University of Stuttgart
Germany

Complex Network Systems

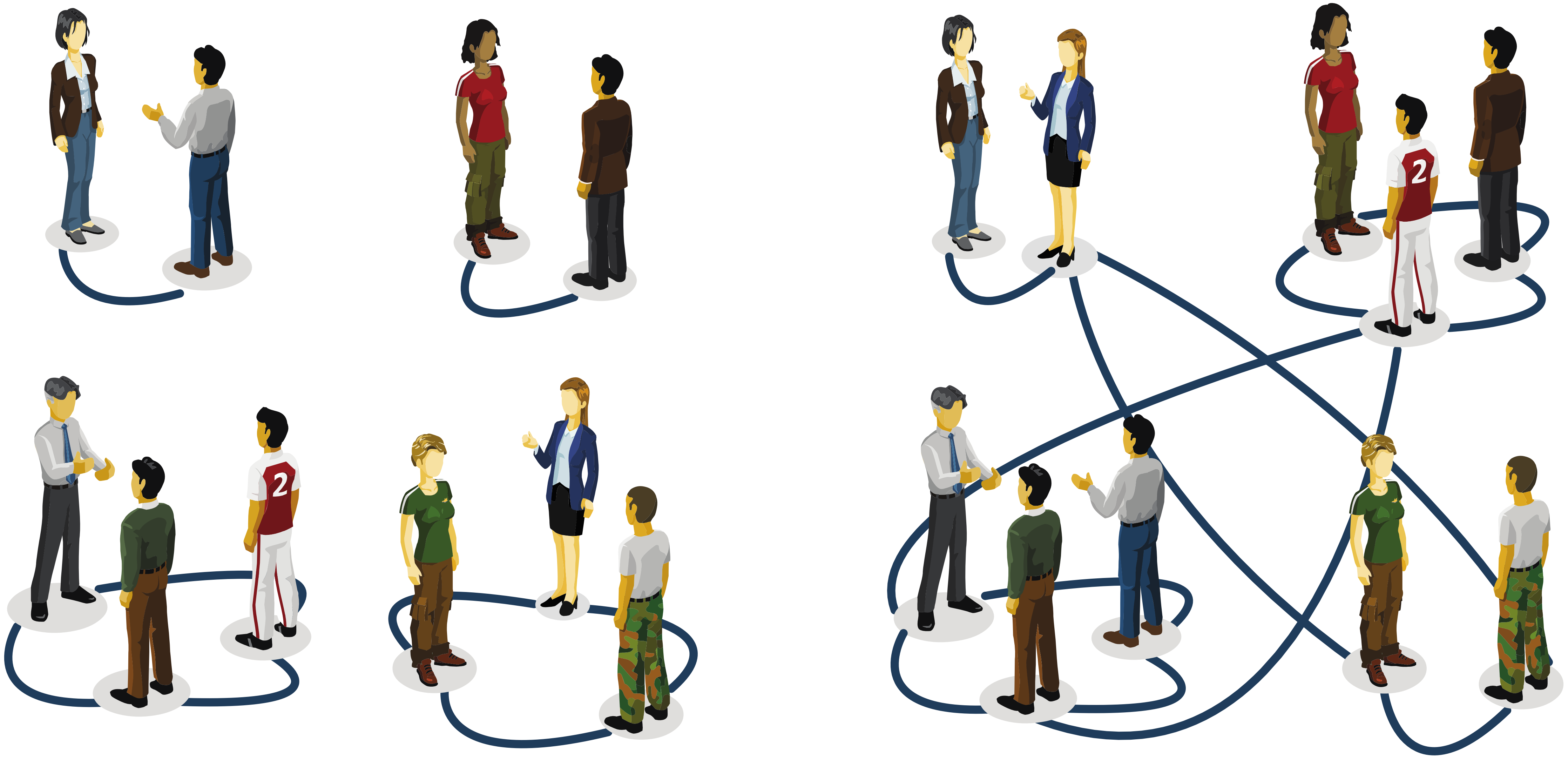
Random graph model

Ilche Georgievski
ilche.georgievski@iaas.uni-stuttgart.de
Room: U38 0.353

2019/2020
Winter

Why should we use network models?

A.-L. Barabási. *Linked: The new science of networks*. Plume Books, 2003.



Can we expect that there would be fine wine left once the guests are gone?

What is a random graph model?

Erdős and Rényi random graph

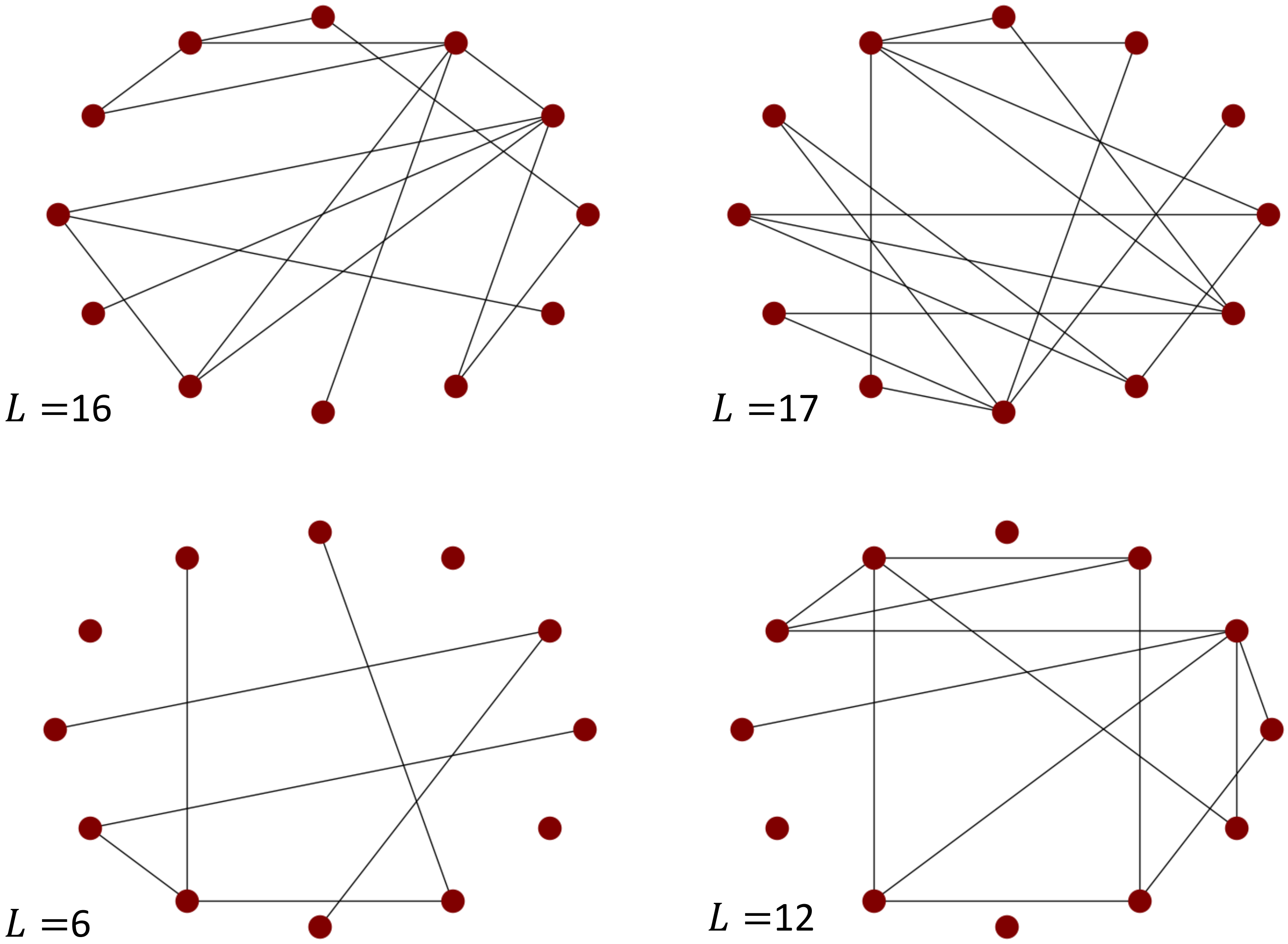
- $G(n, m)$: undirected graph with n nodes and randomly chosen m edges among them
- $G(n, p)$: take the complete graph and associate a unique uniform probability p of existence to all edges

Random graph model procedure

- Start with n isolated nodes
- Select a node pair and generate a random number between 0 and 1
  - If the number exceeds p, connect the selected node pair with a link
  - Otherwise, leave them disconnected
- Repeat the second step for each of the $n(n-1)/2$ node pairs

Erdős-Rényi graph

$G(n, p)$



Do n and p uniquely determine the graph?

