Chapter 5: Network Models part 1

A First Course in Network Science

Tracking Participation

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

O3:O65 =sum(C3:N3)

Participation Slides:https://bcourses.berkeley.edu/courses/1523051/collaborations

Name	Email Address	Lec1	Lec2	Lec3	Lec4	Lec5	Lec6	Lec7	Lec8	Lec9	Lec10	Lec11	Lec12	Total
Chintakrindi, Saroj	sarojchin@berkeley.edu													0
Wang, Sharon	hydroblading@berkeley.edu													0
Rastogi, Ankit	ankit.rastogi.1010@berkeley.edu													0
Sixto, Ileana	ileanas@berkeley.edu													0
Lee, Joanna	joannalee@berkeley.edu													0
Bustami, Jenna	jennabustami@berkeley.edu													0
Kim, Taeheon	kimtaeheon@berkeley.edu													0
Anaya, Daniel	danielanaya@berkeley.edu													0
Schroth, Kendall	kschroth@berkeley.edu													0
Cabrera Guzman, Wilmer	wilmercabrera@berkeley.edu													0
Darmawangsa, Madeline	madelinedm@berkeley.edu													0
McEvoy, Ash	amcevoy@berkeley.edu													0
Shin, Yoon Young	yyshin@berkeley.edu													0
Brown, Alexander	abrown44@berkeley.edu													0
Bobick, Samuel	samuelbobick@berkeley.edu													0

Add a 1pt for your participation in the participation slides of previous lectures

Here: https://docs.google.com/spreadsheets/d/11CVp8CrWftGz1vsWQxaMIUi4OGJ6mCaOXtvIcrHRFrE/edit?usp=sharing

Search for files

0 items selected

+ Folder

Upload

- Data Science for Smart Cities (Spring 202
  - course_image
  - Lectures
    - Lecture 1
    - Lecture 2
    - Lecture 3
  - Syllabus+Installation Instructions
  - Uploaded Media

Name ▲	Date Created	Date Modified	Modified By	Size	Accessibility
Lecture3_Jan31.pdf	1:27pm	1:27pm		4.9 MB	
SW_vs_Models.ipynb	1:27pm	1:27pm		287 KB	
watts-collective_dynamics-nature_1998.pdf	Jan 24, 2023	Jan 24, 2023	Marta Gonzalez	266 KB	

0% of 68.7 GB used

All My Files

Outline

- Degree distributions
- Random networks
- Small-world networks

https://docs.google.com/presentation/d/1s3TWinuhc5-nHp5EHbeCljM0rDlqbmPKZVbnMcA2-kY/edit?usp=sharing

Features of real networks: small-world property

Most real-world networks are small worlds: **short paths**

Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron Email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

Features of real networks: high clustering coefficient

- The **clustering coefficient** of a node is the **fraction of pairs of the node's neighbors that are connected to each other:**

$$C(i) = \frac{\tau(i)}{k_i(k_i - 1)/2} = \frac{2\tau(i)}{k_i(k_i - 1)}$$

where $\boldsymbol{\tau}(i)$ is the number of triangles involving *i.* Note that in this definition, the clustering coefficient is undefined if $k_i < 2$: a node must have at least degree 2 to have any triangles. However NetworkX assumes C=0 if $k=0$ or k=1

Features of real networks: high clustering coefficient

- Many networks have high clustering coefficients

- Other networks, e.g., bipartite and tree-like networks, have low clustering coefficient

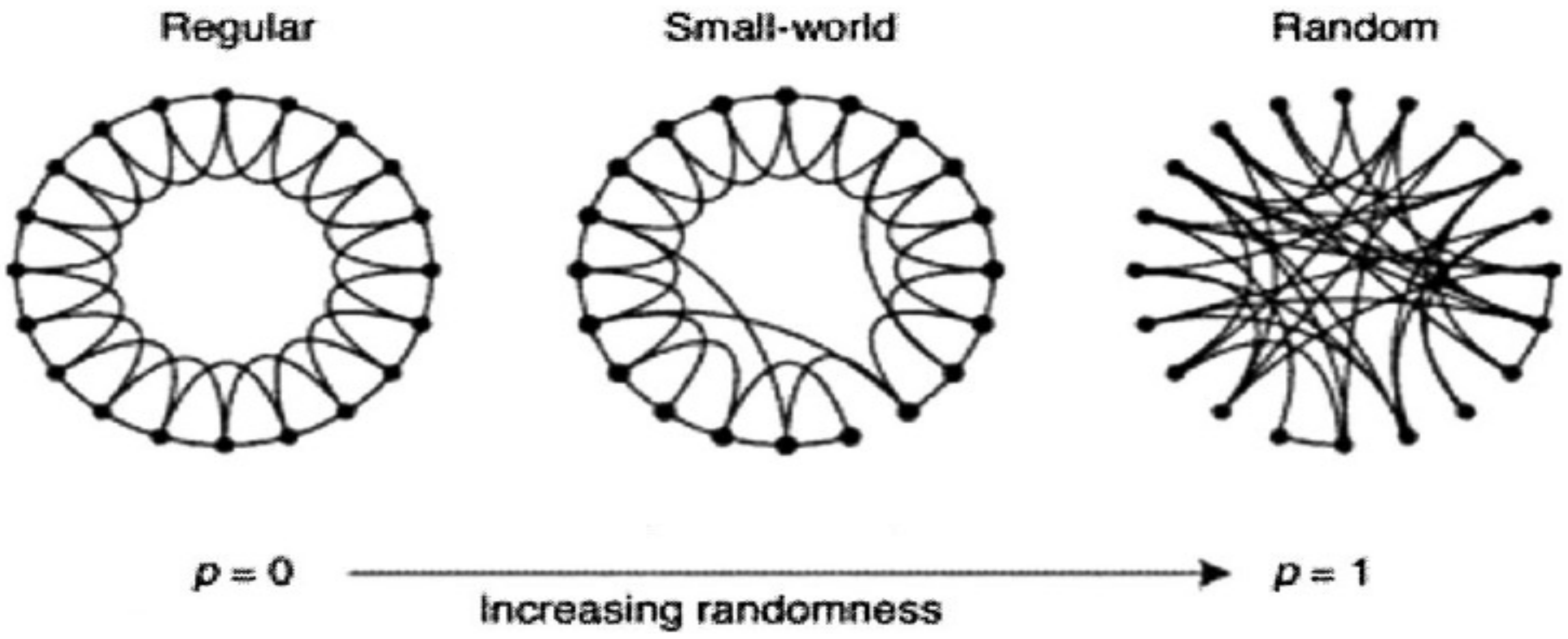
Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron Email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

Models

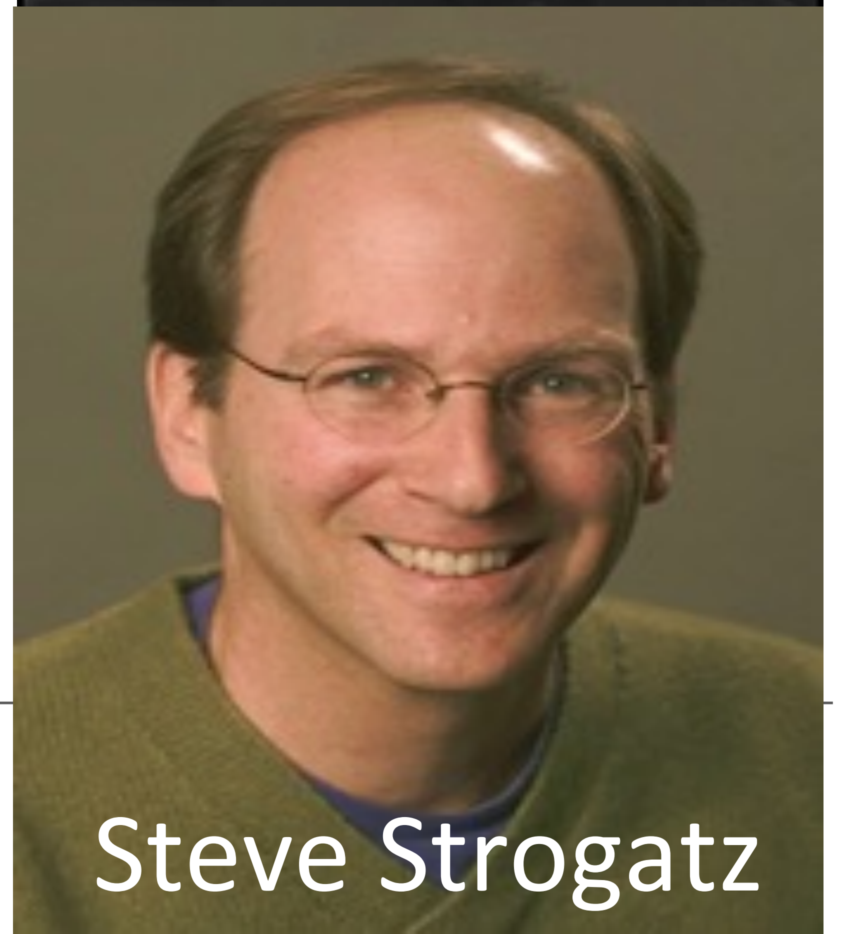
- **Network Model:** set of instructions to build networks
- **Goal:** find models that generate networks with the same characteristics as real-world networks

