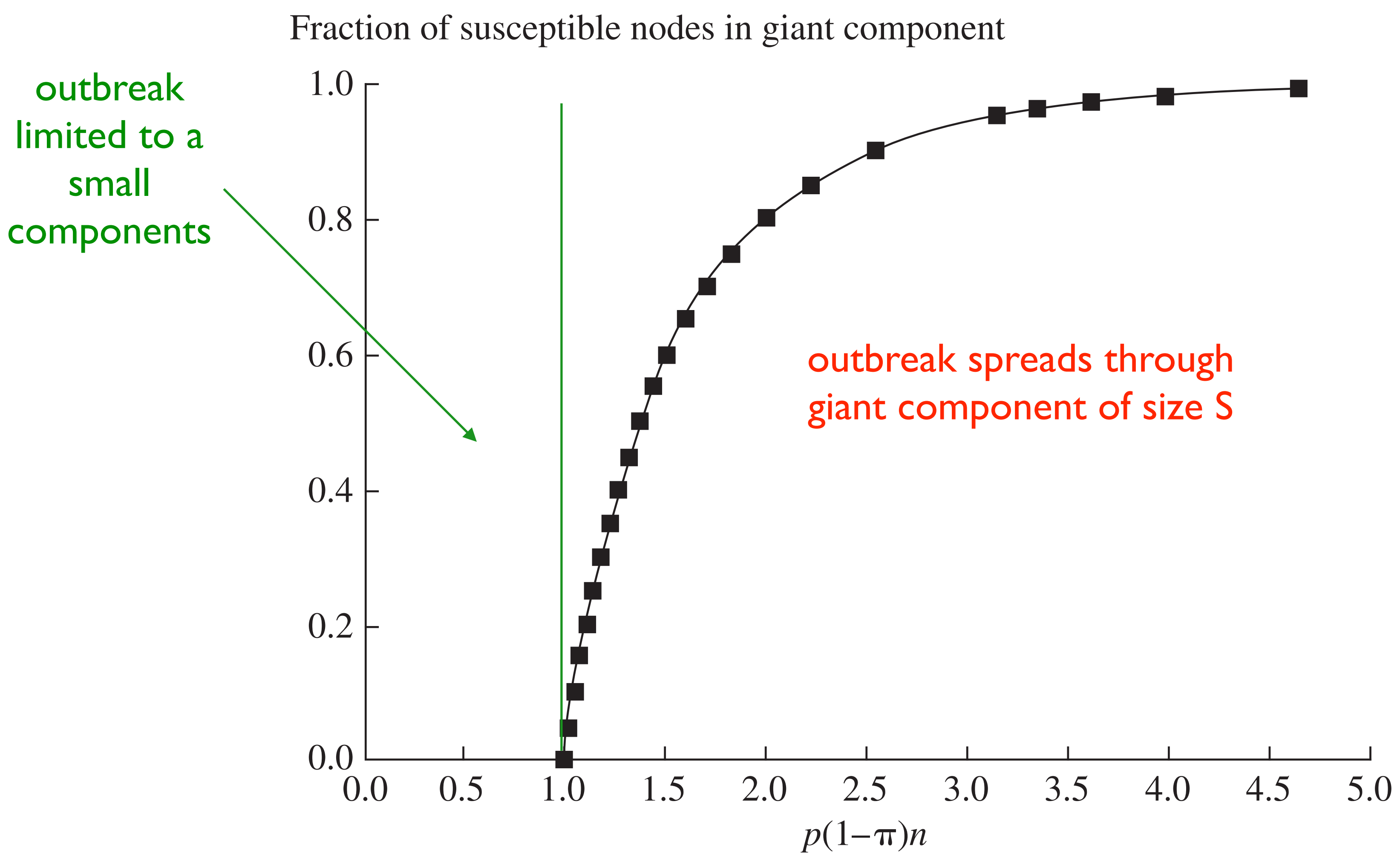


FIGURE 4.8 Fraction of the susceptible population in the largest component of a Poisson random network as a function of the proportion of susceptible nodes $1-\pi$ times the link probability p times the population size n. Barabasi

- Also view as p, N fixed and varying π — "herd immunity"

Most Real Networks are Supercritical

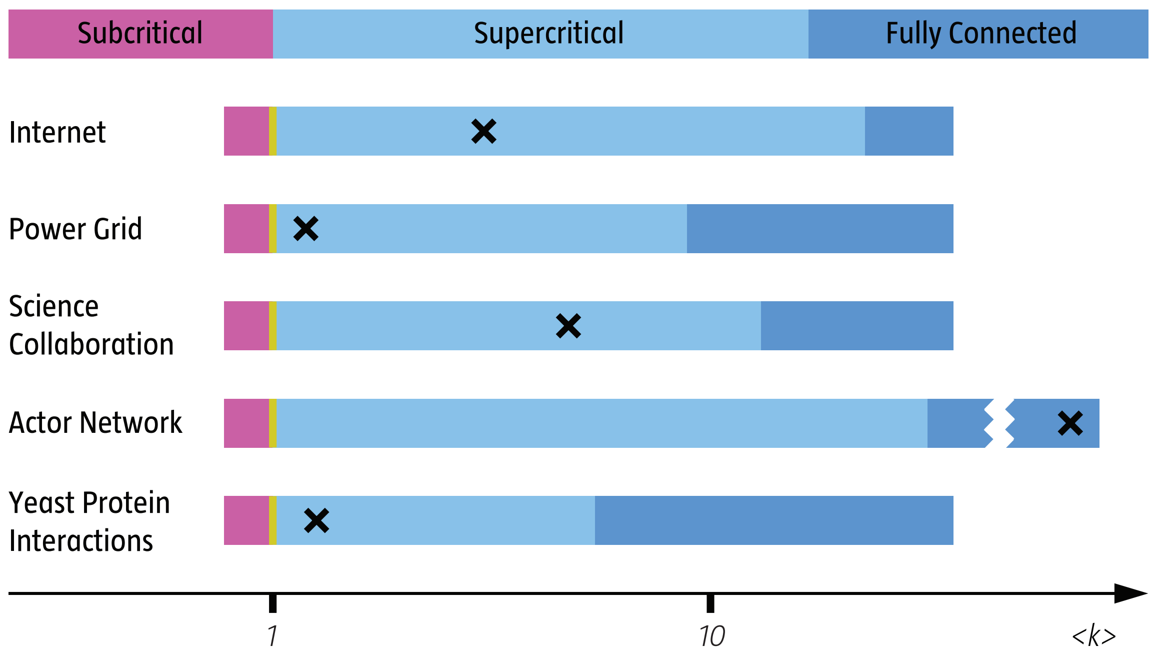
Network	N	L	$\langle k \rangle$	$\ln N$
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	186,936	8.08	10.04
Actor Network	212,250	3,054,278	28.78	12.27
Yeast Protein Interactions	2,018	2,930	2.90	7.61

Table 3.1

Are real networks connected?

