Random Graph Theory

Dr. Natarajan Meghanathan
Professor
Department of Computer Science
Jackson State University, Jackson, MS
E-mail: natarajan.meghanathan@jsums.edu

Introduction

- At first inspection, most real-world networks look as if they are spun randomly.
- To model such networks that are *truly random*, the principle behind "Random Graph Theory" is:
  - Place the links randomly between nodes to reproduce the complexity and apparent randomness of real-world systems.

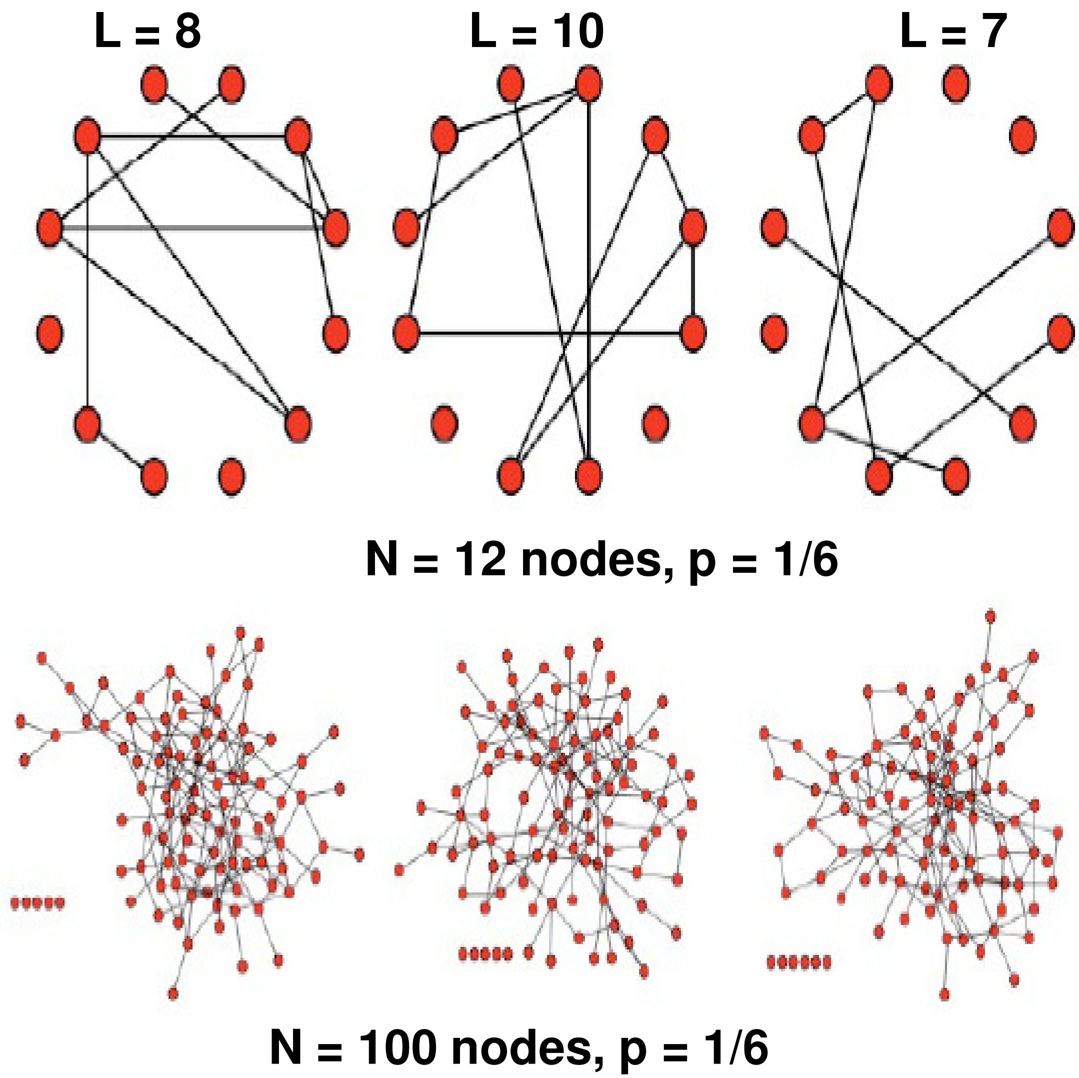
- Two definitions of random networks
  - *G*(*N*, *L*) model: *N* labeled nodes are connected with *L* randomly placed links
  - *G*(*N*, *p*) model: Each pair of *N* labeled nodes are connected with a probability *p*.

- Though the average degree for a node is simply 2L/N in a G(N, L) model, the other key network characteristics are easier to calculate in the G(N, p) model.
  - The construction of the G(N, p) model is closer to the way real systems develop. The total number of links in a network is rarely fixed.

Constructing a G(N, p) Network

- Step 1: Start with N isolated nodes
- Step 2: For a particular node pair (u, v), generate a random number *r*. If $r \leq p$, then, add the link (u, v) to the network.
- Repeat Step 2 for each of the N(N-1)/2 node pairs.

- Each random network we generate with the same parameters (N, p) will look slightly different.
  – The number of links L is likely to be different.



Source: Figure 3.3a: Barabasi

Review of Binomial Distribution

- Let there be N independent experiments with two possible outcomes (in each experiment: success or failure): with the probability of one outcome (say success) is *p* and of the other is 1-*p*.
- The binomial distribution provides the probability p_x that we obtain exactly x successes in a sequence of N experiments.

The binomial distribution has the form $$p_x = \binom{N}{x} p^x (1-p)^{N-x}.$$	Its second moment is $$\langle x^2 \rangle = \sum_{x=0}^{N} x^2 p_x = p(1-p)N + p^2N^2,$$
The mean of the distribution (first moment) is $$\langle x \rangle = \sum_{x=0}^{N} x p_x = Np.$$	providing its standard deviation as $$\sigma_x = \left(\langle x^2 \rangle - \langle x \rangle^2\right)^{\frac{1}{2}} = [p(1-p)N]^{\frac{1}{2}}.$$

C(N, x) = $\binom{N}{x}$ is the different combinations of the results of the N experiments in which there will be X successes and N-X failures.

Links in a G(N, p) Network

- Let L be the number of links arising out of a random network generated according to the G(N, p) model.
- To determine the **Average Number of Links <L>,** we need to model the probability that there will be exactly L links among the total number of node pairs N(N-1)/2 considered to have a link; each node pair has a probability of *p* to form a link. Let Lmax = N(N-1)/2.

$$p_L = \binom{L\max}{L} p^L (1-p)^{L\max - L}$$

$$< L >= \sum_{L=0}^{L\max} L * p_L = (L\max) * p$$

$$< L >= p * \frac{N(N-1)}{2}$$

Average Degree of a Node <k>

$$< K >= \frac{2* < L >}{N}$$

$$= \frac{2 * p * N(N-1)}{2 * N}$$

$$= p * (N-1)$$

Degree Distribution

- For a **random network** of N nodes, each node can have potentially N-1 links.
- The probability p_k that a node has exactly k links is given by the **binomial distribution**:

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

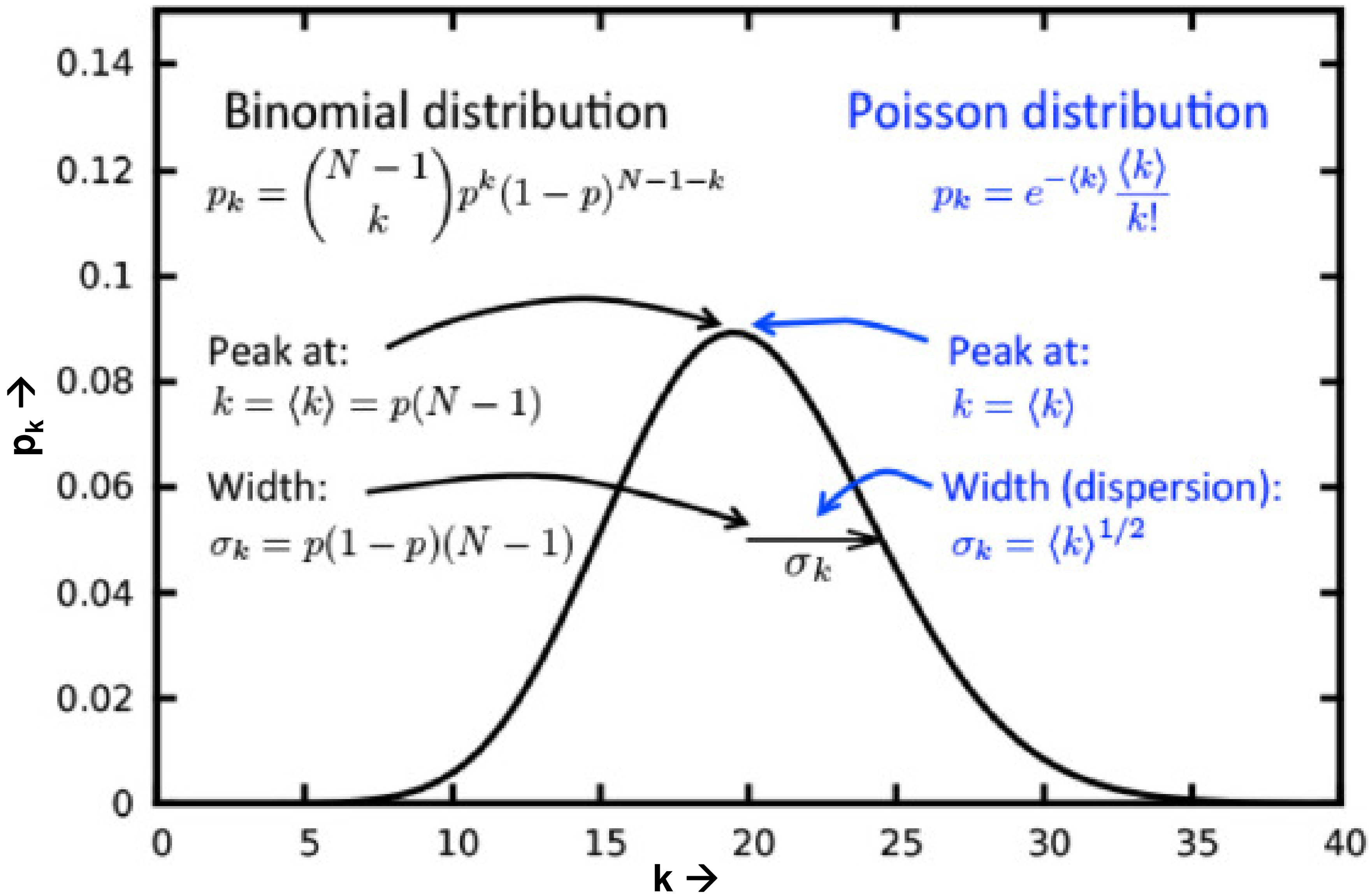
- Using the above binomial distribution to find the average node degree for a random network, we obtain <k> = p*(N-1) and the standard deviation for the node degree is σ_k = p*(1-p)*(N-1).

- **For sparse networks** (for which <k> << N), the probability of finding a node with k neighbors is given by the **Poisson distribution**:

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

- Using the above Poisson distribution to find the average node degree for a random network, we obtain <k> as the mean and the standard deviation for the node degree is $\sigma_k = (\langle k \rangle)^{1/2}$.