$$n = 12$$

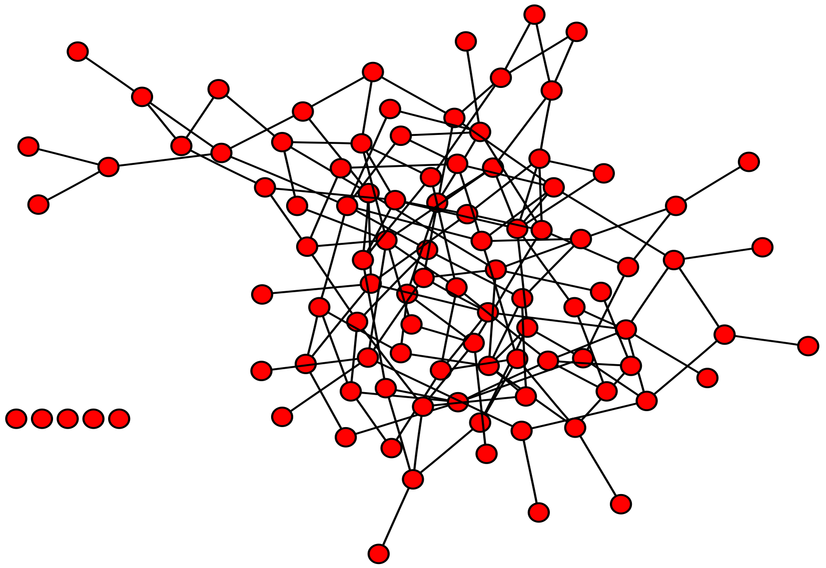
$$p = \frac{1}{6}$$

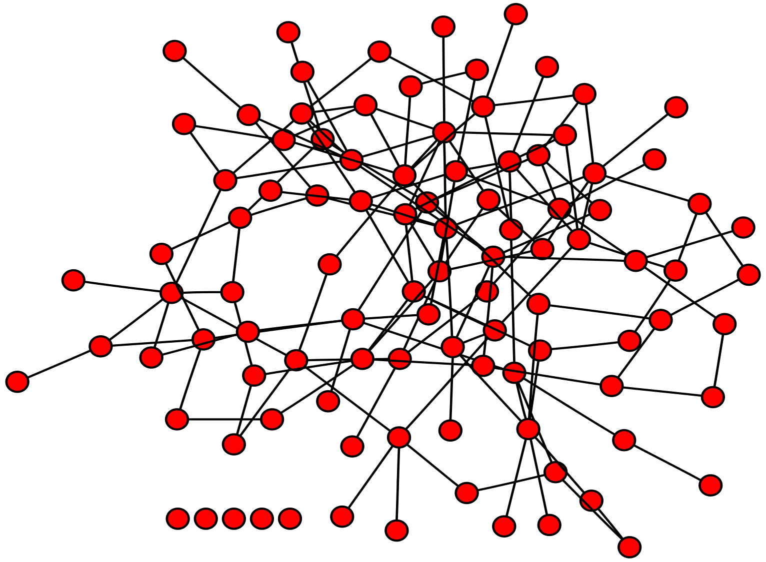
$$\bar{k} = p(n-1)$$

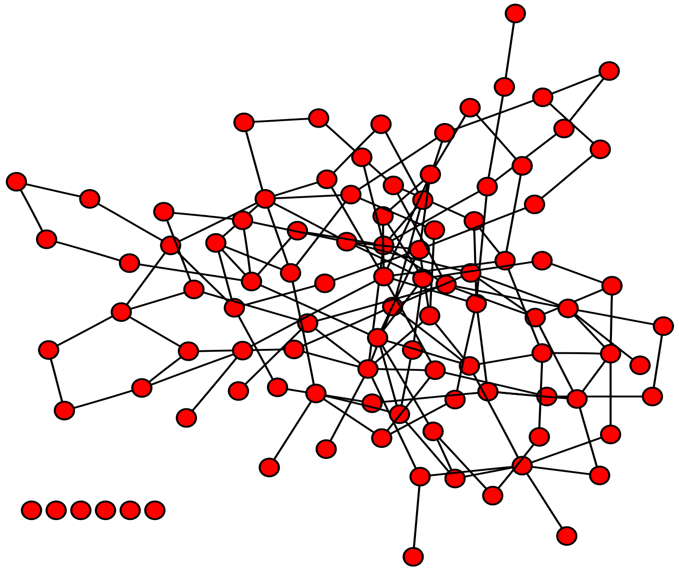
$$\bar{k} = 1.83$$

Erdős-Rényi graph

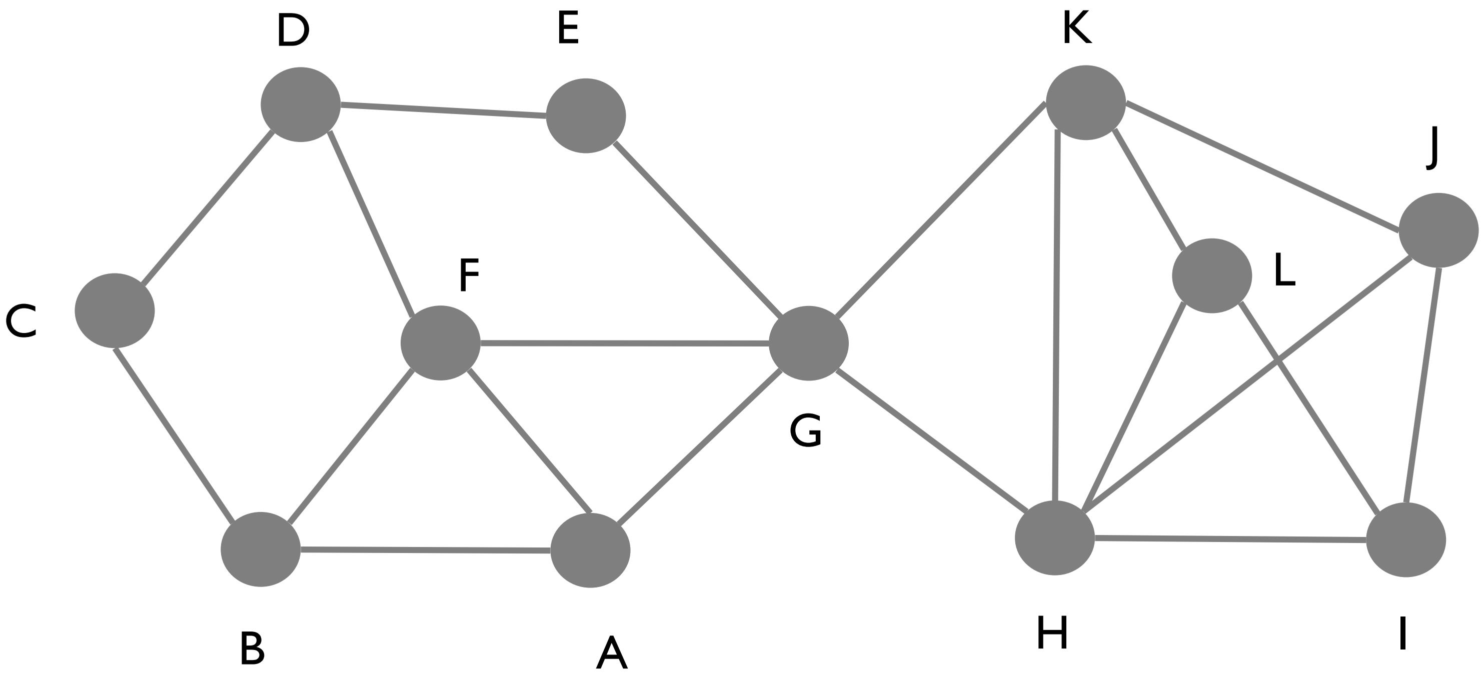
$p = 0.03$
$n = 100$







Degree sequence



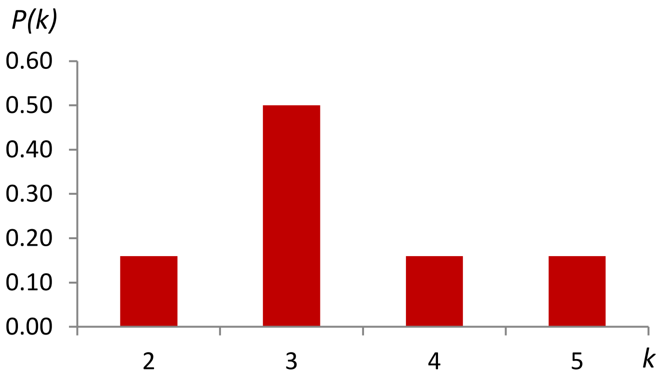
What is the degree sequence?

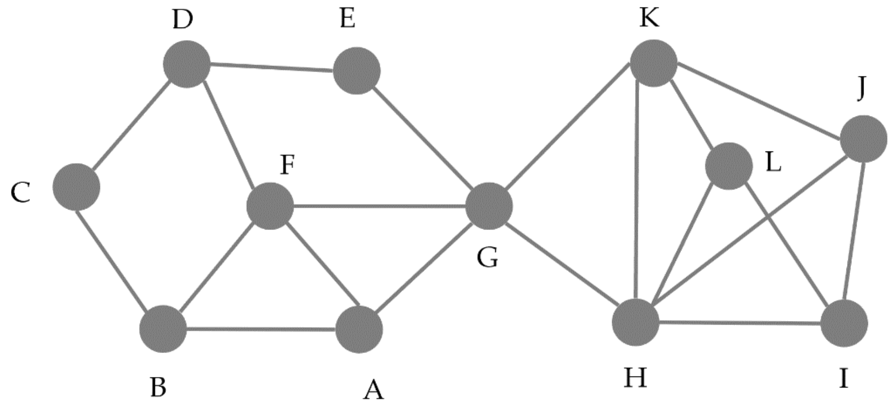
{2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 5, 5}

$C_2 = 2$
$C_3 = 6$
$C_4 = 2$
$C_5 = 2$

Does a degree sequence uniquely specify a graph?

