Lecture 4:
Network Models part 2

CIVENG/CYPLAN C88-LEC-001 > Files > Lectures > Lecture4

→ Instructors: Making Course Materials Accessible (click to expand)

Search for files | 0 items selected | + Folder | Upload

- Data Science for Smart Cities (Sprin
  - course_image
  - Lectures
    - Lecture1
    - Lecture2
    - Lecture3
    - Lecture4
  - unfiled
  - Uploaded Media

Name ▲	Date Created	Date Modified	Modified By	Size
Lecture4_Feb8.pdf	1:25pm	1:25pm		16.8 MB
SW_vs_Models.ipynb	1:25pm	1:25pm		287 KB

0% of 68.7 GB used

All My Files

Outline

- Review from next class (Random Graphs and Small Worlds)
- Participation Slides
- Barabasi Albert Model
- Time to discuss assignment 1, part 2

Features of real networks: small-world property

Most real-world networks are small worlds: they have **short paths** ($l \sim l_{random}$)

They also have **large clustering**

($C \gg C_{random}$)

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

To determine the if a network has a small property we read it as a undirected graph and compare its C and l with a Random Graph of the same number of nodes and links

How do we estimate the properties of a Random Graph with N nodes and L links?

Note: we use lower case l for shortest paths, not confusing it with number of links L

Random networks: summary

A Random Graph defined by (p,N), is built connecting each pair of nodes with probability p. its analytical properties were shown in Lecture 3. Summarized here:

- Degree distribution Bell shaped curve around average, $P(k) = e^{-<k>} \frac{<k>^k}{k!}$
- $<k_{random}>$=p(N-1)=2L/N, where L is the number of links and N and p are input parameters of the model
- Average shortest path length $<l_{random}>$~log(N)/log(k)
- Clustering Coefficient $<C_{random}>$ = p

How do we estimate the properties of a Random Graph with N nodes and L links?

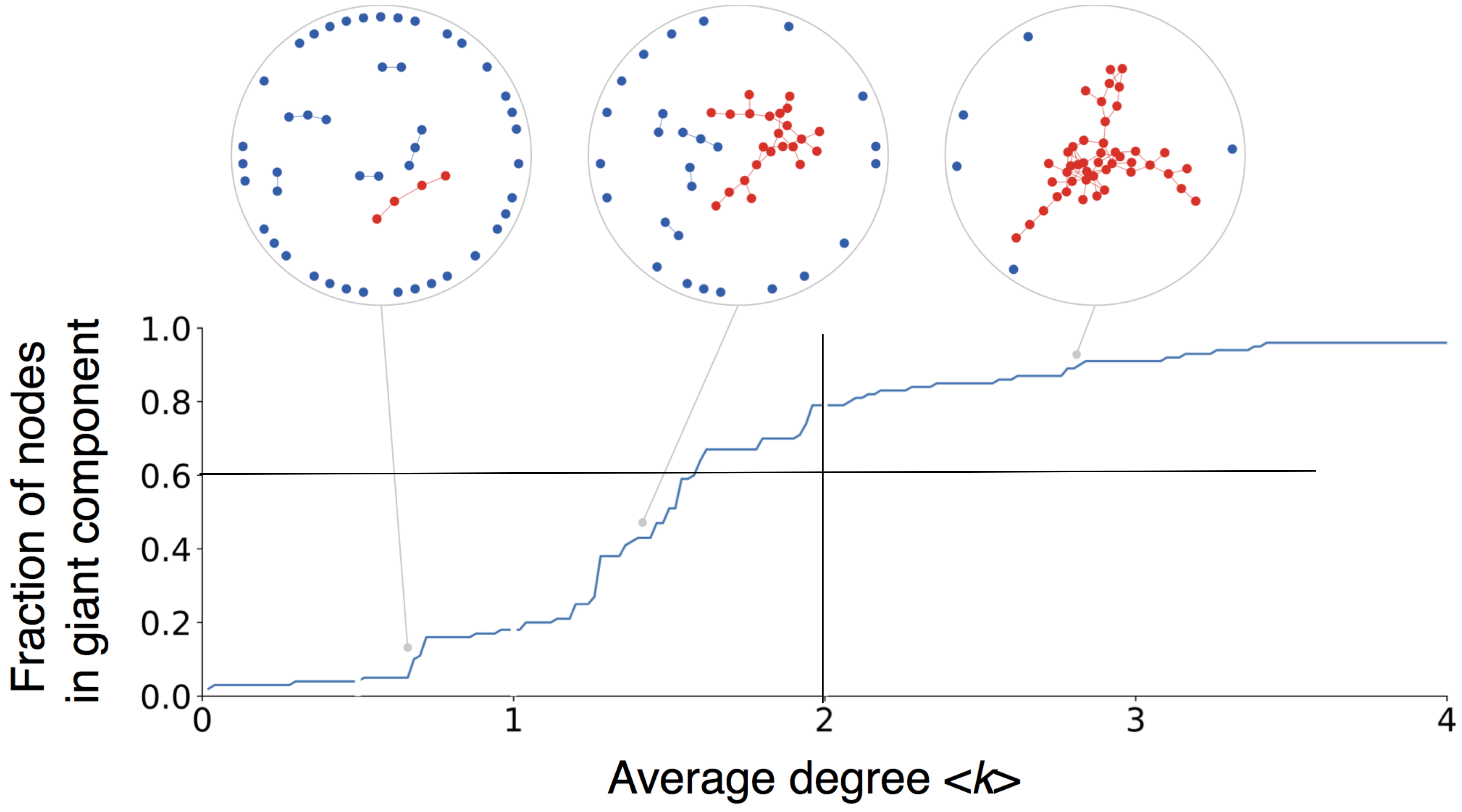
For any network $\langle k \rangle = 2L/N$, where L is the number of links and N the number of nodes

Knowing $\langle k \rangle$ we determine p, and knowing p and k we can determine C_{random} and l_{random}

We calculate p:
$p = \langle k \rangle / (N-1)$

Then:

$\langle C_{random} \rangle = p$
$\langle l_{random} \rangle \sim \log(N)/\log(k)$