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

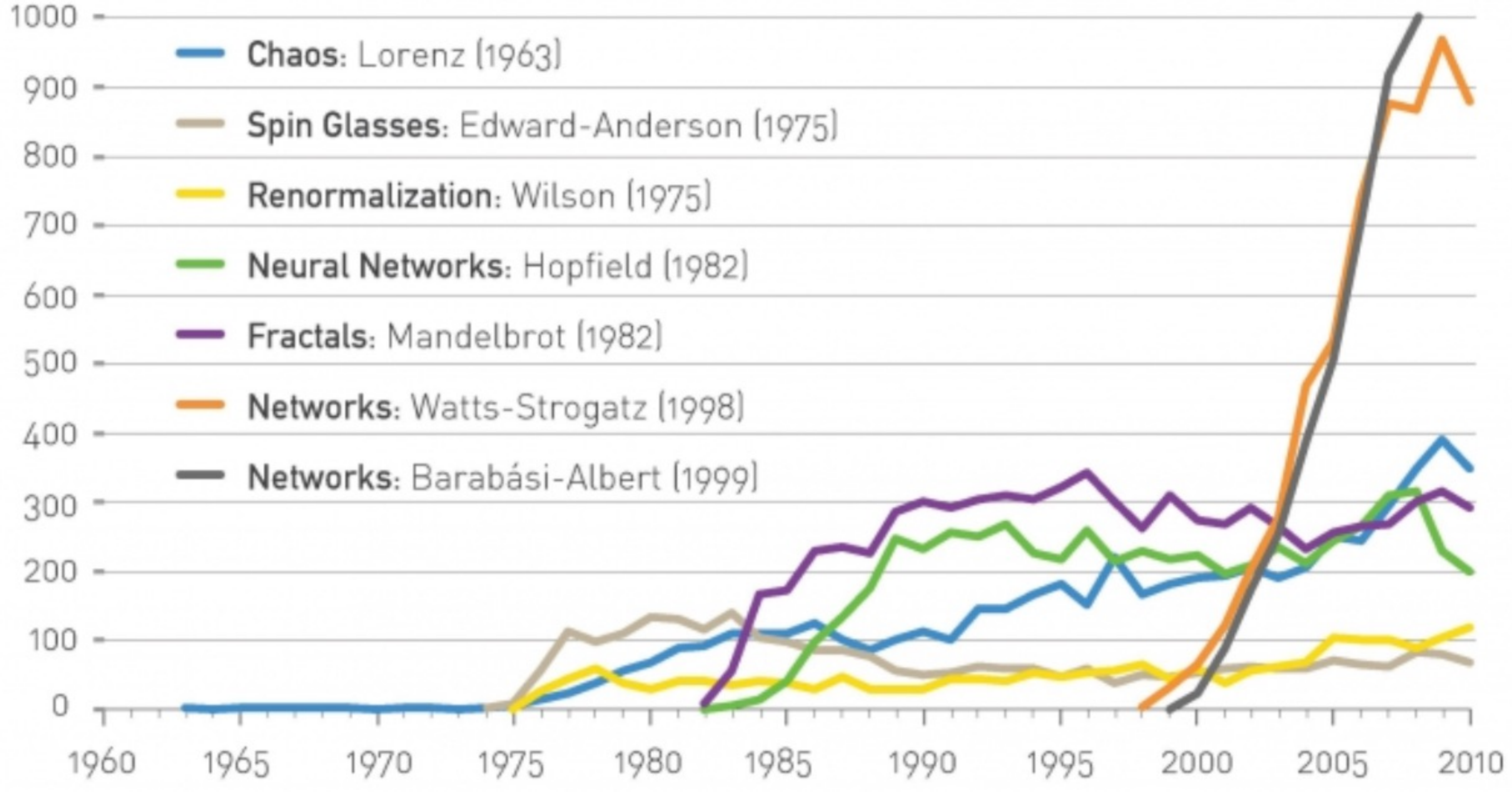


Duncan Watts



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

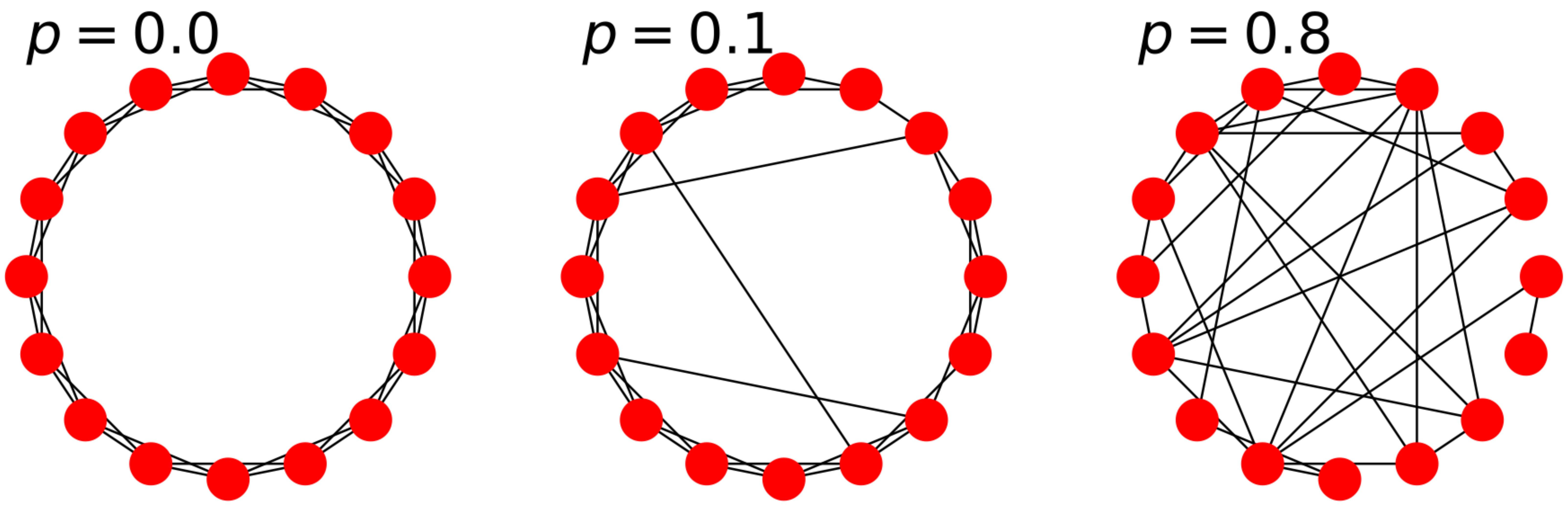
Network Science was born



The Watts-Strogatz model

Start with N nodes form a regular ring lattice, with even degree *k*.

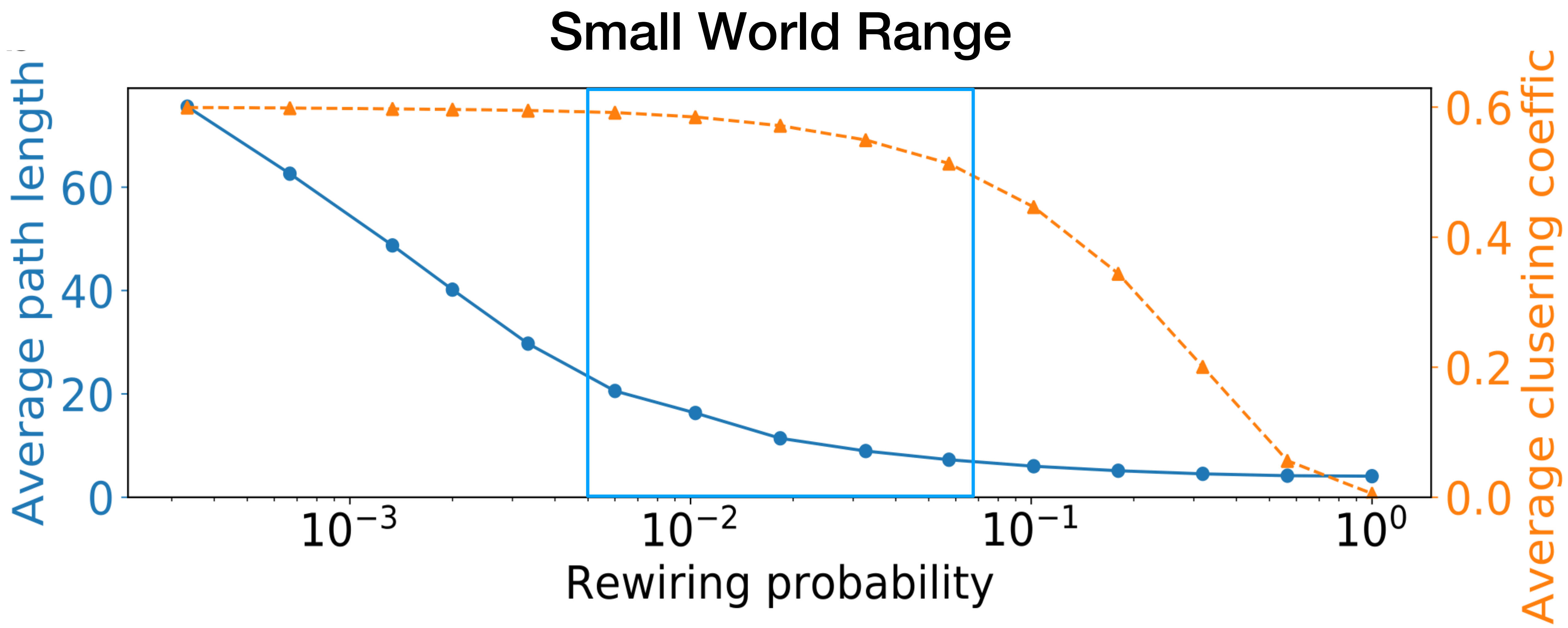
*Then w*ith probability *p*, each link is rewired randomly



Initial state: A ring when each node connected to their k neighbors, k is an even number
Then: visit each of the N nodes and shuffle each of its links with probability p
The figure shows 3 models (N=16,k=4,p=0); (N=16,k=4,p=0.1) and (N=16,k=4,p=0.8)

The Watts-Strogatz model

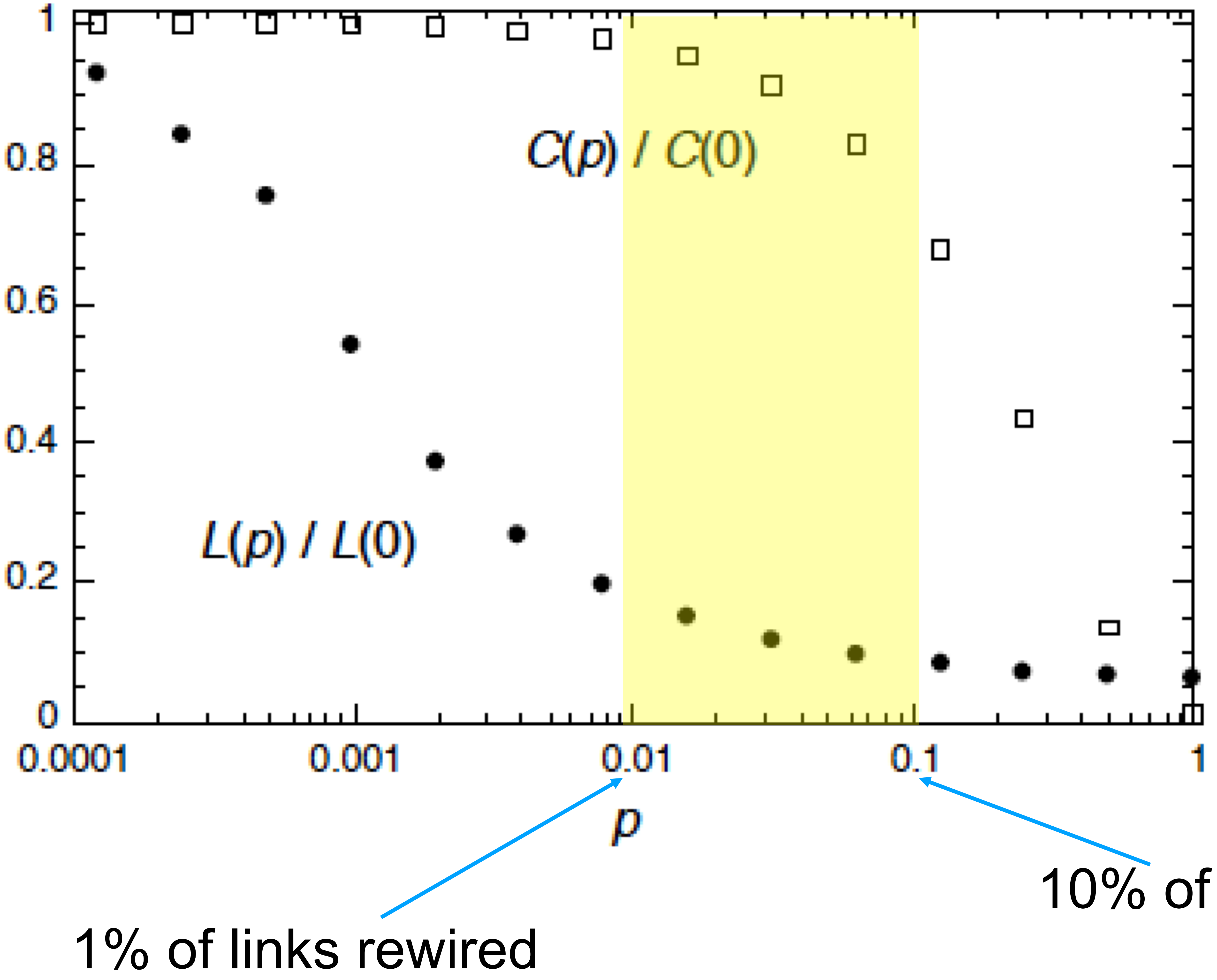
- The expected number of rewired links is $pL = pNk/2$
  - If $p = 0$, no links are rewired: **no change**
  - If p is small, few links are rewired: **the average clustering coefficient stays approximately the same because very few triangles are destroyed, but distances shrink considerably**
  - If $p = 1$, all links are rewired: **the network becomes a random network**



Distances become short already for low values of p; the average clustering coefficient stays high up to large values of p. **There is a range of values of p where the average path length is short and the clustering coefficient is high!**

Remember from last class that <k>=2L/N, L is the number of links N is the number of nodes