Degree distribution

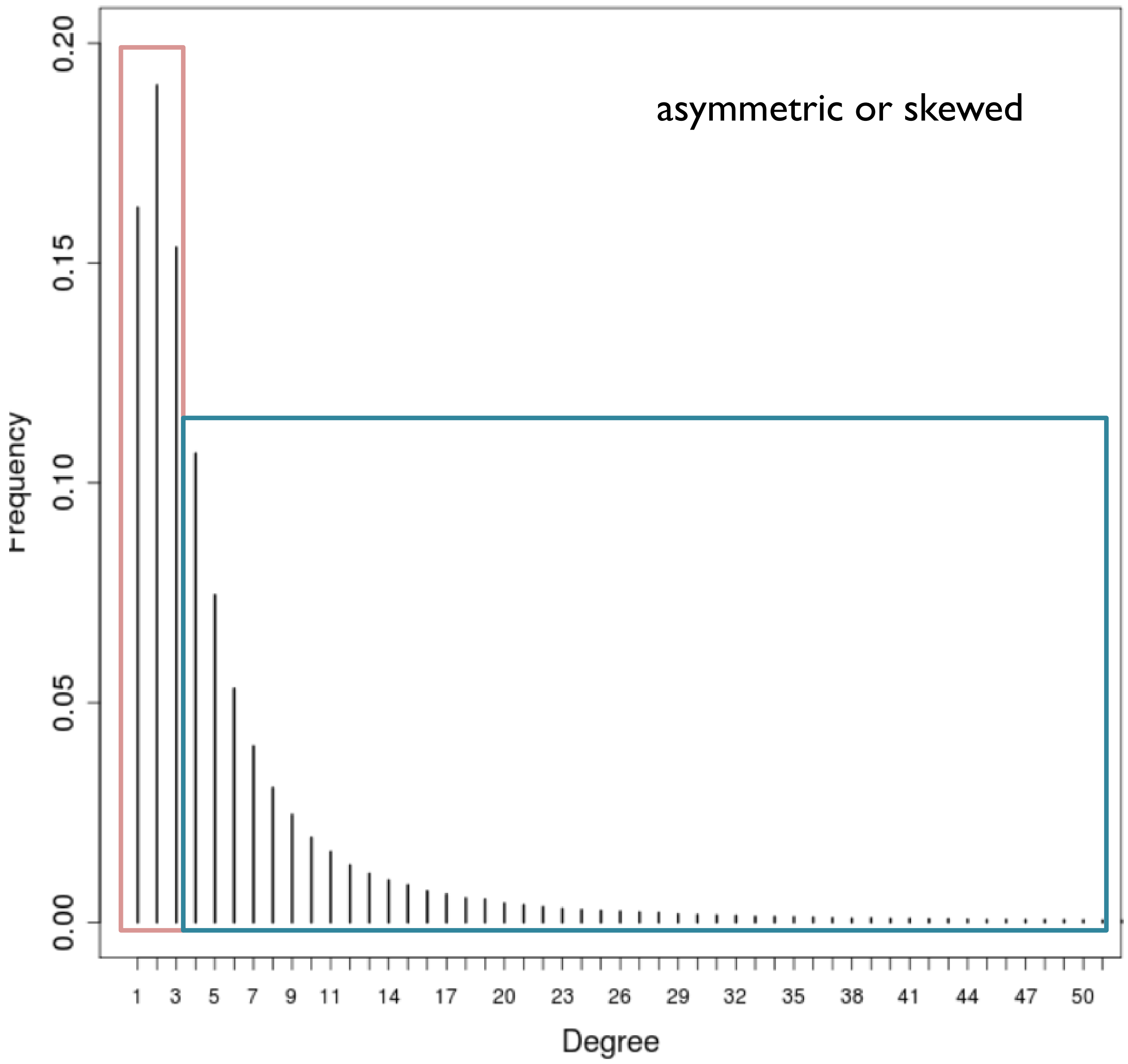




$$p(k) = \frac{C_k}{n}$$

$$\{\frac{1}{6}, \frac{1}{2}, \frac{1}{6}, \frac{1}{6}\}$$ ← degree distribution

Degree distribution in real networks



long tail

Degree distribution of $G(n, p)$

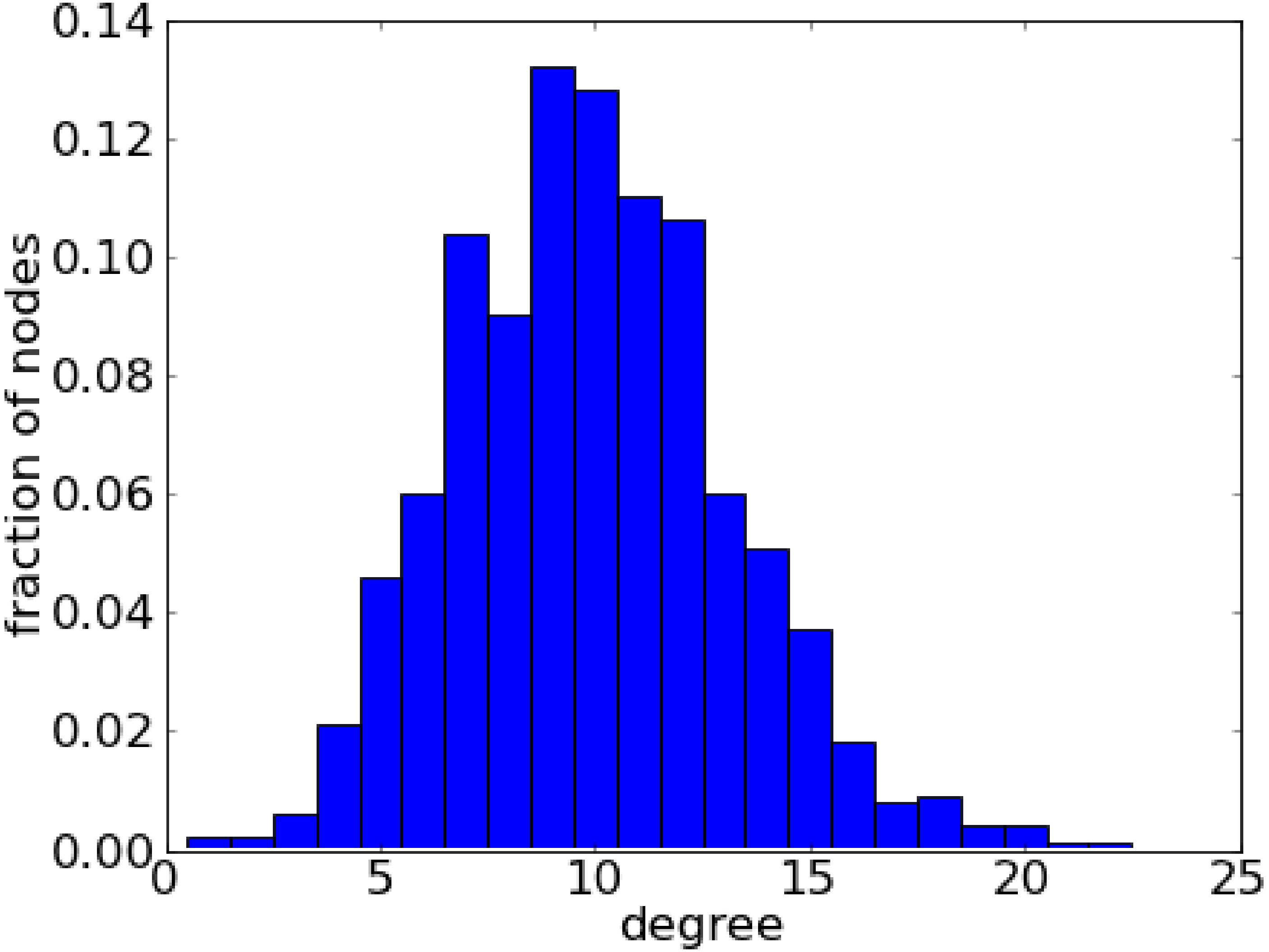
probability of
having k edges

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

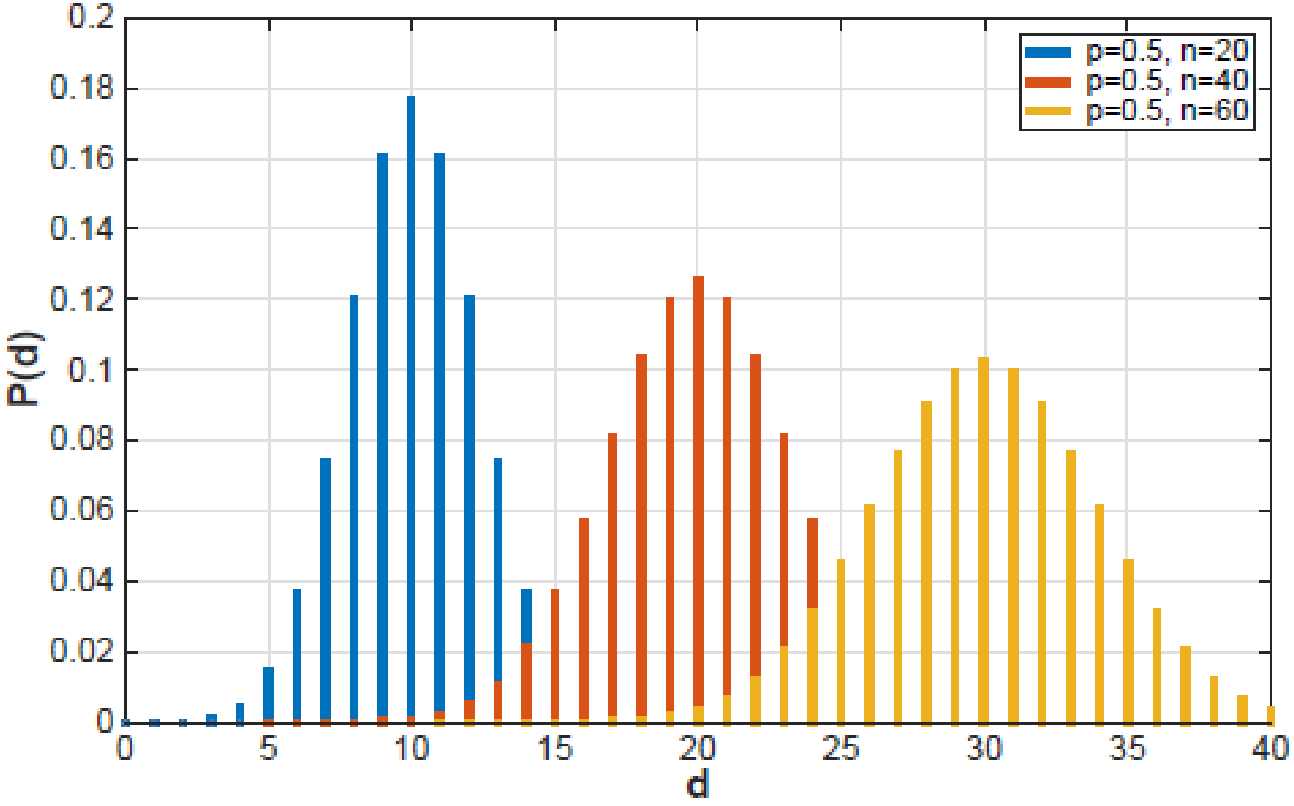
select k nodes
out of $n-1$

probability of
missing the rest of
$n-1-k$ edges