The number of nodes N and links L for several undirected networks, together with $\langle k \rangle$ and $\ln N$. A giant component is expected for $\langle k \rangle > 1$ and all nodes should join the giant component for $\langle k \rangle \geq \ln N$. While for all networks $\langle k \rangle > 1$, for most $\langle k \rangle$ is under the $\ln N$ threshold.

Barabasi



Poisson Random Nets

- Does this model predict the characteristics of real networks (social nets in particular)

Yes
- Giant Component

?
- small world property?

?
- degree distribution?

?
- clustering?

(No)
- homophily (assortative mixing)?

(No)
- degree correlations?

Degree Distribution?

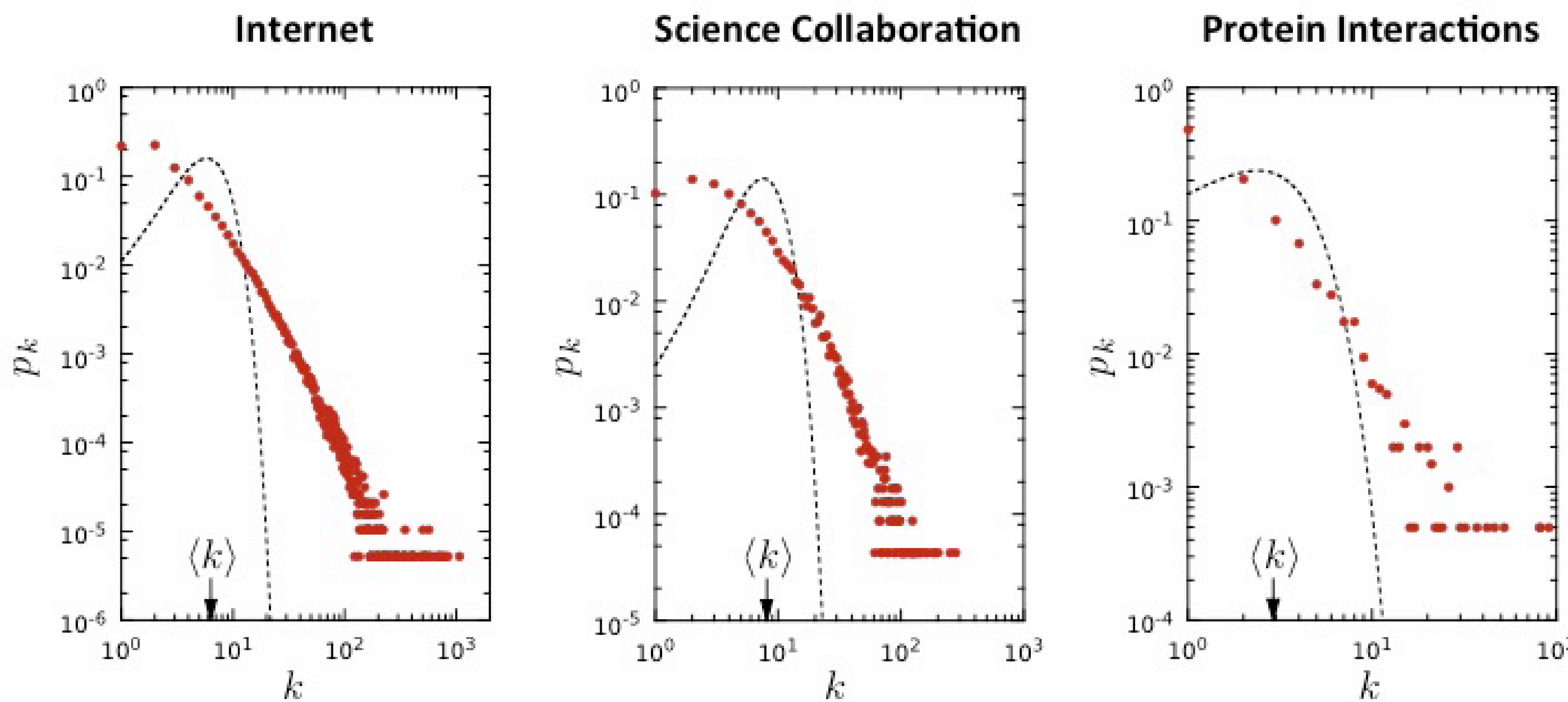


Image 3.5

Degree distribution of real networks.

The degree distribution of the Internet, science collaboration network, and the protein interaction network of yeast (Table 2.1). The dashed line corresponds to the Poisson prediction, obtained by measuring ‹k› for the real network and then plotting Eq. (8). The significant deviation between the data and the Poisson fit indicates that the random network model underestimates the size and the frequency of highly connected nodes, or hubs.

Barabasi

Real networks have degree distributions with heavier tails than Poisson:

a small fraction of nodes have very high degree

Poisson Random Network

- Cluster coefficient
  - Expected value of cluster coefficient is *p=alpha/N*
    - Number of closed triples is binomial(M_i,p) with M_i=L_i choose 2

Poisson Model Predicts:

- For fixed average degree, clustering should fall as 1/N
- Expected local cluster coefficient not a function of node degree
- (data shows both untrue)

