Network Theory and Dynamic Systems

07. Network Models

SOSE 2025

Dr. -Ing. Stefania Zourlidou

Institute for Web Science and Technologies
Universität Koblenz

Recap from Previous Lecture

- Directed Networks
- The Web
- PageRank
- Weighted Networks
- Information and Misinformation Spread
- Co-Occurrence Networks
- Weight Heterogeneity

Objectives of this Lecture

universität koblenz
weiter:denken

- Random Networks
- Small-World Networks

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 (1/2)

- 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

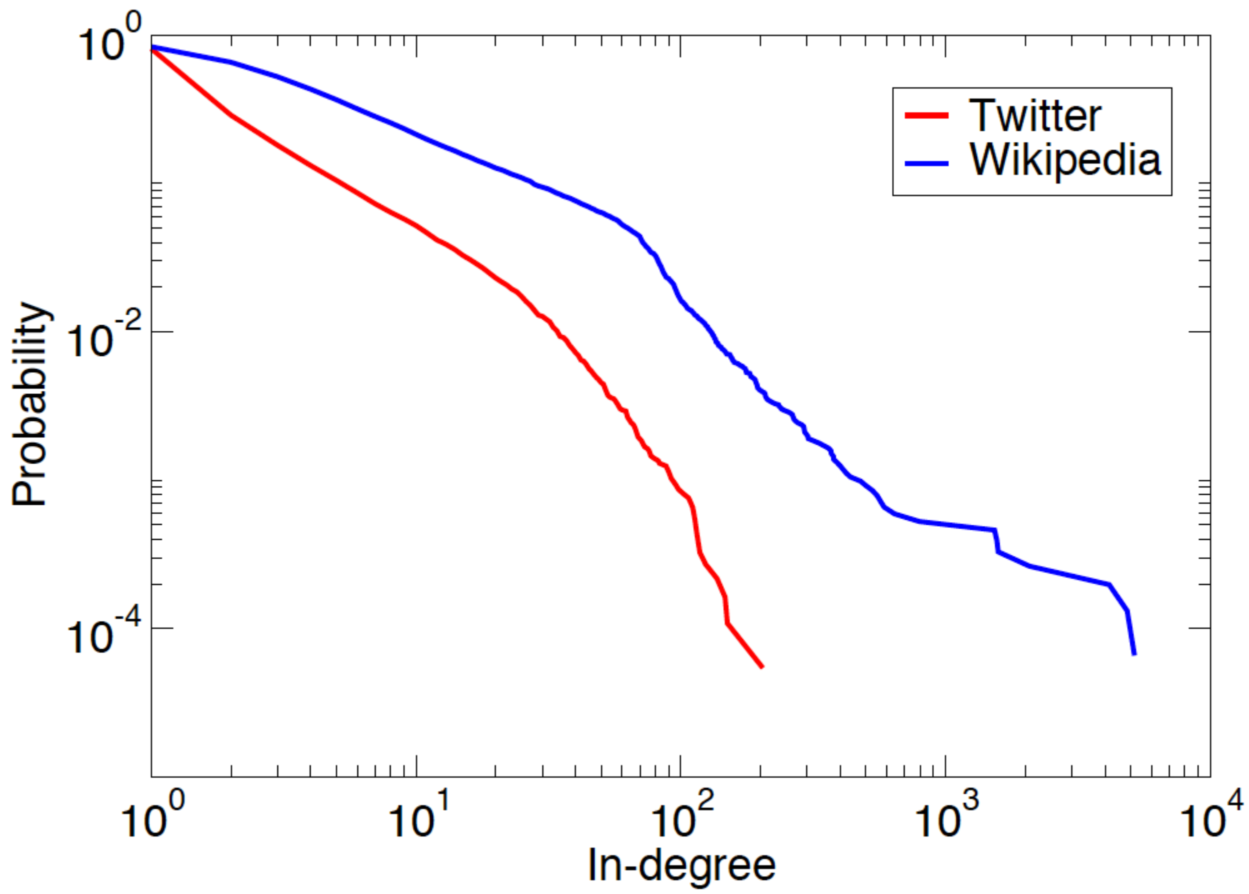
Features of Real Networks: High Clustering Coefficient (2/2)