Degree distribution of $G(n, p)$



Degree distribution of $G(n, p)$



Degree distribution of $G(n, p)$

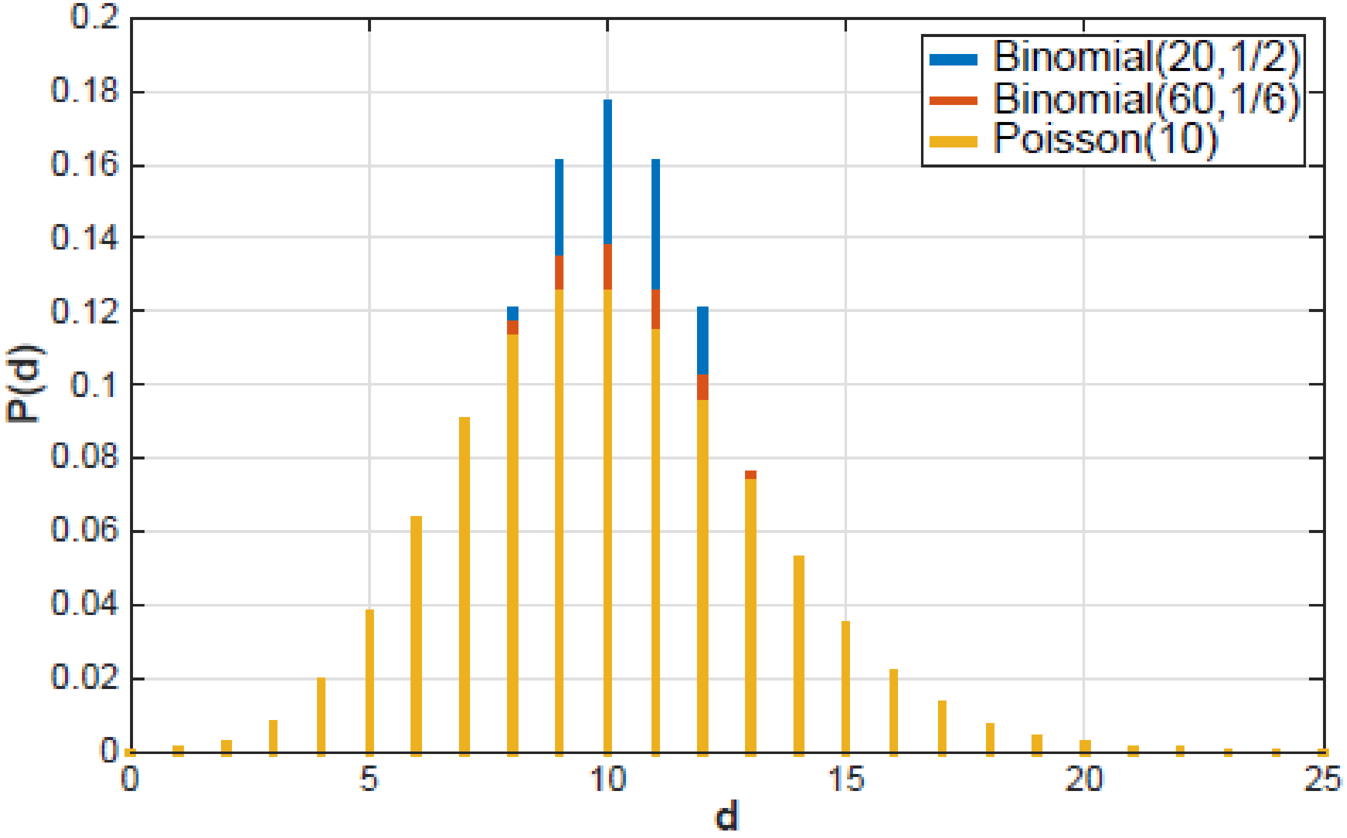
$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Binomial distribution

$$P(k) = e^{-\bar{k}} \frac{\bar{k}^k}{k!}$$

Poisson distribution

Degree distribution of $G(n, p)$



How big are the differences between the node degrees in a particular realisation of a random network?

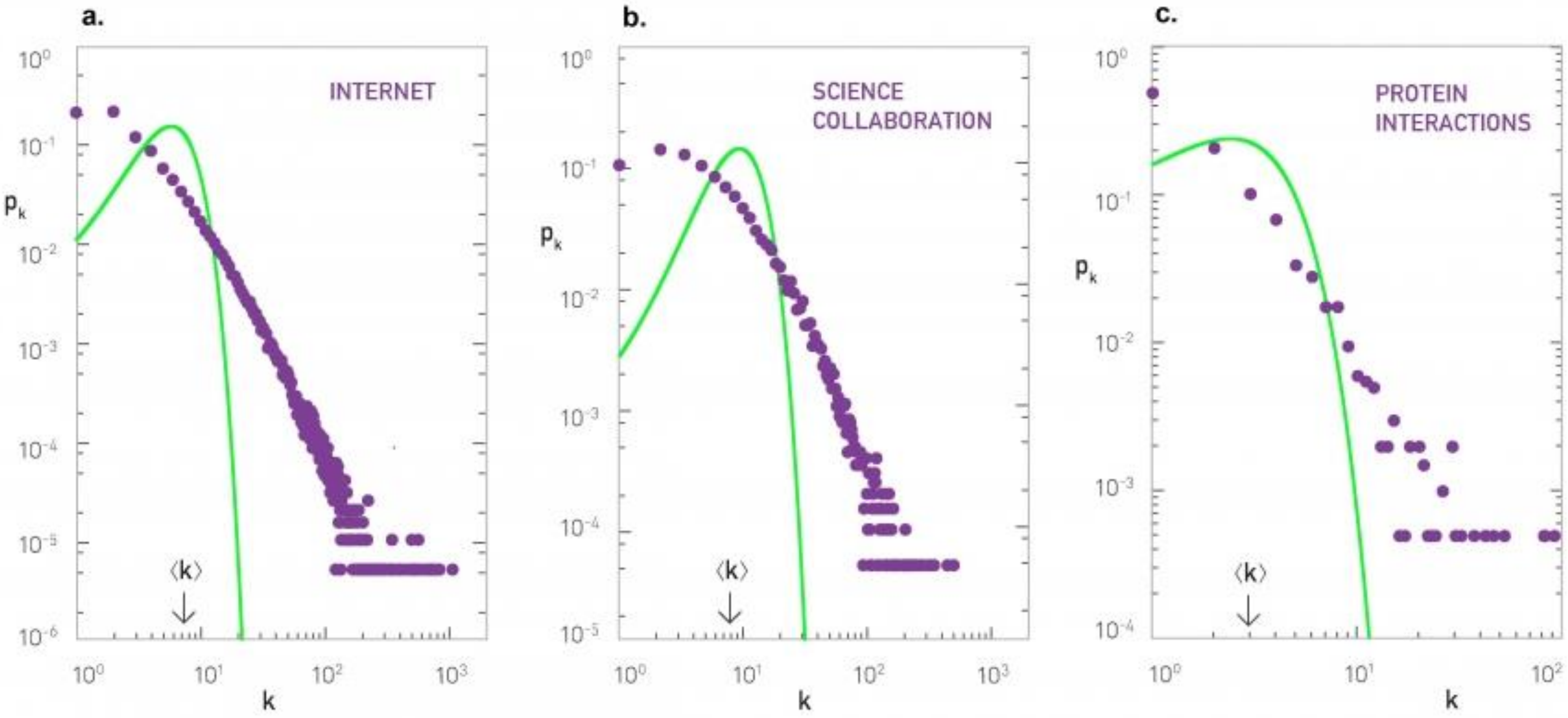
Can high degree nodes coexist with small degree nodes?