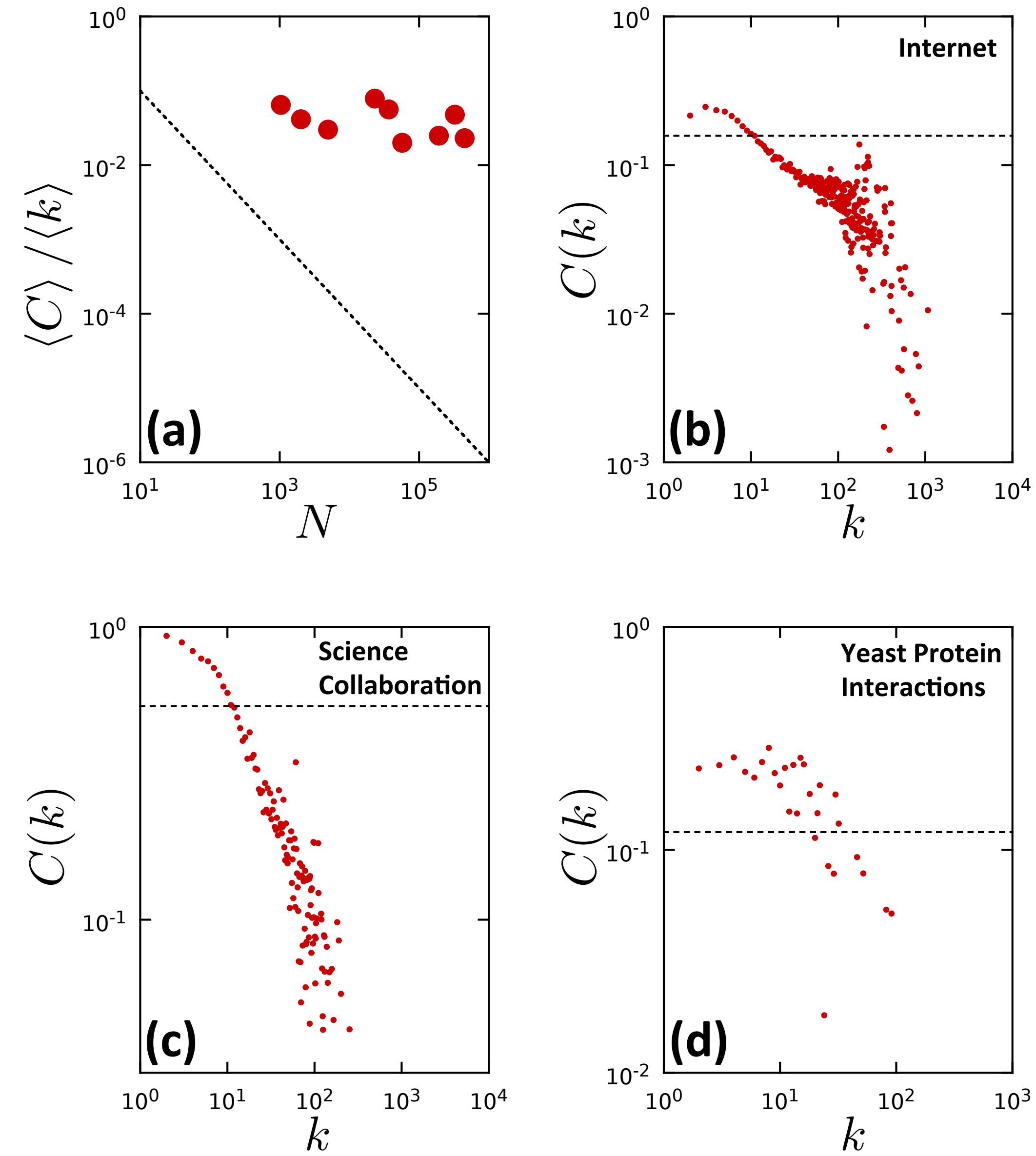
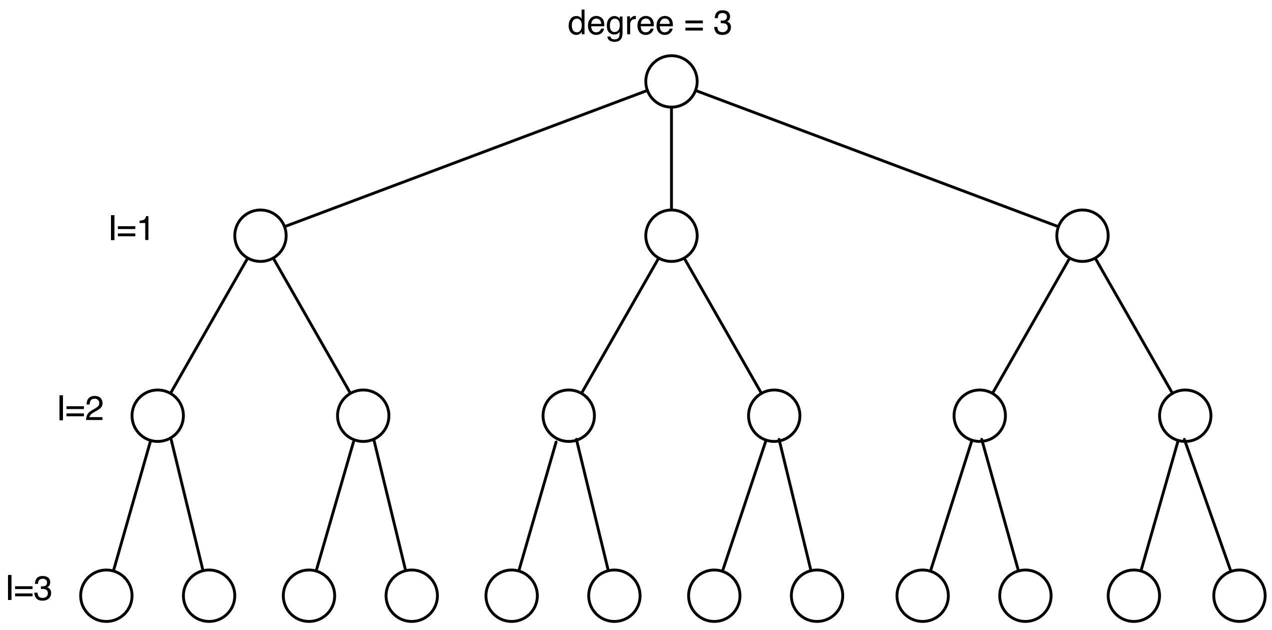


Image 3.16
Clustering in real networks.
Barabasi

$$\mathbb{E}\{C_i\} = \mathbb{E}\{C_{\text{ave}}\} = \mathbb{E}\{C_{\text{global}}\} = p \approx \frac{k_{\text{ave}}}{N}$$

Poisson Random Network

- Does the *small world property* hold?
  - Consider a fixed-degree k and expand from a given node (similar arguments are given with k=average degree)



$k(k-1)^{l-1}$ nodes at level l

$\left(\frac{k}{k-2}\right)((k-1)^l - 1)$ nodes reached by level l

$D \approx 2\frac{\log(N-1)}{\log(k-1)}$ diameter

practically: $d_{\text{ave}} \approx \frac{\log(N)}{\log(k)}$

Poisson Random Network

- Does the *small world property* hold?
  - More complicated/accurate methods for computing the diameter and/or shortest path based on this idea
    - How fast does the graph "expand" as you do a BFS from a node?
    - How many levels in the BFS to reach all nodes?

practically: $$d_{\mathrm{ave}} \sim \frac{\ln(N)}{\ln(k_{\mathrm{ave}})}$$

- More complex analysis provides "corrections" to the above…

Poisson Random Network

Network Name	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\log N}{\log \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.59
WWW	325,729	1,497,134	4.60	11.27	93	8.32
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	186,936	8.08	5.35	15	4.81
Actor Network	212,250	3,054,278	28.78	-	-	-
Citation Network	449,673	4,707,958	10.47	11.21	42	5.55
E Coli Metabolism	1,039	5,802	5.84	2.98	8	4.04
Yeast Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Table 3.2

Six degrees of separation.

The average distance $\langle d \rangle$ and the maximum distance d_{max} of the ten networks explored in this book. The last column provides $\langle d \rangle$ predicted by Eq. (19), indicating that it offers a reasonable approximation to $\langle d \rangle$. Yet, the agreement is not perfect - we will see in the next chapter that for many real networks Eq. (19) needs to be adjusted. For directed networks we list the average out-degree $\langle k_{out} \rangle$ and the path lengths are measured only along the direction of the links.