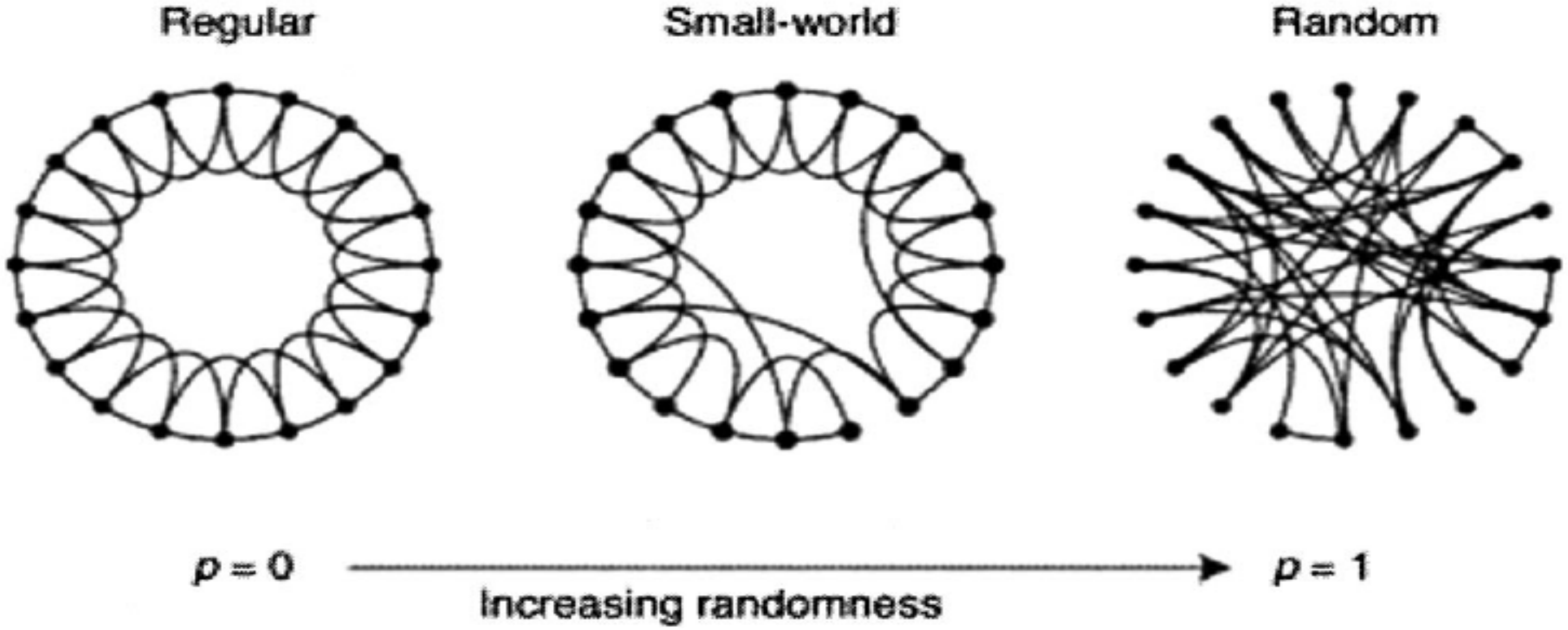


Note that for average degree larger than 1 (<k> > 1), the largest component also known as giant component contains a large fraction of the nodes.

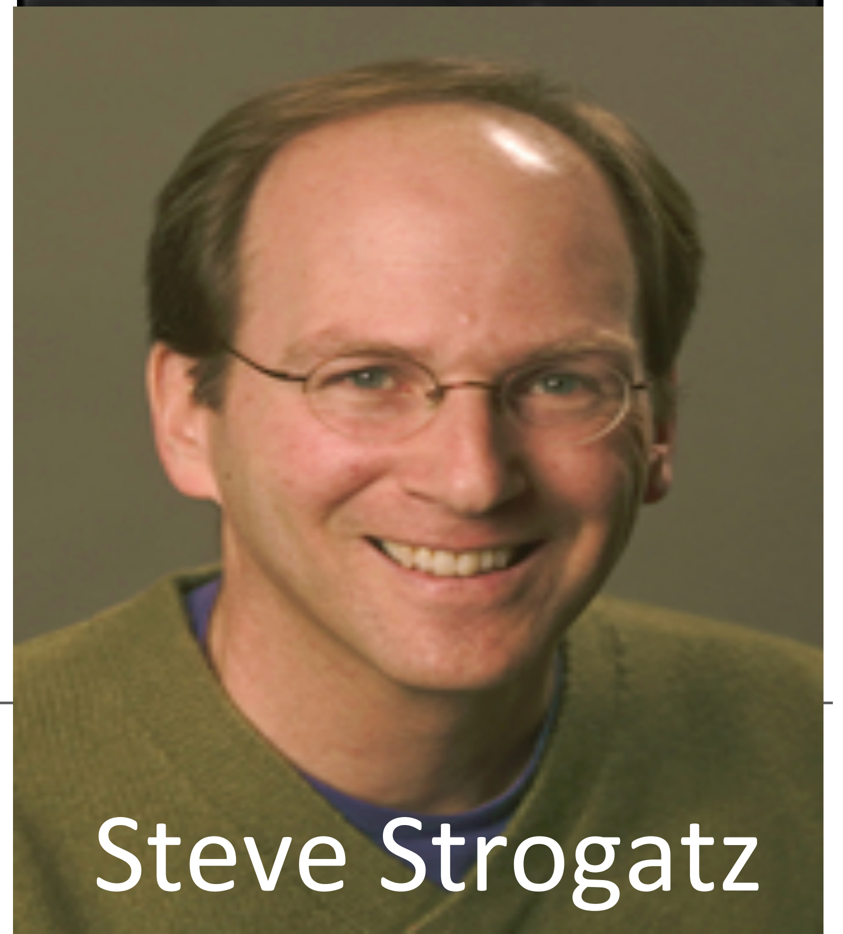
Collective dynamics of 'small-world' networks

letters to nature

Duncan J. Watts* & Steven H. Strogatz

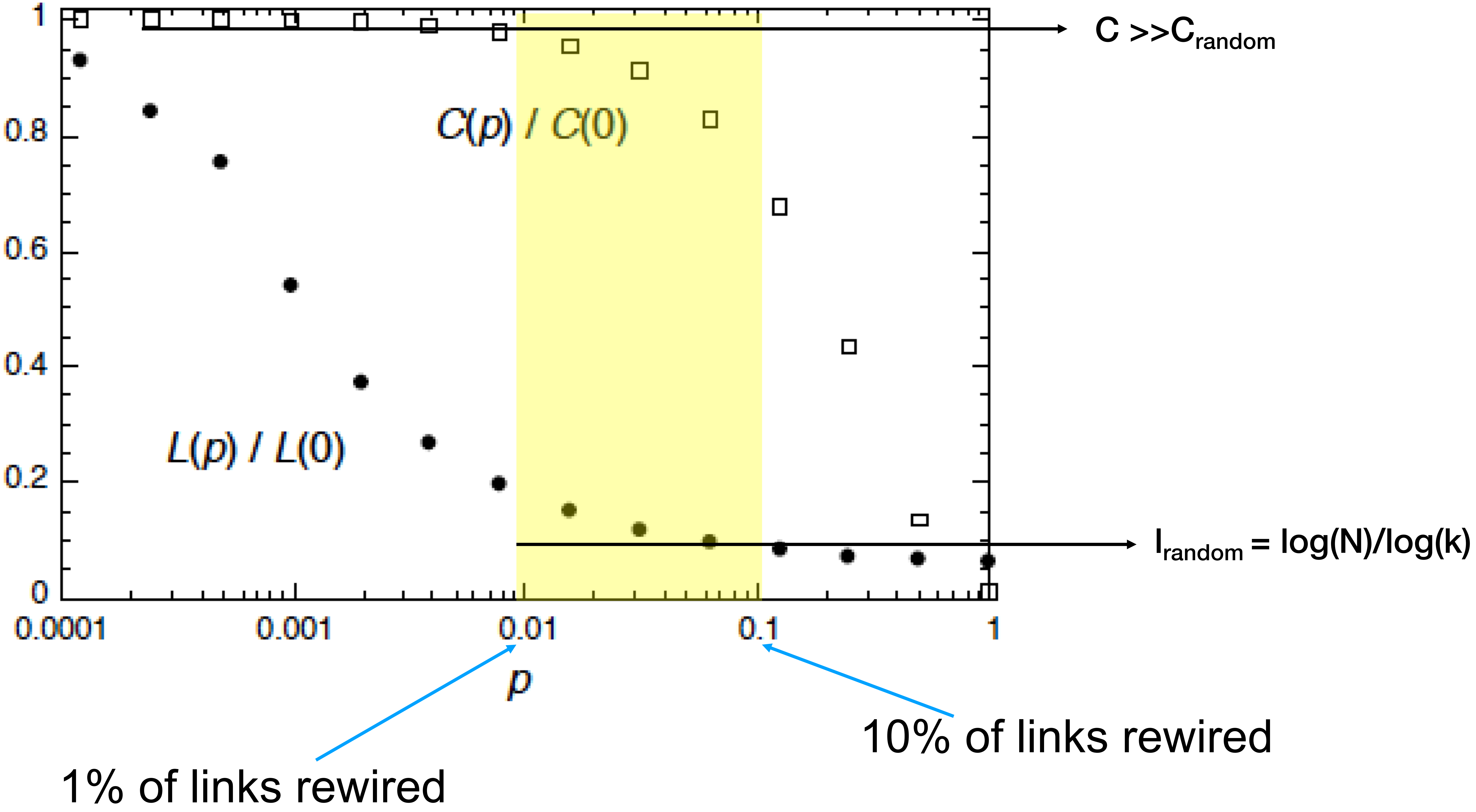


Duncan Watts



Parameters (N,p,k), number of nodes, number of neighbors in ring, rewiring probability