- 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

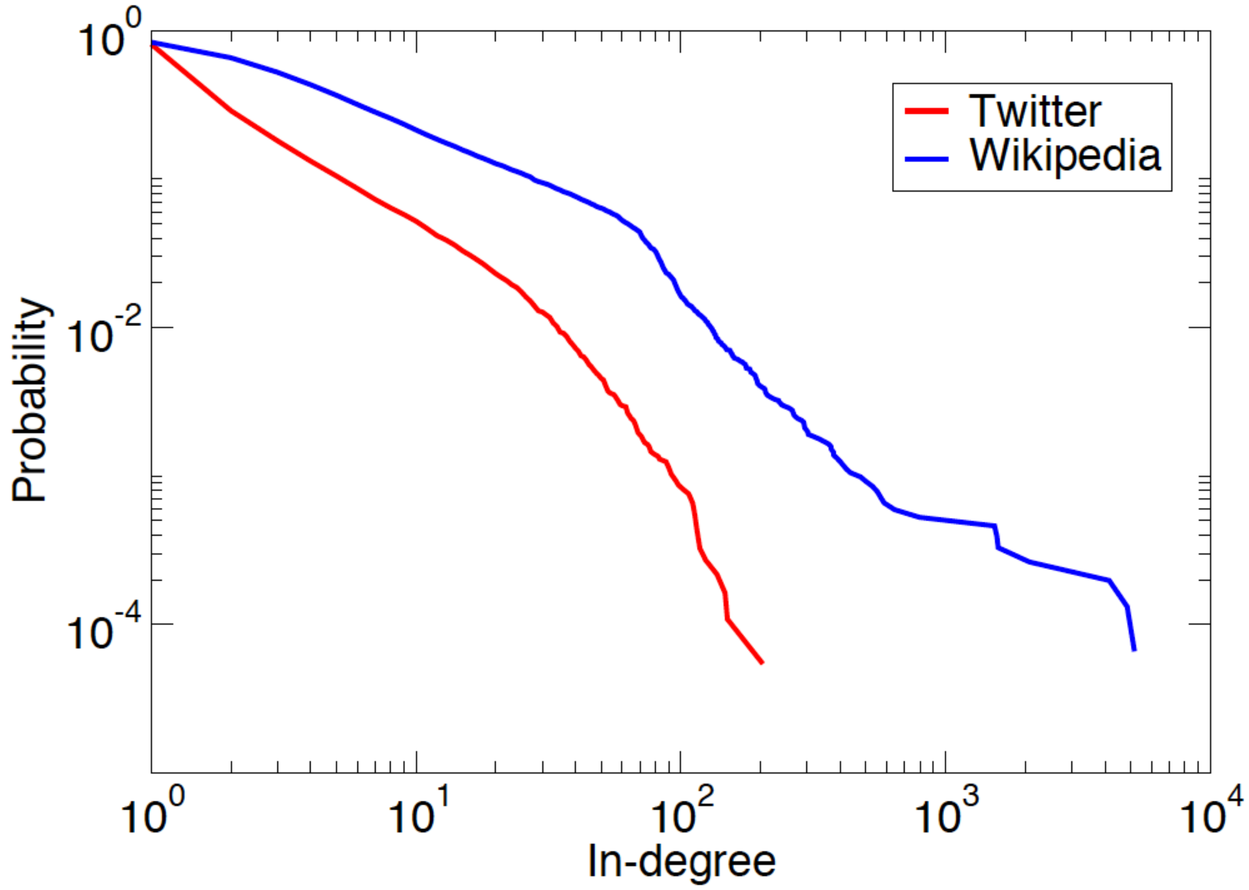
Features of Real Networks: Heterogeneity (1/6)

- **Heavy-tail distributions**: the variable goes from small to large values
- **Hubs**: nodes with high degree



Features of Real Networks: Heterogeneity (2/6)

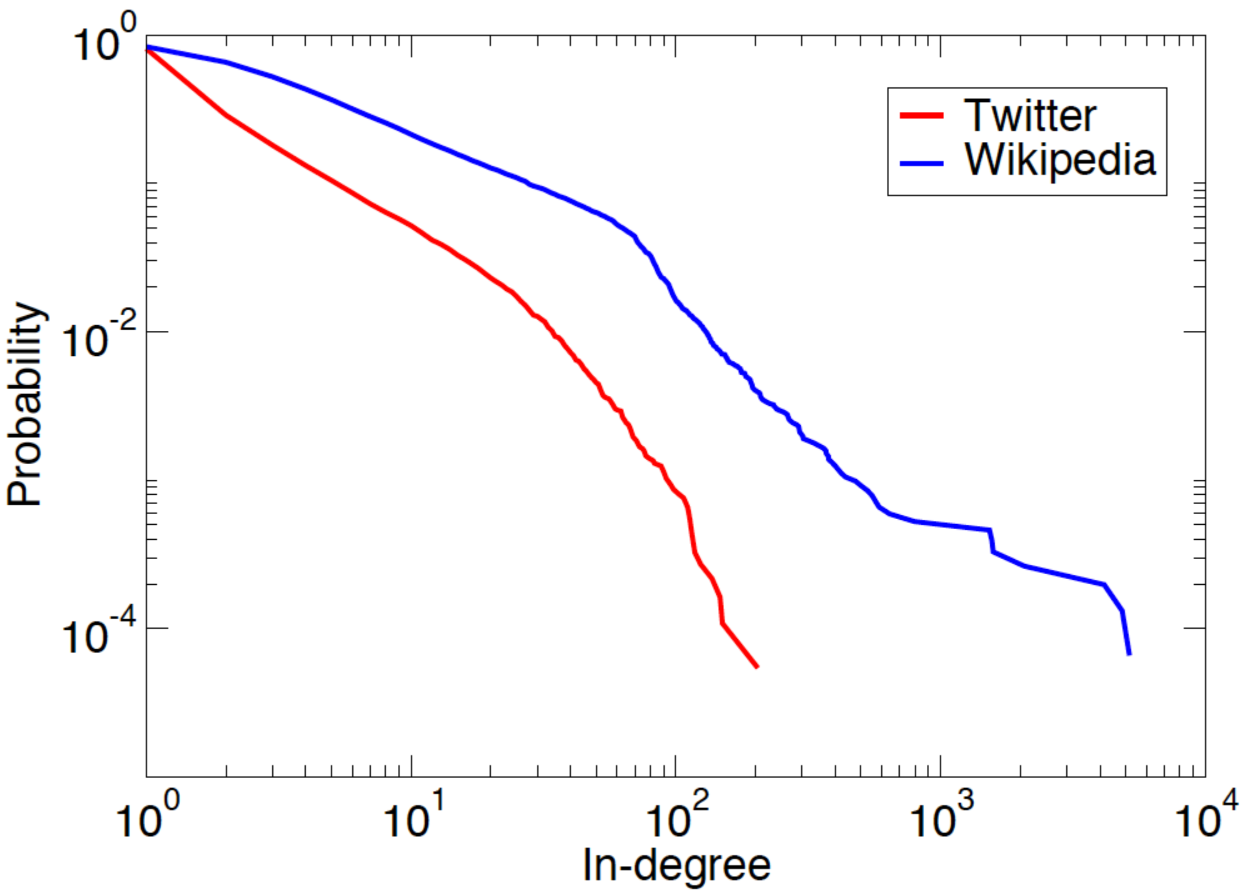
- **Heterogeneity**: variability or diversity in the connectivity of nodes
  - Real-world networks often exhibit significant heterogeneity, meaning that different nodes have vastly different numbers of connections