Barabasi

- Poisson networks do have the *small world property!*

Poisson Random Network

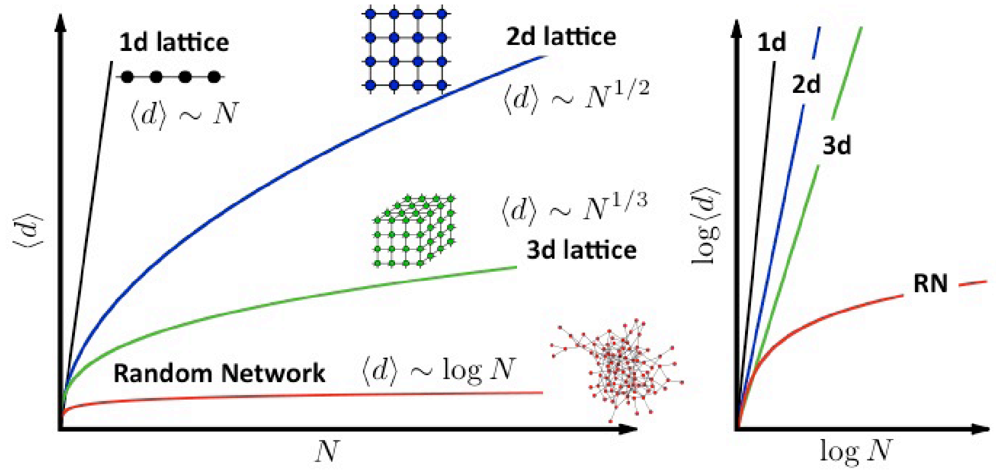


Image 3.10

Why are small worlds surprising?

Much of our intuition about distance is based on our experience with regular lattices, which do not display the small world phenomenon. Indeed,

- For a one-dimensional lattice (a line of length *N*) the diameter and the average path length scale linearly with N: $d_{max} \sim \langle d \rangle \sim N$.
- For a square lattice $d_{max} \sim \langle d \rangle \sim N^{1/2}$.
- For a cubic lattice $d_{max} \sim \langle d \rangle \sim N^{1/3}$.
- In general, for a *d*-dimensional lattice we have $d_{max} \sim \langle d \rangle \sim N^{1/d}$.

Such polynomial dependence predicts a much faster increase with *N* than Eq. (19), indicating that in regular lattices the path lengths are significantly longer than in a random network. The figure shows the predicted *N*-dependence of $\langle d \rangle$ for regular and random networks on a linear (left) and on a log-log (right) scale. If the social network would form a regular 2d lattice, where each individual knows only its nearest neighbors, the average distance between two individuals would be roughly $(7 \times 10^9)^{1/2} = 83{,}666$. Even if we correct for the fact that a person has about 1,000 acquaintances, not four, the average separation will be orders of magnitude larger than predicted by Eq. (19).

Barabasi

Our intuition tends to make us think in terms of regular networks

Some networks are regular and do not exhibit small world properties (non-social)

Poisson Random Network

At a glance: Random networks

- *Definition*: N nodes, where each node pair is connected with probability p.
- *Average degree:* $\langle k \rangle = p(N-1)$
- *Average number of links:* $\langle L \rangle = \frac{p(N-1)}{2}$
- *Degree distribution:* $p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$.

For sparse networks $(k \ll N)$, P_k has the Poisson form

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}.$$

- *Giant component* (N_G):

$\langle k \rangle < 1$: no giant component $(N_G \sim \ln N)$

$1 < \langle k \rangle < \ln N$: one giant component and disconnected clusters

$$\left(N_G \sim N^{\frac{2}{3}} \right)$$

$\langle k \rangle > \ln N$: all nodes join the giant component $N_G \sim (p - p_c) N$

- *Average distance:* $\langle d \rangle \propto \frac{\log N}{\log \langle k \rangle}$,
- *Clustering coefficient:* $C = \frac{\langle k \rangle}{N}$.

Box 3.8

Barabasi

- Summary (Poisson predicts real?)
  - Degree distribution No
  - Giant Component Yes
  - Small World Yes
  - Clustering No

Overview

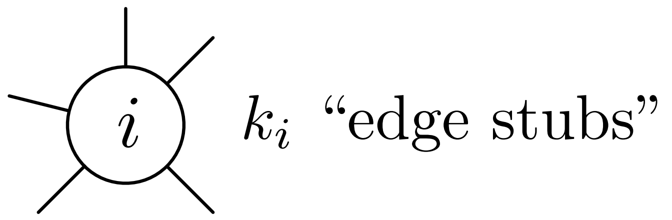
- Probability Review
- Static Random Models
  - Poisson random graphs (completely random)
  - The configuration model (specify the degree distribution)
    - *Power-law degree distribution*
  - The small world model (Watts-Strogatz)
  - Exponential and Markov models
- Growth models for random graphs

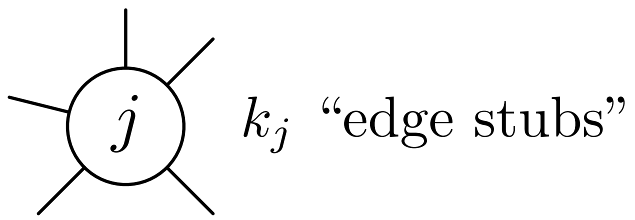
Configuration Model

- Since many real-world networks do not have a Poisson degree distribution, generate a random network with a specific degree sequence (implies degree distribution)
    - Several models to approach this or similar
        - *Configuration model*
        - *Expected Degree model (Jackson 4.1.5) and Hidden Parameter model (Barabasi 4.8)*
        - *Degree-preserving randomization (Barabasi 4.8)*

Configuration Model

- Given a specific degree sequence {k_i}, set up nodes with each node having the specified number of degree stubs
  - Randomly pick two stubs and connect with link (these are no longer stubs)
  - Repeat until all stubs are links
    - Allows for self-links and multi-links, but the probability of these go to zero as N goes to infinity





Provides random graph with specified degree distribution

(Poisson random network is a special case)

Configuration Model

- Condition for the emergence of the Giant Component

$$\mathbb{E}\left\{K^2\right\} - 2\mathbb{E}\left\{K\right\} > 0$$

THRESHOLD condition for giant component to exist asymptotically

$$(1-S) = \sum_{k=0}^{\infty}(1-S)^k p_K(k)$$

S = fraction of nodes in the GC when above threshold is met

Note that for Poisson distribution with mean alpha:

$$m_K = \alpha$$

$$\sigma_K^2 = \alpha$$

$$\longrightarrow \mathbb{E}\left\{K^2\right\} = \sigma_K^2 + m_K^2 = \alpha + \alpha^2 \longrightarrow$$

$$\mathbb{E}\left\{K^2\right\} - 2\mathbb{E}\left\{K\right\} > 0$$

$$\Updownarrow$$

$$\alpha > 1$$

(This may have some issues since it allows for nodes to connect to themselves)