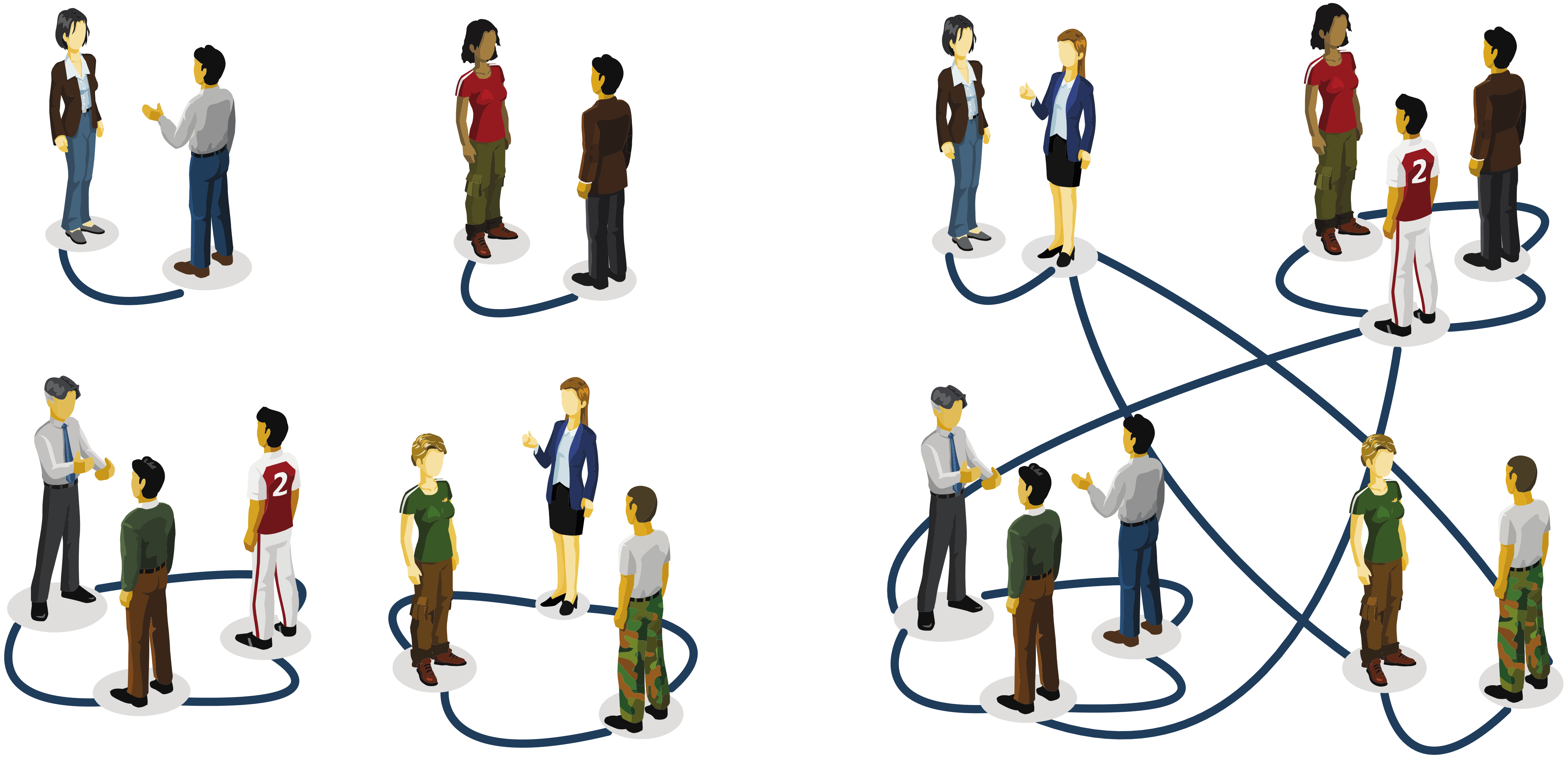
Social network as a random network

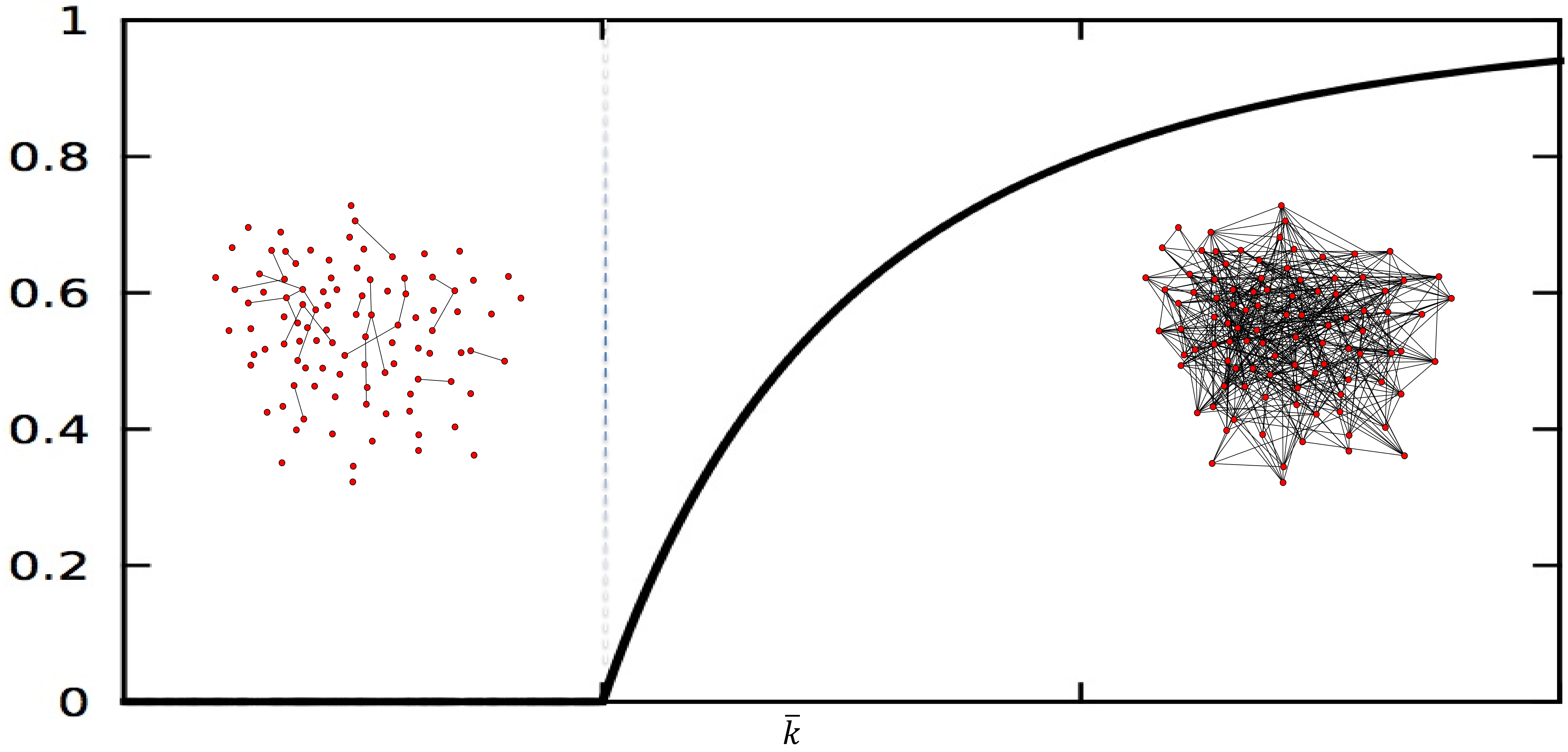
- Typical person knows about 1000 individuals on a first name basis
  - $\bar{k} \approx 1000$
  - $n \approx 7 \times 10^9$
  - $k_{max} = 1185$
  - $k_{min} = 816$
  - $\sigma_k = 31.62$ for $\bar{k} = 1000$
- Typical person has between 968 and 1032 friends
- All individuals are expected to have a comparable number of friends

In a large random network, the degree of most nodes is in the narrow vicinity of $\bar{k}$