Features of Real Networks: Heterogeneity (3/6)

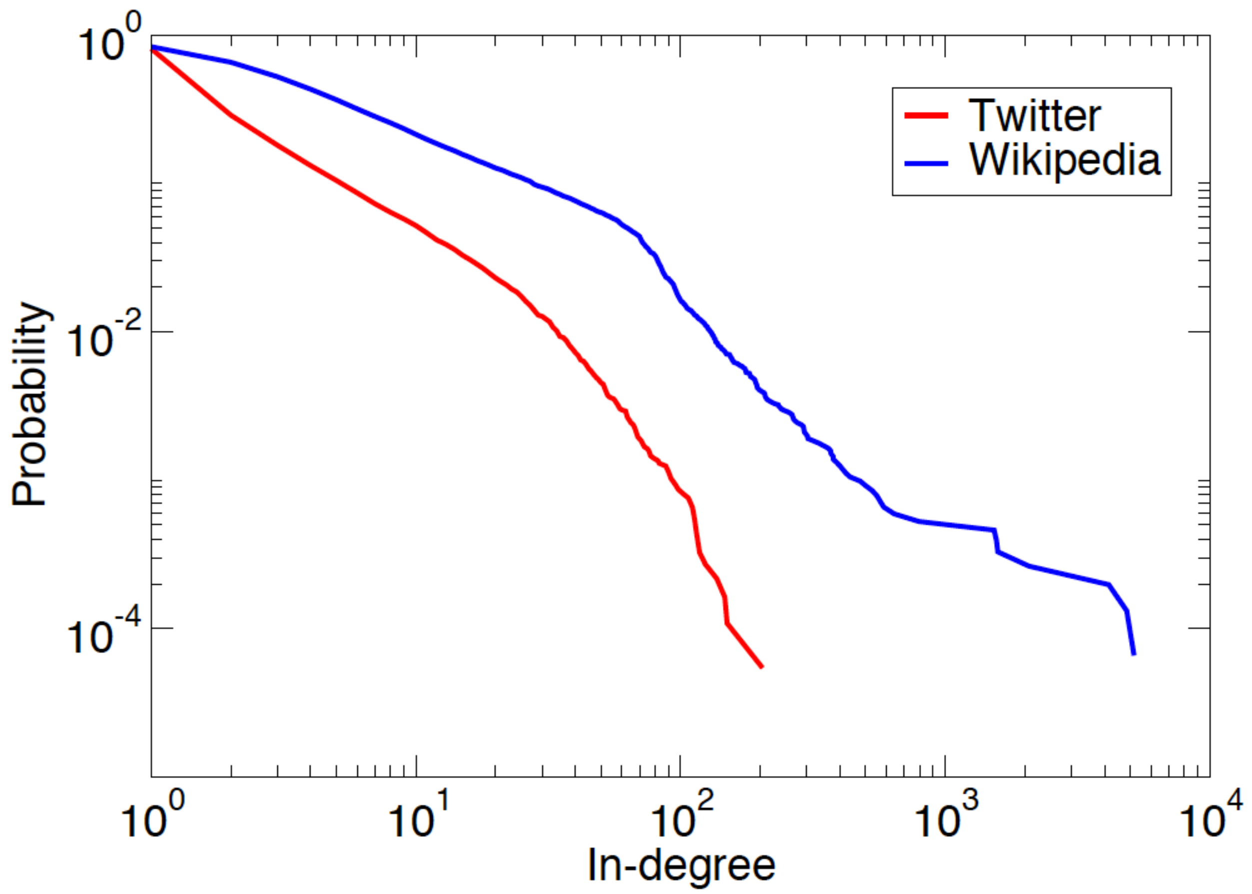
universität koblenz
weiter:denken

- **Heavy-tail Distributions**: These distributions have a long tail, indicating that while most nodes have a relatively small number of connections, there are a few nodes with a very high number of connections
    - This is typical in real networks and is a key aspect of their heterogeneity