Degree Distribution

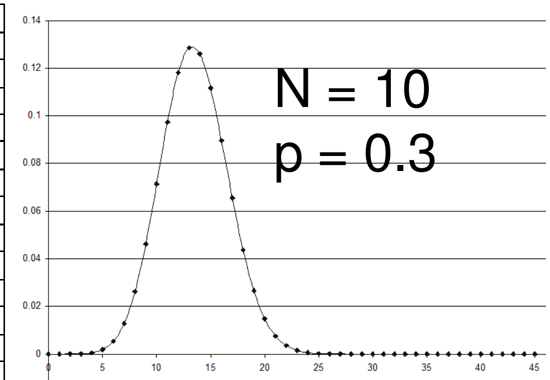


N = 10, p = 0.3

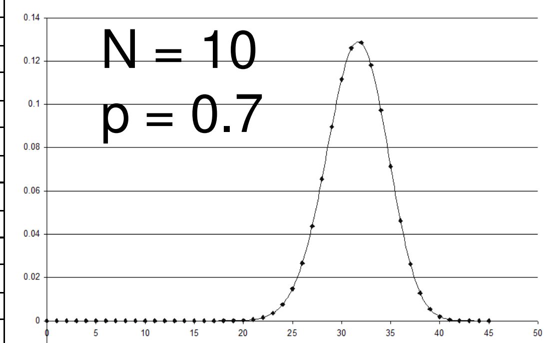
L	C(45, L)	(0.3)^L	(1-0.3)^(45-L)	Prob(L)
0	1	1	1.07007E-07	1.07007E-07
1	45	0.3	1.52867E-07	2.0637E-06
2	990	0.09	2.18381E-07	1.94578E-05
3	14190	0.027	3.11973E-07	0.000119526
4	148995	0.0081	4.45676E-07	0.000537869
5	1221759	0.00243	6.36681E-07	0.001890225
6	8145060	0.000729	9.09544E-07	0.005400642
7	45379620	0.000219	1.29935E-06	0.01289541
8	2.16E+08	6.56E-05	1.85621E-06	0.026251371
9	8.86E+08	1.97E-05	2.65173E-06	0.046252415
10	3.19E+09	5.9E-06	3.78819E-06	0.071360869
11	1.02E+10	1.77E-06	5.4117E-06	0.097310275
12	2.88E+10	5.31E-07	7.73099E-06	0.118162477
13	7.3E+10	1.59E-07	1.10443E-05	0.128550387
14	1.67E+11	4.78E-08	1.57775E-05	0.12592691
15	3.45E+11	1.43E-08	2.25393E-05	0.111535263
16	6.47E+11	4.3E-09	3.21991E-05	0.089626551
17	1.1E+12	1.29E-09	4.59987E-05	0.065525293
18	1.72E+12	3.87E-10	6.57124E-05	0.043683529
19	2.44E+12	1.16E-10	9.38748E-05	0.026604254
20	3.17E+12	3.49E-11	0.000134107	0.01482237
21	3.77E+12	1.05E-11	0.000191581	0.007562434
22	4.12E+12	3.14E-12	0.000273687	0.003535683
23	4.12E+12	9.41E-13	0.000390982	0.001515293
24	3.77E+12	2.82E-13	0.000558546	0.000595294
25	3.17E+12	8.47E-14	0.000797923	0.000214306
26	2.44E+12	2.54E-14	0.00113989	7.06502E-05
27	1.72E+12	7.63E-15	0.001628414	2.13072E-05
28	1.1E+12	2.29E-15	0.002326305	5.87035E-06
29	6.47E+11	6.86E-16	0.003323293	1.47482E-06
30	3.45E+11	2.06E-16	0.004747562	3.37101E-07
31	1.67E+11	6.18E-17	0.006782231	6.99058E-08
32	7.3E+10	1.85E-17	0.009688901	1.31073E-08
33	2.88E+10	5.56E-18	0.013841287	2.21293E-09
34	1.02E+10	1.67E-18	0.019773267	3.34728E-10
35	3.19E+09	5E-19	0.028247525	4.50859E-11
36	8.86E+08	1.5E-19	0.040353607	5.36737E-12
37	2.16E+08	4.5E-20	0.05764801	5.59532E-13
38	45379620	1.35E-20	0.0823543	5.04841E-14
39	8145060	4.05E-21	0.117649	3.88339E-15
40	1221759	1.22E-21	0.16807	2.49647E-16
41	148995	3.65E-22	0.2401	1.30477E-17
42	14190	1.09E-22	0.343	5.32561E-19
43	990	3.28E-23	0.49	1.59237E-20
44	45	9.85E-24	0.7	3.10203E-22
45	1	2.95E-24	1	2.95431E-24

N = 10, p = 0.7

L	C(45, L)	(0.3)^L	(1-0.3)^(45-L)	Prob(L)
0	1	1	2.95431E-24	2.95431E-24
1	45	0.7	9.84771E-24	3.10203E-22
2	990	0.49	3.28257E-23	1.59237E-20
3	14190	0.343	1.09419E-22	5.32561E-19
4	148995	0.2401	3.6473E-22	1.30477E-17
5	1221759	0.16807	1.21577E-21	2.49647E-16
6	8145060	0.117649	4.05256E-21	3.88339E-15
7	45379620	0.082354	1.35085E-20	5.04841E-14
8	2.16E+08	0.057648	4.50284E-20	5.59532E-13
9	8.86E+08	0.040354	1.50095E-19	5.36737E-12
10	3.19E+09	0.028248	5.00315E-19	4.50859E-11
11	1.02E+10	0.019773	1.66772E-18	3.34728E-10
12	2.88E+10	0.013841	5.55906E-18	2.21293E-09
13	7.3E+10	0.009689	1.85302E-17	1.31073E-08
14	1.67E+11	0.006782	6.17673E-17	6.99058E-08
15	3.45E+11	0.004748	2.05891E-16	3.37101E-07
16	6.47E+11	0.003323	6.86304E-16	1.47482E-06
17	1.1E+12	0.002326	2.28768E-15	5.87035E-06
18	1.72E+12	0.001628	7.6256E-15	2.13072E-05
19	2.44E+12	0.00114	2.54187E-14	7.06502E-05
20	3.17E+12	0.000798	8.47289E-14	0.000214306
21	3.77E+12	0.000559	2.8243E-13	0.000595294
22	4.12E+12	0.000391	9.41432E-13	0.001515293
23	4.12E+12	0.000274	3.13811E-12	0.003535683
24	3.77E+12	0.000192	1.04604E-11	0.007562434
25	3.17E+12	0.000134	3.48678E-11	0.01482237
26	2.44E+12	9.39E-05	1.16226E-10	0.026604254
27	1.72E+12	6.57E-05	3.8742E-10	0.043683529
28	1.1E+12	4.6E-05	1.2914E-09	0.065525293
29	6.47E+11	3.22E-05	4.30467E-09	0.089626551
30	3.45E+11	2.25E-05	1.43489E-08	0.111535263
31	1.67E+11	1.58E-05	4.78297E-08	0.12592691
32	7.3E+10	1.1E-05	1.59432E-07	0.128550387
33	2.88E+10	7.73E-06	5.31441E-07	0.118162477
34	1.02E+10	5.41E-06	1.77147E-06	0.097310275
35	3.19E+09	3.79E-06	5.9049E-06	0.071360869
36	8.86E+08	2.65E-06	0.000019683	0.046252415
37	2.16E+08	1.86E-06	0.00006561	0.026251371
38	45379620	1.3E-06	0.0002187	0.01289541
39	8145060	9.1E-07	0.000729	0.005400642
40	1221759	6.37E-07	0.00243	0.001890225
41	148995	4.46E-07	0.0081	0.000537869
42	14190	3.12E-07	0.027	0.000119526
43	990	2.18E-07	0.09	1.94578E-05
44	45	1.53E-07	0.3	2.0637E-06
45	1	1.07E-07	1	1.07007E-07



L vs. Prob(L)



L vs. Prob(L)

If N = 10,

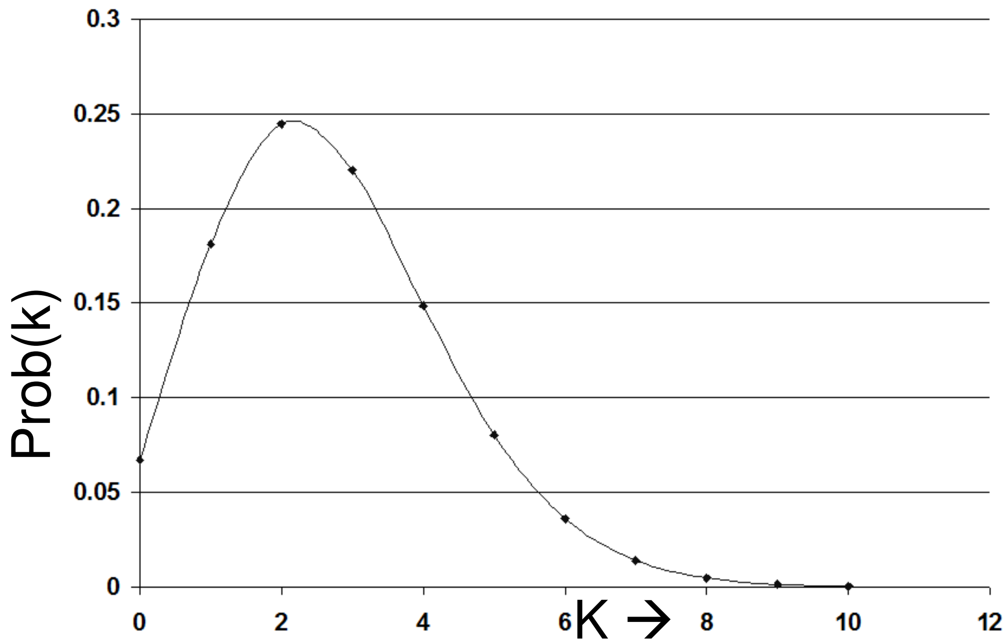
Lmax = N(N-1)/2 = 45

N = 10, p = 0.3

$\Rightarrow \langle k \rangle = p(N-1)$

$\Rightarrow \langle k \rangle = 0.3*9 = 2.7$

k	e^(-<k>)	<k>^k	k!	Prob(k)
0	0.067091	1	1	0.067091
1	0.067091	2.7	1	0.181146
2	0.067091	7.29	2	0.244547
3	0.067091	19.683	6	0.220092
4	0.067091	53.1441	24	0.148562
5	0.067091	143.4891	120	0.080223
6	0.067091	387.4205	720	0.036101
7	0.067091	1046.035	5040	0.013925
8	0.067091	2824.295	40320	0.0047
9	0.067091	7625.597	362880	0.00141
10	0.067091	20589.11	3628800	0.000381