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



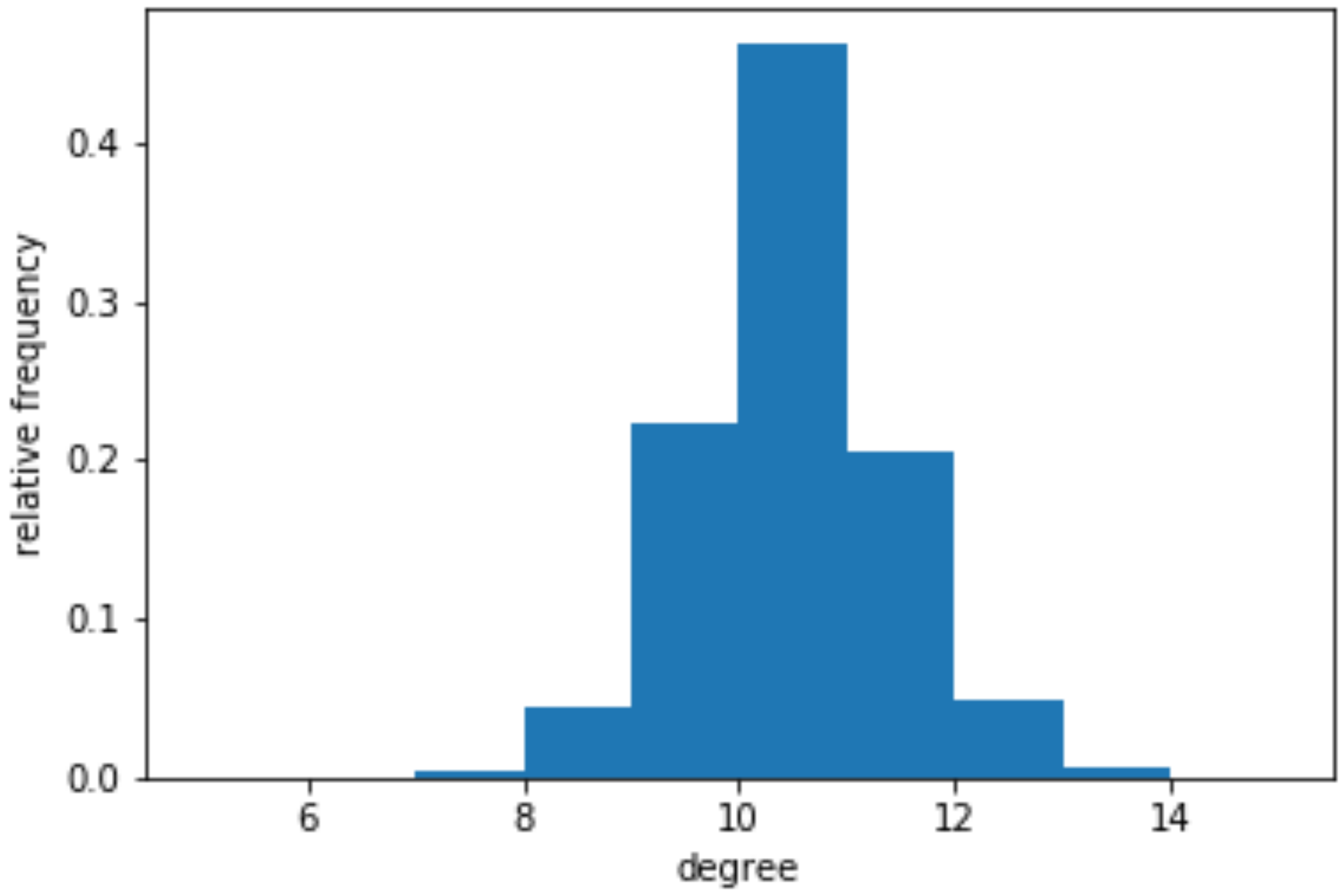
Between 1% and 10% rewired the shortest path decreases and the C stays high!!

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 <l>~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

See this result demonstrated in SW_vs_Models.ipynb

(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))

The Watts-Strogatz model: summary

- A regular lattice whose links are randomly rewired, with some probability *p*
- There is a range of values of the rewiring probability *p* for which distances between pairs of nodes are short (small-world property) and the average clustering coefficient is high: **good!**
- The nodes have approximately the same degree, there are no hubs: **bad!**

```
# small-world model network
G = nx.watts_strogatz_graph(N,k,p)
```

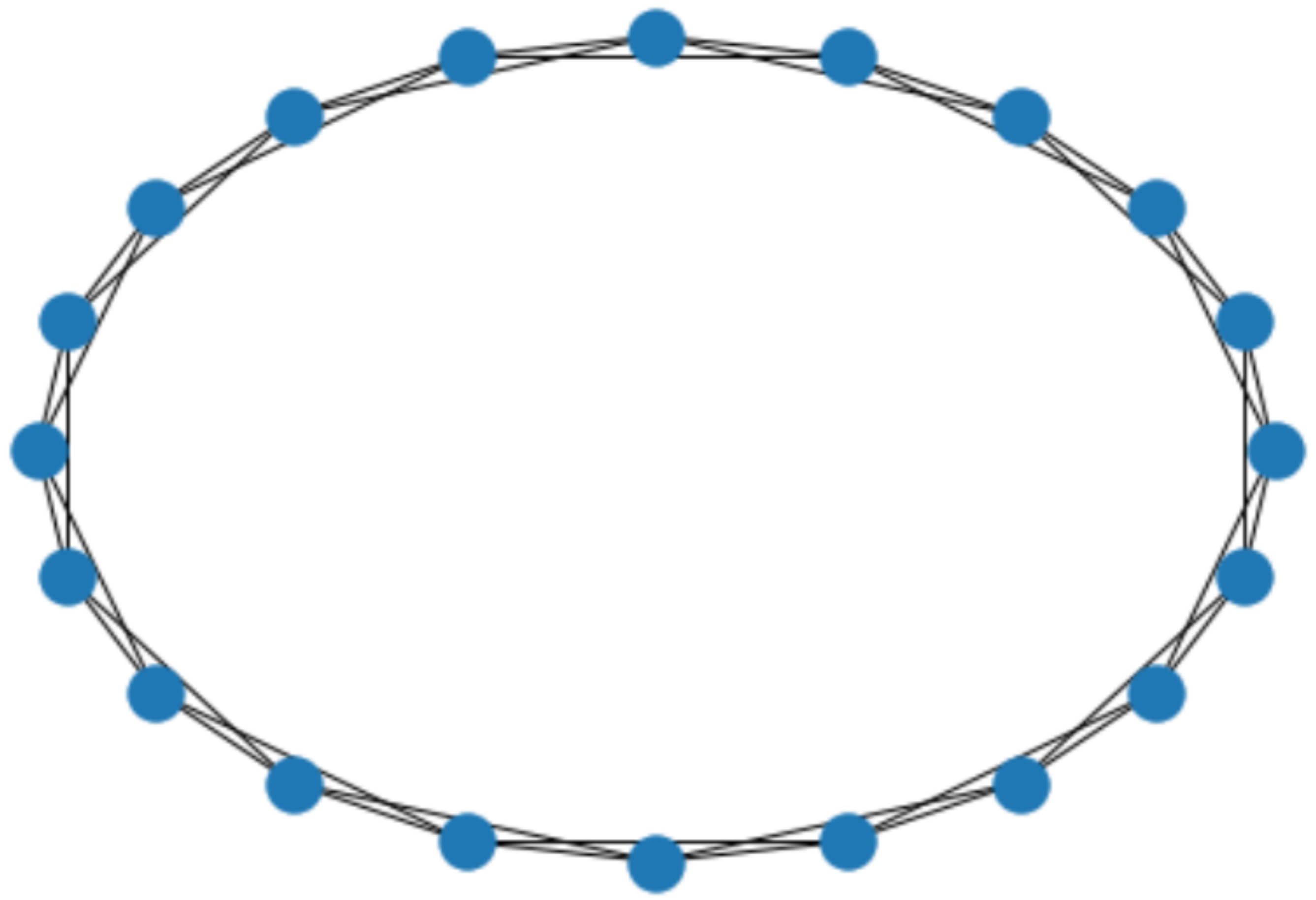
Lab: SW_vs_Models.ipynb

Parameters are p, k and N

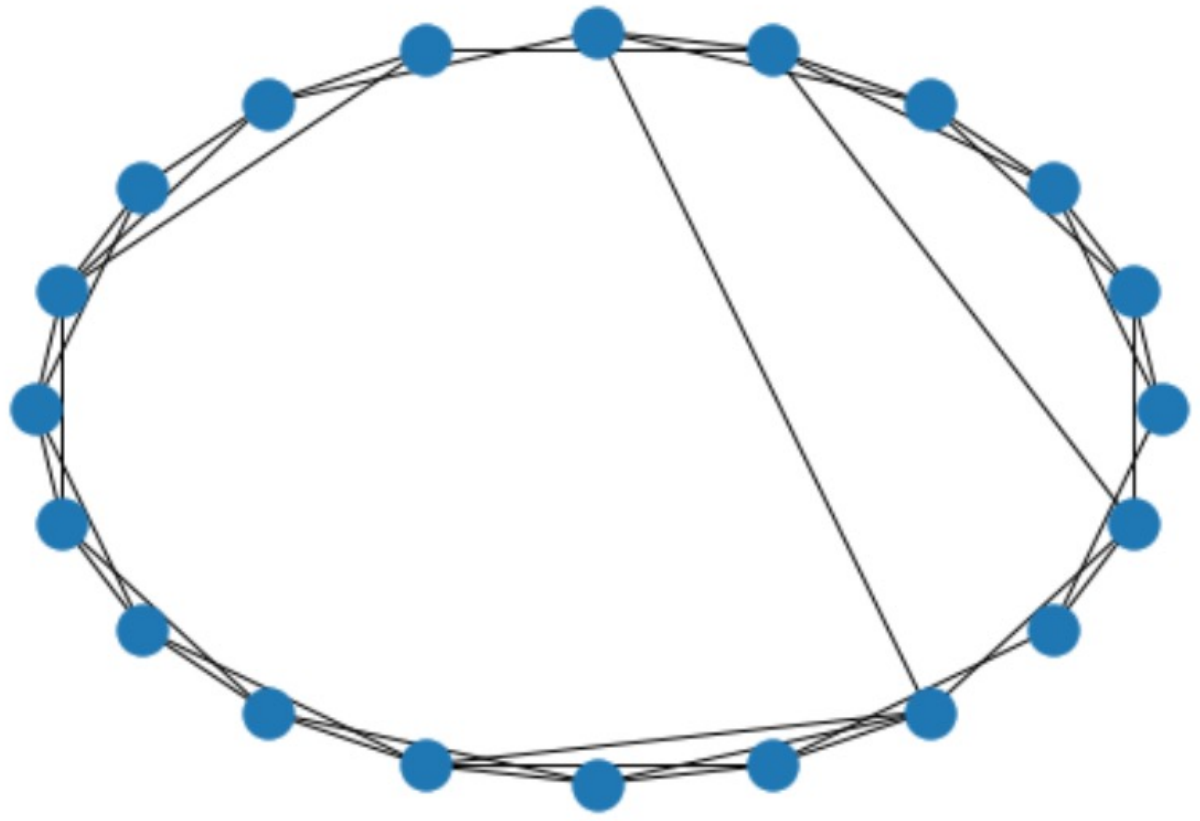
Let's reproduce Fig 1 of the article

```
G_s=nx.watts_strogatz_graph(20,4,0)
```

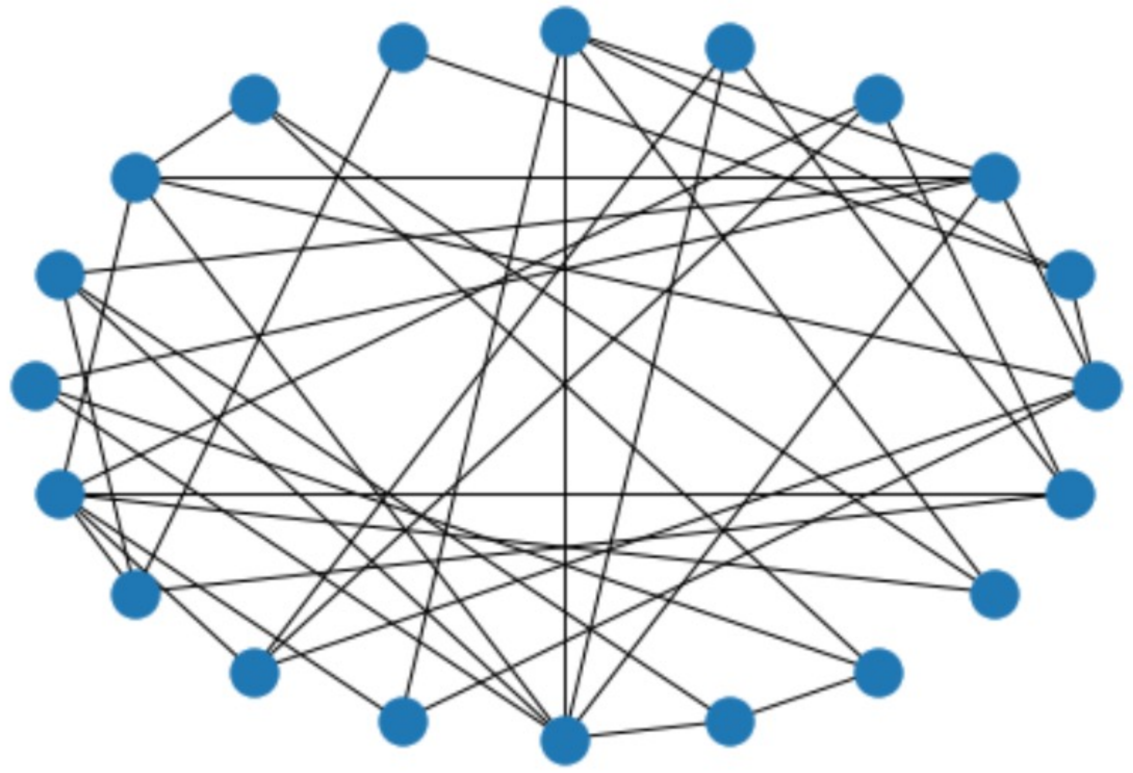
```
nx.draw_circular(G_s)
```



```
nx.draw_circular(nx.watts_strogatz_graph(20,4,0.05,seed=123))
```



```
nx.draw_circular(nx.watts_strogatz_graph(20,4,1.0))
```