Watts/Strogatz model:
Change in clustering coefficient and average path length as a function of the proportion of rewired edges



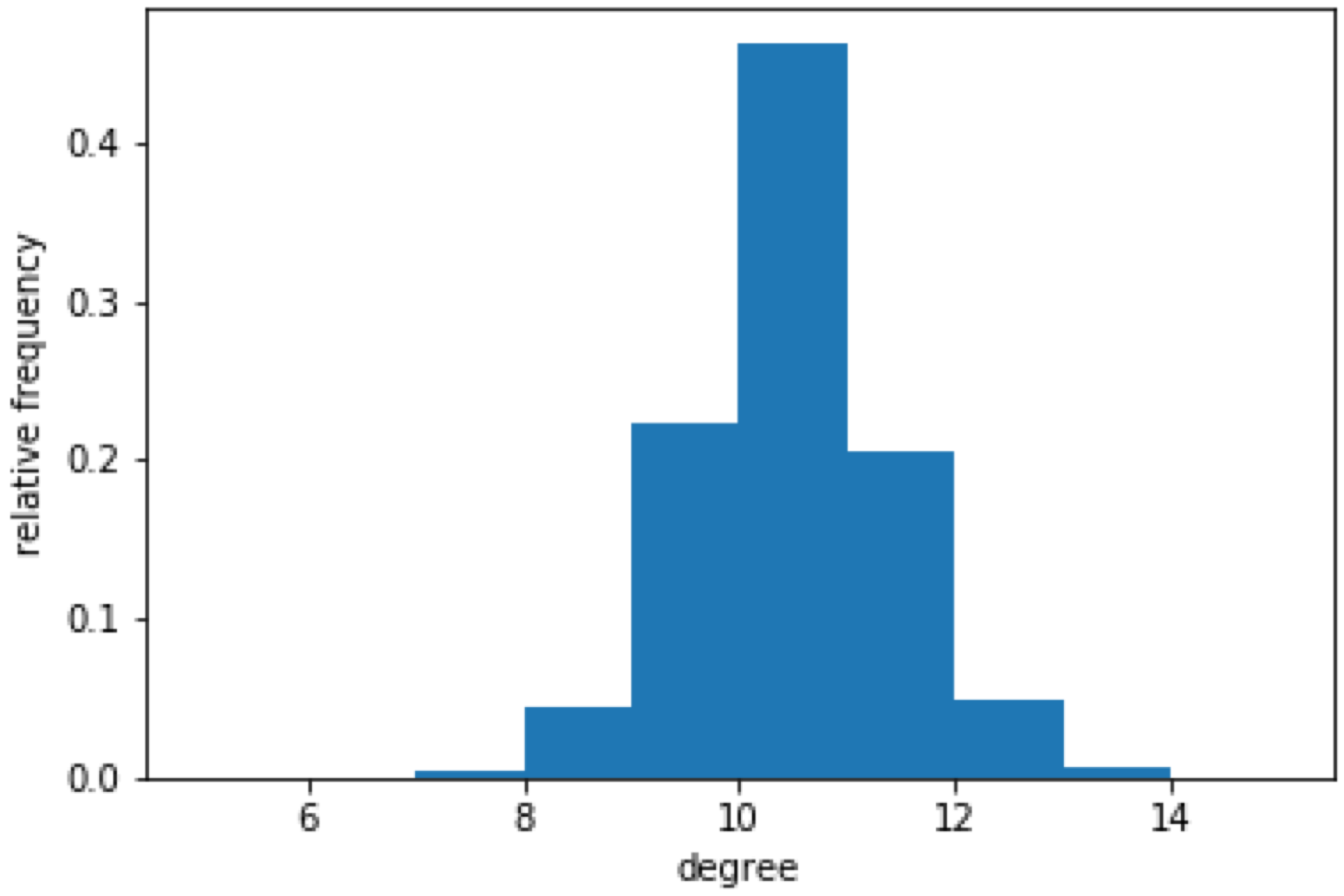
Between 1% and 10% rewired the shortest path decreases and the C stays high!!

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

The Watts-Strogatz model: degree distribution

- The degree distribution is peaked as most nodes have the same degree: **no hubs!**
- The Watts-Strogatz model fails to reproduce the broad degree distributions observed in many real-world networks

***N* = 10000, *k* = 10, *p* = 0.1**



Small World Model: summary

- Degree distribution Delta shaped curve around average
- <k> is an input parameter
- Average shortest path length $\langle l \rangle \sim \log(N)/\log(k)$ for p in the range (0.01,0.1), e.g. when we rewire a small fraction of the links
- Clustering Coefficient $C(p) = C(0)(1-p)^3$

(where C(0) is the clustering coefficient with p=0, that property was demonstrated by A. Barrat, M. Weigt. On the properties of small-world networks. *The European Physical Journal B* **13**, 547–560 (2000))

Let’s go to the participation slides

https://docs.google.com/presentation/d/1VflYeaaQqpa8Wys_i91q6KodpkEgwkGFowNM8R0_SS8/edit?usp=sharing

Network growth

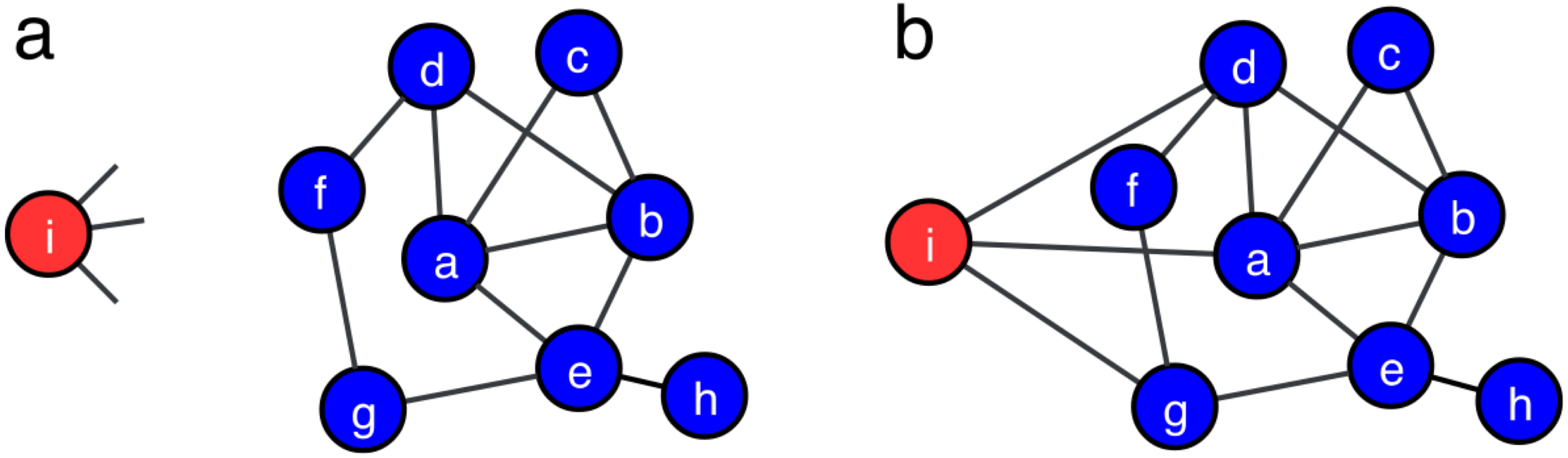
- **Note:** Real-world networks are dynamic!
- **Examples:**
  - The Web in 1991 had a single node, today there are trillions
  - Citation networks of scientific articles and collaboration networks of scientists keep growing due to the publication of new papers
  - The collaboration network of actors keeps growing due to the release of new movies
  - The protein interaction network has been growing over the course of 4 billion years: from a few genes to over 20,000