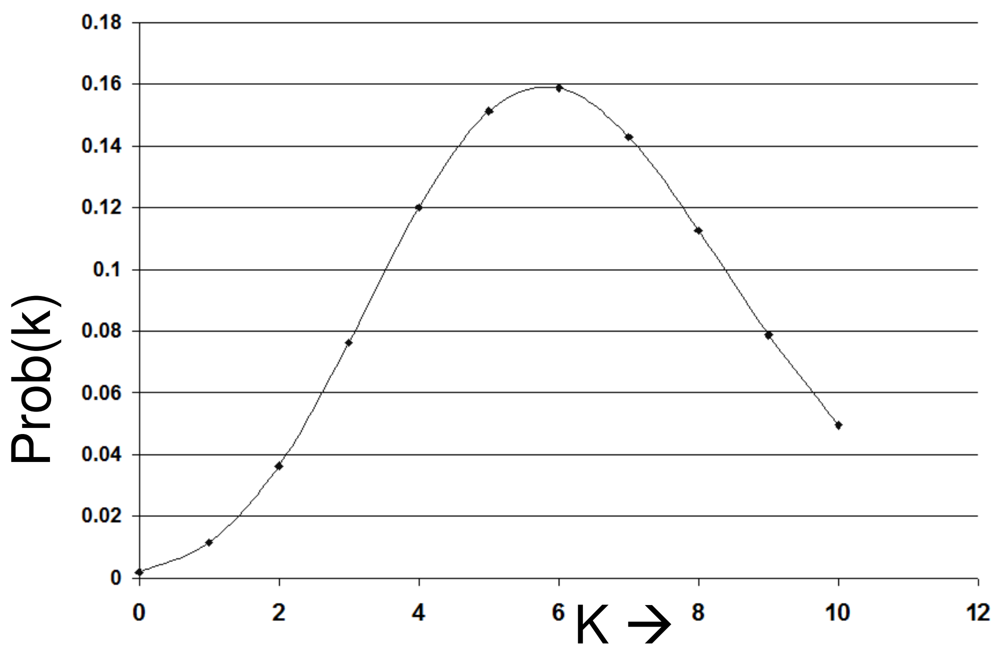


N = 10, p = 0.7

$\Rightarrow \langle k \rangle = p(N-1)$

$\Rightarrow \langle k \rangle = 0.7*9 = 6.3$

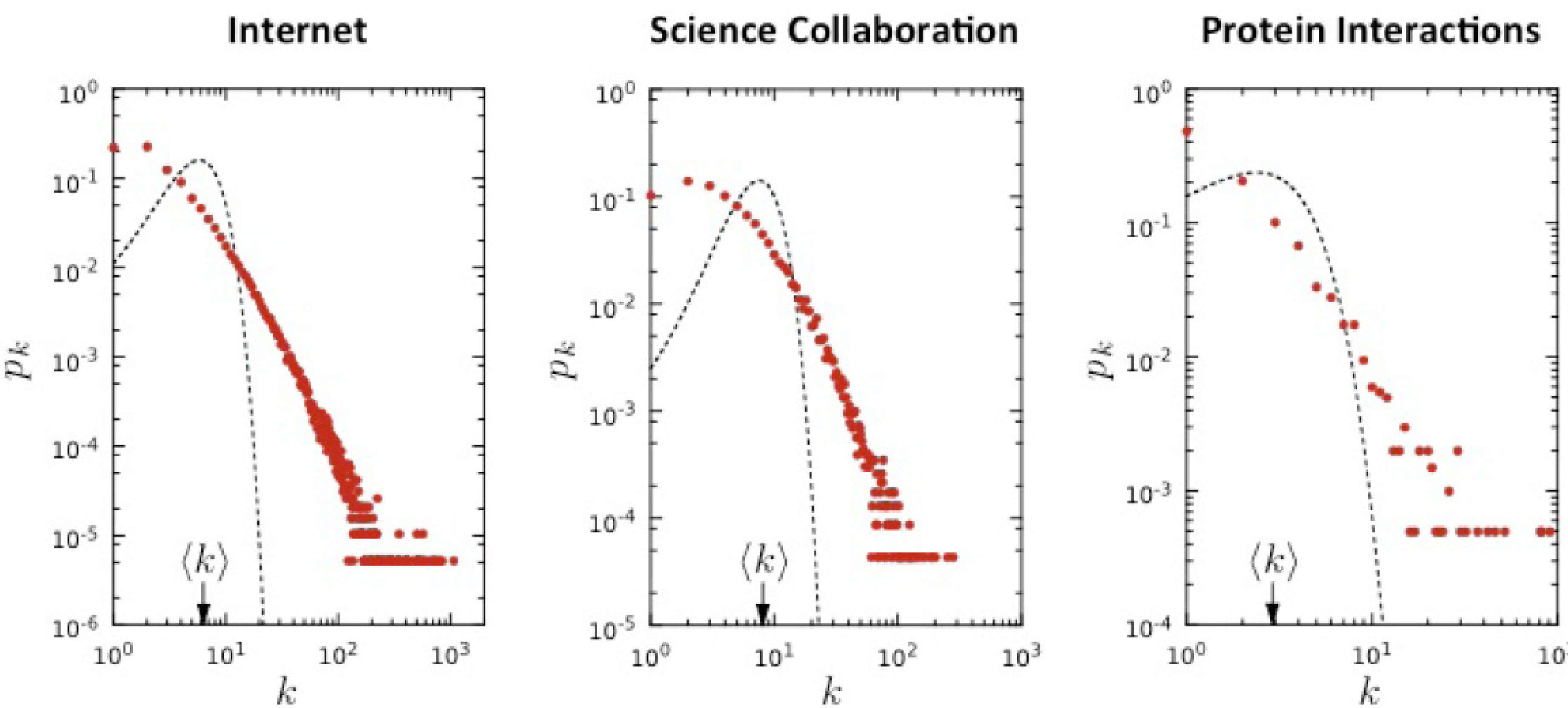
k	e^(-<k>)	<k>^k	k!	Prob(k)
0	0.001829	1	1	0.001829
1	0.001829	6.3	1	0.011523
2	0.001829	39.69	2	0.036297
3	0.001829	250.047	6	0.076223
4	0.001829	1575.296	24	0.120051
5	0.001829	9924.365	120	0.151265
6	0.001829	62523.5	720	0.158828
7	0.001829	393898.1	5040	0.142945
8	0.001829	2481558	40320	0.112569
9	0.001829	15633814	362880	0.078798
10	0.001829	98493029	3628800	0.049643



Real Networks do not have a Poisson degree distribution

- Let us assume that the world's social network (typically, $N = 10^9$ nodes and average node degree $\langle k \rangle = 1000$) follows a random network model.
- Using the results obtained for random networks, the above values for the global social network corresponds to:
  - Dispersion (std. dev.) = $\langle k \rangle^{1/2} = 31.62$.
- The above results indicate that in the global social network, the degree of most nodes is in the vicinity of $\langle k \rangle$.
  - However, we have people with number of contacts significantly larger than 1000 and significantly lower than 1000 too.
- The random network cannot be used to model a network with few extremely popular individuals (hubs) and networks with large differences in node degrees.

Degree Distribution of Real Networks



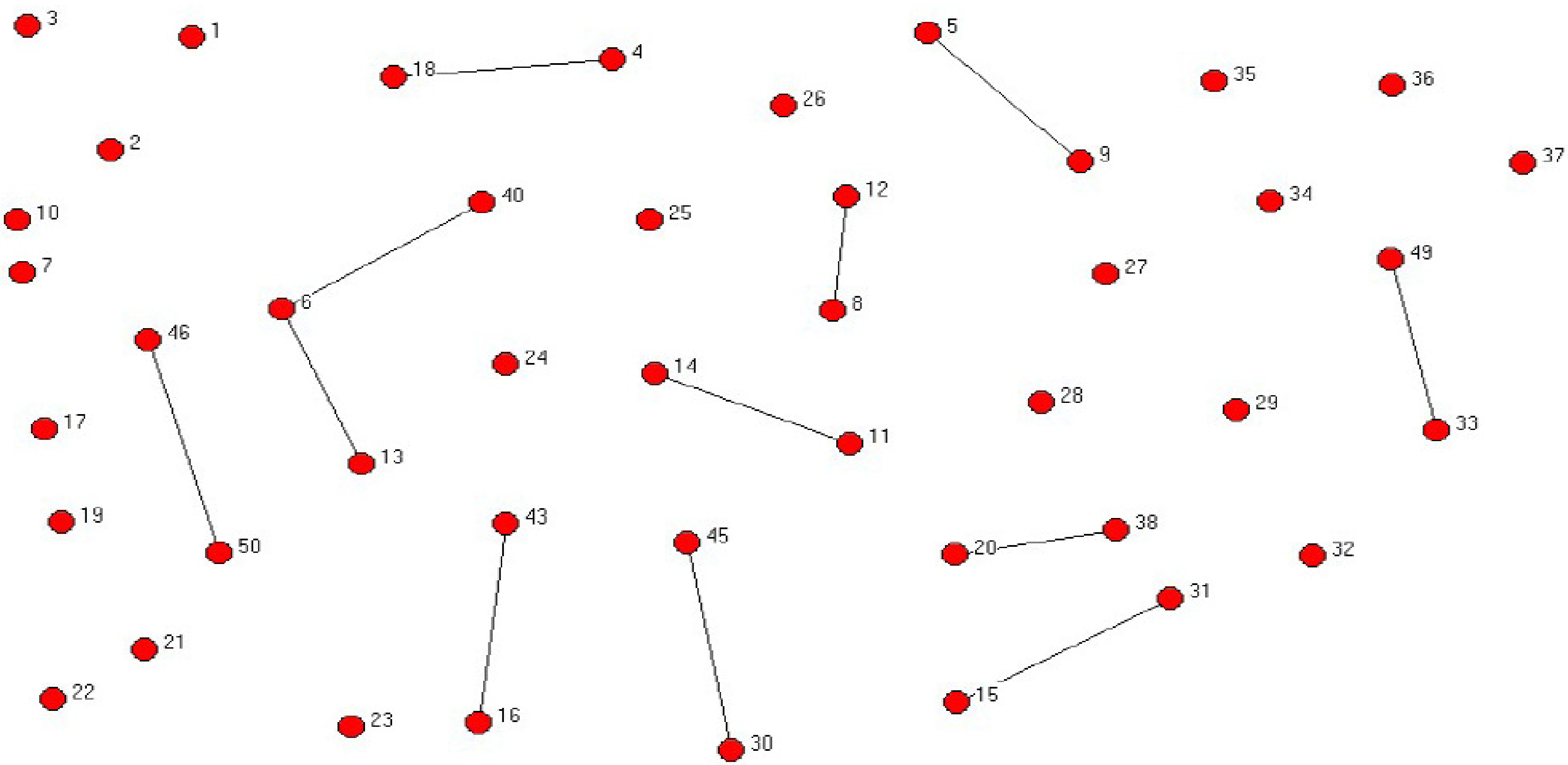
Source: Figure 3.5: Barabasi

The Poisson distribution underestimates the presence of nodes with larger degrees. For example, the maximum degree for a node in the Internet (according to the random model) is expected to be 20; there are nodes with degrees close to 1000. Likewise, the dispersion predicted under the random model is 2.52 (much smaller than the measured value of 14.44).