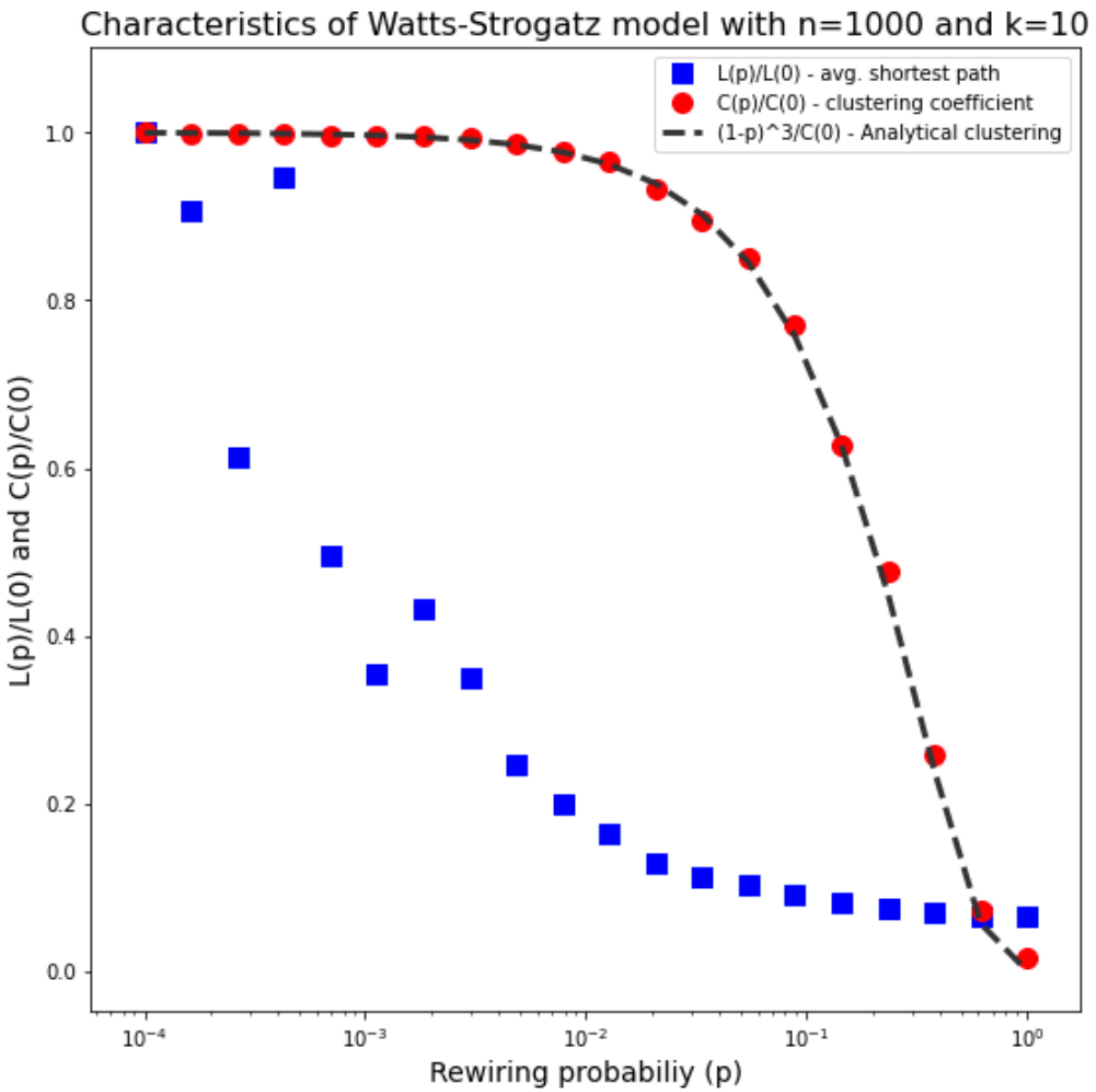
```
Average clustering coefficient =  0.5
Average shortest path length =  2.8947368421052633
```

```
Average clustering coefficient =  0.1780952380952381
Average shortest path length =  2.110526315789474
```

Closer to p=0

```
Average clustering coefficient =  0.38666666666666666
Average shortest path length =  2.3
```

Closer to p=1


Characteristics of Watts-Strogatz model with n=1000 and k=10
L(p)/L(0) - avg. shortest path
C(p)/C(0) - clustering coefficient
(1-p)^3/C(0) - Analytical clustering
L(p)/L(0) and C(p)/C(0)
Rewiring probabiliy (p)
0.0
0.2
0.4
0.6
0.8
1.0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}

Note: By construction of the model k is an even number

```
##### Watts-Strogatz model parameters #####

# Number of nodes (graph vertices)
n=1000
# Degree of the initial ring-lattice
k=10
# Resolution (number of different values of p, the rewiring probability)
r = 20

# Generate a set of r logarithmically spaced values of p
p_range=[0] + np.logspace( -4, 0, r ).tolist()

###### For each p, compute the characterstic measures of the Watt-Strogatz model (n,k,p) ####
clustering = []
sp_length  = []

for p in p_range:

    # Create a Watts-Strogatz graph
    graph = nx.watts_strogatz_graph( n, k, p )

    # Compute the average shortest-path length L(p)
    sp_length.append( nx.average_shortest_path_length( graph ) )

    # Compute the clustering coefficient C(p)
    clustering.append( nx.average_clustering( graph ) )
```

Step 2

Test the analytical values of $C(p)$:

- The clustering coefficient has the following analytical value:

$$C(p) = C(0) \times (1 - p)^3$$

see references in: http://www.scholarpedia.org/article/Small-world_network

:

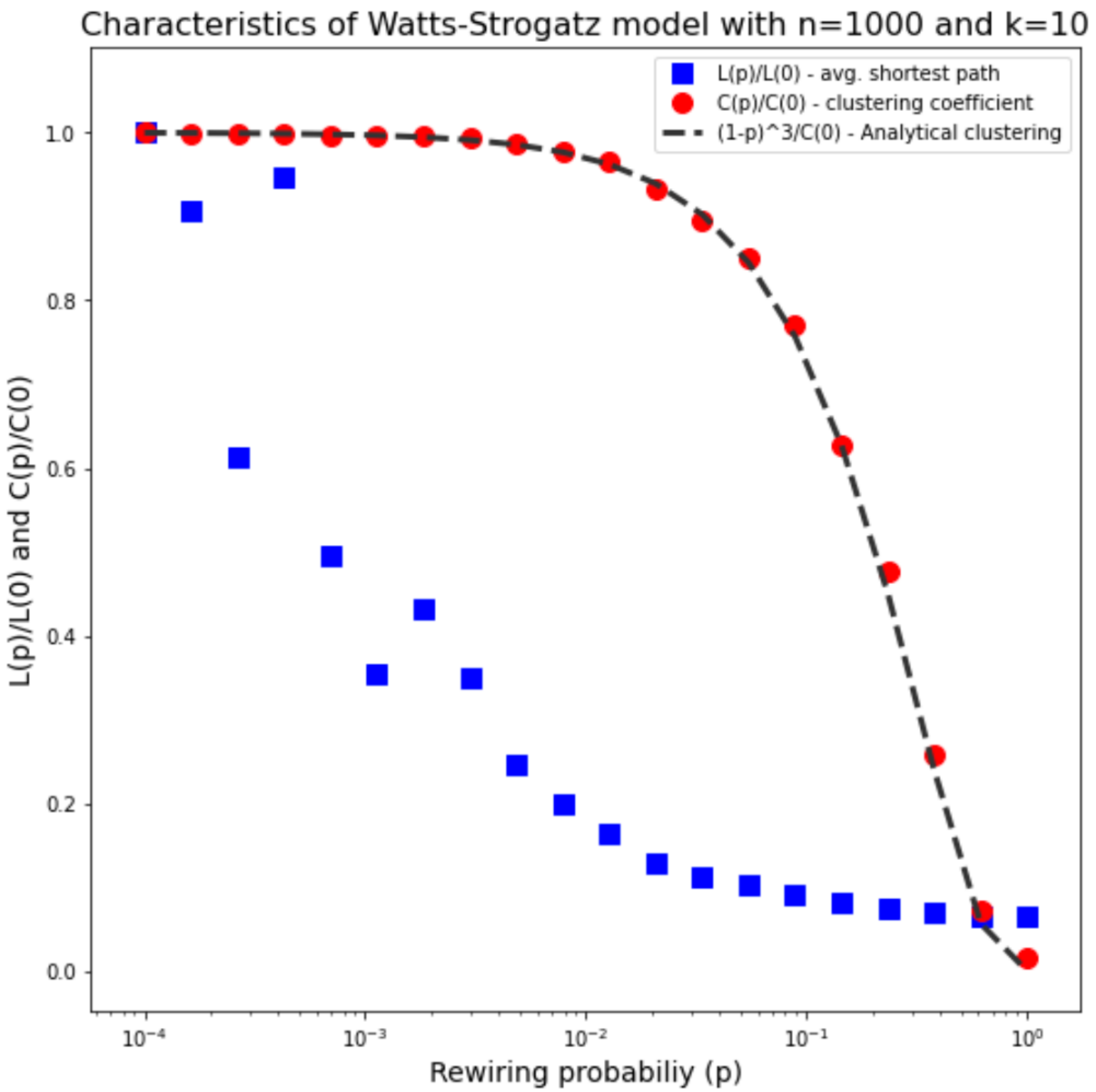
```

##### Plot L(p)/L(0) and C(p)/C(0) #####
plt.figure(1,figsize=(9,9))

plt.semilogx(p_range[1:] , [L / sp_length[0]  for L in sp_length[1:]]  , 'sb', markersize=10, label='L(p)/L(0) - av
plt.semilogx(p_range[1:] , [C / clustering[0] for C in clustering[1:]] , 'or', markersize=10, label='C(p)/C(0) - cl
plt.semilogx(p_range[1:] , [ (1-p)**3.0 for p in p_range[1:]] , '--', color=[0.2,0.2,0.2], linewidth=3, markersize=

# Extend the plot region to see the points at both ends
lims = plt.xlim()
plt.xlim([lims[0]*0.9, lims[1]*1.1])
lims = plt.ylim()
plt.ylim([lims[0]*0.95, lims[1]*1.05])


plt.title( "Characteristics of Watts-Strogatz model with n=%d and k=%d"%(n,k) , fontsize=16)
plt.xlabel( "Rewiring probabiliy (p)" , fontsize=14)
plt.ylabel( "L(p)/L(0) and C(p)/C(0)" , fontsize=14)
plt.legend( loc='best' )
plt.show()
```



Characteristics of Watts-Strogatz model with n=1000 and k=10
L(p)/L(0) - avg. shortest path
C(p)/C(0) - clustering coefficient
(1-p)^3/C(0) - Analytical clustering
L(p)/L(0) and C(p)/C(0)
Rewiring probabiliy (p)

Let's try to generate a Small World network model with 1000 links 300 nodes and clustering coefficient of 0.5

k=2*1000/300 = 6.666, we can select k=6 or 8, if you select 7 the watts_strogatz function round it to the lowest even number.

```
C0 =  nx.average_clustering(nx.watts_strogatz_graph( 300, 6, 0 ))
```

$C(p) = C(0) \times (1 - p)^3$

```
p=1-pow(0.5/C0,1/3)
```

```
p
```

```
0.05896397111896978
```

```
gs =  nx.watts_strogatz_graph( 300, 6, p )
```

```
print('Average clustering coefficient = ',nx.average_clustering(gs))
print('Average shortest path length = ',nx.average_shortest_path_length(gs))
print ("Number of nodes:" , gs.number_of_nodes())
print ("Number of edges:" , gs.number_of_edges())
print ("2*Edges/N:" , 2*gs.number_of_edges()/gs.number_of_nodes())
print ("Average degree of  directed graph = ",np.mean(list(dict(gs.degree()).values())))
```

```
Average clustering coefficient =  0.5121904761904749
Average shortest path length =  5.797279821627647
Number of nodes: 300
Number of edges: 900
2*Edges/N: 6.0
Average degree of  directed graph =  6.0
```