Network growth

- **General procedure:**

1. A new node comes with a given number of stubs, indicating the number of future neighbors of the node (degree)

2. The stubs are attached to some of the old nodes, according to some rule



Preferential attachment

- **Note:** Nodes prefer to link to the more connected nodes
- **Examples:**
    - Our knowledge of the Web is biased towards popular pages, which are highly linked, so it is more likely that our website points to highly linked Web sites
    - Scientists are more familiar with highly cited papers (which are often the most important ones), so they will tend to cite them more often than poorly cited ones in their own papers
    - The more movies an actor makes, the more popular they get and the higher the chances of being cast in a new movie

Which model?

- Our network model should have the following features:
  - **Growth:** the number of nodes grows in time following the addition of new nodes. The models considered so far are **static**
  - **Preferential attachment:** new nodes tend to be connected to the more connected nodes. The models considered so far set links among pairs of random nodes, regardless of their degree

Preferential attachment

- "***For to every one who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away***"

  —Gospel of Matthew 25:29

- **Take-home message:** the rich gets richer and the poor gets poorer!

The Barabási-Albert model

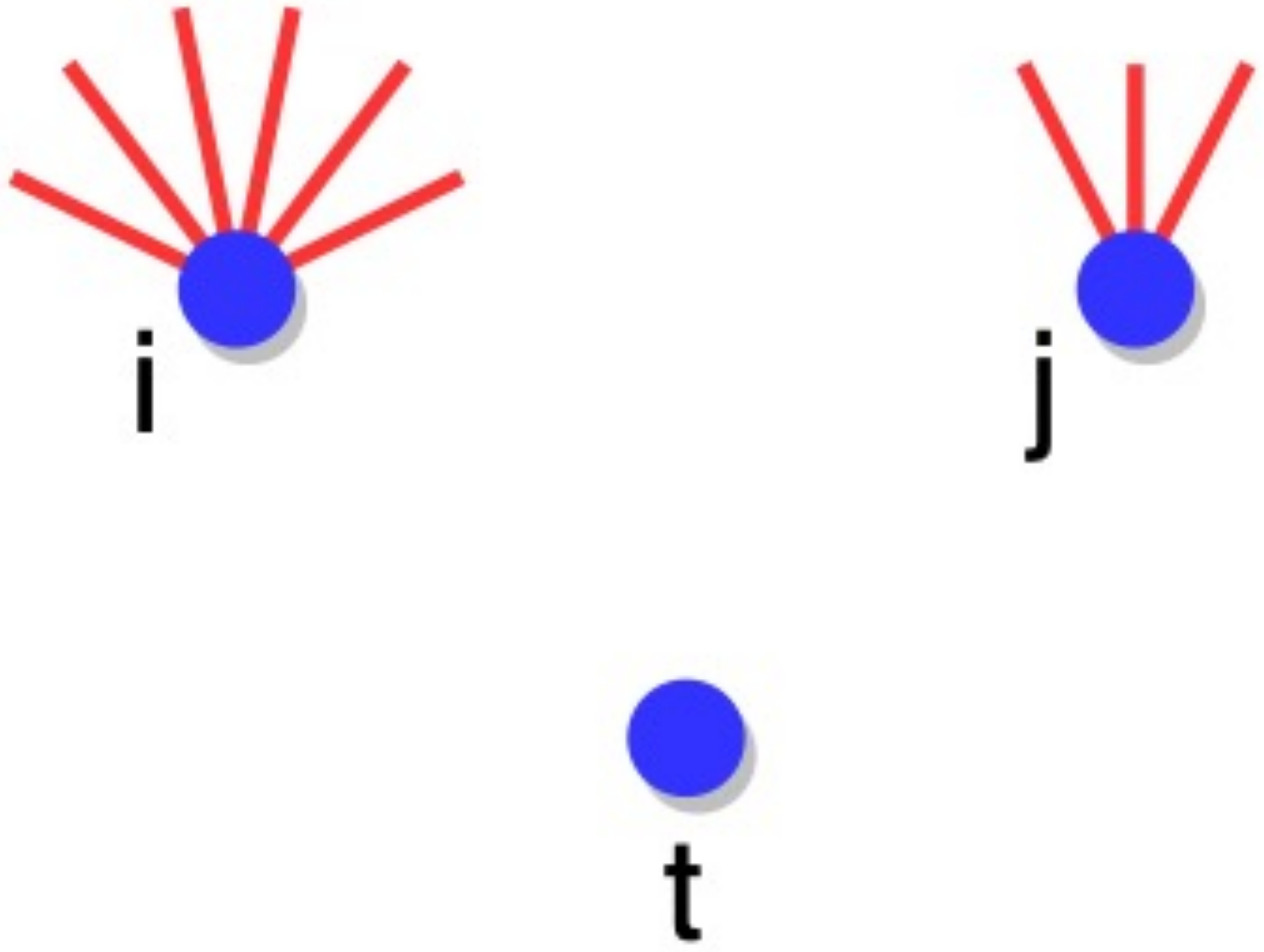
- **Procedure:**
  - Start with a group of m_0 nodes, usually fully connected (clique)
  - At each step a new node *i* is added to the system, and sets *m* links with some of the older nodes ($m \leq m_0$)
  - The probability that the new node *i* chooses an older node *j* as neighbor is **proportional to the degree k_j of *j*:**

$$\Pi(i \leftrightarrow j) = \frac{k_j}{\sum_l k_l}$$

  - The procedure ends when the given number *N* of nodes is reached

The Barabási-Albert model

Example: if *t* has to choose between node *i*, with degree 6, and node *j*, with degree 3, the probability of choosing *i* is twice the probability of choosing *j*



Picking nodes in Python

- **Question:** how do we pick nodes with a given probability?
- **Answer:** use the **random** module
- **Example:** picking list elements (*e.g.*, nodes) with the same probability, *i.e.*, completely at random:

```
import random
nodes  = [1 ,2, 3, 4]
selected_node = random.choice(nodes)
```