Phase Transitions in Random Networks

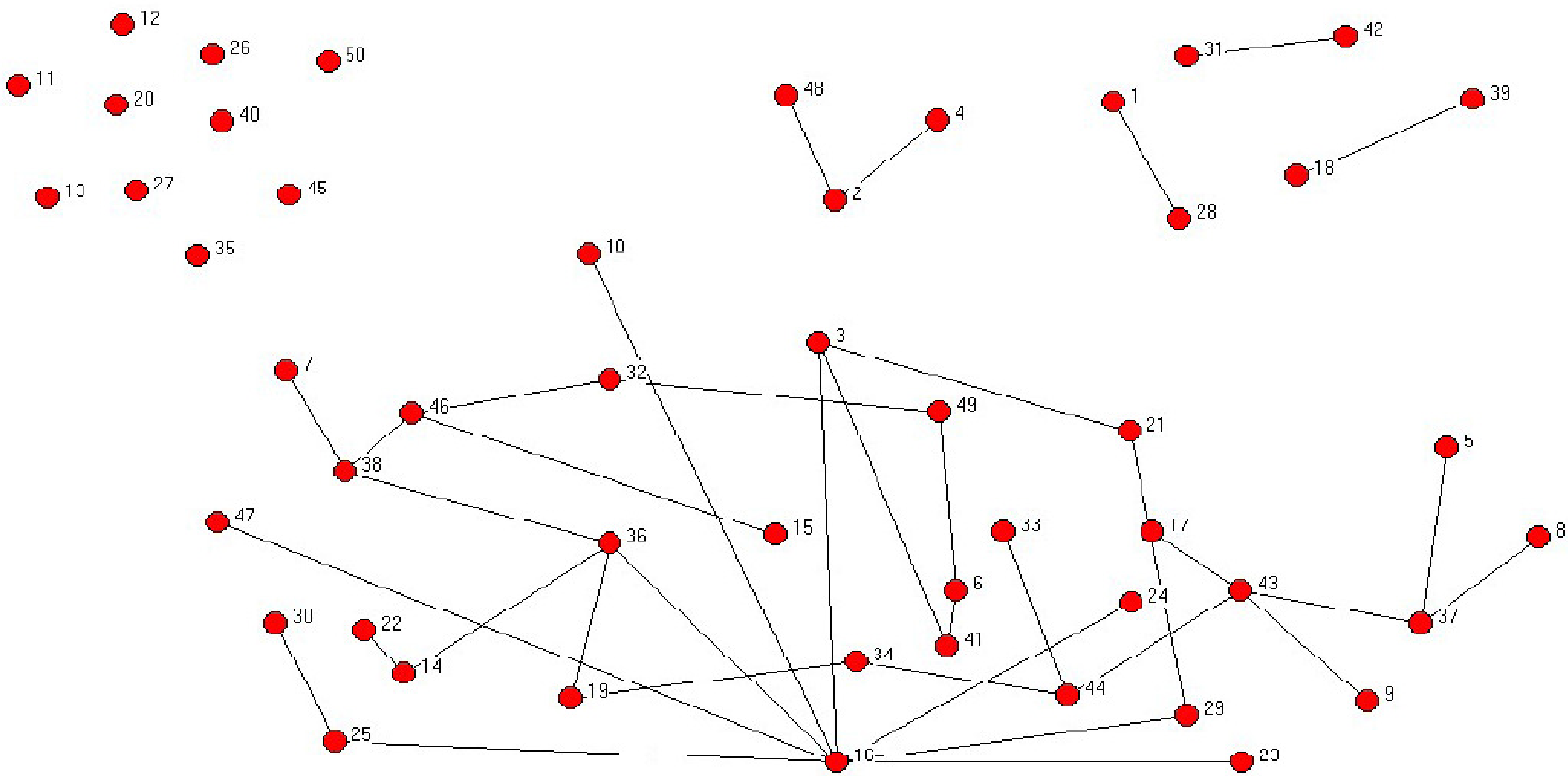
- If $p \geq 1/n^2$, the network has some links (avg. deg. $1/n$)
- If $p \geq 1/n^{3/2}$, the network has a component with at least three links (avg. deg. $1/n^{1/2}$)
- If $p \geq 1/n$, the network has a cycle; the network has a unique giant component: a component with at least n^a nodes (for some fixed $a < 1$); (avg. deg. 1)
- If $p \geq \log(n)/n$, then the network is connected; (avg. deg. $\log(n)$)

Phase Transitions in Random Networks
p = 0.01; 50 nodes (1)

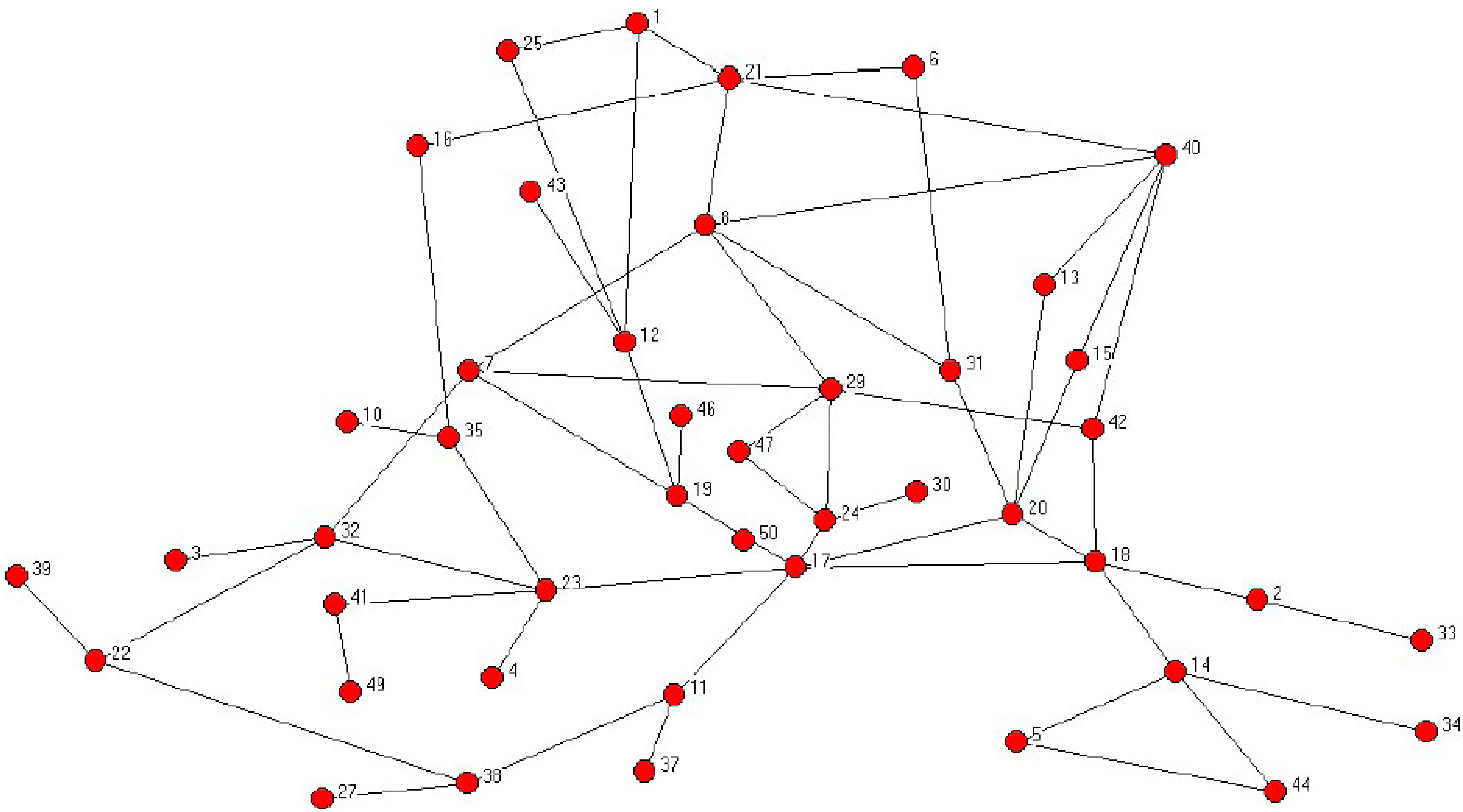


Phase Transitions in Random Networks p = 0.03; 50 nodes (2)

p = 0.02 for the emergence of a cycle and a giant component



Phase Transitions in Random Networks p = 0.1; 50 nodes (3)



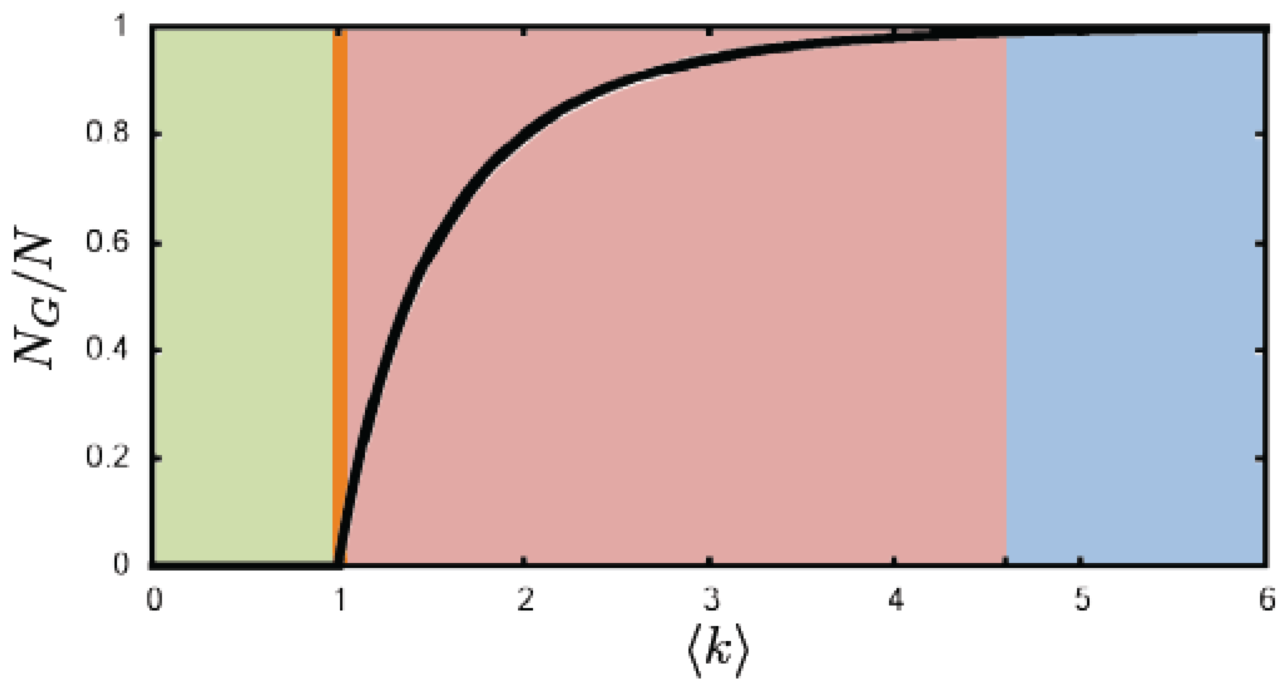
Evolution of a Random Network

- Giant component is the largest cluster within the network.
- The size of the giant component (N_G) varies with the average degree <k>.
  - For p = 0, we have <k> = 0. Hence, we observe only isolated nodes. Hence, $N_G = 1$ and $N_G/N \rightarrow 0$ for large N.
  - For p = 1, we have <k> = N-1. Hence, the network is a complete graph and all nodes belong to a single cluster. Hence, $N_G = N$ and $N_G/N = 1$.
- One would expect that the giant component will grow gradually from $N_G = 1$ to $N_G = N$ if we increase <k> from 0 to N-1.
  - However, as observed from theoretical analysis studies, N_G/N remains 0 for small <k>. Once <k> exceeds a critical value (1), N_G/N increases rapidly signaling the emergence of a giant component.
    - We have a giant component if and only if when each node has on average more than one link.

Evolution of a Random Network

- We know that <k> = p(N-1).
- For the critical value of <k> = 1 when the giant component emerges, $p_c(N-1) = 1$.
  - $p_c = 1/(N-1) \approx 1/N$.
  - This indicates: Larger the network, the smaller the value of *p* for the emergence of a giant component.