Note: For the Watts Strogatz Model we can fix the clustering coefficient with p and the average degree is given by the parameter k

Note: For the Watts Strogatz Model we can fix the clustering coefficient with p and the average degree is given by the parameter k

Let's generate a Erdos-Renyi network model with 300 nodes and average degree 6

prg(n-1)=k=2L/(n)

```
prg = 6/(299)
```

```
prg
```

```
0.020066889632107024
```

```
Ger = nx.erdos_renyi_graph(300, prg, seed=None, directed=False)
```

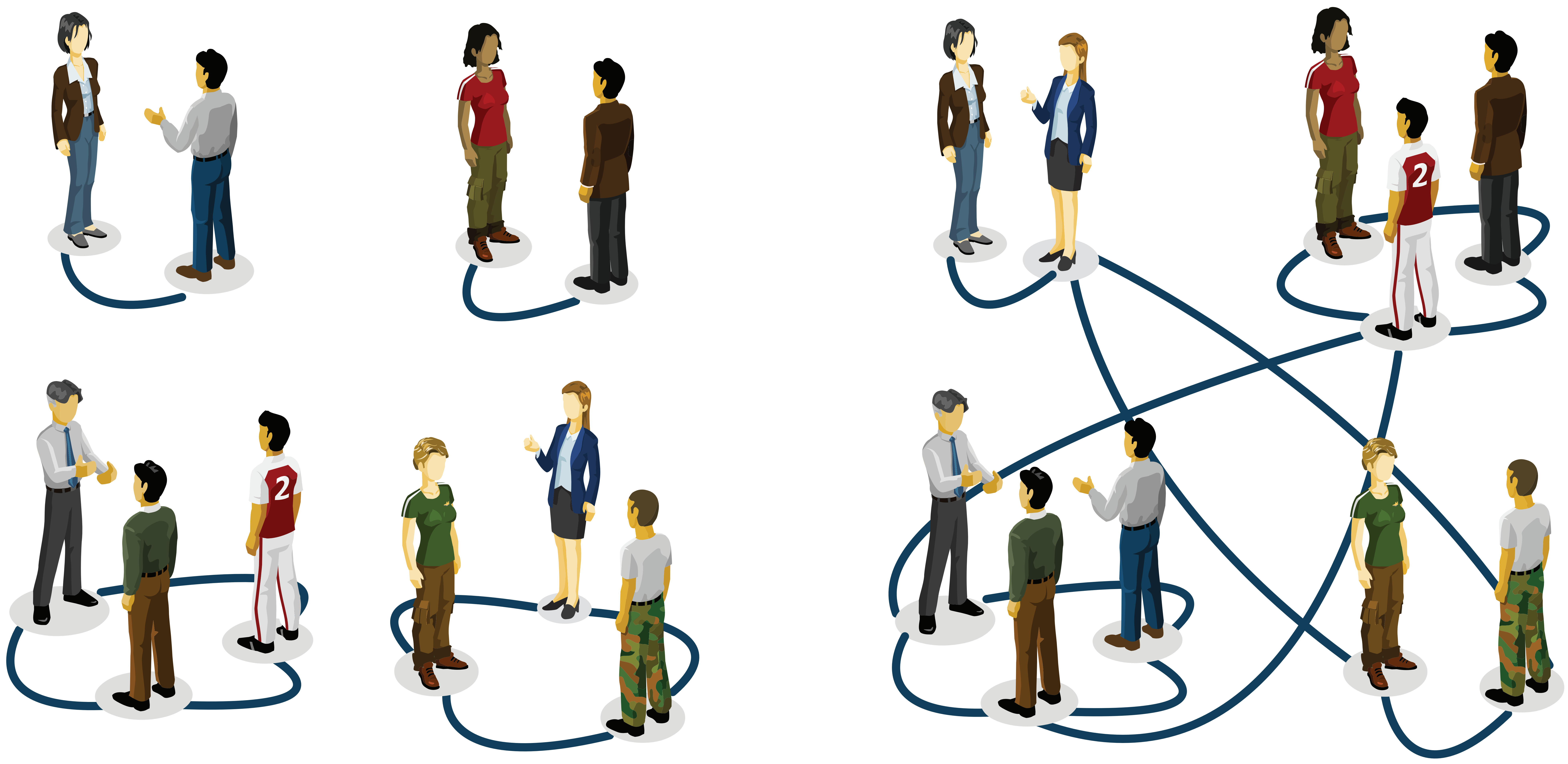
```
print('Average clustering coefficient = ',nx.average_clustering(Ger))
print('Average shortest path length = ',nx.average_shortest_path_length(Ger))
print ("Number of nodes:" , Ger.number_of_nodes())
print ("Number of edges:" , Ger.number_of_edges())
print ("2*Edges/N:" , 2*Ger.number_of_edges()/Ger.number_of_nodes())
print ("Average degree of the graph is = ",np.mean(list(dict(Ger.degree()).values())))
```

```
Average clustering coefficient =  0.017697783697783683
Average shortest path length =  3.2805574136008917
Number of nodes: 300
Number of edges: 948
2*Edges/N: 6.32
Average degree of the graph is =  6.32
```

Note: For the ER graph we fix the links or average degree, and the clustering coefficient C~p

Random networks

RANDOM NETWORK MODEL



RANDOM NETWORK MODEL

Definition:

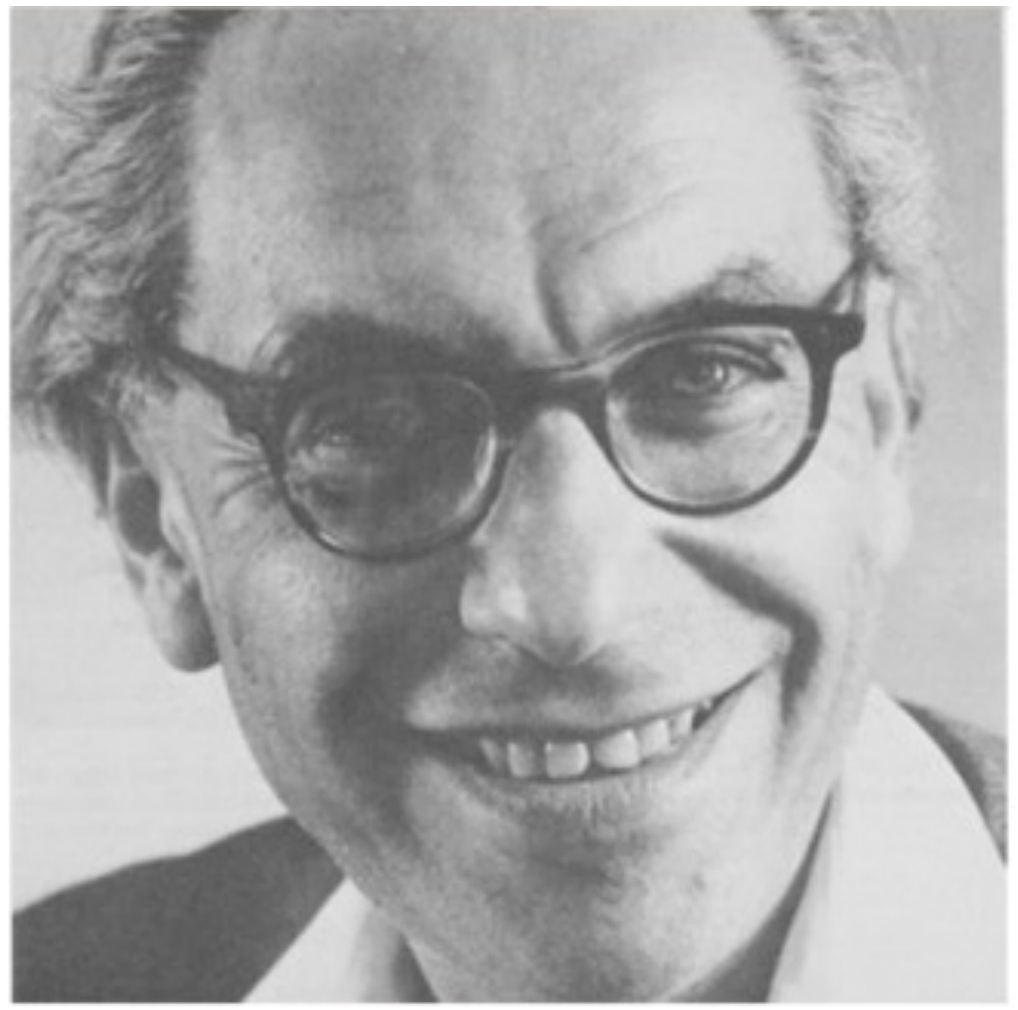
A **random graph** is a graph of N nodes where each pair of nodes is connected by probability **p**.

Note s random graph with parameters (N,p)
has pN(N-1)/2 links by definition

```
G = nx.erdos_renyi_graph(N,p) # Erdos-Renyi random graph in Networkx
```

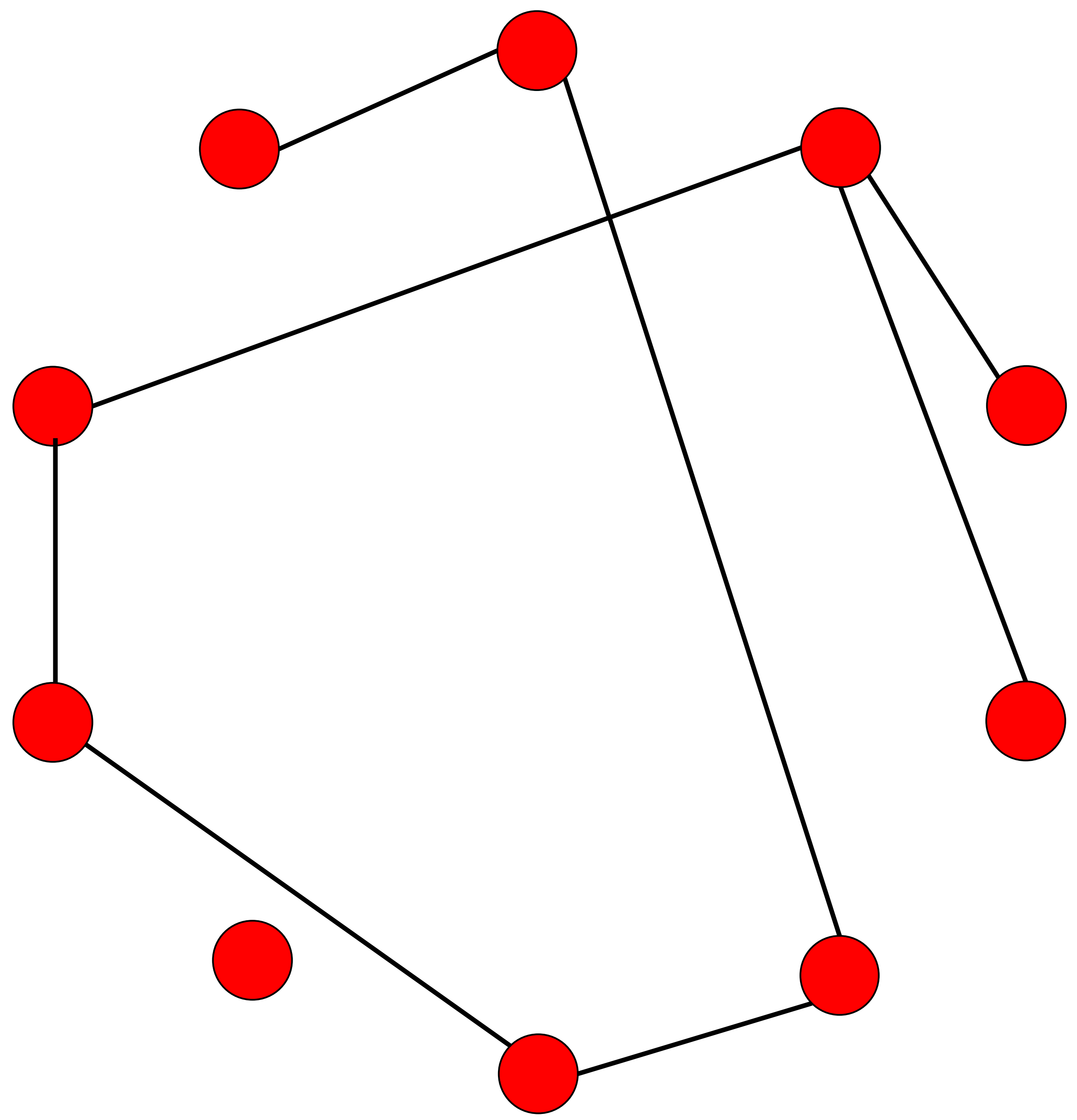
RANDOM NETWORK MODEL

Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)