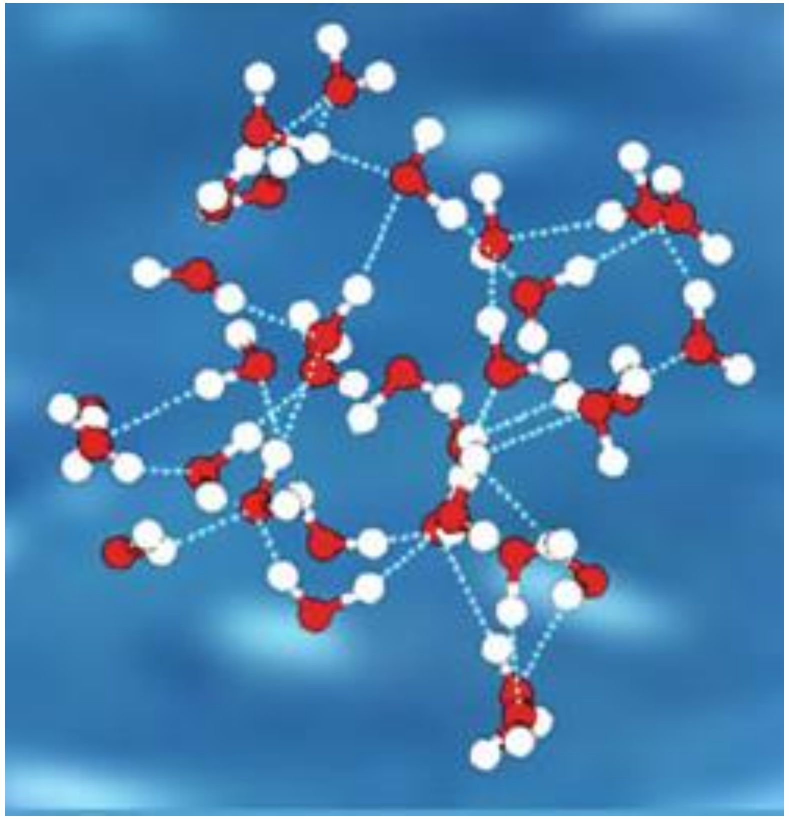
Real networks are not Poisson



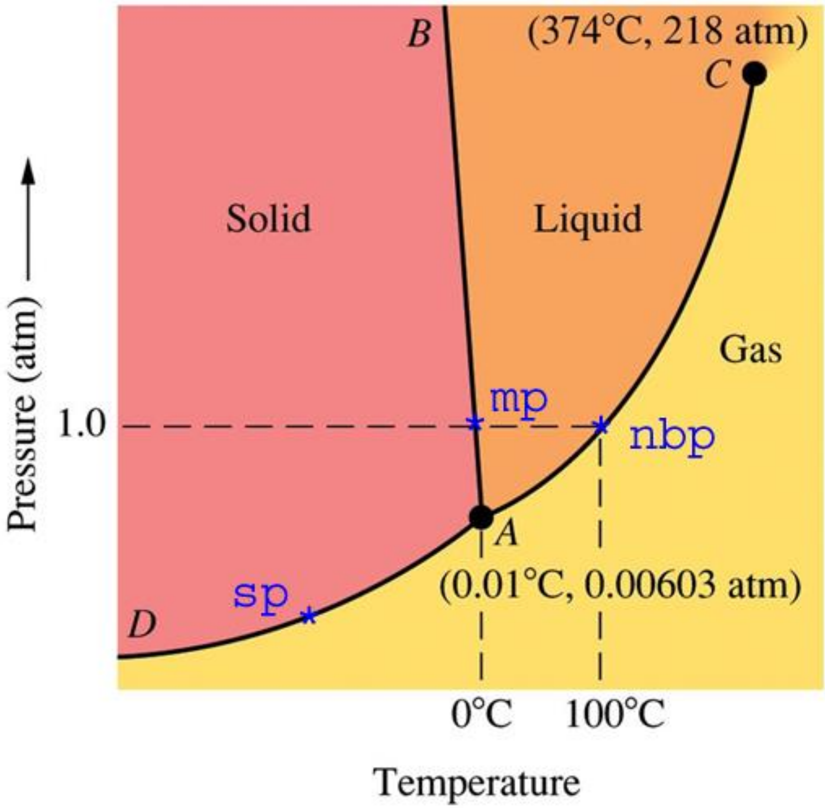




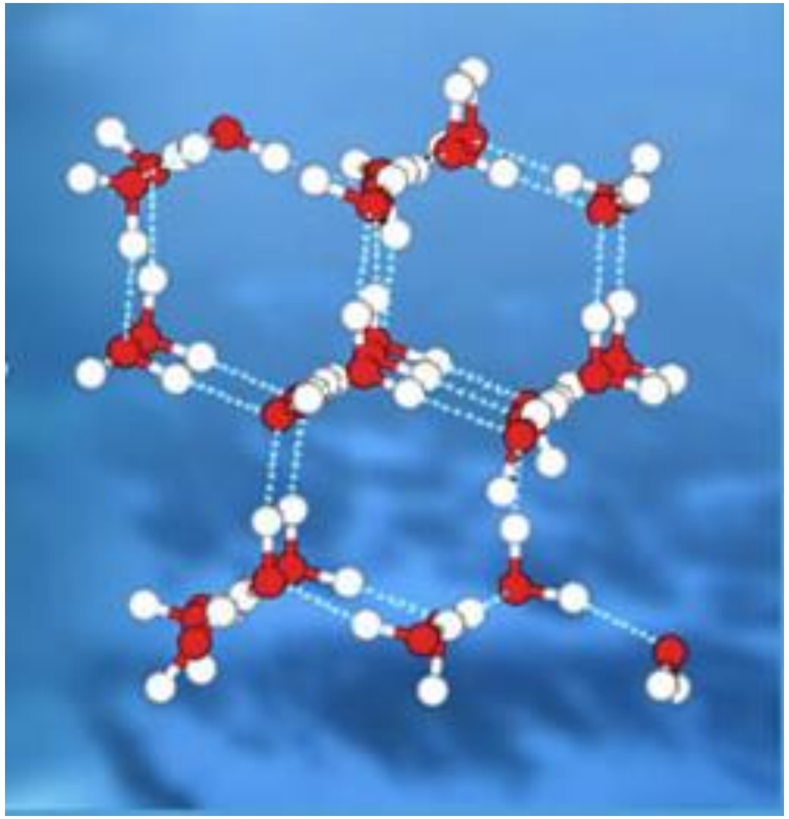
How does this transition happen?



Water

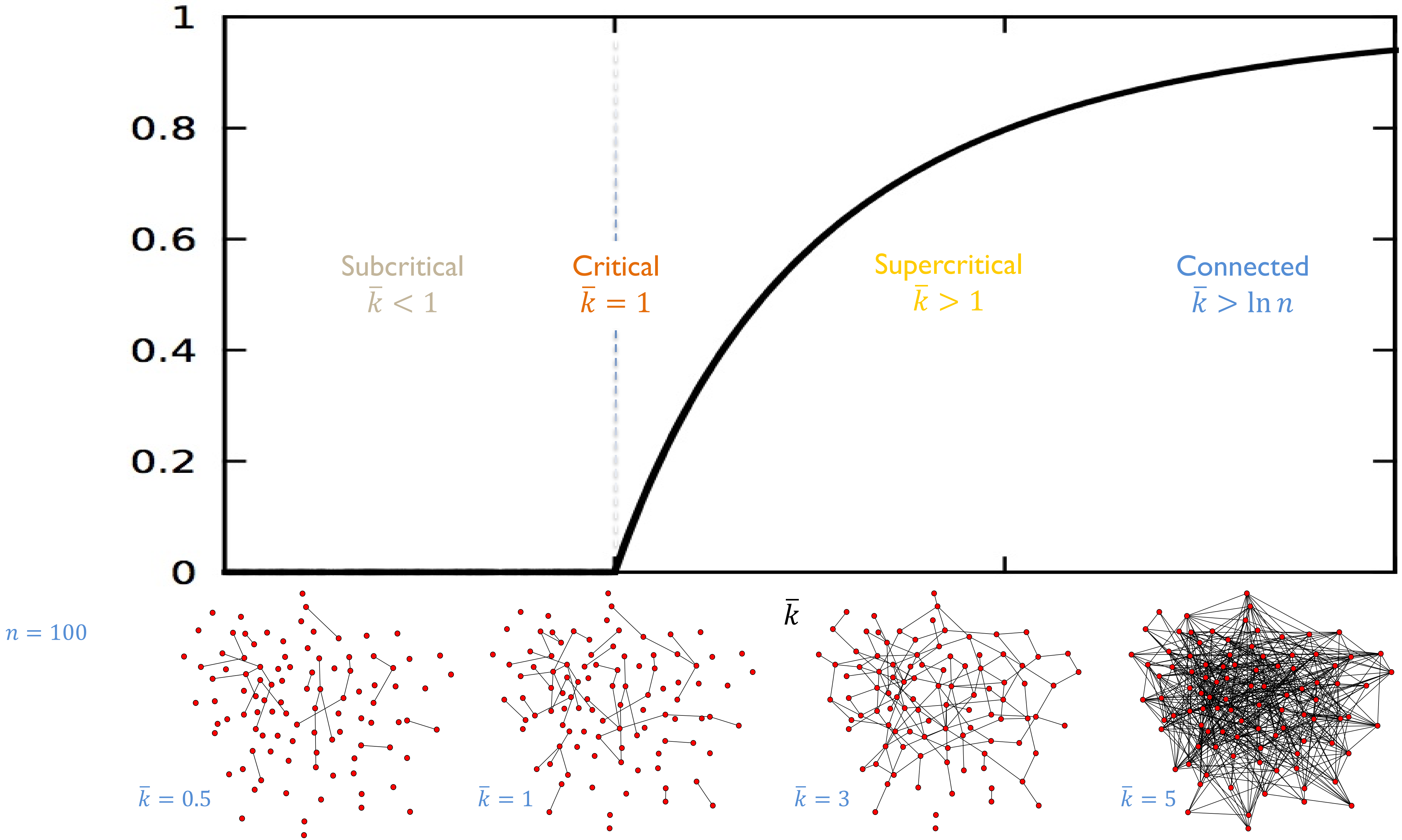


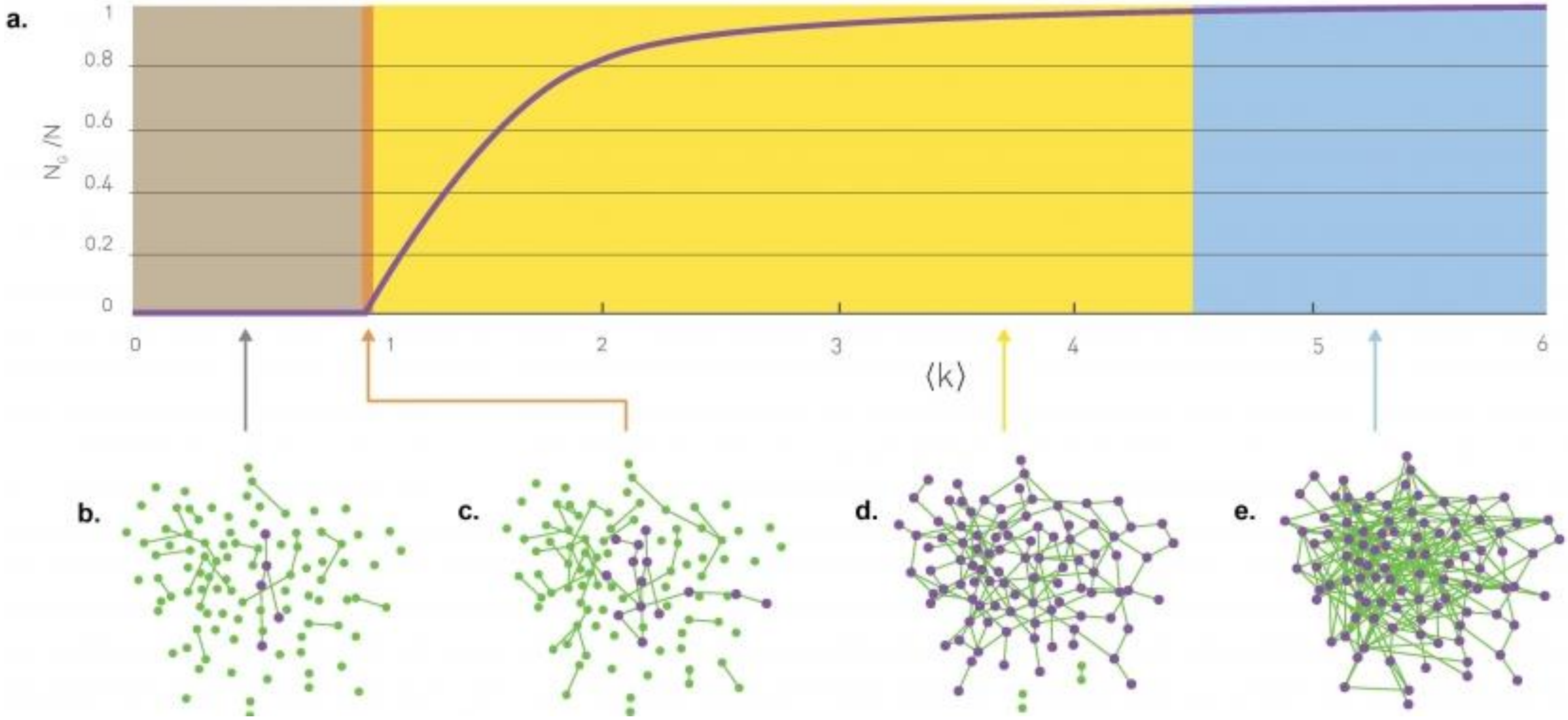
Solid
Liquid
Gas
B
(374°C, 218 atm)
C
mp
nbp
1.0
Pressure (atm)
A
(0.01°C, 0.00603 atm)
sp
D
0°C
100°C
Temperature