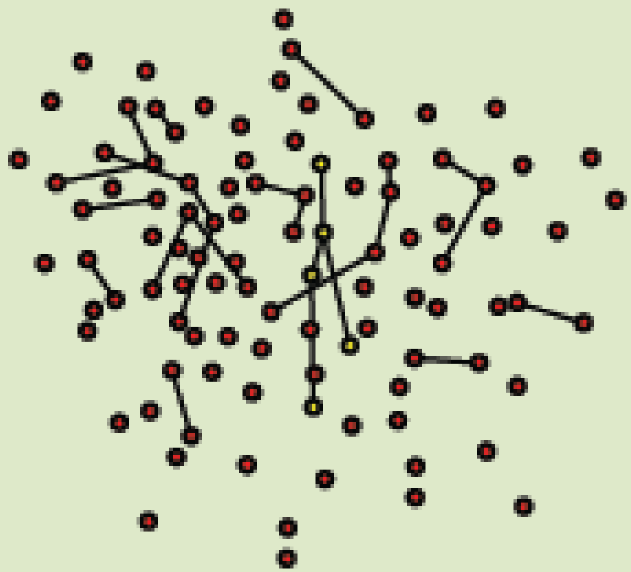
where $S = N_G/N$



Source: Figure 3.6a Barabasi

Evolution: Topological Transitions

Consider creating a random network according to the G(N, L) model



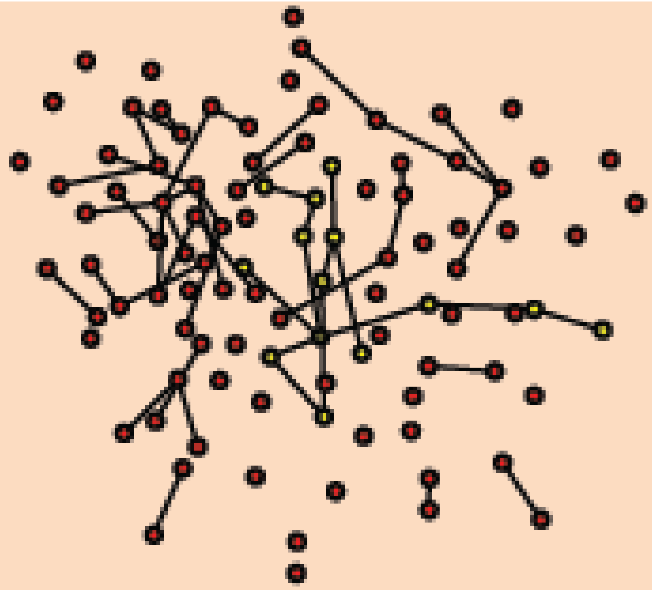
$\langle k \rangle < 1$

<u>Subcritical regime</u>

- 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.

- <u>Subcritical regime:</u>
  - $0 < \langle k \rangle < 1$ and $p < 1/N$
  - The largest cluster is expected to be a tree with lnN nodes. Hence, $N_G/N = \ln N / N \rightarrow 0$ in the $N \rightarrow \infty$ limit, indicating that the largest component is tiny compared to the size of the network.
  - Components have comparable sizes, lacking a clear winner to be designated as a giant component.

Evolution: Topological Transitions



$\langle k \rangle = 1$

<u>Critical point</u>

- No giant component.
- Cluster size distribution:

$$P(s) \sim s^{-3/2}$$

- Size of the largest cluster:

$$N_G \sim N^{2/3}$$

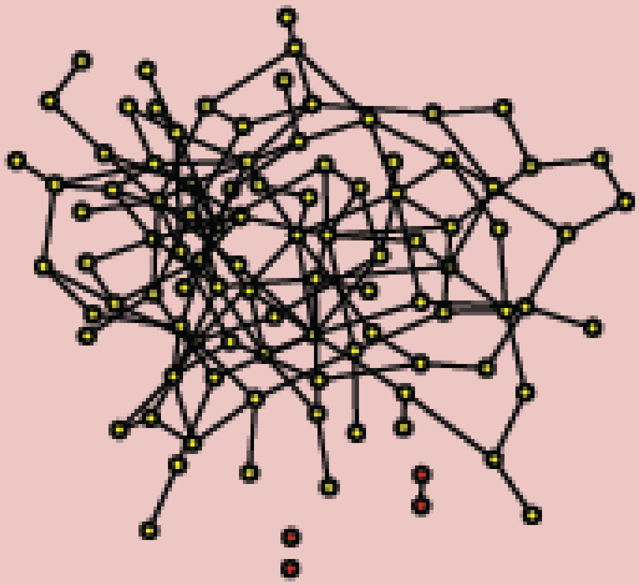
- The clusters may contain loops.

– **<k>** = 1 and p = 1/N

– The largest cluster is expected to be of size $N^{2/3}$ and contain loops, while the smaller clusters are typically trees.

– The largest cluster is still tiny compared to the network size. $N_G/N = N^{(-1/3)} \rightarrow 0$ as $N \rightarrow \infty$.

Evolution: Topological Transitions

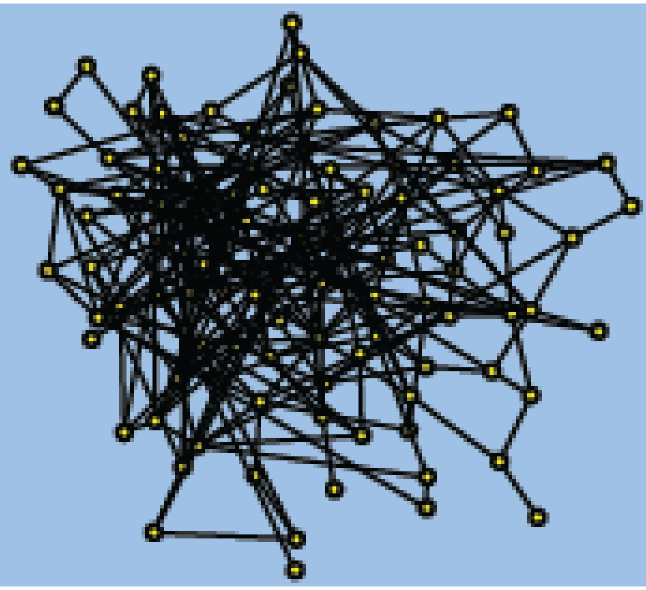


$\langle k \rangle > 1$

p > 1/N

Supercritical regime

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



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

p ≥ (ln N)/N

Fully connected regime

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

Prediction of Random Network Theory: Real Networks are Supercritical

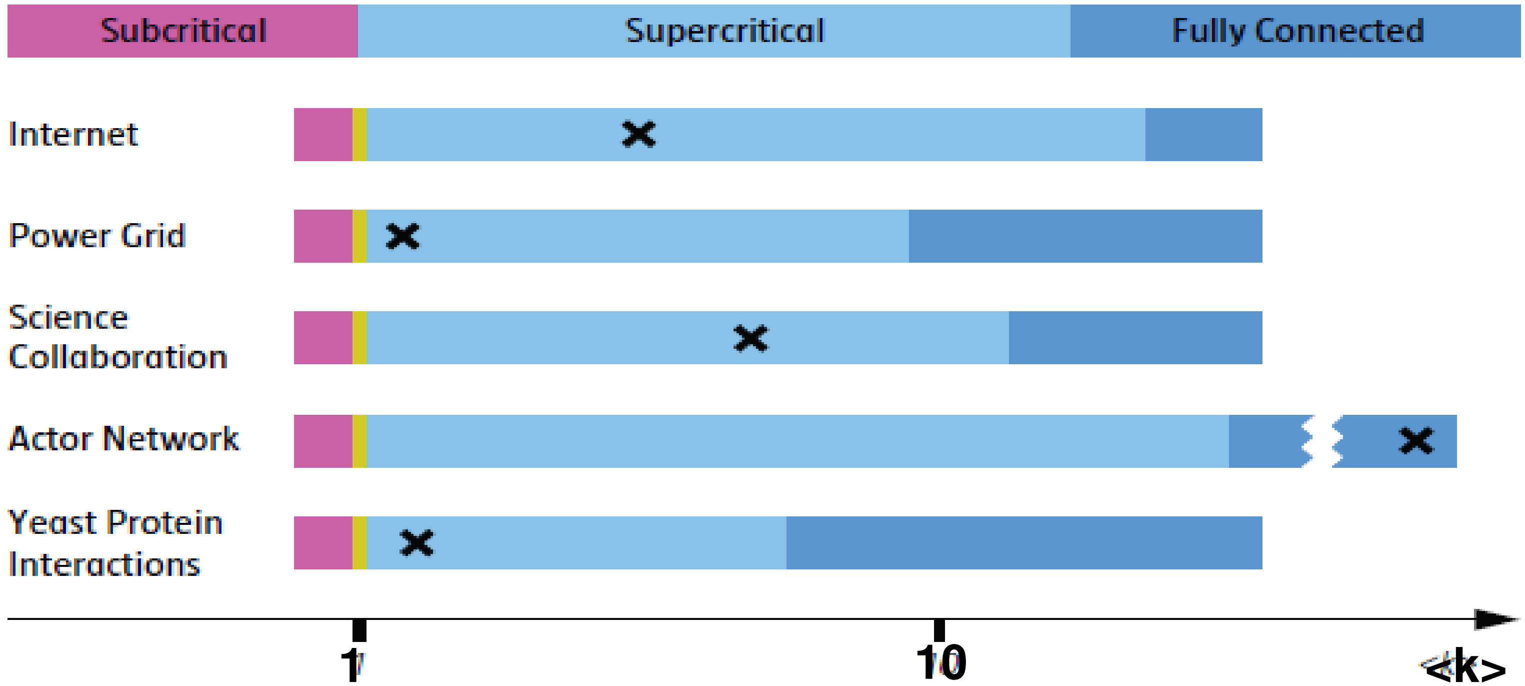
- The theoretical thresholds uncovered for random networks are:
  - For $\langle k \rangle > 1$, a giant component emerges that contains a finite fraction of all nodes.
  - For $\langle k \rangle > \ln N$, all components are absorbed by the giant component, resulting in a single connected network.

Network	N	L	$\langle k \rangle$	$\ln N$	$\frac{\ln N}{\ln \langle k \rangle}$
Internet	192,244	609,066	6.34	12.17	**6.59**
Power Grid	4,941	6,594	2.67	8.51	**8.67**
Science Collaboration	23,133	186,936	8.08	10.04	**4.81**
Actor Network	212,250	3,054,278	28.78	12.27	**3.65**
Yeast Protein Interactions	2,018	2,930	2.90	7.61	**7.15**

Source:
Table 3.1
Barabasi

Prediction of Random Network Theory: Real Networks are Supercritical

- Just based on the N and L values for the real networks, we could predict (according to the principles of Random Network Theory) that:
  - All real networks should have a giant component (since their <k> exceeds 1)
  - For most real networks (except the actor network), the giant component does not absorb all the nodes (components) as their <k> value is less than lnN. Hence, most real networks according to Random Network theory are in the supercritical topology regime.

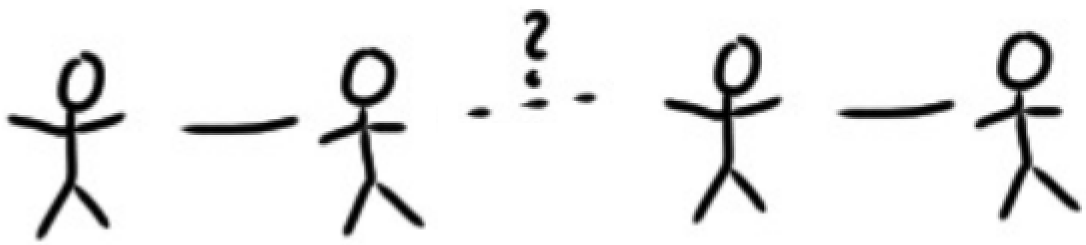


Source:
Figure 3.8
Barabasi

Giant Components: Intuitive Idea

If your friend starts getting connected to someone other than yourself, then you are more likely to belong to a larger component.