A random graph with parameters (N,p) has pN(N-1)/2 links by definition



Erdös-Rényi model (1960)

Connect with probability p

p=1/6 N=10

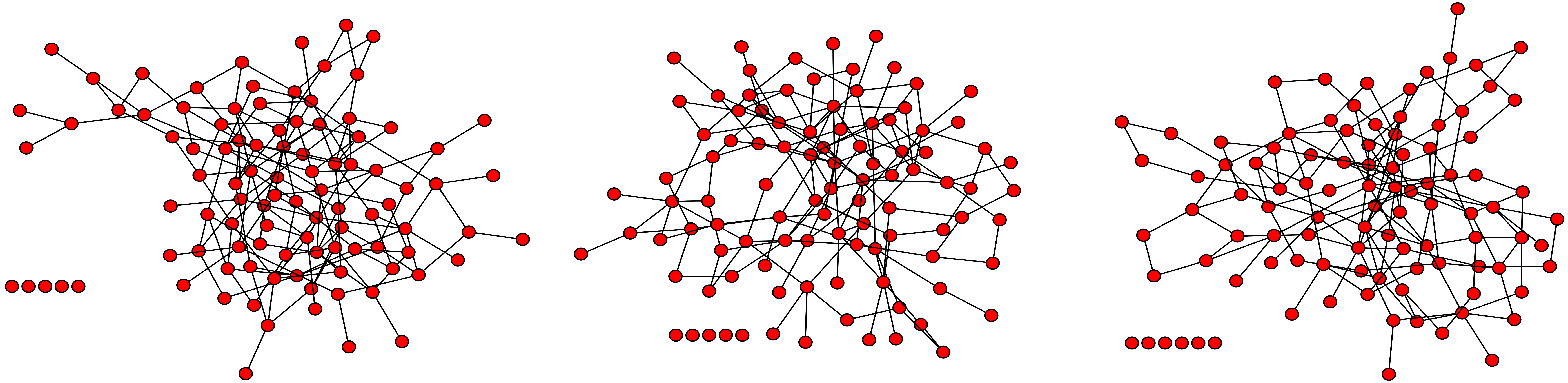
<k> = (3+1+1+2+2+0+2+2+1+2)/10=1.6

<k>=2L/N = 2p(N)(N-1)/2N → **Estimated Value**

<k> = 0.1666 (9) = 1.5

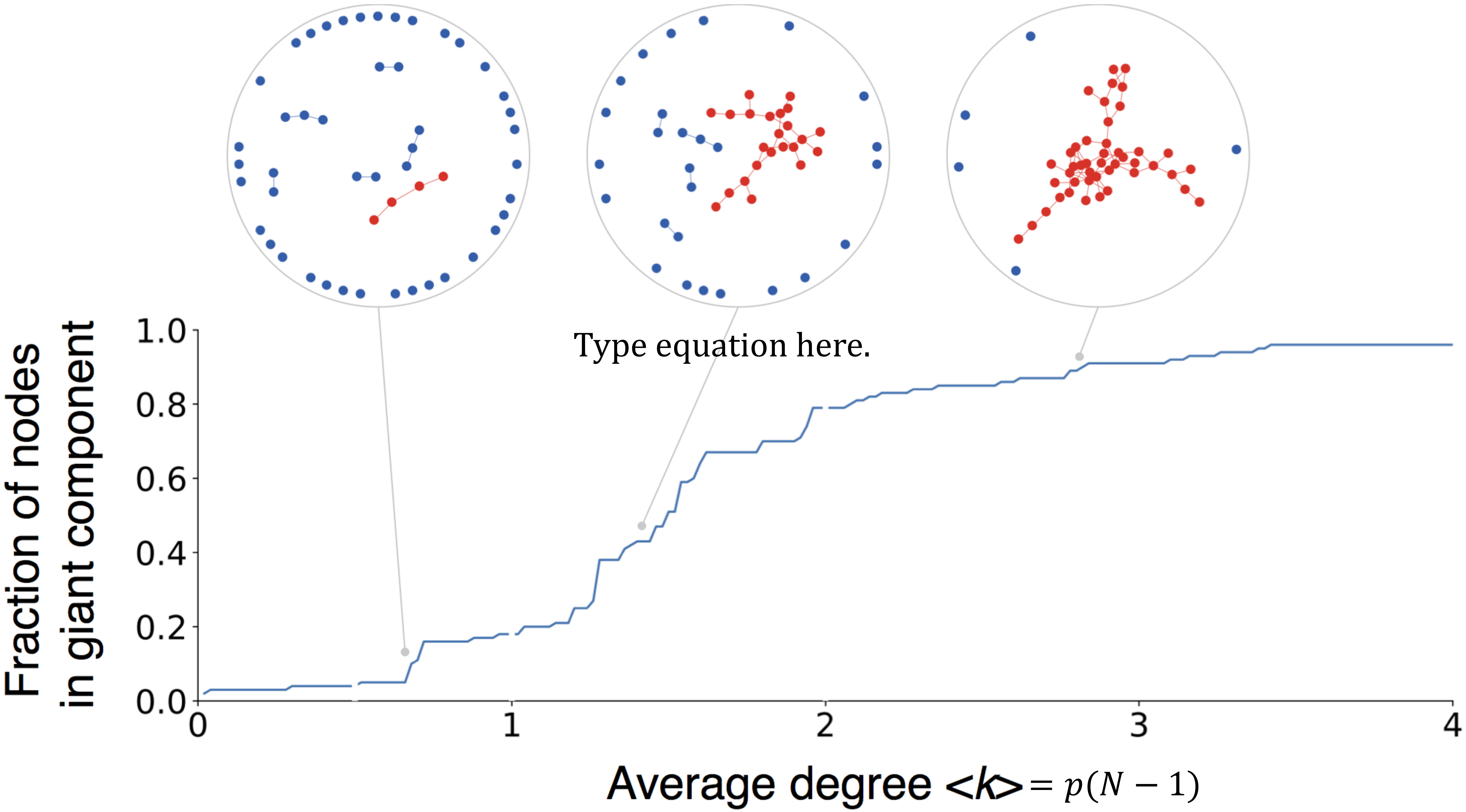
RANDOM NETWORK MODEL

p=0.03
N=100



It is not deterministic, every realization gives us a different configuration

Random networks: evolution



If <k> is less than 2 the network is broken in various components

Random networks Number of Links

- **(Expected) number of links <*L*> of a random network with *N* nodes:** number of "heads" with probability of yielding heads equal to *p* and the number of trials *t* equal to the number of all node pairs of the network:

$$t = N(N-1)/2 \quad \rightarrow \quad \langle L \rangle = \frac{pN(N-1)}{2}$$

Random networks average degree

(Expected) average degree <*k*> of a random network with *N* nodes: number of "heads" with probability of yielding heads equal to *p* and the number of trials *t* equal to the number of potential neighbors of a node:

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

Note: this is referred as the analytical property of the average degree of a Random Graph

Random networks: degree distribution

- **Question:** what is the probability that a node has *k* neighbors?

- **Back to coin tossing problem:** what is the probability that a coin that yields heads with probability *p* results in *k* heads out of *N*–1 (independent) trials?

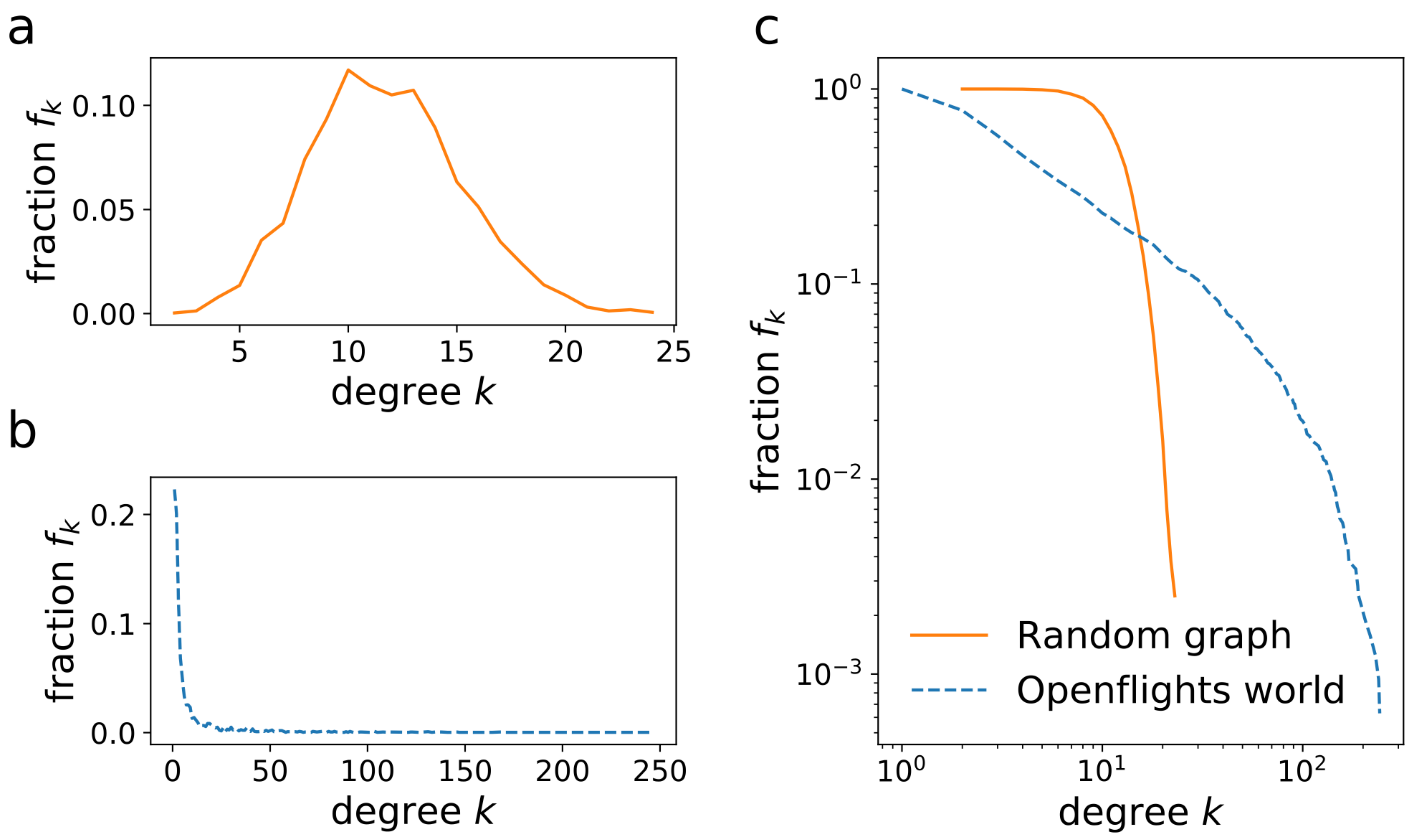
- **Binomial distribution:**

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

- For small *p* and large *N* the binomial distribution is well approximated by a bell-shaped curve ⇒ **most degree values are concentrated around the peak, so the average degree is a good descriptor of the distribution**

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

Random networks: degree distribution

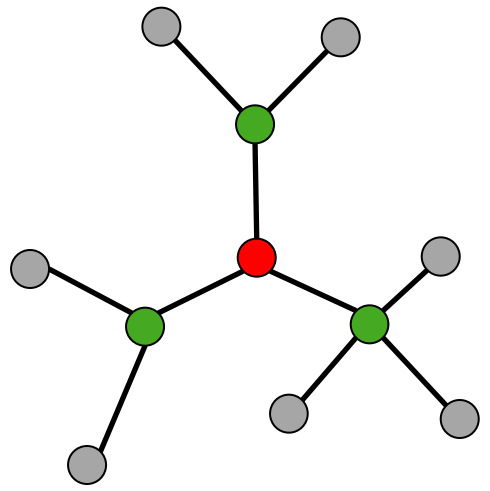


The degree distribution of random networks **is very different** from the broad distributions of most real-world networks!