Configuration Model

- Expected value of average/global cluster coefficient

$$\mathbb{E}\left\{C_i\right\} = \mathbb{E}\left\{C_{\text{ave}}\right\} = \mathbb{E}\left\{C_{\text{global}}\right\} = \frac{k_{\text{ave}}}{N}\left[\frac{\mathbb{E}\left\{K^2\right\} - k_{\text{ave}}}{k_{\text{ave}}^2}\right]^2$$

deviation from Poisson

(bracketed term is 1 for Poisson)

$$\frac{\mathbb{E}\left\{K^2\right\}}{(\mathbb{E}\left\{K\right\})^2}$$

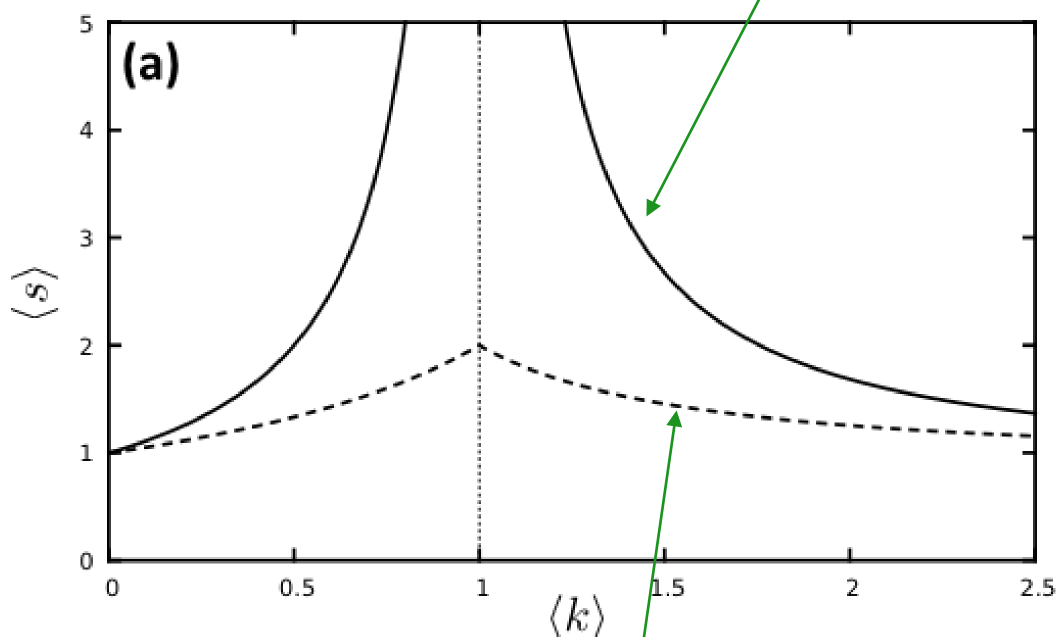
non-negligible correction term depending on the degree distribution and the size of the network

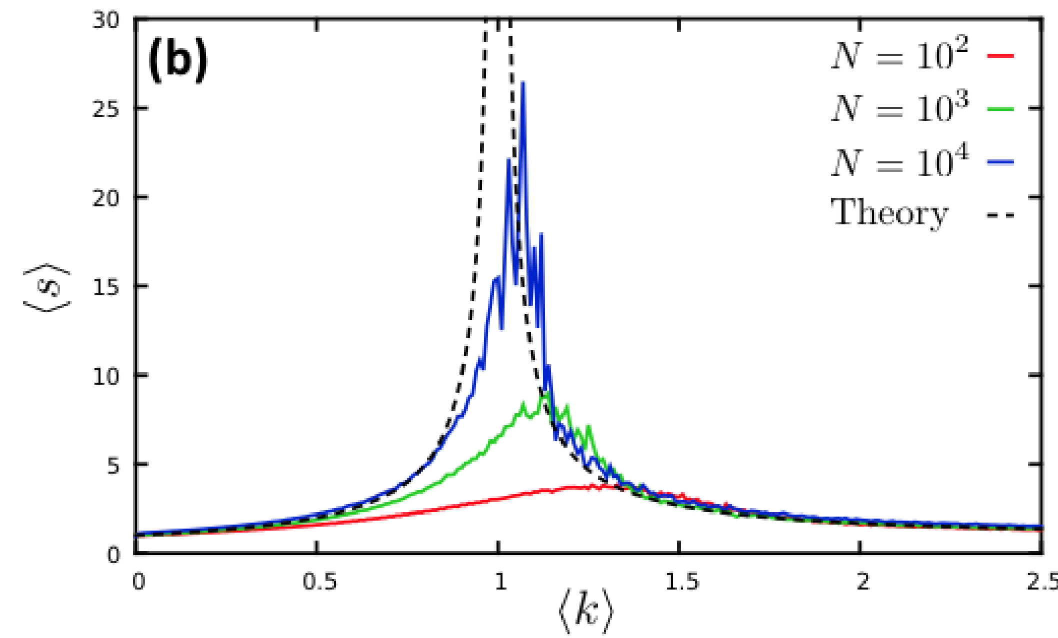
Configuration Model

- Some more detailed derivations for the configuration model (Poisson is a special case)
  - *Probability distribution of degree of node connected to randomly selected edge (Jackson 4.2.1)*
  - Threshold for giant component (Jackson 4.2.6)
  - *Sub-threshold (GC) component size distribution (Jackson 4.4)*
    - Implies Threshold for GC
    - Implies mean size of sub-threshold components
  - Probability distribution of maximum degree in finite network Newman (III.C.2)

Poisson Model Mean Component Size

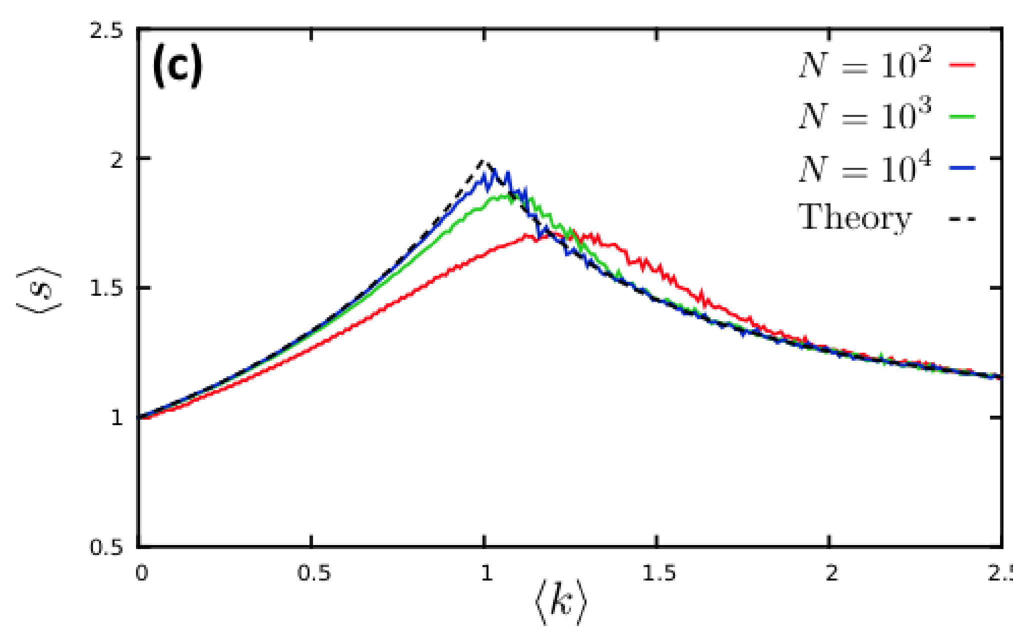
randomly selected node
$$\frac{1}{1 - k_{\text{ave}} + k_{\text{ave}} S}$$