Ice

Evolution of $G(n,p)$


1
0.8
0.6
0.4
0.2
0
Subcritical
$\bar{k} < 1$
Critical
$\bar{k} = 1$
Supercritical
$\bar{k} > 1$
Connected
$\bar{k} > \ln n$
$\bar{k}$
$n = 100$
$\bar{k} = 0.5$
$\bar{k} = 1$
$\bar{k} = 3$
$\bar{k} = 5$



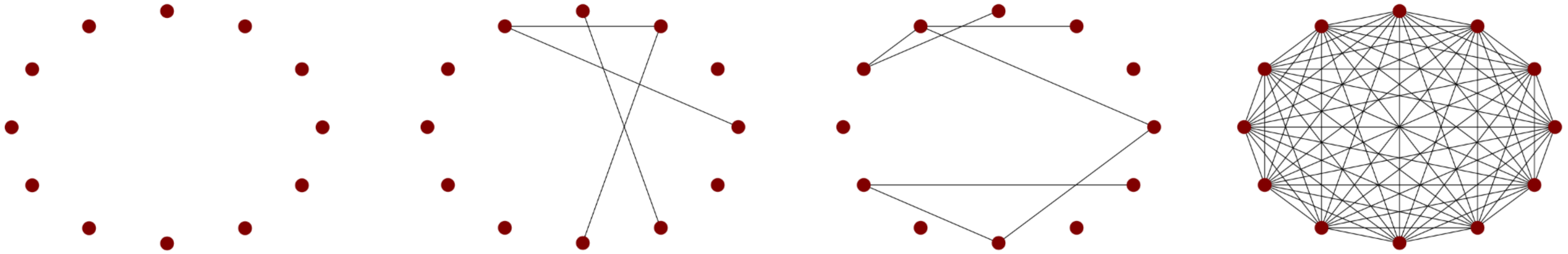
No giant component
Size of the largest cluster: $\sim \ln n$
Clusters are trees

No giant component
Size of the largest cluster: $\sim n^{2/3}$
Clusters may contain loops

Unique giant component
Size of the giant component: $\sim (p - p_c)n$
Small clusters are trees
Giant component has loops

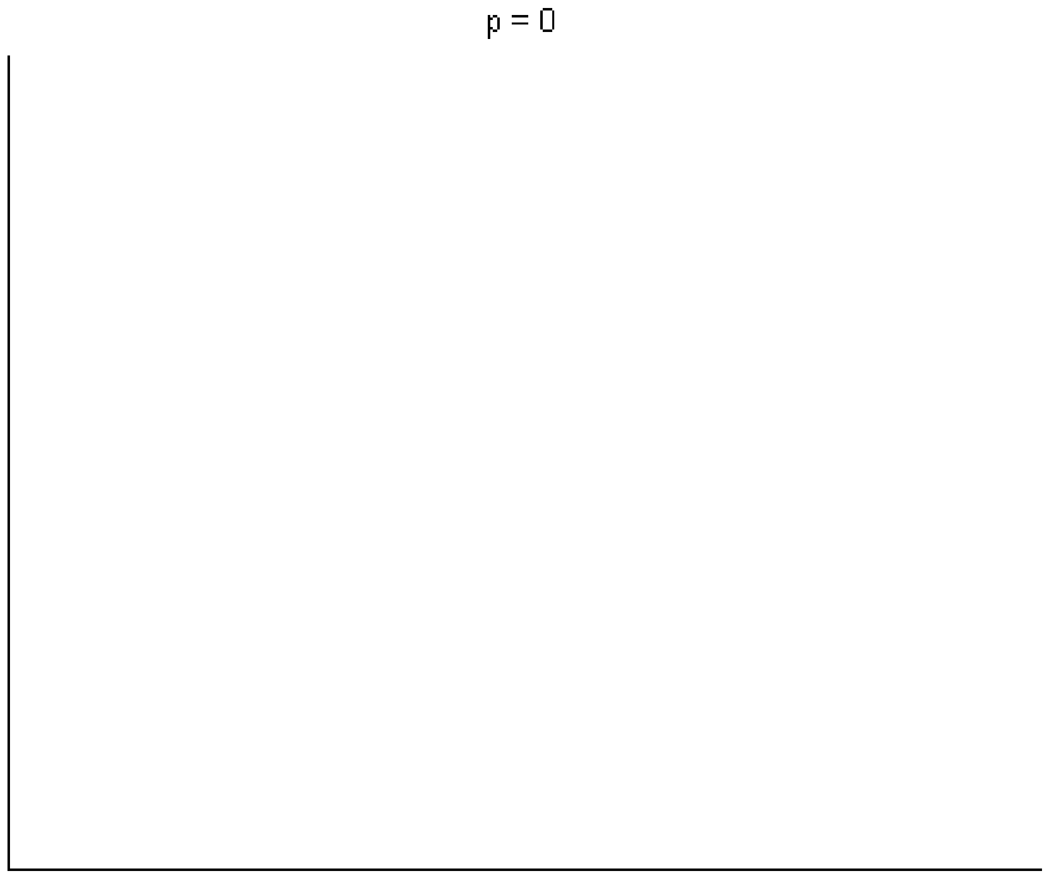
Unique giant component
Size of the giant component: n
No isolated nodes or clusters
Giant component has loops

Evolution of $G(n, p)$

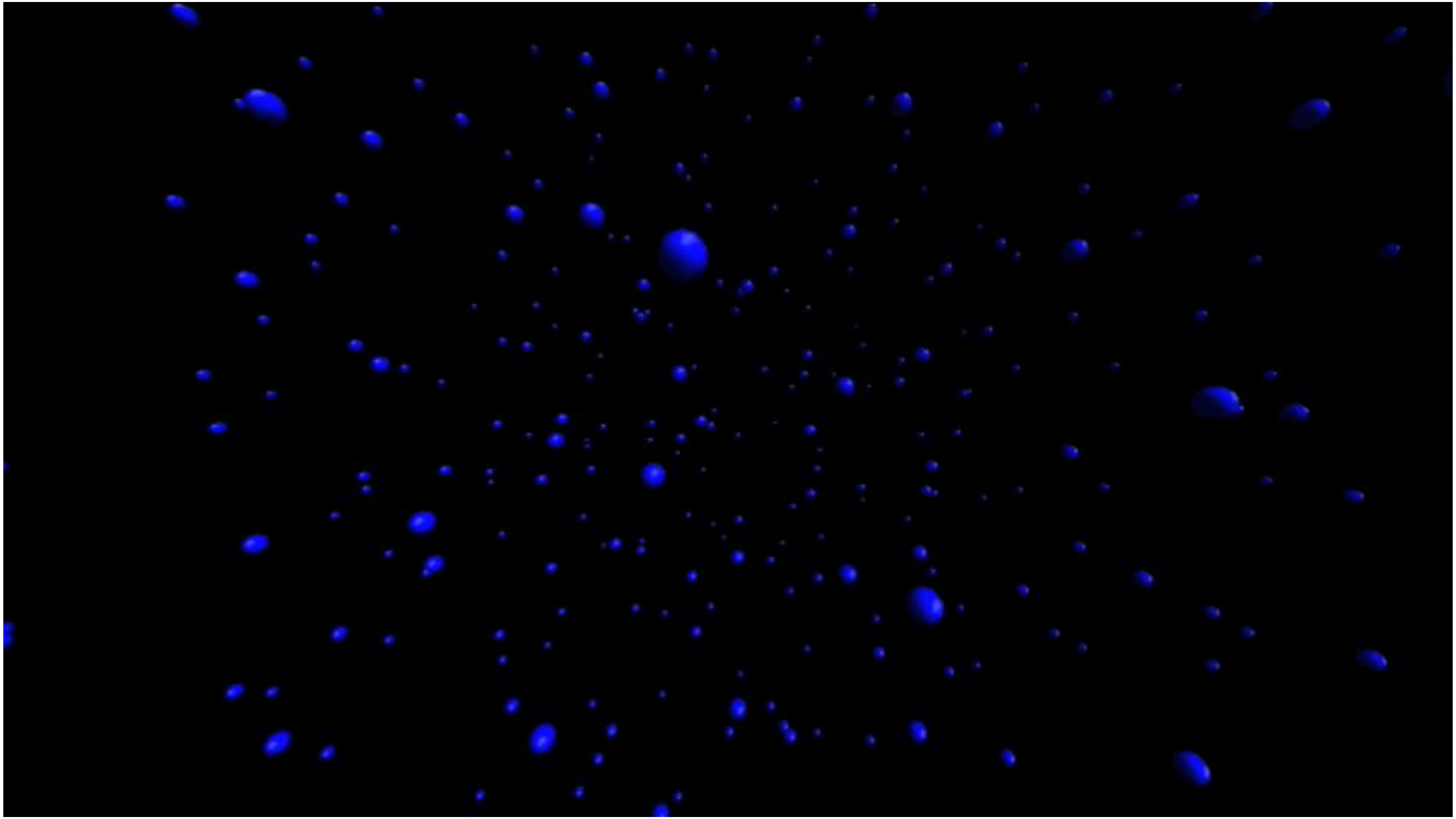


Probability (p)	0.0	0.045	0.095	1.0
Average degree	0.0	0.6667	1.1667	11.0
Diameter	0	3	6	1
Giant component size	0	3	7	66
Average shortest path length	0.0	1.6667	2.7142	1.0