Random networks: average shortest path and clustering coefficient

Random networks: small-world property

- **Question:** how many nodes are there (on average) *d* steps away from any node?
- **Premise:** since nodes have approximately the same degree, let us assume they have all exactly the same degree *k*
  - At distance $d = 1$ there are k nodes
  - At distance $d = 2$ there are $k(k-1)$ nodes
  - At distance d=3 there are $k(k-1)^2$
  - …
  - At distance d there are $k(k-1)^{d-1}$ nodes



- If *k* is not too small, the **total number of nodes within a distance *d*** from a given node is approximately:

$$N_d \sim k(k-1)^{d-1} \sim k^d$$

Random networks: small-world property

- **Question:** how many steps does it take to cover the whole network?

$$N \sim k^{d_{max}}$$

$$\log N \sim d_{max} \log k$$

$$d_{max} \sim \frac{\log N}{\log k}$$

- The diameter of the network **grows like the logarithm** of the network size

- **Example:** N = 7,000,000,000, k =150 (Dunbar's number)

$$d_{max} = 4.52$$

Random networks: clustering coefficient

- The clustering coefficient of a node *i* can be interpreted as the probability that two neighbors of *i* are connected

$$C_i = \frac{\text{number of pairs of connected neighbors of i}}{\text{number of pairs of neighbors of i}}$$

- **Question:** what is the probability that two neighbors of a node are connected?

- **Answer:** since links are placed independently of each other, it is just the probability *p* that any two nodes of the graph are connected:

$$C_i = p = \frac{\langle k \rangle}{N-1} \sim \frac{\langle k \rangle}{N}$$

- Since <*k*> is usually a small number, the average clustering coefficient of random networks with realistic values for <*k*> and *N* is **much smaller** than the ones observed in real-world networks

Random networks: summary analytical properties

- Degree distribution Bell shaped curve around average
- $<k>=p(N-1)=2L/N$, where L is the number of links and N and p are input parameters of the model
- Average shortest path length $<l>\sim\log(N)/\log(k)$
- Clustering Coefficient $<C> = p$

```
G = nx.erdos_renyi_graph(N,p) # Erdos-Renyi random graph
```

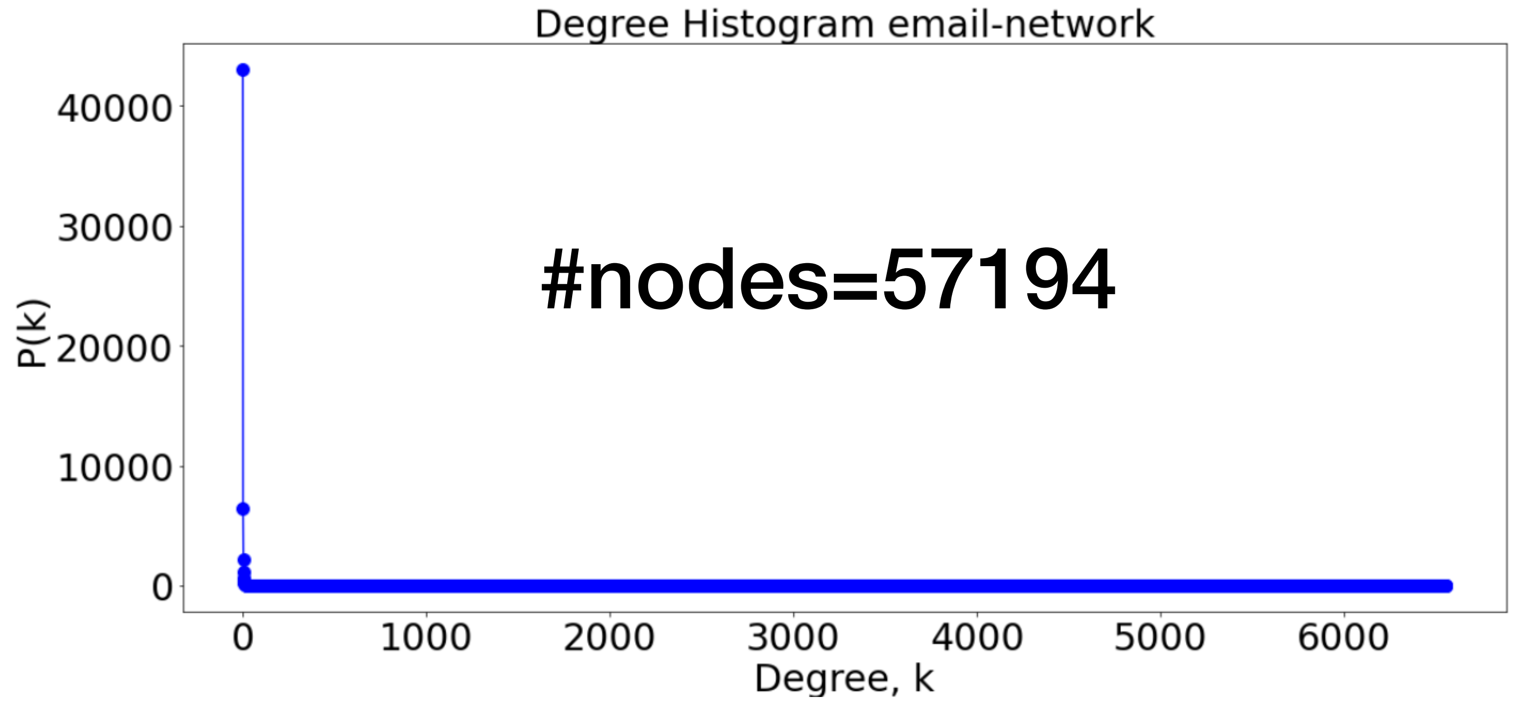
Random networks vs Empirical Networks

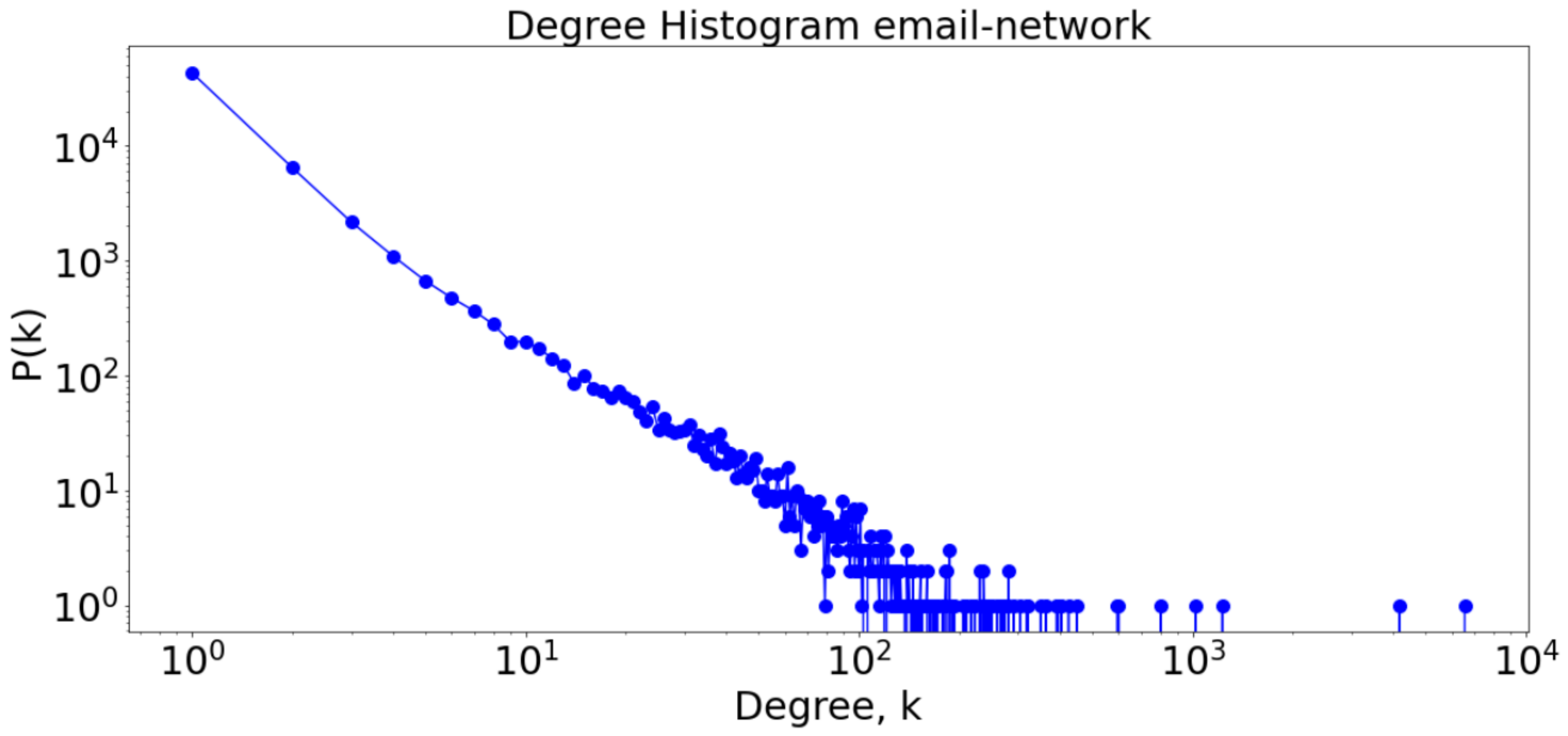
- Links are placed at random, independently of each other
- Distances between pairs of nodes are short (small-world property): **good!**
- The average clustering coefficient is much lower than on real networks of the same size and average degree: **bad!**
- The nodes have approximately the same degree, there are no hubs: **bad!**
- **Conclusion:** the random network is not a good model of many real-world networks

Features of real networks: heterogeneity

- **Heavy-tail distributions:** the variable goes from small to large values

- **Hubs:** nodes with high degree





Probability Density Function in log-log

Probability density function or Distribution Function is the relative likelihood to find the value of the variable within the bin width

This probability is given by the integral of this variable's PDF over that range —that is, it is given by the area under the density function

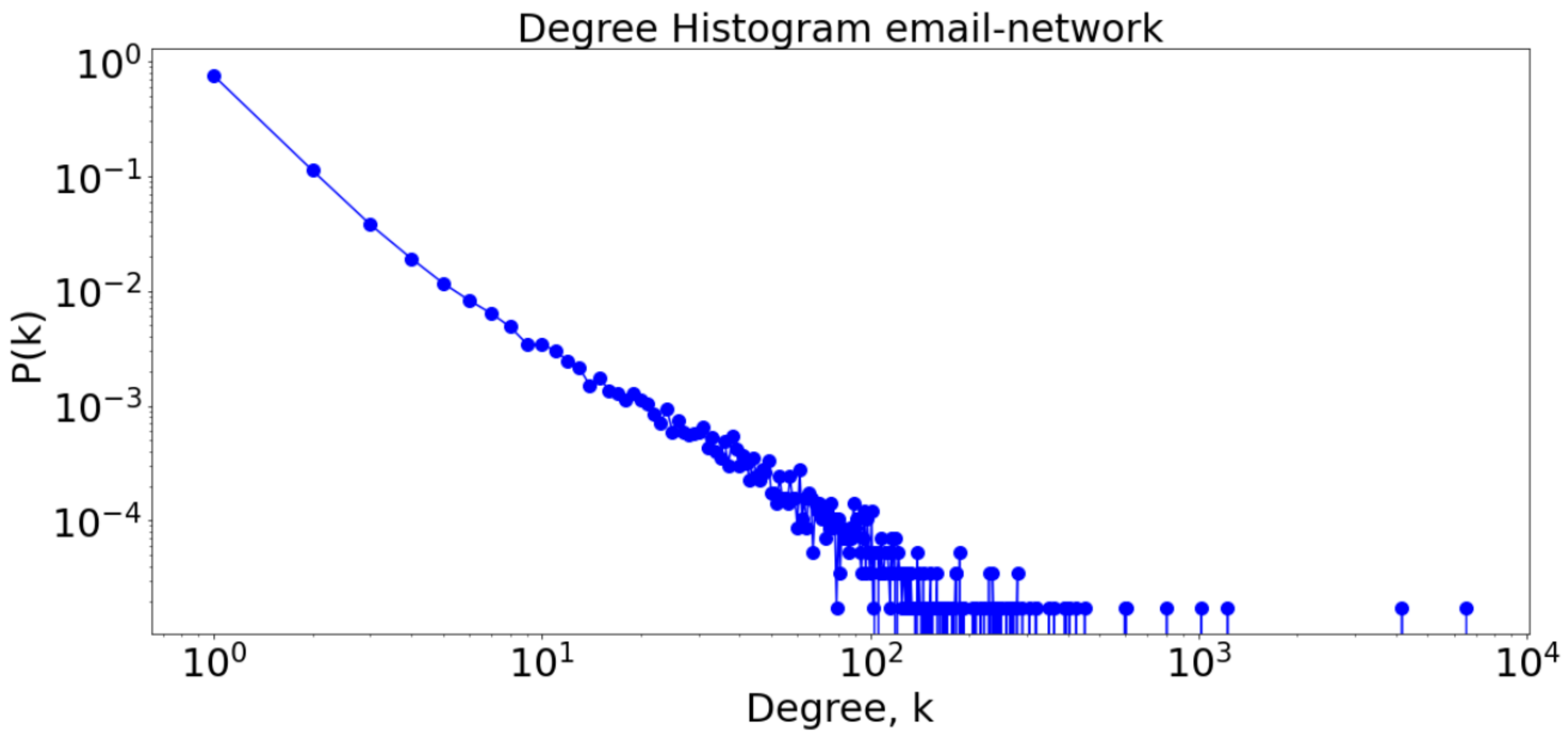
density = counts / (sum(counts) * np.diff(bins)), so that the area under the histogram integrates to 1 (**np.sum(density * np.diff(bins)) == 1**).

```
1 n0, bins0 = np.histogram(degs0, bins = list(range(min(degs0), max(degs0)+1, 1)),density=True)
```