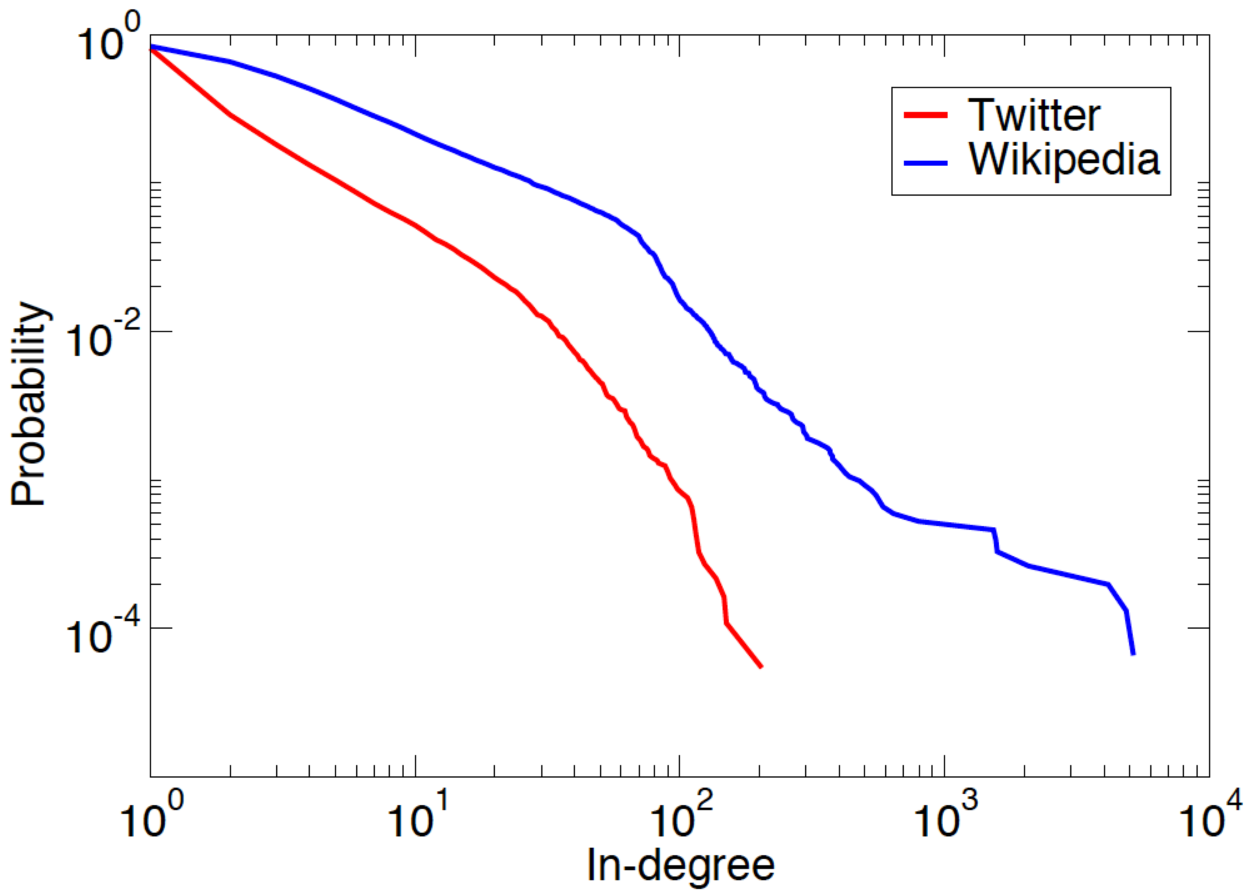
Features of Real Networks: Heterogeneity (4/6)

- **Hubs**: These are nodes with a high degree of connections
  - In a heterogeneous network, a small number of hubs will have many more connections compared to the average node



Features of Real Networks: Heterogeneity (5/6)

- Both Twitter and Wikipedia exhibit a heavy-tail distribution, meaning that while most nodes have a low in-degree, a few nodes (hubs) have a very high in-degree
  - This is evident from the long tails of the curves extending to high in-degree values on the x-axis
- The blue line (Wikipedia) extends further along the x-axis compared to the red line (Twitter), indicating that Wikipedia has some nodes with significantly higher in-degrees than Twitter
  - This suggests that Wikipedia's network has more extreme hubs than Twitter's network



Features of Real Networks: Heterogeneity (6/6)

- Heterogeneity parameter (see Lecture 4: Hubs): a measure of how broad the degree distribution is

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

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

Why Models Are Needed

- To understand and reproduce the heterogeneity seen in real networks, we need models that generate networks with realistic degree distributions
- **Goal:** Reproduce this heterogeneity in *synthetic* networks
  - Helps to understand underlying mechanisms (e.g., preferential attachment)
  - Enables simulation and prediction (e.g., spreading processes, robustness)
  - Assists in designing better interventions and infrastructure

Beyond Heterogeneity: Why Else Do We Need Network Models?

1. Mechanism Discovery
- Reveal generative rules behind network formation (e.g., growth, rewiring)
- Example: Preferential attachment explains the emergence of hubs

2. Controlled Experiments
- Test how structural features (e.g., clustering, modularity) affect dynamics
- Run simulations in a safe, synthetic environment

3. Predictive Insights
- Anticipate network evolution or failure points
- Guide policy decisions (e.g., vaccination strategies in contact networks)

4. Design of Artificial Systems
- Engineer networks (e.g., peer-to-peer, transport) with desired properties like resilience or efficiency

5. Understanding Function from Structure
- In biological or neural networks, structure constrains function
- Models help bridge the gap between topology and behavior