Evolution of $G(n, p)$



From David Gleich, Purdue University



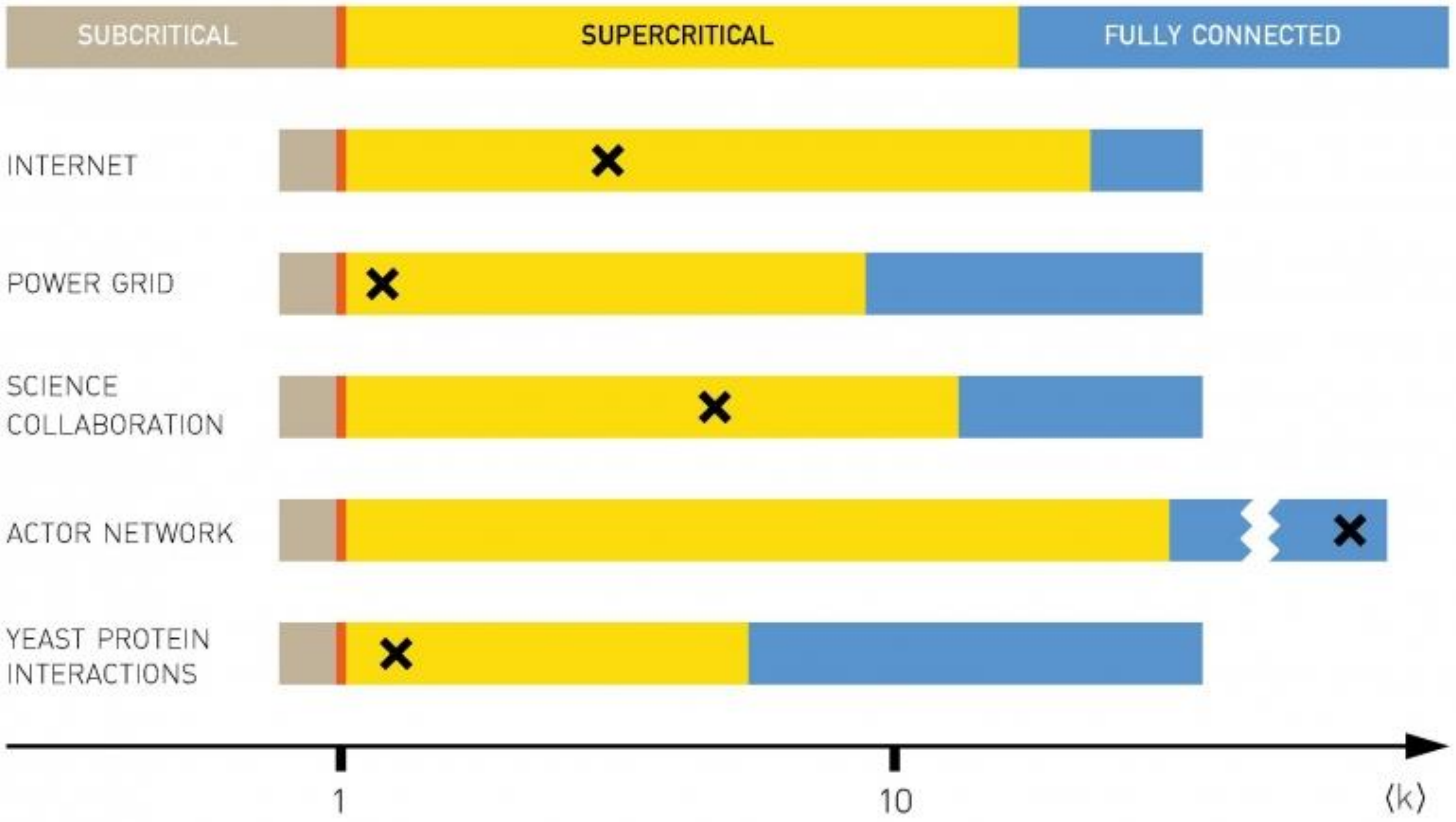
Do real networks satisfy the criteria for the existence of a giant component?

Will this giant component contain all nodes for $\bar{k} > \ln n$, or will there still be some disconnected nodes and components?

Real networks are supercritical

Network	n	L	$\bar{k}$	$\ln n$
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,437	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	2.90	7.61

Real networks are supercritical



Clustering coefficient of $G(n,p)$

Local clustering coefficient

Global clustering coefficient

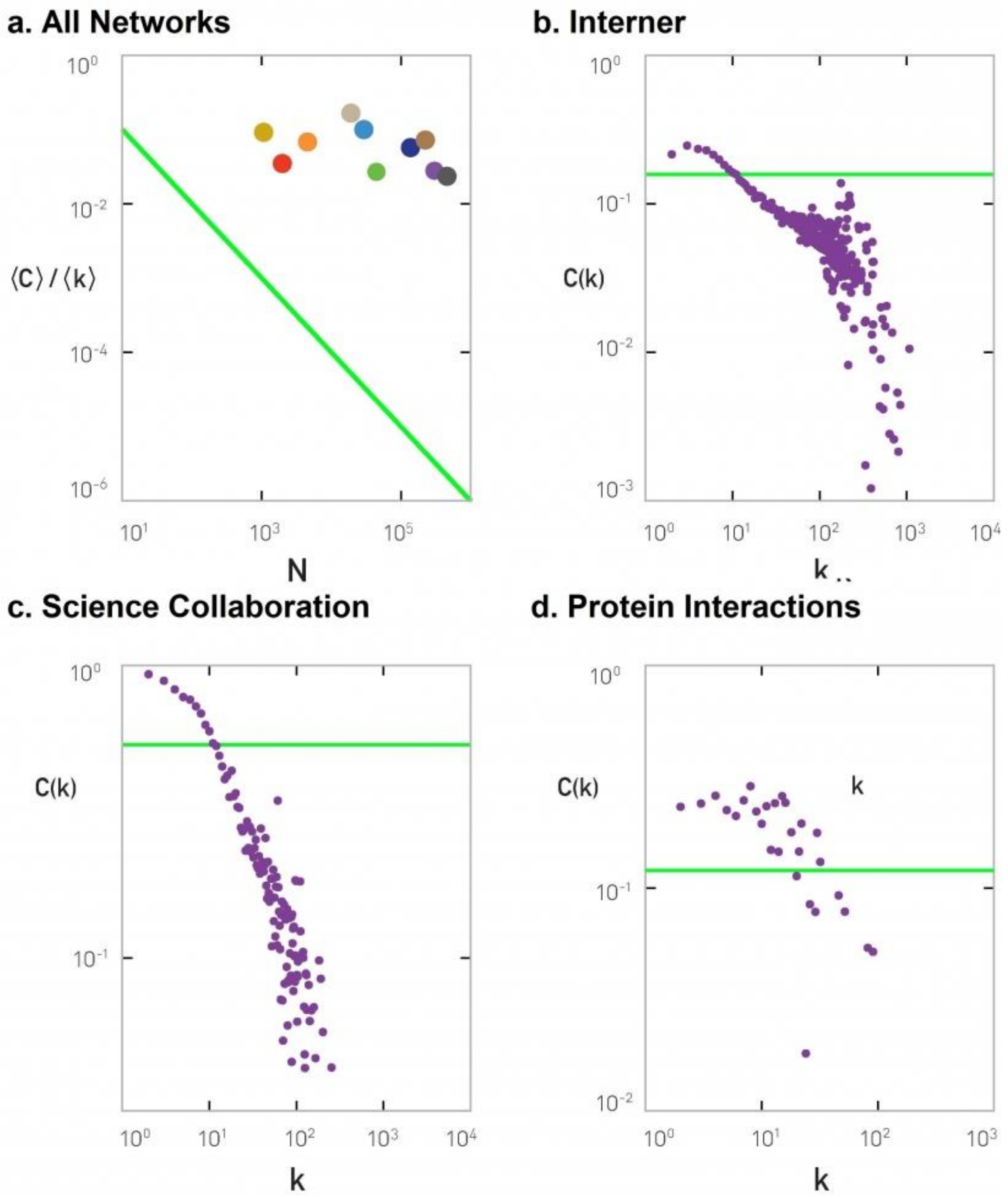
$E(C_i) = p$

$C = p$

expected

generation probability

Is this mirroring the clustering coefficient of real networks?


a. All Networks
b. Interner
c. Science Collaboration
d. Protein Interactions
⟨C⟩ / ⟨k⟩
N
C(k)
k
10^0
10^{-2}
10^{-4}
10^{-6}
10^1
10^3
10^5
10^{-1}
10^{-3}
10^2
10^4

Summary

- Real networks are not random
- Degree to which random networks describe or not real systems can be decided using
  - Degree distribution
    - Random networks have binomial distribution
    - Poisson distribution does not capture degree distribution of real networks
  - Connectedness
    - Giant component for $\bar{k} = 1$
    - Most real networks are not fragmented
  - Average path length
    - Accounts for the emergence of the small-world phenomenon
  - Clustering coefficient
    - In random networks, independent of a node's degree and depends on the system size
    - In real networks, it decreases with a node's degree and is largely independent of the system size
- Small-world phenomenon is the only property reasonably explained by the random network model

If real networks are not random, why do we study the random network model?

Exercise

Create three random networks ($G(n,p)$) with $n = 250$ using NetworkX.

a) Plot the networks
b) Give a table with measurements of the following properties:
   - Average degree
   - Average shortest path length
   - Number of connected components
   - Clustering coefficient
c) For which probability is the average degree ~ 1? What is the size of the Giant component at the phase transition?

Sources

- Leskovec, J. Analysis of Networks, CS224W, Stanford University (2018), http://web.stanford.edu/class/cs224w/
- Mateos, G. Degrees, Power Laws and Popularity, University of Rochester, 2018.
- Zafarani, R., Abbasi, M. A. and Liu, H. *Social Media Mining: An Introduction*, Cambridge University Press, 2014.
- Barabási, A. Network Science, http://networksciencebook.com