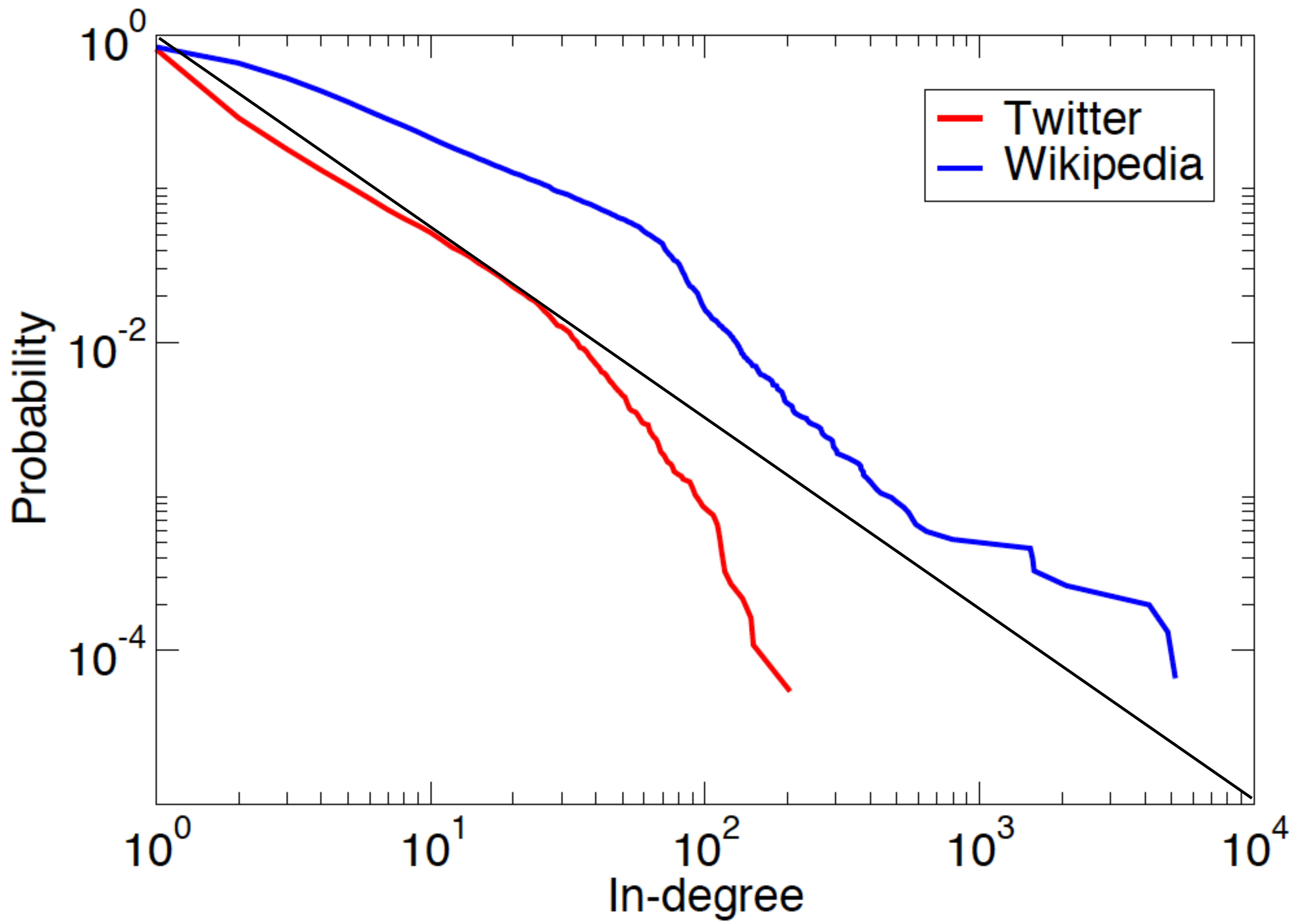
```
1 np.sum(n0 * np.diff(bins0))
```

```
0.9999999999999998
```

plt.loglog(bins0[:-1],n0)



Heterogenous distributions are use Log-Bined LogLog plots



Summary

- Properties of Random Graph
- Properties of the Small World (Watts & Strogatz Model)

This week

- Finalize participation slides (LAB)
- Finalize Assignment 1, part 1 (due Jan 31st + 48 hours grace period)

Lab Small World and Random Graph Models

We review it on Lecture 4 on February 8th

Answers here:
https://docs.google.com/presentation/d/1CqgviVVaYe612XNsYyyKCDr_YTfbwMsqSeCNhmEPk2s/edit?usp=sharing

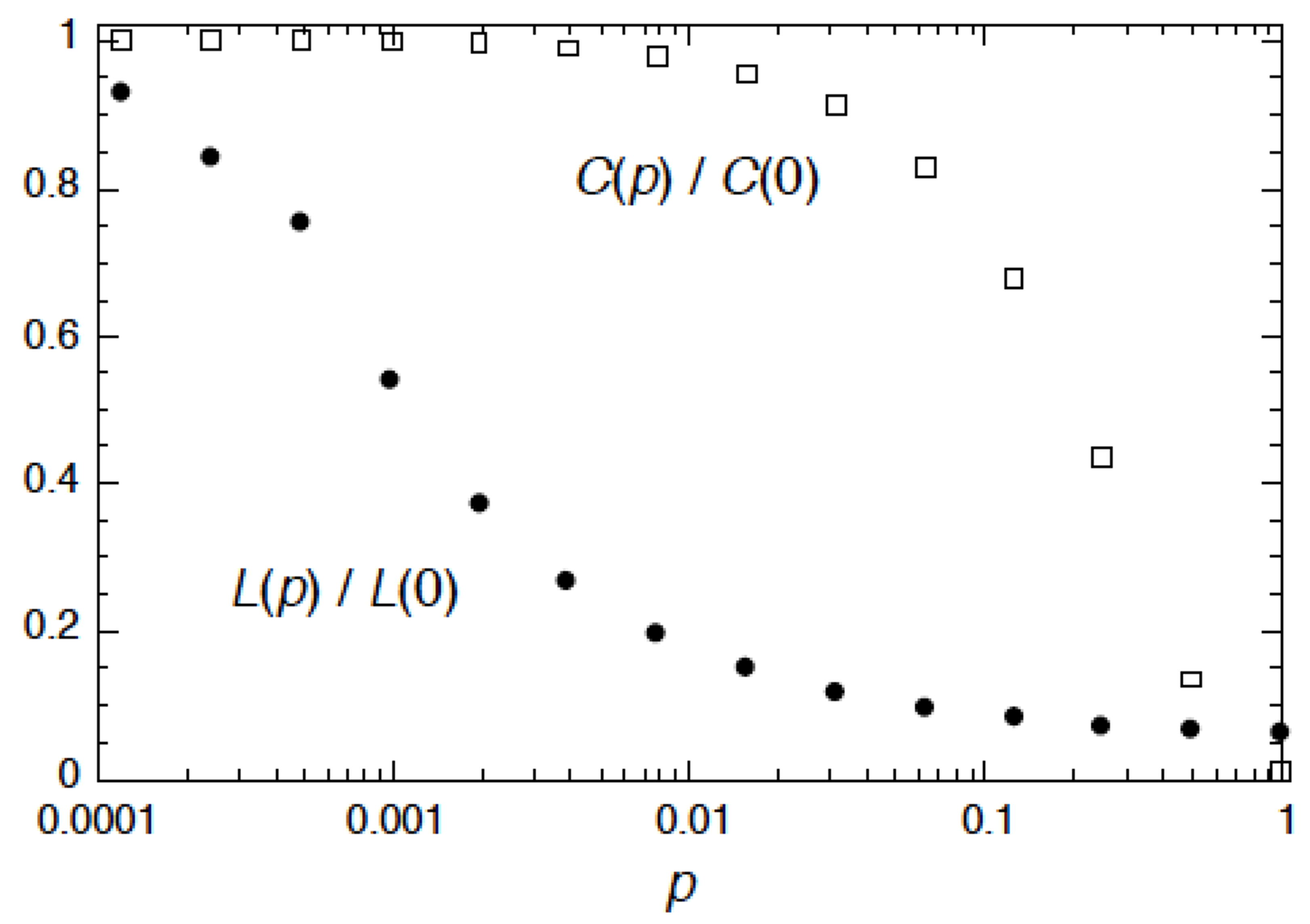
Note: this serves is a study guide Assignment 1 part 2 and Lecture 3

Question 1: Modify the script SW_vs_Models.ipynb provided in Lecture 3 and create a Watts Strogatz graph g=nx.`watts_strogatz_graph`*(n, k, p, seed=None)* with N=10, p=0.0, and k=4

Show the code and solution to draw the network, and to calculate the Clustering Coefficient and the Average Shortest Path Length

Question 2: For a Watts Strogatz graph g=nx.`watts_strogatz_graph`(*n*, *k*, *p*, *seed=None*) with N=100, and k=8, report the Clustering Coefficient C and Average Path Length with various values of p=0, 0.01, 0.1, 0.5 and 1

Question 3: If you were to reproduce Fig 2 of the Article (below) with networkx, what values of N, k would you use? (as reported in the article). Use those values with g=nx.`watts_strogatz_graph`(*n*, *k*, *p*, *seed=None*)
Show the results of C(0) and L(0), which are the Clust. Coef. and Avg. Path. Length of the graph g with p=0



(See the caption of the Figure in the Article to choose n and k)

Continuation

Question 4: What is the analytical expression of the Clustering coefficient of a random graph as a function of its parameters N and p

Question 5: What is the analytical expression of the average degree of a random graph as a function of its parameters N and p

Question 6: Generate a Watts-Strogatz graph with 1000 nodes, average degree 10 and clustering coefficient 0.5. What is the average shortest path?

Question 7: Based on the article of Lecture 3, when a network has Small World property?

Question 8: If an empirical network has 488,337 nodes, 586,004 links, average shortest path length <l>=12 and clustering coefficient C=0.24. Does it have the Small World property?

Question 9: Does the network in protein.edgelist.txt is a Small World?

Record your participation here:

https://docs.google.com/spreadsheets/d/11CVp8CrWftGz1vsWQxaMIUi4OGJ6mCaOXtvIcrHRFrE/edit?usp=sharing

Office Hours

- Marta: Tuesdays 4:00-5:30 PM @Room 406c and @ Zoom

  Thursdays 9:00-11:00 by appointment via email (20 min)

  https://berkeley.zoom.us/j/4124372482

- Giuseppe: Wednesdays 11:10-1pm, @Davis 305A or

  link: https://berkeley.zoom.us/my/perona

- Martin: Thursdays 1-3pm, @Davis 305A or

  https://berkeley.zoom.us/j/7628895377

- Register to the piazza site:

  https://piazza.com/berkeley/spring2023/ce88cp88