Picking nodes in Python

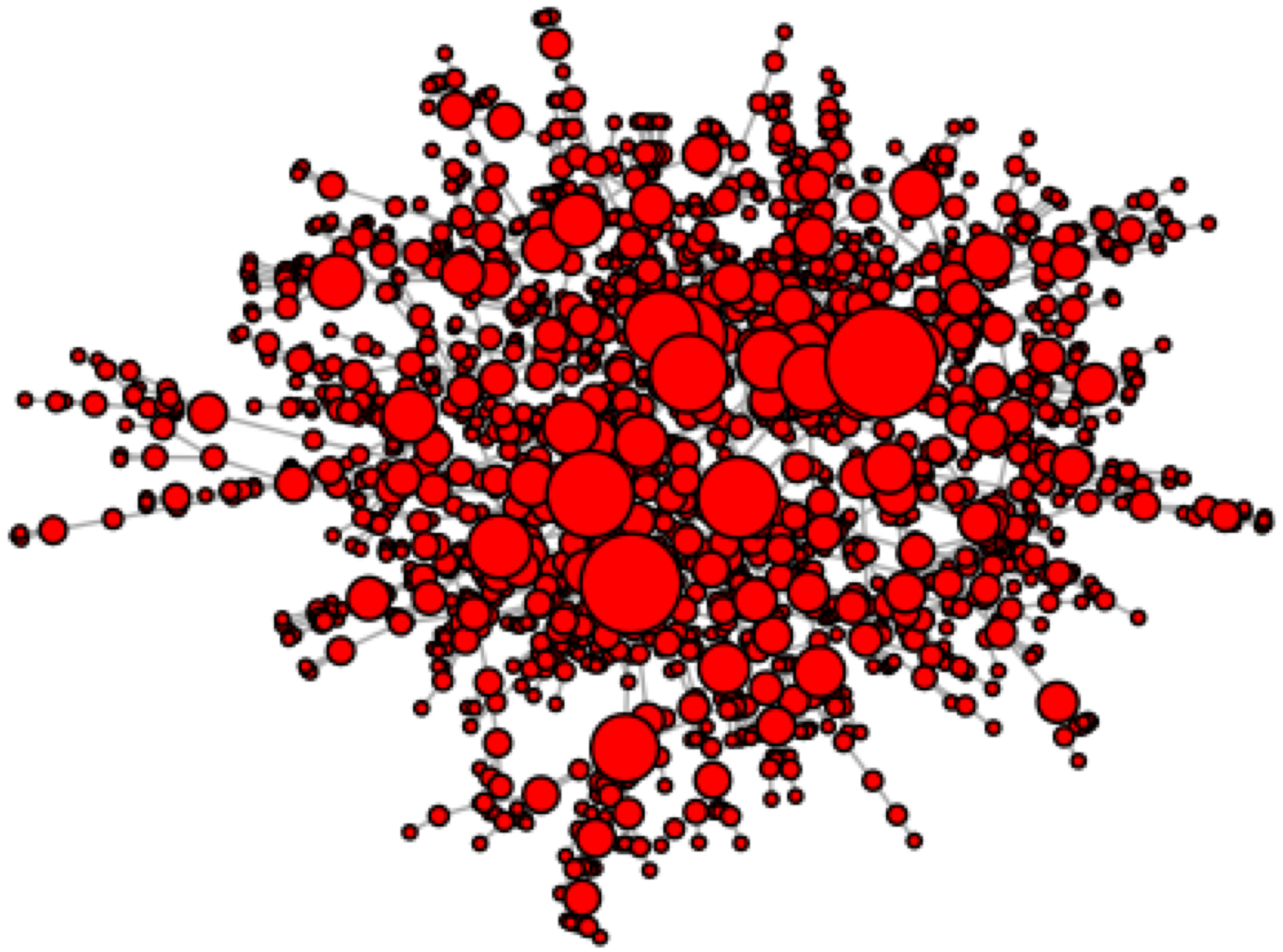
- **Question:** what if nodes have to be picked with different probabilities?
- **Answer:** we need to provide a second list, whose elements are the weights associated with the nodes
- **Note:** weights are used to calculate probabilities, but do not have to be integers or add up to one
- **Example:** picking nodes with probability proportional to their degrees, as in preferential attachment:

```
import random
nodes   = [1 ,2, 3, 4]
degrees = [3, 1, 2, 2]
selected_node = random.choice(nodes, degrees)
```

The Barabási-Albert model

- **Rich-gets-richer phenomenon:** due to preferential attachment, the more connected nodes have higher chances to acquire new links, which gives them a bigger and bigger advantage over the other nodes in the future!

- This is how **hubs** are generated



```
# BA model network
G =
nx.barabasi_albert_graph(N,Kmin)
```

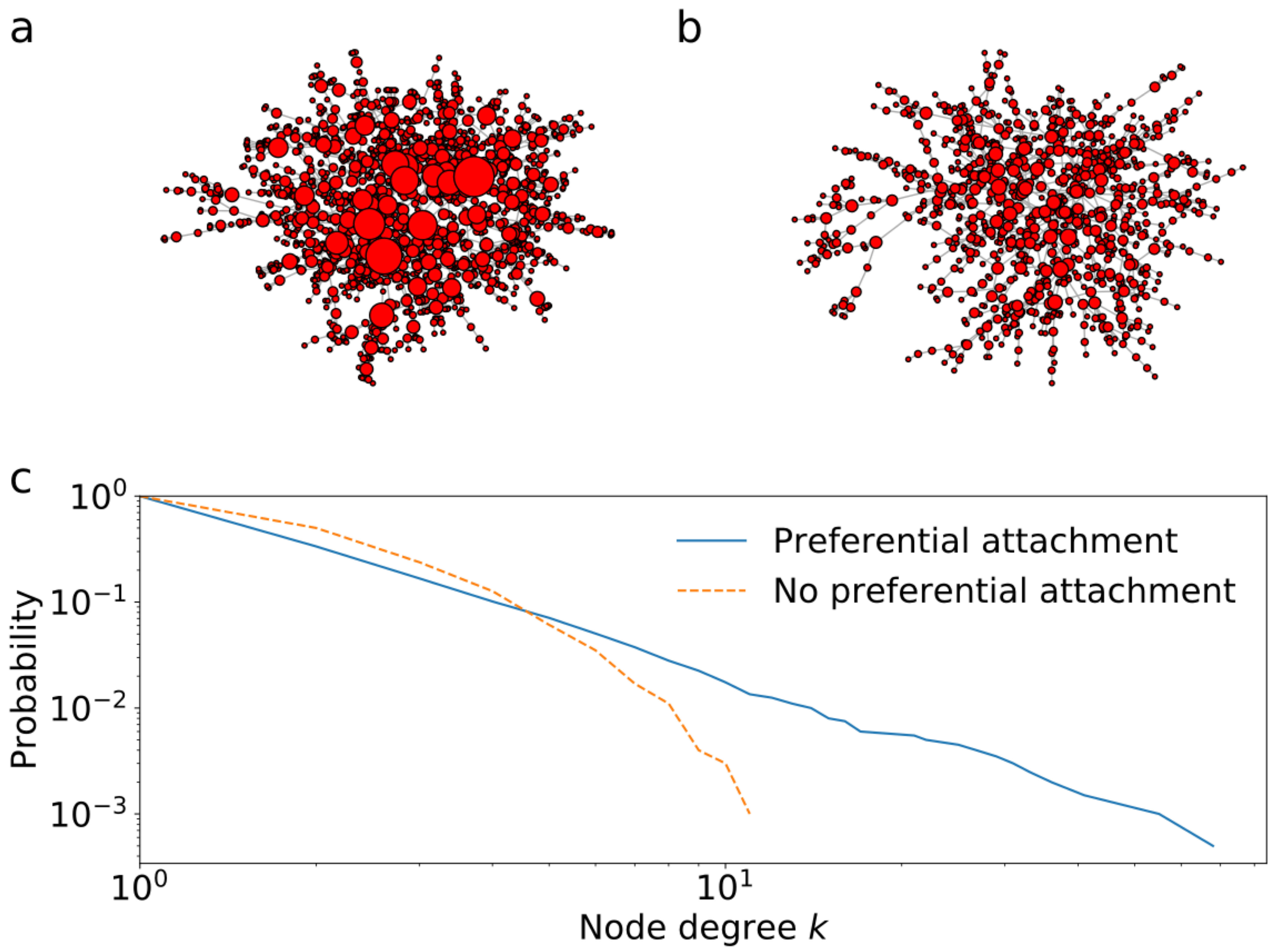
The Barabási-Albert model

- Hubs are the **oldest** nodes: they get the initial links and acquire an advantage over the other nodes, which increases via preferential attachment

- **Question:** if old nodes have an advantage over newer nodes anyway, do we need preferential attachment at all? Can we explain the existence of hubs just because of growth?