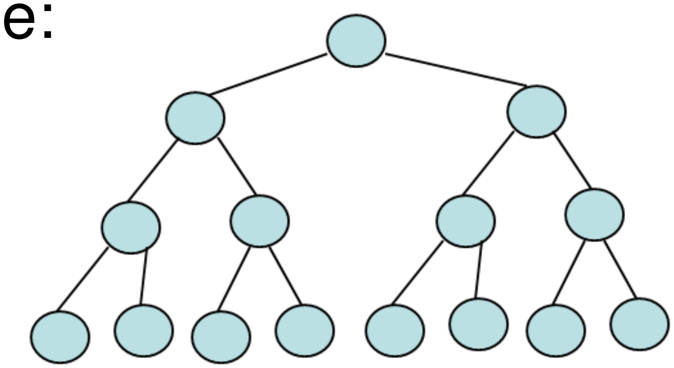
The emergence of the giant component sets in when each node has degree of at least 1. Any new edge added to the network is more likely to merge two disconnected groups. Hence, the giant component is very likely to emerge if the average degree of a node exceeds 1.



As the network evolves, there cannot be two giant components. The addition of new edges is likely to merge two giant components and evolve them as one single giant component.

Small World Property

- Distance between two randomly chosen nodes in a random network is surprisingly short.
- Consider a random network with average degree <k>. A node in this network has on average:
  - $<k>$ nodes at distance one ($d = 1$).
  - $<k>^2$ nodes at distance two ($d = 2$).
  - $<k>^3$ nodes at distance three ($d = 3$).
  - ….
  - $<k>^d$ nodes at distance d.



- The expected number of nodes up to distance d from the starting node is:
  - $$N(d) = 1 + <k> + <k>^2 + \ldots + <k>^d = \frac{<k>^{d+1} - 1}{<k> - 1}$$

Small World Property

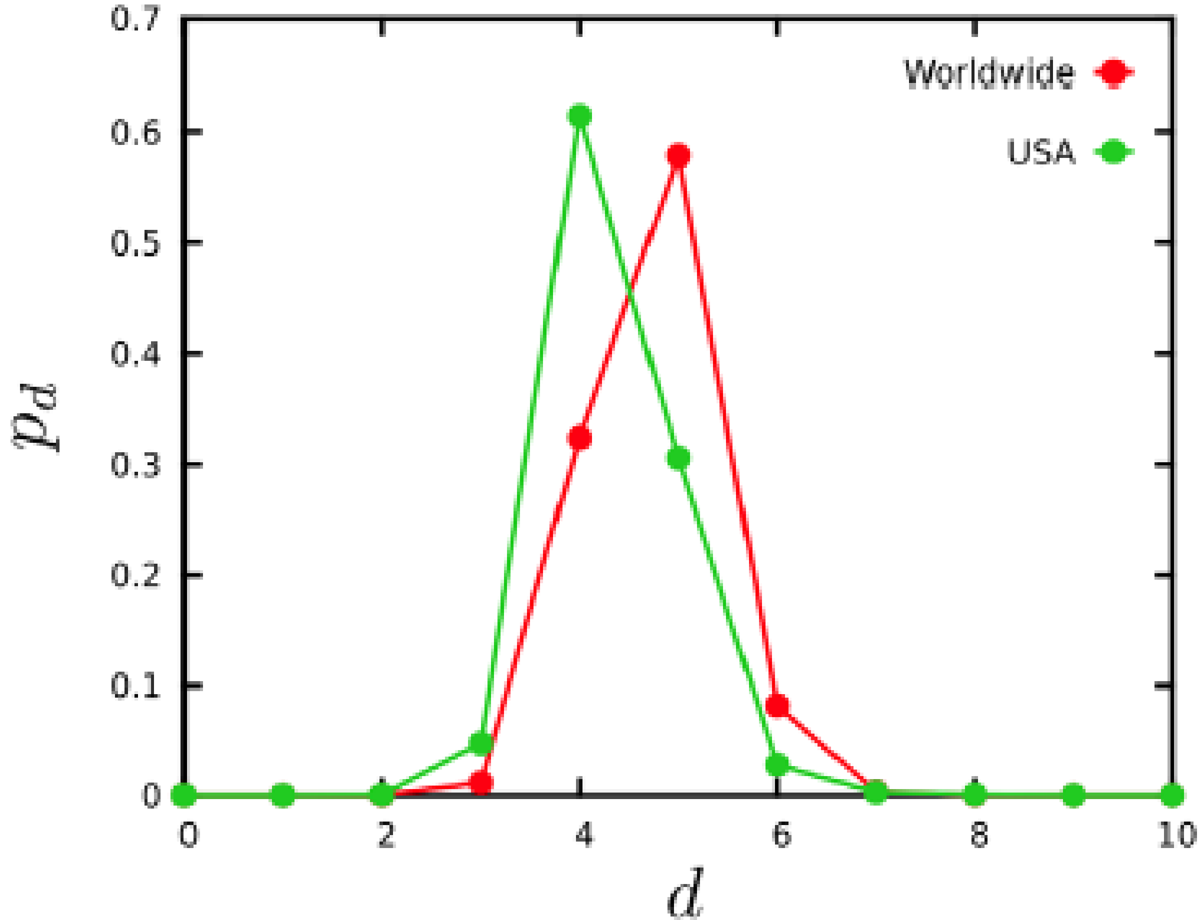
- Let d_{max} denote the maximum distance (the network diameter) at which N(d) reaches N. That is, N(d_{max}) = N.
- Assuming that <k> >> 1,
  - $\langle k \rangle^{dmax} \approx N$.
  - dmax = ln N / ln <k>
- As seen from the results for real networks, lnN/ln<k> approximates more better for the average distance between two randomly chosen nodes.
  - This is because dmax is often dominated by a few extreme paths, while <d> is averaged over all node pairs, a process that diminishes the fluctuations.
- Thus, the average distances <d> in a random network are proportional to lnN, rather than N.
- The 1/(ln<k>) term implies that denser the network, the smaller is the distance between the nodes.

Small World Property

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

Source: Table 3.2: Barabasi

Small World Property: Facebook



For Facebook,
N = 7*10^9 users
<k> = 1000

$$\langle d \rangle = \frac{\ln 7 \times 10^9}{\ln(10^3)} = 3.28$$

Based on the actual Facebook connections,
<d> = 4.74

Clustering Coefficient

- The local clustering coefficient Ci captures the density of links in node i's immediate neighborhood.
  - $C_i = 0$ implies there are no links between i's neighbors
  - $C_i = 1$ implies that each of node i's neighbors link to each other.
- Let k_i be the degree of node i.
- Max. number of possible links between the k_i neighbors of node i are $k_i(k_i – 1)/2$.
- <u>Actual Local Clustering Coefficient of node</u> i is the actual number of links between the neighbors of i divided by the maximum number of possible links between the neighbors.
- If *p* is the probability that any two nodes in a network are connected, then the expected number of links between the k_i neighbors of node i is:

$$\langle L_i \rangle = p\frac{k_i(k_i - 1)}{2}.$$

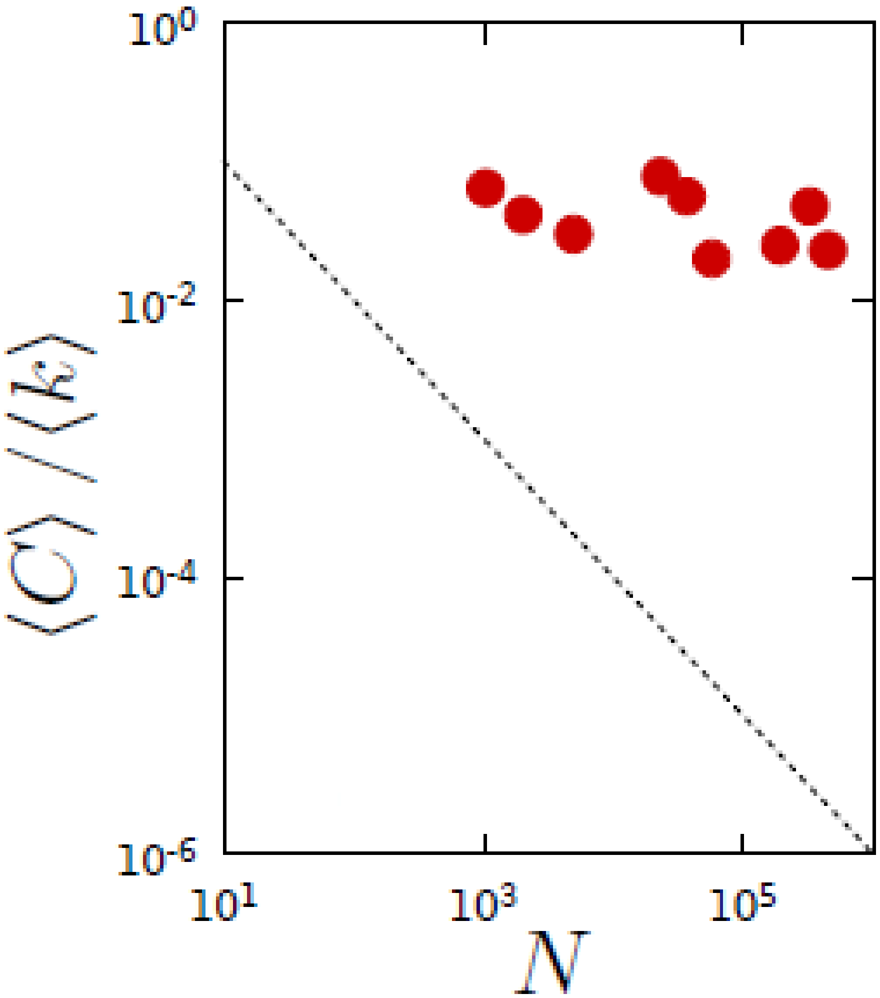
- <u>Expected Local clustering coefficient of node</u> i:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

Clustering Coefficient

- Observations based on Random Network Theory
- For fixed <k>, the larger the network, the smaller is a node's expected clustering coefficient.
  - Thus, the network's average clustering coefficient <C> is expected to decrease as 1/N.
- The expected local clustering coefficient of a node is independent of the node's degree

Clustering Coefficients for Real Networks



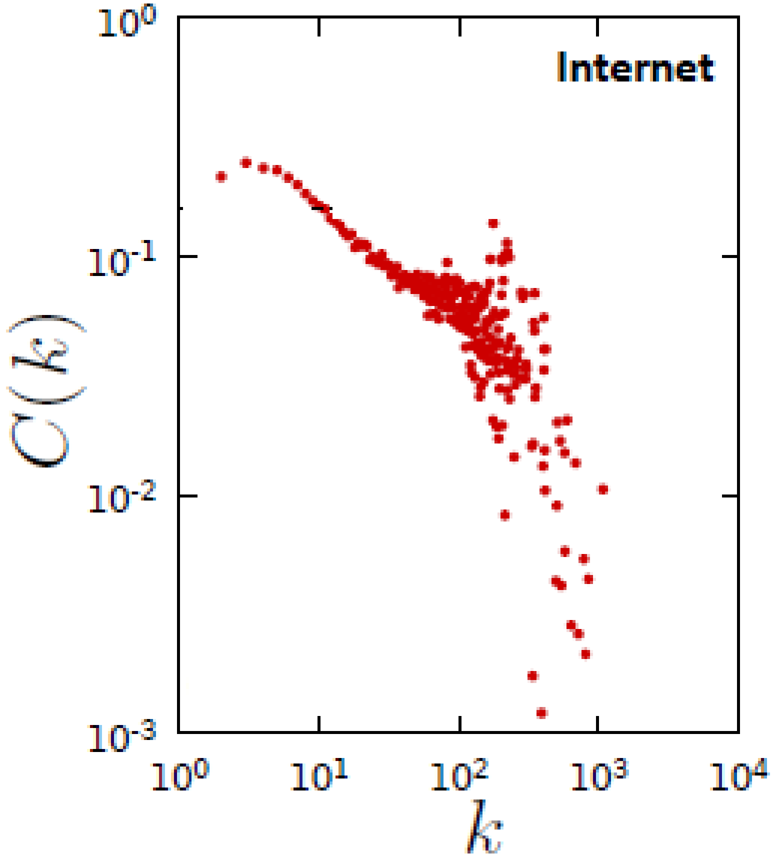
Each circle corresponds to a real network.