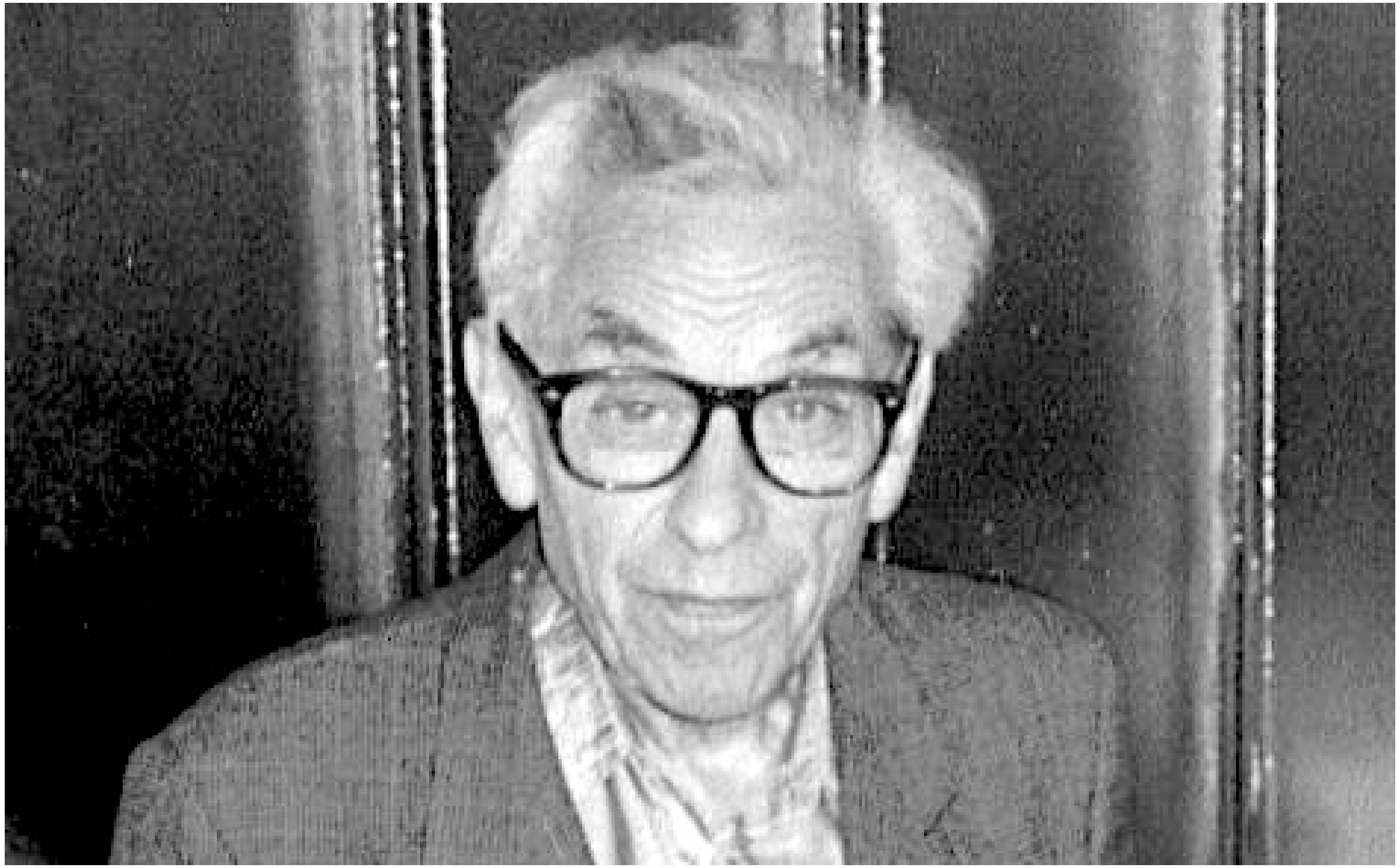
Network Models

universität koblenz
weiter:denken

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

1. Random Networks

Random Networks (1/2)



Paul Erdős (1913-1996)

Random Networks (2/2)

- Simple idea: placing links at random between pairs of nodes
- **Algorithm** (Gilbert random network model (parameters: nodes N and link probability p), equivalent to Erdős-Rényi model):

  1. Start with N nodes and zero links

  2. Go over all pairs of nodes; for each pair of nodes i and j, generate a random number r between 0 and 1

     1. If $r < p \Rightarrow i$ and j get connected

     2. If $r > p \Rightarrow i$ and j remain disconnected

- Erdős-Rényi model: the number of links of the network is fixed
- Gilbert random model: the number of links of the network is variable

Random Networks: Evolution (1/8)

Focus: Connected Components

- In random networks, we examine how components *evolve* as we increase the link probability p

Boundary Cases:

- For p=0:
  - No links at all → Each of the N nodes is isolated
  - Result: N disconnected components (each node is its own component)
- For p=1:
  - All possible links exist → Every node connected to every other
  - Result: 1 fully connected component (a complete graph)

What happens in between p=0 and p=1?

Random Networks: Evolution (2/8)