(randomly selected node)

randomly selected component
$$\frac{2}{2 - k_{\text{ave}} + k_{\text{ave}} S}$$



(randomly selected component)

Barabasi

Power Law Degree Distribution

- Heavy-tailed degree distribution that matches many real-world network observed degree distributions

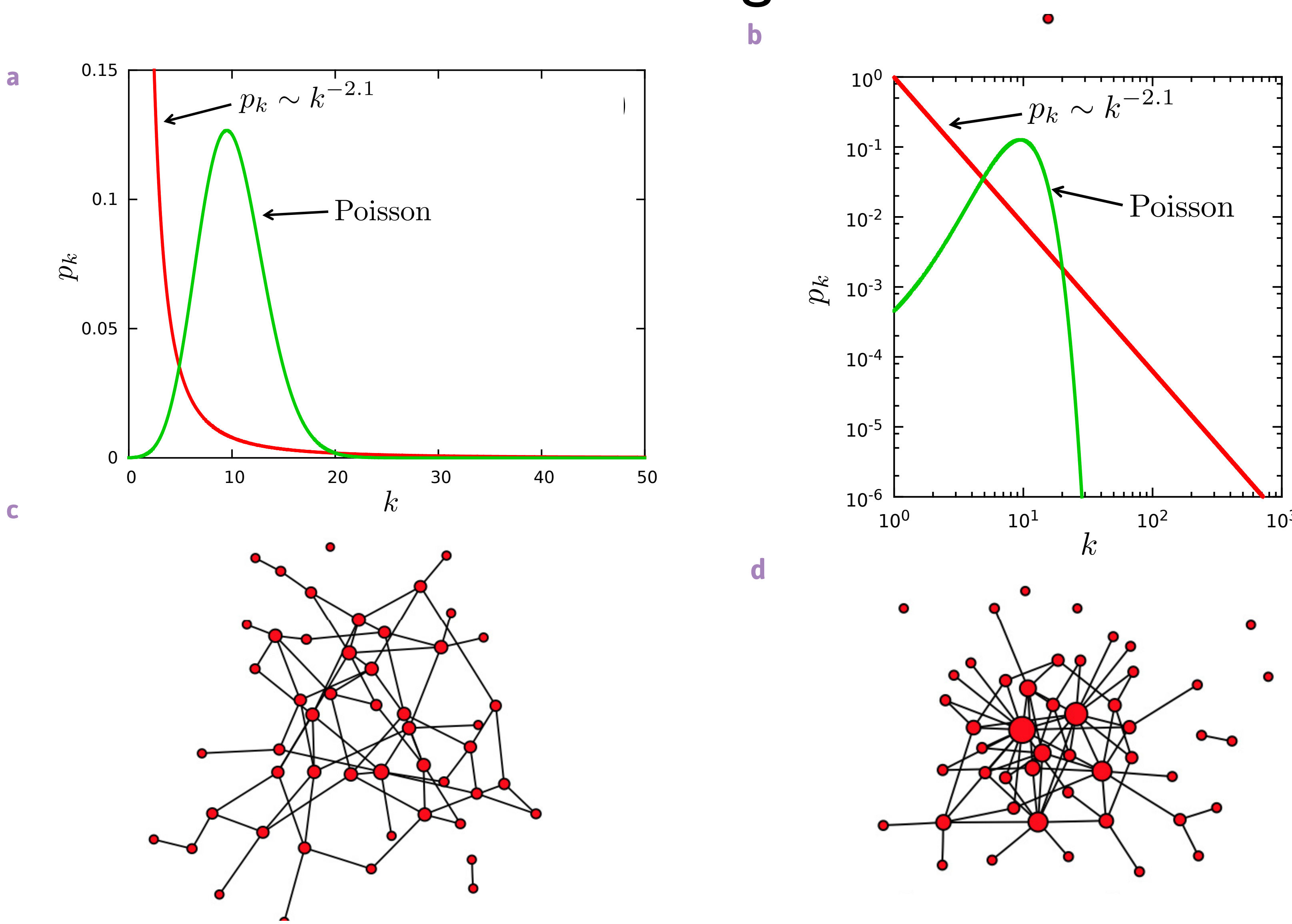


Figure 4.4
Poisson vs. power-law distributions

(a) A Poisson function and a power-law function with γ= 2.1. Both distributions have $\langle k \rangle$= 10.
(b) The curves in (a) shown on a log-log plot, offering a better view of the difference between the two functions in the high-k regime.
(c) A random network with $\langle k \rangle$= 3 and N = 50, illustrating that most nodes have comparable degree $k \simeq \langle k \rangle$.
(d) A scale-free network with $\langle k \rangle$= 3, illustrating that numerous small-degree nodes coexist with a few highly connected hubs.