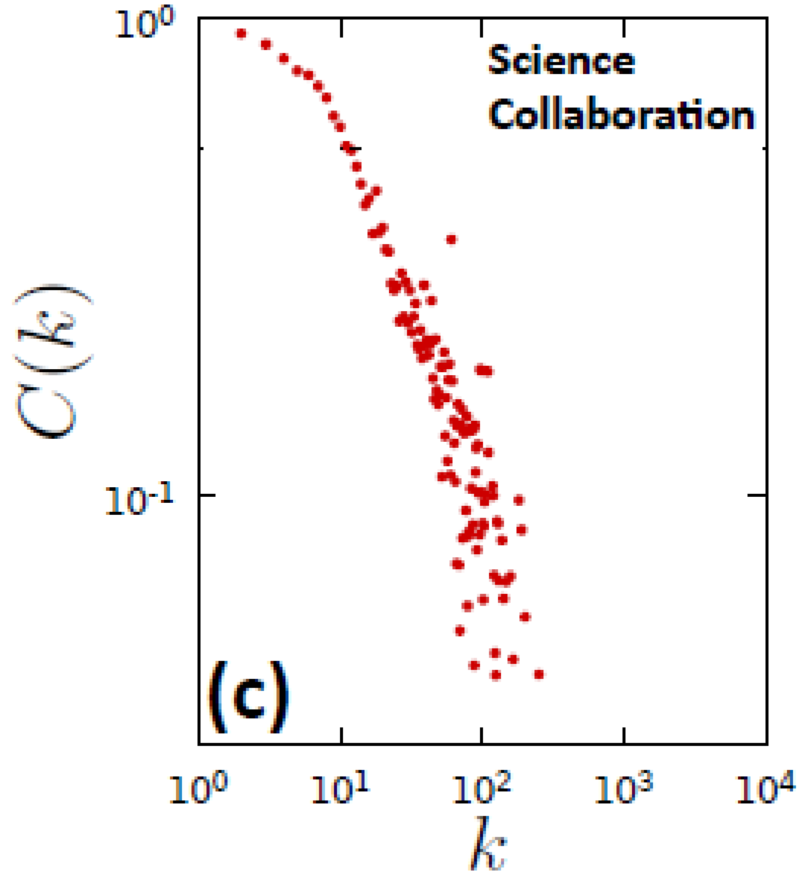
Directed networks were made undirected to calculate C.

For random networks, the average clustering coefficient decreases as 1/N. In contrast, for real networks, <C> has only a weak dependence on N.

Real networks have a much higher Clustering coefficient than expected for a random network of similar N and L.

Clustering for Real Networks





C(k) is measured by averaging the local clustering coefficient of all nodes with the same degree k.

According to the Random Network theory model, C(k) is independent of the individual node degrees. However, we find that C(k) decreases as k increases. Nodes with fewer neighbors have larger local clustering coefficients and vice-versa

Clustering Coeff. Real Networks

Networks	Actual	Random, G(n, p)
– Prison		
Friendships	0.31	0.0134
Co-authorships		
Math	0.15	0.00002
Biology	0.09	0.00001
Economy	0.19	0.00002
WWW		
Web links	0.11	0.002

Real Networks are not Random

- Degree distribution:
  - Random networks – binomial distribution, in general, and Poisson distribution for k << N.
    - Highly connected nodes (hubs) are effectively forbidden.
  - Real networks: More highly connected nodes, compared to that predicted with random model.
- Connectedness:
  - Random networks: One single giant component exists only if <k> > ln N.
  - Real networks: One single giant component exists for several networks with <k> < ln N.
- Average Path Length (small world property):
  - For both random and real networks, the average path length scales as log N / log <k>.
- Clustering coefficient:
  - Random model: Local clustering coefficient is independent of the node's degree and <C> depends on the system size as 1/N.
  - Real networks: C decreases with node degrees and is largely independent of the system size.

Real Networks are not Random

- Except for the small world property, the properties observed for real-world networks are not matching with that observed for random networks.

- Then why study random graph theory?
- If a certain property is observed for real-world networks, we can refer to the random graph theory and analyze whether the property is observed by chance (like the small world property).
- If the property observed does not coincide with that of the random networks (like the local clustering coefficient), we need to further analyze the real-world network for the existence of the property because it did not just happen by chance.
- Establish useful benchmarks (e.g., for component structure, diameter, degree distribution, clustering, etc)

Simulating a Random Network ER Model

- Let S be the set of all node pairs
- Until S gets empty
  - Pick a node *u* randomly in the network.
    - If this node has at least one node in the set S that it is not yet considered for a possible edge, then randomly select a node v among these candidate nodes.
      - Generate a random number *r*
      - If the value of r <= p, the probability for an edge, then connect the two nodes u-v.
      - Else do not connect them
    - Either way, remove the node pair u-v from set S

Realistic Variations of the Random Network Model

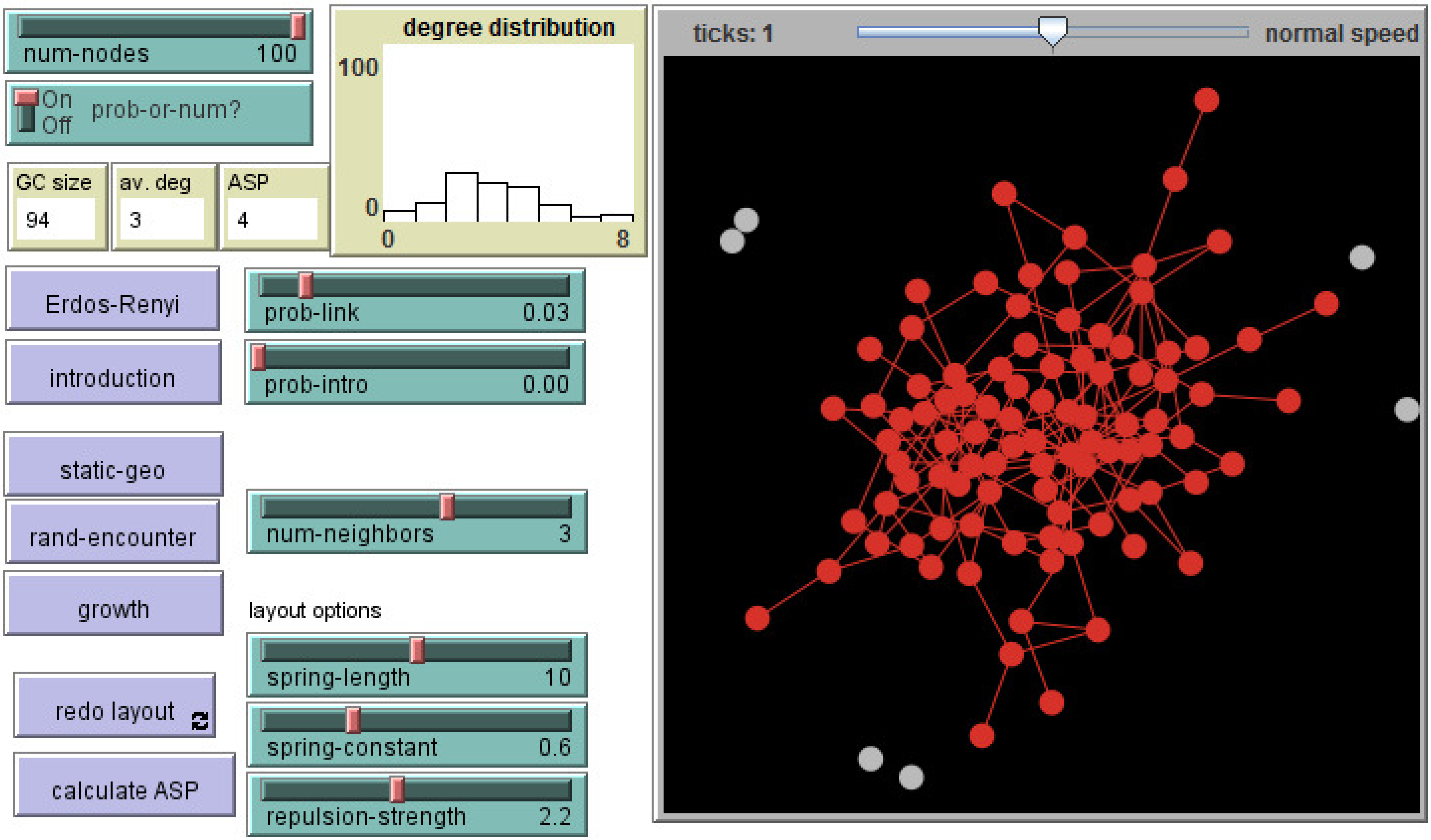
- **Introduction Model:** A node has higher chances of establishing a link with a neighbor of its neighbor (e.g., with the friend of a friend) rather than with an arbitrarily selected node.
  - Operate with a probability, p-intro, the probability that a node prefers to connect to the neighbor of a neighbor node.