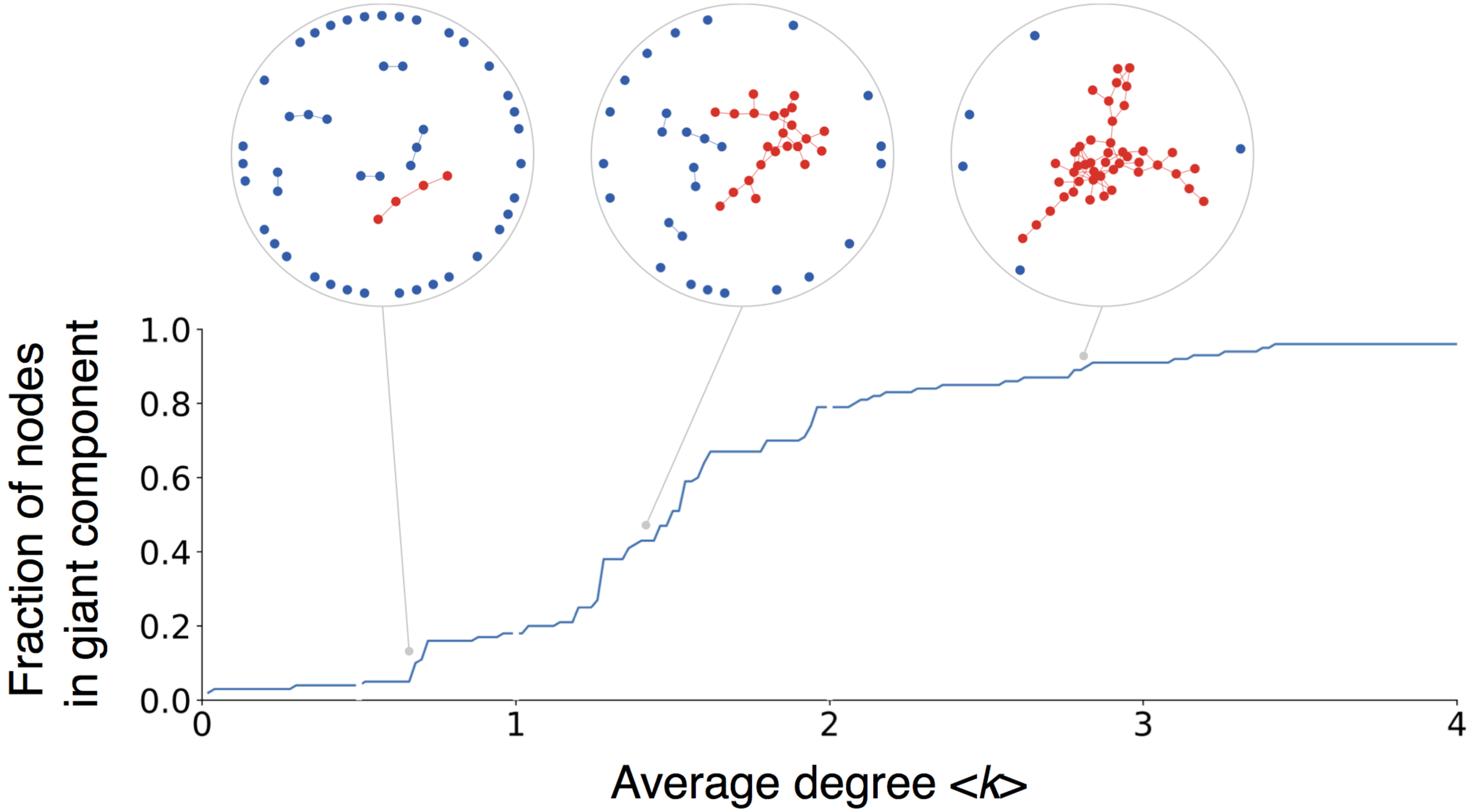
- **Question:** what happens as we add links to the network?
  - **Naïve expectation**
    - The size of the largest connected component **increases gradually**
    - The growth is **smooth and continuous**, like water filling a container

Random Networks: Evolution (3/8)

- **Question:** what happens as we add links to the network?
  - **Naïve expectation**
    - The size of the largest connected component **increases gradually**
    - The growth is **smooth and continuous**, like water filling a container
  - **Wrong:** there is an abrupt increase for a given value of the link probability p

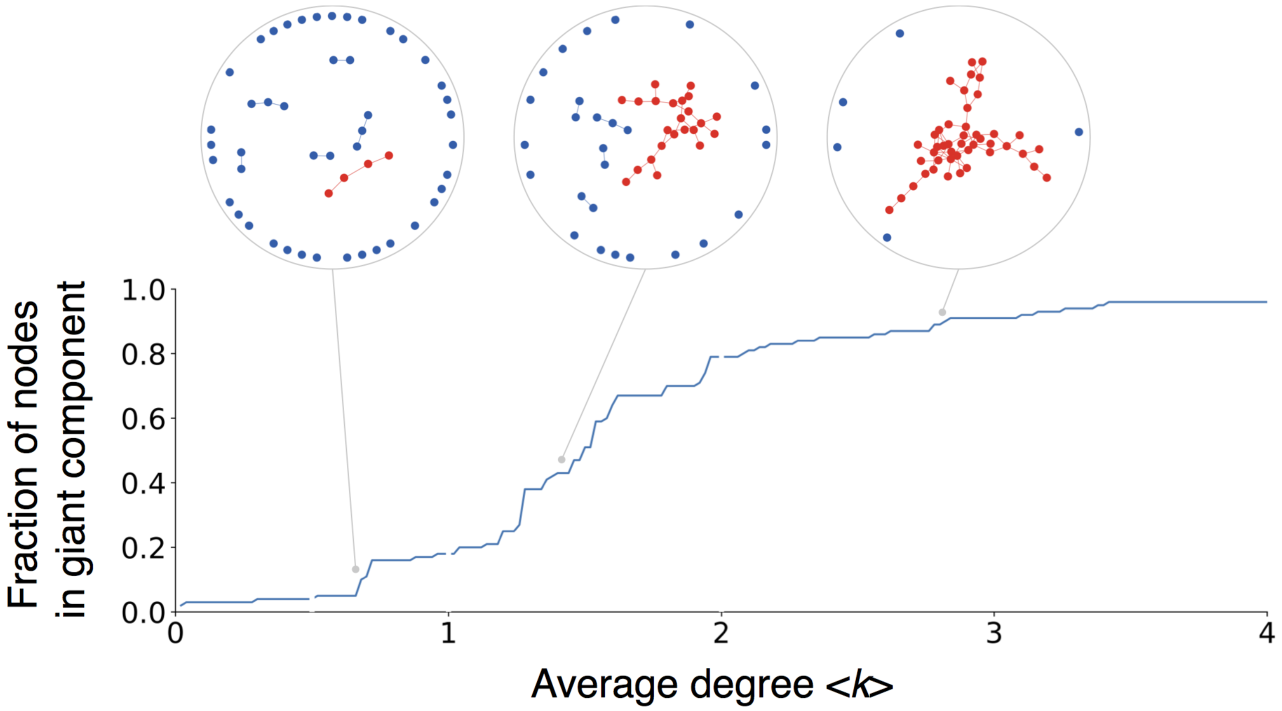
Random Networks: Evolution (4/8)