Barabasi

Power Law Degree Distribution

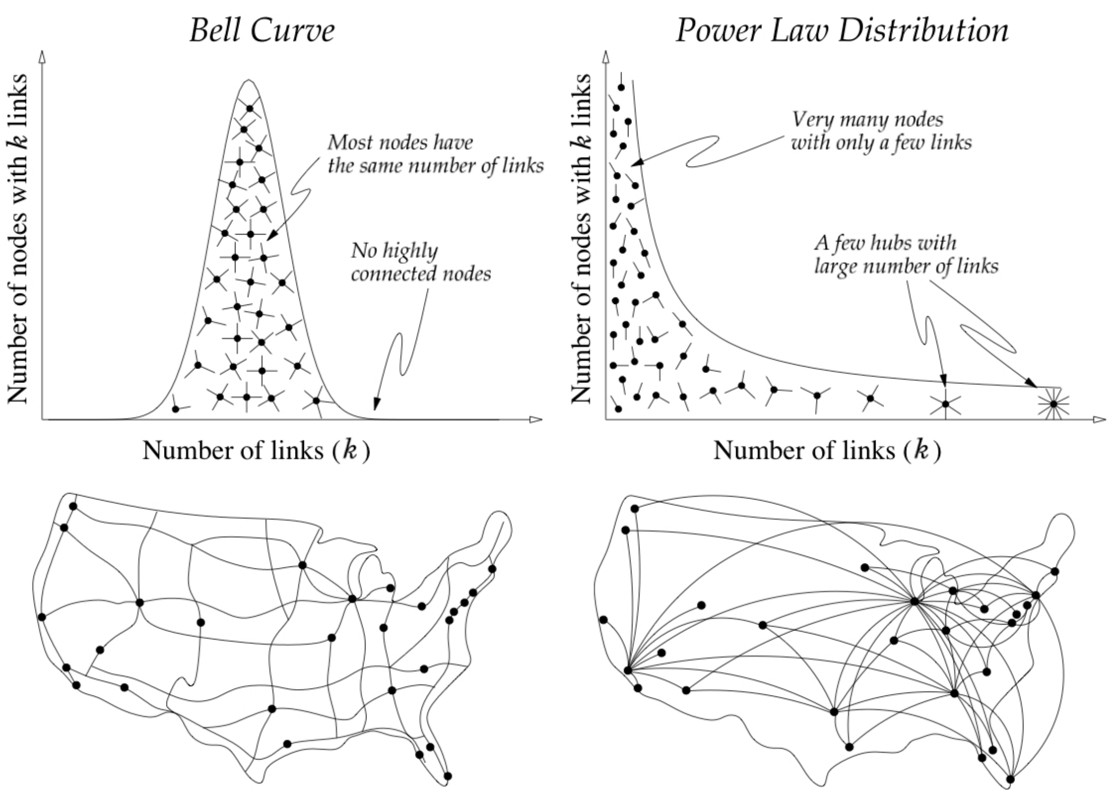


Figure 4.6
Random versus scale-free networks

Left column: the degrees of a random network follow a Poisson distribution, which is rather similar to the Bell curve shown in the figure. This indicates that most nodes have comparable degree. Hence nodes with a large number of links are absent (top panel). Consequently a random network looks a bit like a national highway network in which nodes are cities and links are the major highways connecting them (bottom panel). Indeed, there are no major cities with hundreds of highways and no city is disconnected from the highway system.

Right column: In a network with a power-law degree distribution most nodes have only a few links. These numerous small nodes are held together by a few highly connected hubs (top panel). Consequently a scale-free network looks a bit like the air-traffic network, whose nodes are airports and links are direct flights between them. Most airports are tiny, with only a few flights linking them to other airports. Yet, we can also have few very large airports, like Chicago or Atlanta, that hold hundreds of airports together, acting as major hubs (bottom panel).

Once hubs are present, they change the way we navigate the network. For example, if we travel from Boston to Los Angeles by car, we must drive through many cities (nodes). On the airplane network, however, we can reach most destinations via a single hub, like Chicago.

After [4].

Barabasi

- AKA "scale-free" networks

Power Law Degree Distribution

- Special case of configuration network
- Discussed in detail in Barabasi, chapter 4
- Networks can be grown with the Barabasi-Albert model for preferential attachment
  - Yields the power law degree distribution
  - No longer has independent degrees for different nodes
    - Model incorporates degree correlation explicitly

Cover power-law and preferential attachment later (soon)

Exponential, p*, Markov Random Nets

- These models are more complicated than the Poisson and configuration models
  - Attempt to capture correlation properties to come degree
    - Degree correlation, clustering, etc.
- There is a trade-off between more accurate models and more difficulty in analyzing the resulting models
  - The references point to these models as interesting/ promising, but not widely utilized

Small-World Models (Watts-Strogatz)

- Simple model that captures
  - Clustering (triadic closure, transitivity)
  - Small world property
- Basic idea
  - Begin with a regular lattice (local, regular connections)
    - provides clustering
  - Add/rewire a subset of links randomly
    - provides small-world properties

Small-World Models (Watts-Strogatz)

starting point

random
rewire

random add
wires

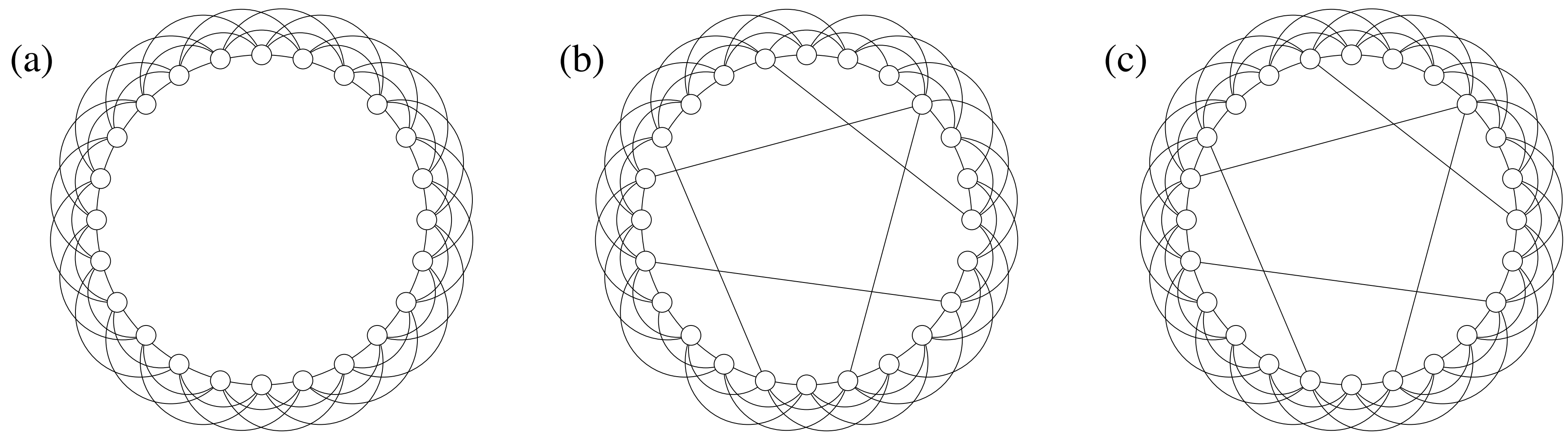
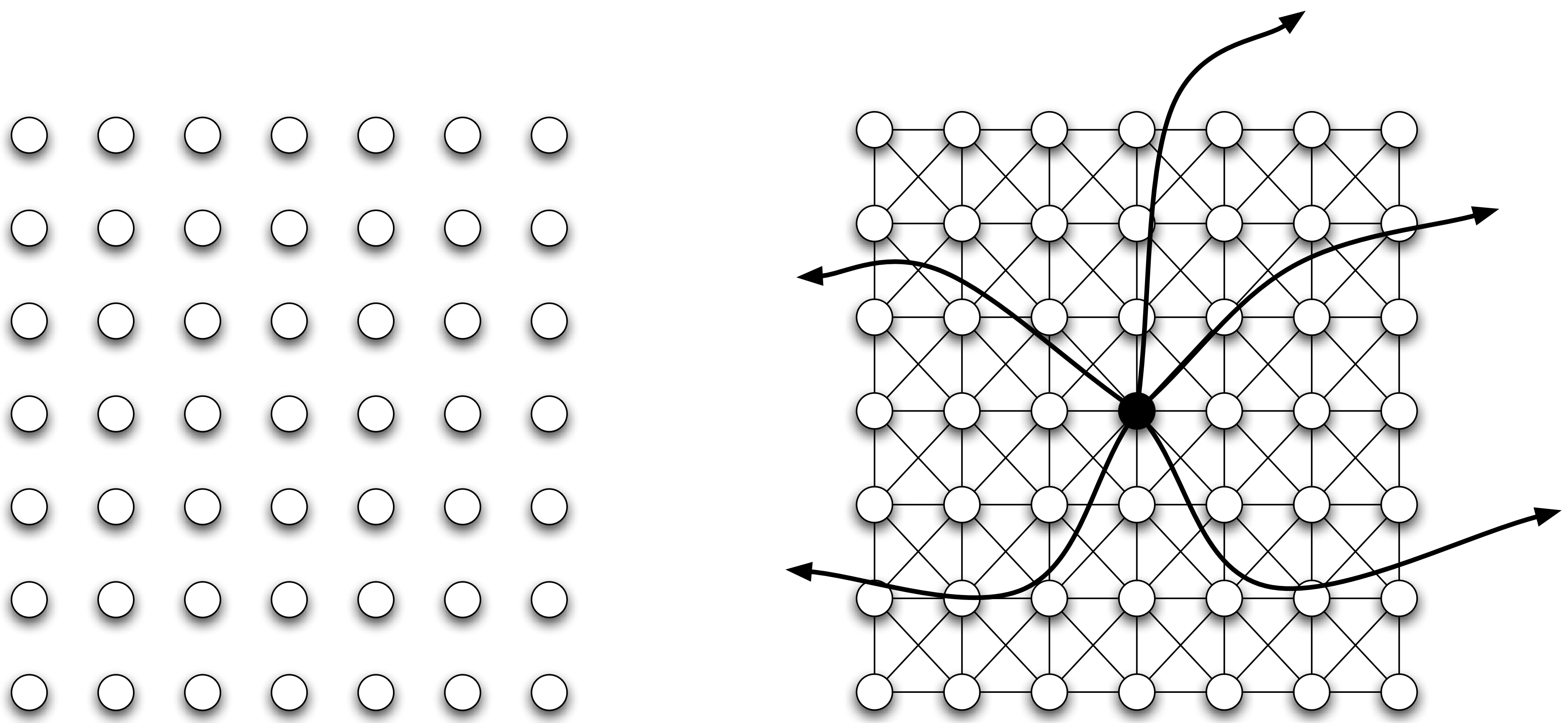