- **Alternative model:** each new node chooses its neighbors at random, not with probability proportional to their degree

The Barabási-Albert model



Conclusion: growth + random attachment does not generate hubs.

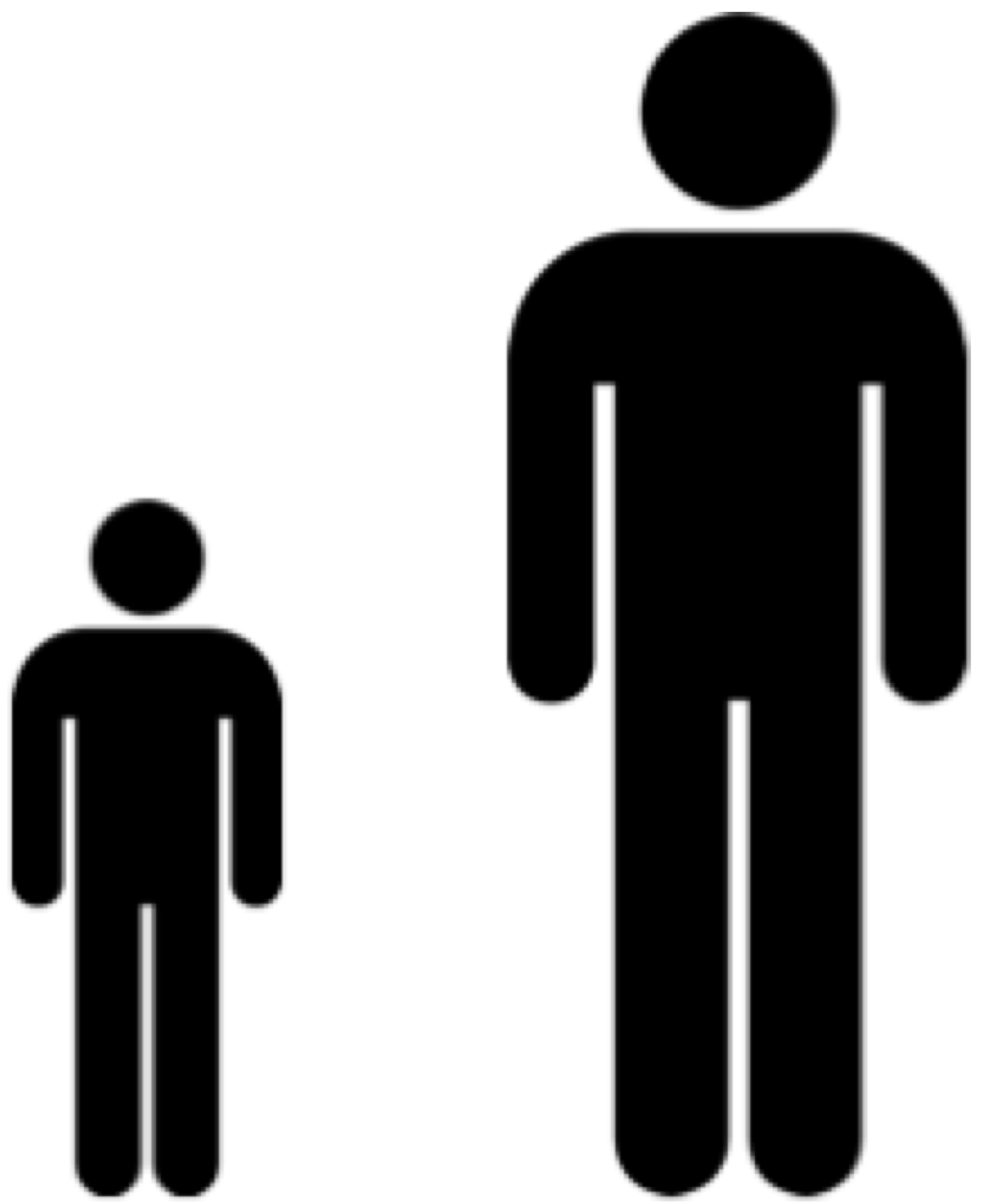
Preferential attachment is necessary!

Extensions of the BA model: Rank model

- **Pitfall of preferential attachment:** BA model implies that nodes have a perception of how important other nodes are, *i.e.*, how large is their degree

- **Objection:** in the real world there is no such perception of the absolute value of things, **it is far easier to perceive the relative value!**

- **Solution:** ranking!



Extensions of the BA model: Rank model

- **Procedure:**
  - Nodes are ranked based on a property of interest (*e.g.*, age, degree). The rank of node i is R_i
  - Start with a group of m_0 nodes, usually fully connected (clique)
  - At each step a new node i is added to the system, and sets m links with some of the older nodes ($m \leq m_0$)
  - The probability that the new node i chooses an older node j as neighbor is **proportional to a power of the rank of *j*:**