universität koblenz
weiter:denken



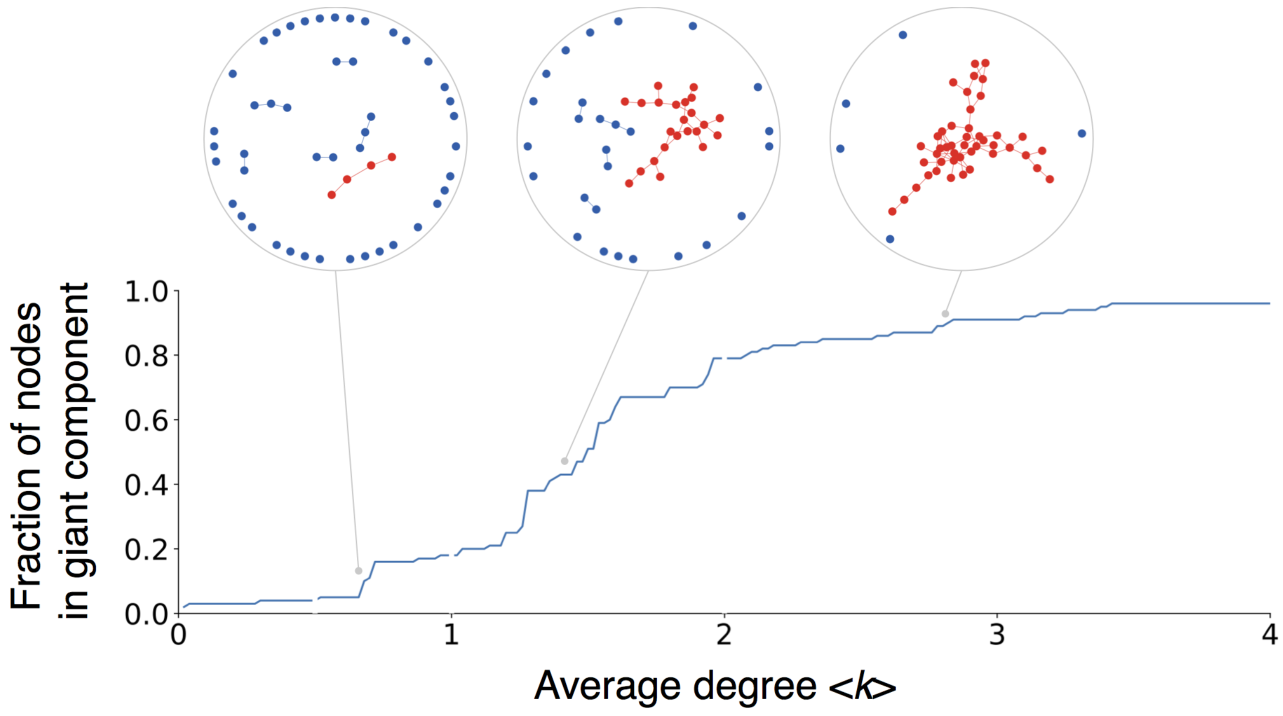
Random Networks: Evolution (5/8)

- **Phase Transition** at ⟨k⟩=1: One of the key findings by Erdős and Rényi is the phase transition that occurs when the average degree ⟨k⟩ reaches 1
  - When ⟨k⟩ is less than 1, the graph is composed mostly of small, disconnected components (as illustrated by the first circle on the left)