FIG. 11 (a) A one-dimensional lattice with connections between all vertex pairs separated by k or fewer lattice spacing, with $k = 3$ in this case. (b) The small-world model [412, 416] is created by choosing at random a fraction p of the edges in the graph and moving one end of each to a new location, also chosen uniformly at random. (c) A slight variation on the model [289, 324] in which shortcuts are added randomly between vertices, but no edges are removed from the underlying one-dimensional lattice.

Newman

Small-World Models (Watts-Strogatz)



(a) *Nodes arranged in a grid* (b) *A network built from local structure and random edges*

Figure 20.2: The Watts-Strogatz model arises from a highly clustered network (such as the grid), with a small number of random links added in.

Easley & Kleinberg

Concept is applicable to any regular lattice (1D, 2D, 3D, etc.)

Small-World Models (Watts-Strogatz)

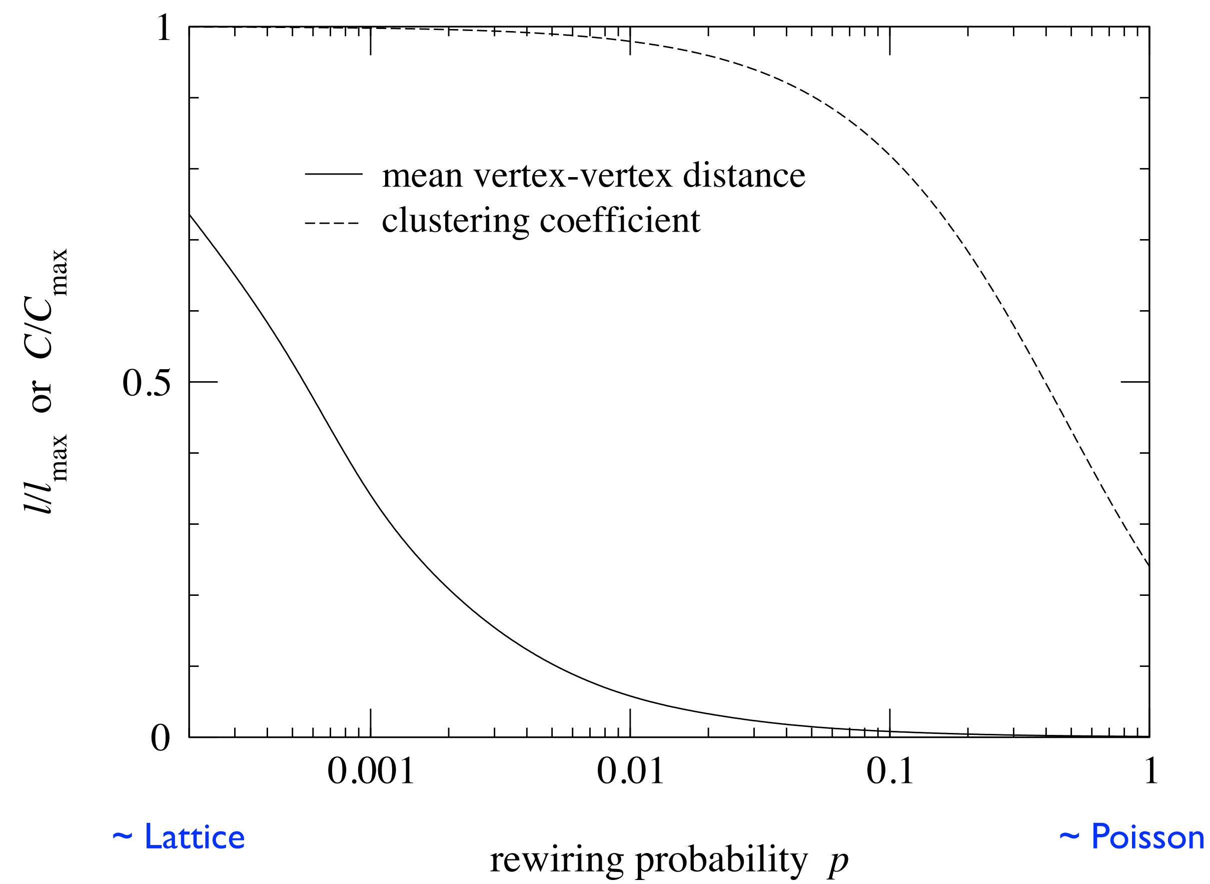


FIG. 12 The clustering coefficient C and mean vertex–vertex distance ℓ in the small-world model of Watts and Strogatz [416] as a function of the rewiring probability p. For convenience, both C and ℓ are divided by their maximum values, which they assume when $p = 0$. Between the extremes $p = 0$ and $p = 1$, there is a region in which clustering is high and mean vertex–vertex distance is simultaneously low.

Newman