$$\Pi(i \leftrightarrow j) = \frac{R_j^{-\alpha}}{\sum_l R_l^{-\alpha}}$$

SW_vs_Models.ipynb

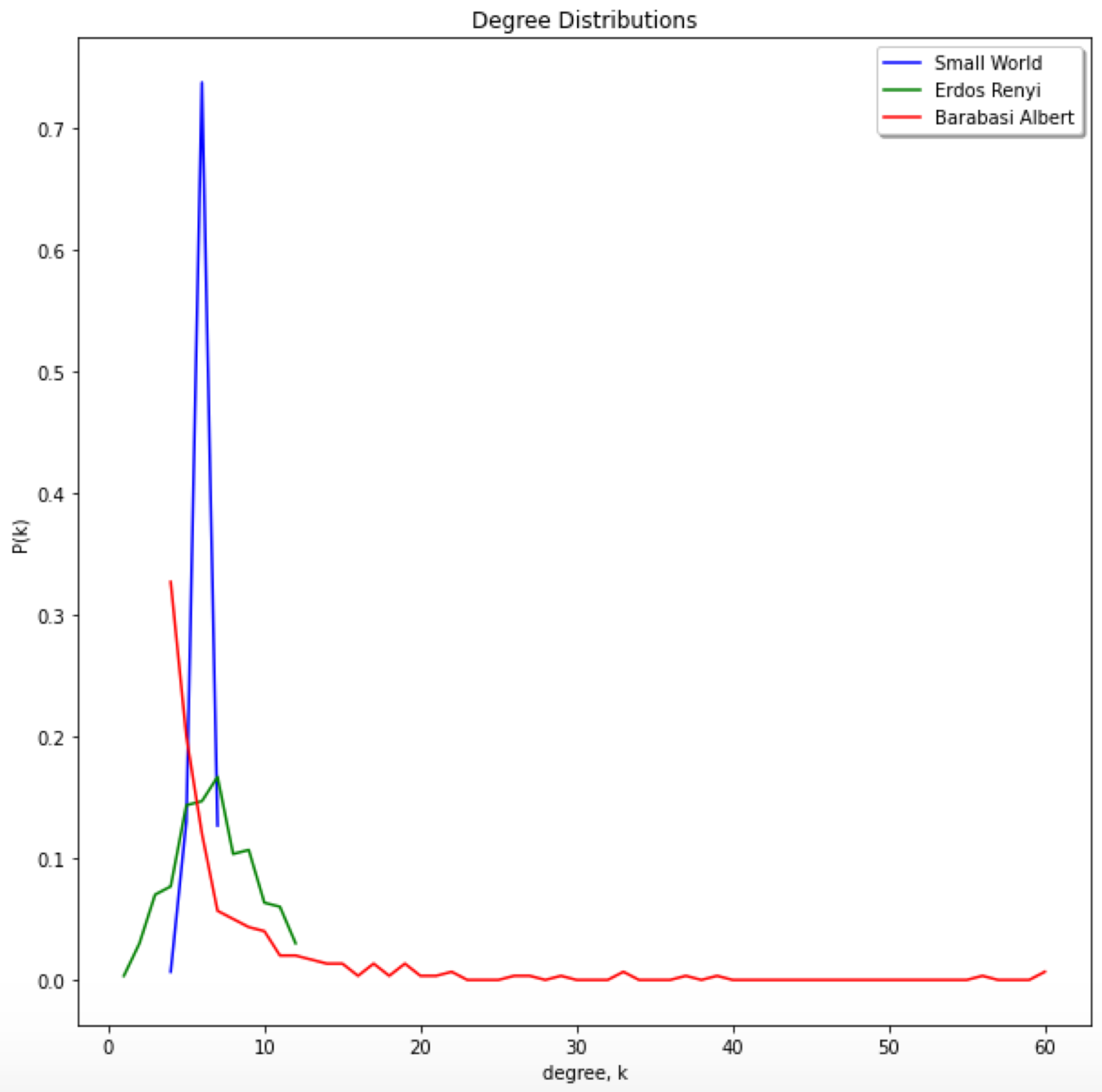
Let's try to generate a Barabasi-Albert network model with 1000 links 300 nodes and clustering coefficient of 0.5

In [30]:
```
Gba=nx.barabasi_albert_graph(300, 4,seed=123)
```

Let's compare the degree distributions

In [152]:
```
#G1:
degs1 = list(dict(nx.degree(gs)).values())
n1, bins1 = np.histogram(degs1, bins = list(range(min(degs1), max(degs1)+1, 1)), density="True")

#G2:
degs2 = list(dict(nx.degree(Ger)).values())
n2, bins2 = np.histogram(degs2, bins = list(range(min(degs2), max(degs2)+1, 1)), density="True")

#G3:
degs3 = list(dict(nx.degree(Gba)).values())
n3, bins3 = np.histogram(degs3, bins = list(range(min(degs3), max(degs3)+1, 1)), density="True")

#to plot:
plt.figure(figsize=(10,10)) #use once and set figure size

plt.plot(bins1[:-1],n1,'b-', markersize=10, label="Small World")
plt.plot(bins2[:-1],n2,'g-', markersize=10, label="Erdos Renyi")
plt.plot(bins3[:-1],n3,'r-', markersize=10, label="Barabasi Albert")
plt.legend(loc='upper right', shadow=True)
plt.title('Degree Distributions')
plt.xlabel('degree, k')
plt.ylabel('P(k)')
```

Out[152]:
```
Text(0, 0.5, 'P(k)')
```



The degree distribution is a negative power of k.

They are also called power law networks

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

Let's find C of the Degree Distribution

Continuum Formalism
In analytical calculations it is often convenient to assume that the degrees can take up any positive real value:

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

$$\int_{k_{\min}}^{\infty} p(k)dk = 1$$

$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$

Note: For the Barabasi Albert model γ=3

$$p(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}.$$

$$\int_{k_1}^{k_2} p(k)dk$$

Divergences in scale-free distributions

$$P(k) = Ck^{-\gamma} \qquad k = [k_{\min}, \infty) \qquad \int_{k_{\min}}^{\infty} P(k)dk = 1 \qquad C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma}\, dk} = (\gamma - 1)k_{\min}^{\gamma-1}$$

$$P(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}$$

$$< k^m >= \int_{k_{\min}}^{\infty} k^m P(k)dk \qquad < k^m >= (\gamma - 1)k_{\min}^{\gamma-1} \int_{k_{\min}}^{\infty} k^{m-\gamma}\, dk = \frac{(\gamma - 1)}{(m - \gamma + 1)} k_{\min}^{\gamma-1} \left[k^{m-\gamma+1} \right]_{k_{\min}}^{\infty}$$

If m-γ+1<0: $$< k^m >= -\frac{(\gamma - 1)}{(m - \gamma + 1)} k_{\min}^{m}$$

If m-γ+1>0, the integral diverges.

For a fixed γ this means that all moments with m>γ-1 diverge.

DIVERGENCE OF THE HIGHER MOMENTS

$$< k^m >= (\gamma - 1) k_{\min}^{\gamma-1} \int_{k_{\min}}^{\infty} k^{m-\lambda}\, dk = \frac{(\gamma-1)}{(m-\gamma+1)} k_{\min}^{\gamma-1} \left[k^{m-\gamma+1} \right]_{k_{\min}}^{\infty}$$

For a fixed λ this means all moments m>γ-1 diverge.