- **Key Observations:**
  - Smaller Giant Component Size for smaller p;
  - Larger average shortest path length;
  - Uneven node degree distribution;

Simulating a Random Network Introduction Model

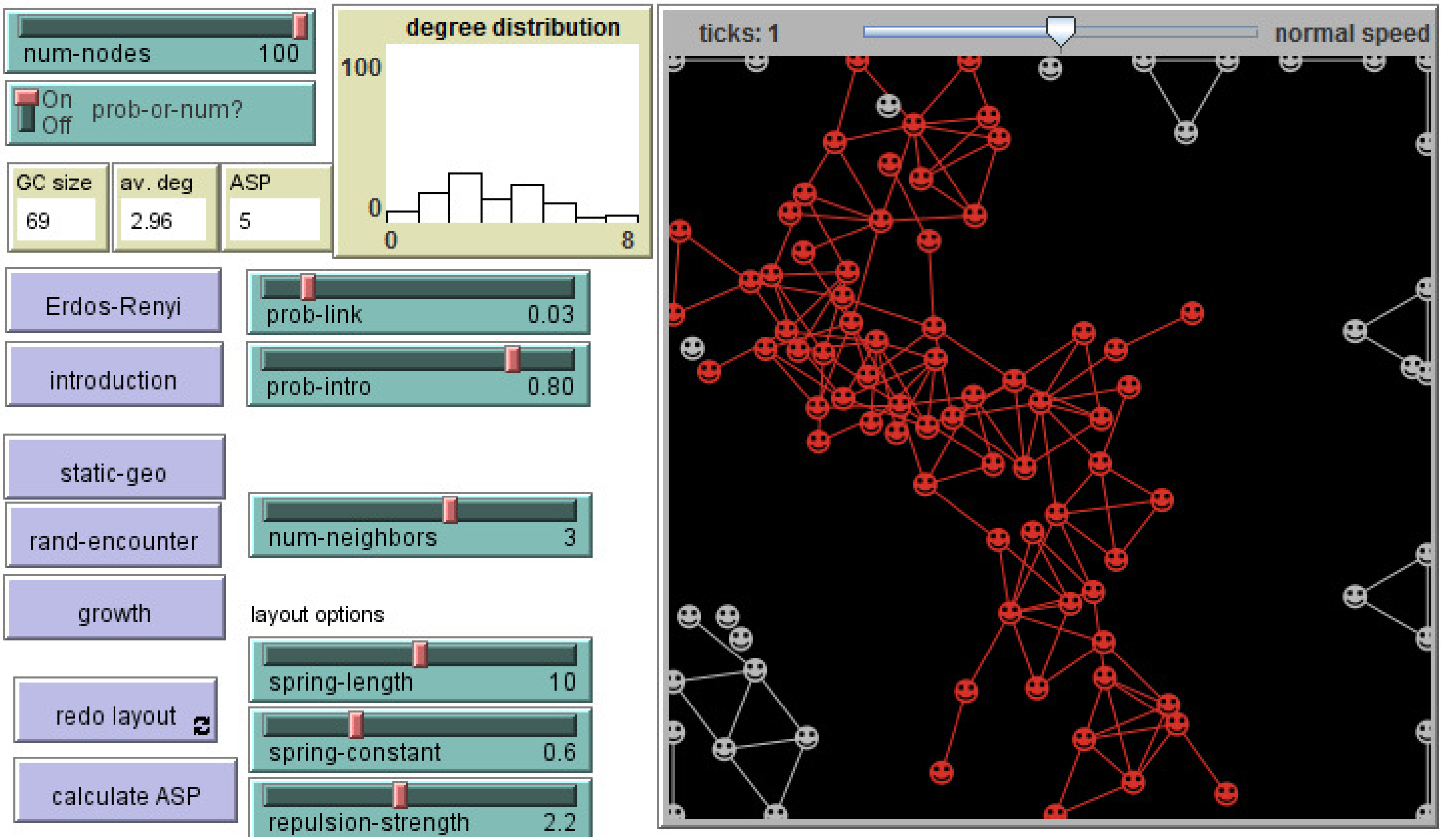
- Let S be the set of all node pairs
- Until S gets empty
  - Pick a node u randomly in the network.
    - If this node has at least one node in the set S that it is not yet considered for a possible edge
      - Generate a random number r-intro.
      - If r-intro <= p-intro, the set of candidate nodes that are chosen for connection are the unconnected neighbors of neighbor nodes.
      - Else, the set of candidate nodes are all the unconnected nodes in the network.
      - Among the chosen candidate nodes, the node connects to a randomly chosen node *v* with a probability p.
        - » Generate a random number *r*
        - » If the value of r <= p, the probability for an edge, then connect the two nodes u-v.
        - » Else do not connect them
      - Either way, remove the node pair u-v from set S

ER Model



Num Nodes = 100; p = 0.03; p-intro = 0
GC Size – 94; Avg. Degree = 3; Avg. Shortest Path Length = 4

Introduction Model



Num Nodes = 100; p = 0.03; p-intro = 0.80
GC Size – 69; Avg. Degree = 2.96; Avg. Shortest Path Length = 5

Problem Example 1

- Consider a random network generated according to the G(N, p) model where the total number of nodes is 12 and the probability that there are links between any two nodes is 0.20. Determine the following:
  - The probability that there are exactly 60 links in the network
  - The average number of links in the network
  - The average node degree
  - The standard deviation of the node degree
  - The average path length (distance between any two nodes in the network)
  - The average local clustering coefficient for any node in the network.
  - The expected local clustering coefficient for a node that has exactly 5 neighbors.

Problem Example 1: Solution (1)

- There are N = 12 nodes
- Prob[link between any two nodes] = p = 0.2

Max. possible number of links between any two nodes is
$(N)(N-1)/2 = (12*11/2) = 66$

(1) Prob[there are exactly 60 links in the network]
$= C(66, 60) * p^{60} * (1-p)^{(66-60)}$

$C(66, 60) = 66! / (60! * 6!)$

$= 60! *61*62*63*64*65*66 / (60!*1*2*3*4*5*6)$

$= 90858768$

Prob[there are exactly 60 links in the network]
$= 90858768 * (0.2)^{60} * (0.8)^{6}$
$= 2.75 * 10^{-35}$

Problem Example 1: Solution (2)

- There are N = 12 nodes
- Prob[link between any two nodes] = p = 0.2

Max. possible number of links between any two nodes is (N)(N-1)/2 = (12*11/2) = 66

(2) The average number of links in the network = p * N(N-1)/2

= 0.2 * 66 = 13.2

(3) Average node degree = p*(N-1) = 0.2 * 11 = 2.2

(4) Standard deviation of node degree =

$$\sqrt{p(1-p)*(N-1)}$$

= sqrt(0.2*0.8*11) = 1.33

(5) Average path length = ln N / ln <k> = ln(12) / ln(2.2) = 3.15

(6) Avg. Local clustering coefficient for any node in the network = p = 0.2.

(7) The expected local clustering coefficient for a node in a random network is independent of its number of neighbors. Hence, the answer is 0.2

Problem Example 2

- Consider the evolution of a random network according to the G(N, L) model, where the total number of nodes is 100. Consider adding (randomly) one link at a time to the network. The total number of links added is sufficiently large enough to create one single connected component of the entire network. Determine the following:
  - The critical value of the probability (of the number of links) that a giant component emerges for the above network and the average size of the giant component at that value?
  - The minimum value of the average degree per node in the giant component of the fully connected regime.
  - The maximum value for the average path length between any two nodes in the giant component that encompasses all the nodes in the network.

Problem Example 2: Solution (1)

- There are N = 100 nodes
  - The critical value of the probability (of the number of links) that a giant component emerges for the above network? $p_c = 1/N = 1/100 = 0.01$

    Critical regime $N_G = N^{2/3} = 100^{(2/3)} = 21.88$

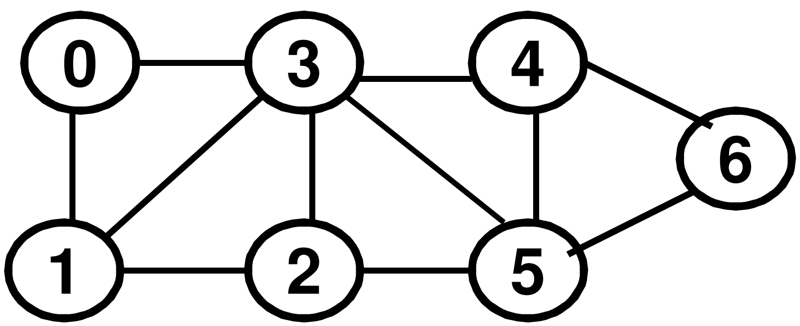
  - In the fully connected regime, the average node degree has to be at least lnN. That is, $\langle k \rangle \geq \ln N$.
  - Min $\langle k \rangle = \ln N = \ln(100) = 4.61$

  - The average path length is given by: $\ln N / \ln\langle k \rangle$
  - Using the minimum value of $\ln\langle k \rangle$ in the above expression, we obtain the max. average path length to be: $\ln(100)/\ln(4.61) = 3.01$.

Example 3: Degree Distribution Analysis

- For the graph given, find whether or not the links happened by chance? For this, do the following:
- a) Find the frequency (probability) distribution of the degree of the vertices in the actual graph.
- b) Find the average degree of the graph and use it as a parameter to determine a probability distribution of the vertices in a random graph.
- c) Compare the probability distributions of (a) and (b) and arrive at your conclusion for the overall question posed above.

Use a threshold value of 0.15 for the root mean square difference between the two Probability distributions.

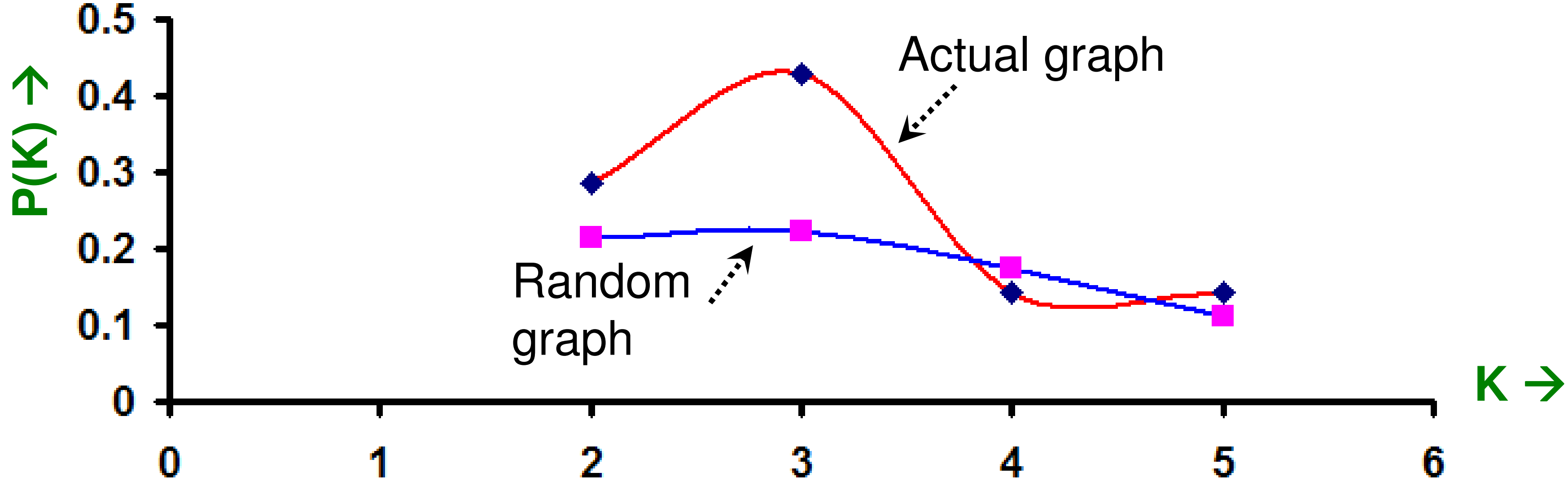


Node	Degree
0	2
1	3
2	3
3	5
4	3
5	4
6	2

Degree	# nodes	P(k)
2	2	2/7
3	3	3/7
4	1	1/7
5	1	1/7

Avg. Degree
2*2/7 + 3*3/7 + 4*1/7 + 5*1/7 = 3.14

Example:
$P(2) = \exp(-3.14) * (3.14)^2 / 2! = 0.2154$



When modeled as a Poisson Distribution,

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

K	P(K)-Poisson	P(K)-Actual	Sq. Diff.
2	0.21338	0.28571	0.005233
3	0.22333	0.42857	0.042123
4	0.17532	0.14286	0.001054
5	0.11009	0.14286	0.001073
		Mean Sq. Diff	0.012371
	Square Root (Mean Sq. Diff)		0.11122 < 0.15

Hence, the links could be considered to have happened by chance.