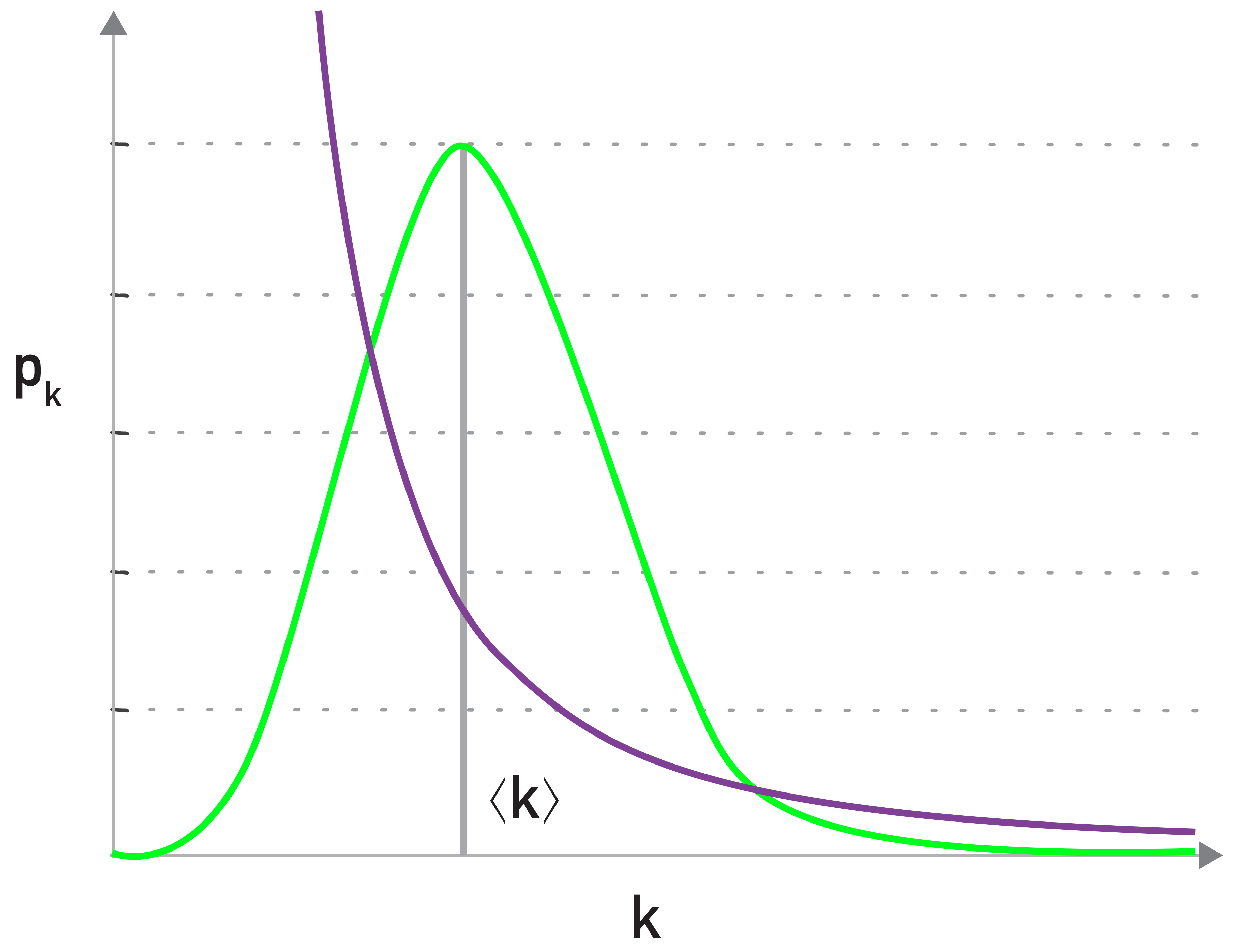
Network	Nodes (N)	Links (L)	Average degree ($\langle k \rangle$)	Maximum degree (k_{max})	Heterogeneity parameter (κ)
Facebook Northwestern Univ.	10,567	488,337	92.4	2,105	1.8
IMDB movies and stars	563,443	921,160	3.3	800	5.4
IMDB co-stars	252,999	1,015,187	8.0	456	4.6
Twitter US politics	18,470	48,365	2.6	204	8.3
Enron Email	36,692	367,662	10.0	1,383	14.0
Wikipedia math	15,220	194,103	12.8	5,171	38.2
Internet routers	190,914	607,610	6.4	1,071	6.0
US air transportation	546	2,781	10.2	153	5.3
World air transportation	3,179	18,617	11.7	246	5.5
Yeast protein interactions	1,870	2,277	2.4	56	2.7
C. elegans brain	297	2,345	7.9	134	2.7
Everglades ecological food web	69	916	13.3	63	2.2

Many degree exponents are smaller than 3

→ $<k^2>$ diverges in the N→∞ limit!!!

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

THE MEANING OF SCALE-FREE



Random Network

Randomly chosen node: $k = \langle k \rangle \pm \langle k \rangle^{1/2}$

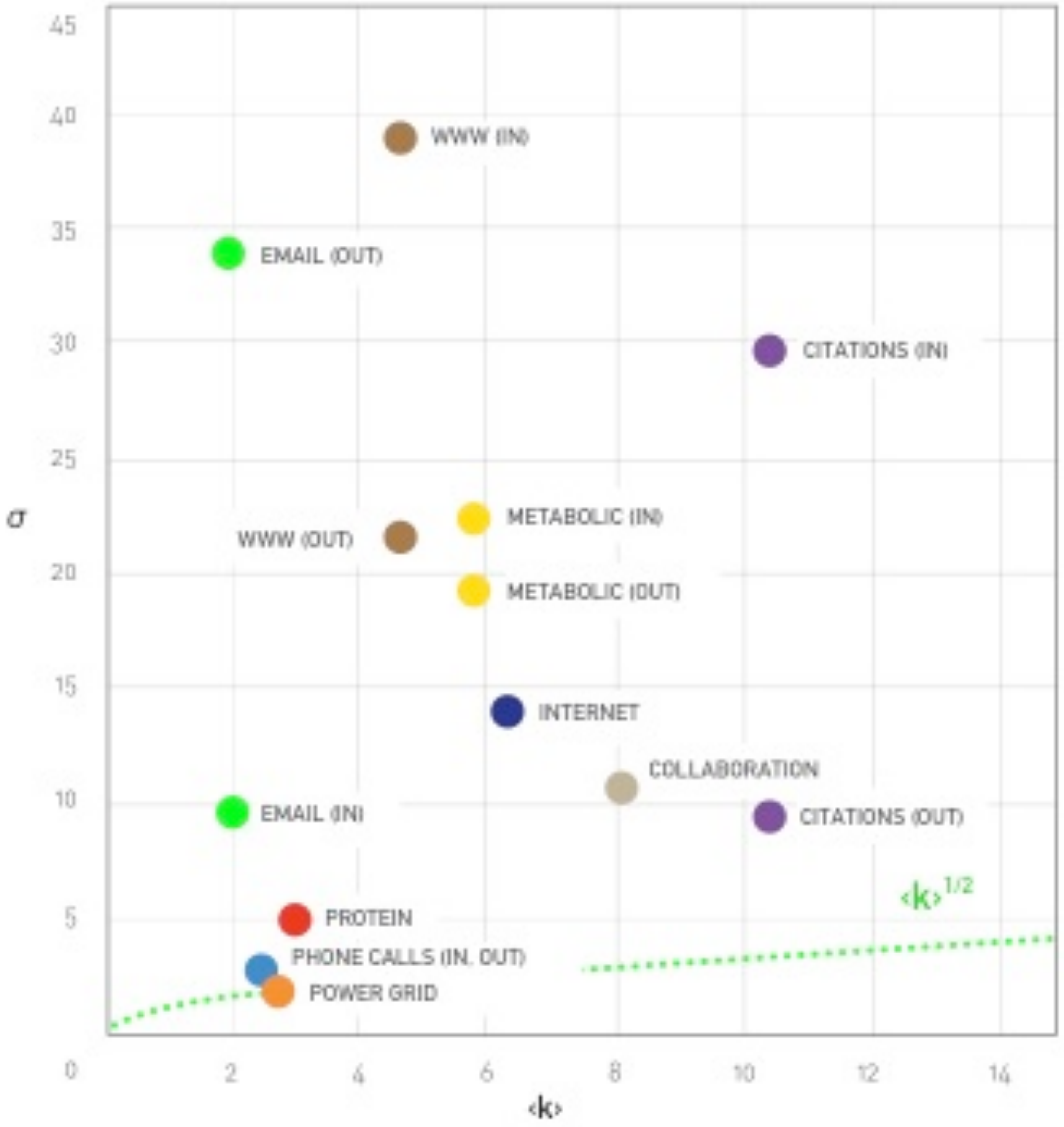
Scale: $\langle k \rangle$

Scale-Free Network

Randomly chosen node: $k = \langle k \rangle \pm \infty$

Scale: none

THE MEANING OF SCALE-FREE



$$k = \langle k \rangle \pm \sigma_k$$

Barabasi Model: summary

- Input parameters are (N, k_{min})
- $\langle k \rangle = 2k_{min}$ (from slide 27)
- Degree distribution power law $P(k) = 2k_{min}k^{-3}$
- Average shortest path length $\langle l \rangle \sim \log(N)/\log\langle k \rangle$
- Clustering Coefficient $C \sim \langle k \rangle / N$

(where C(0) is the clustering coefficient with p=0, that property was demonstrated by A. Barrat, M. Weigt. On the properties of small-world networks. *The European Physical Journal B* **13**, 547–560 (2000))

For next class

- Do assignment 1, part 2 (Due Feb 15th)

https://bcourses.berkeley.edu/courses/1511587/assignments/8402098