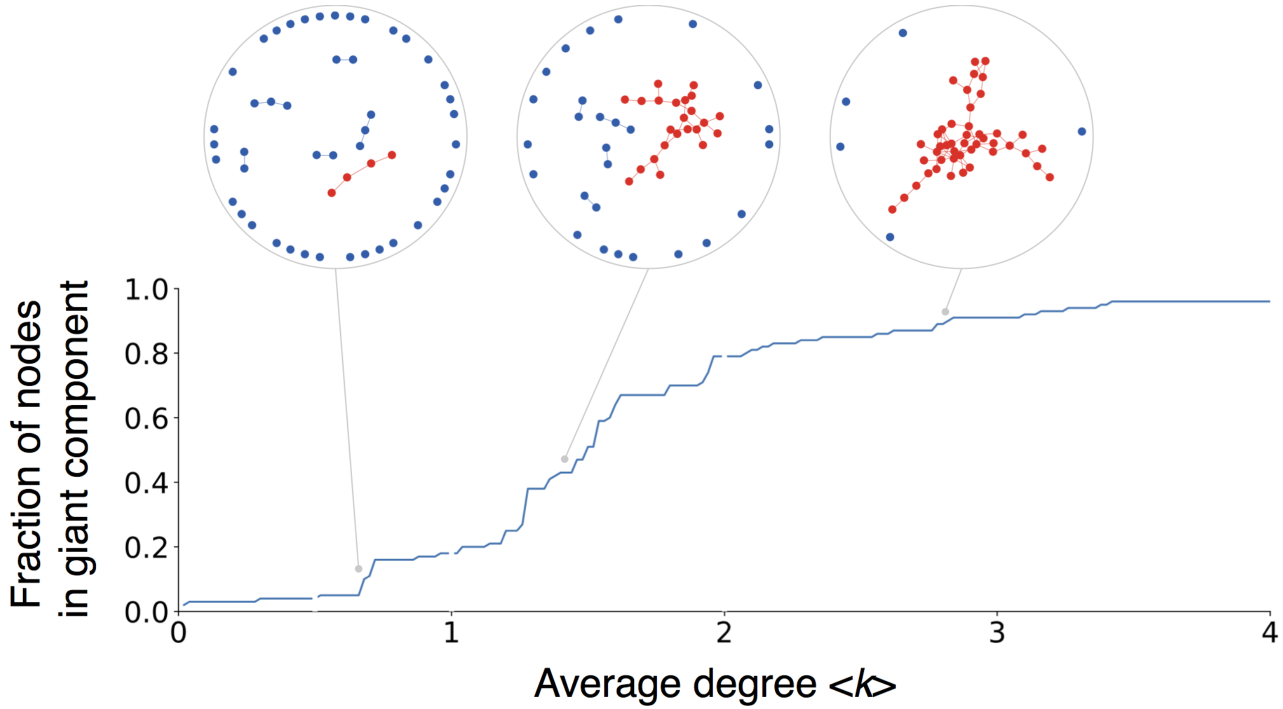
Random Networks: Evolution (6/8)

- **Emergence of Giant Component:** As ⟨k⟩ increases and surpasses 1, a giant component begins to form
  - This is illustrated by the second circle from the left, where we start seeing a more connected structure



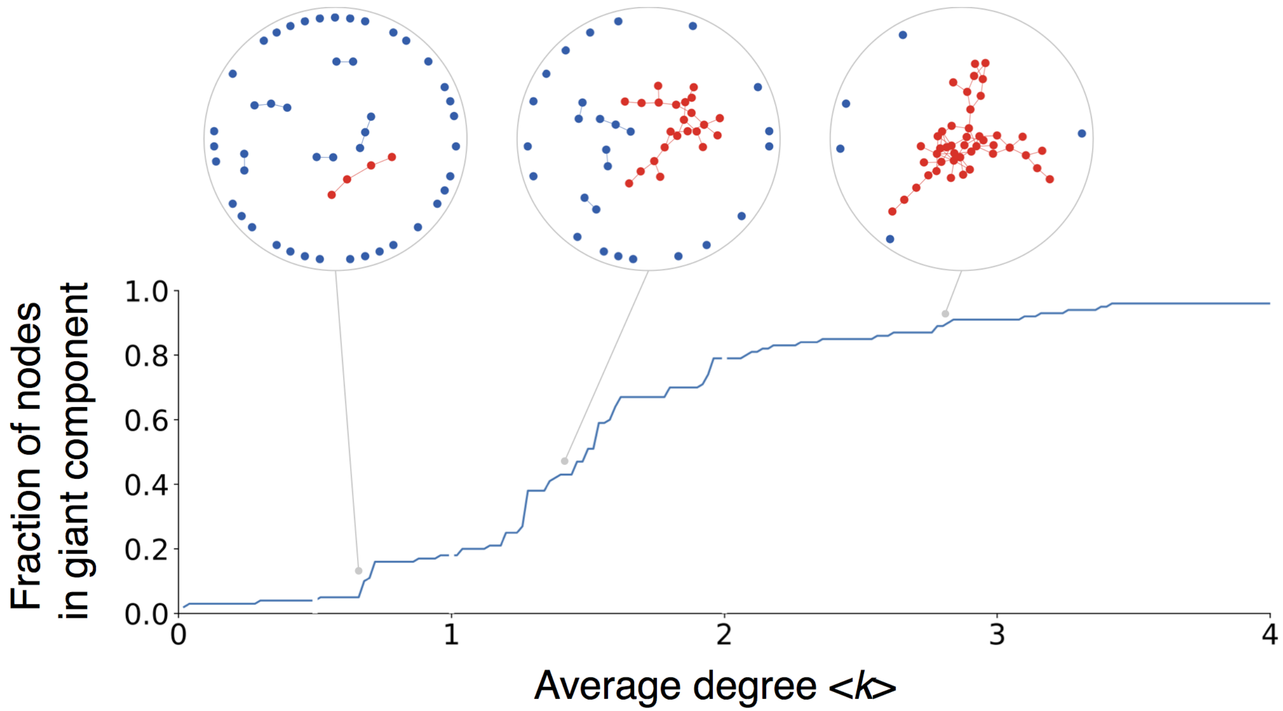
Random Networks: Evolution (7/8)

- **Growth of Giant Component:** As ⟨k⟩ continues to increase beyond 1, the size of the giant component grows rapidly, encompassing more nodes
  - The third circles show the network becoming more interconnected with a larger giant component. By the time ⟨k⟩ is around 3, most nodes are part of this giant component



Random Networks: Evolution (8/8)

- The structure of random networks **changes qualitatively,** not just quantitatively, at specific thresholds
- This makes understanding phase transitions **crucial** in network science



Random Networks: Number of Links, Density, Average Degree (1/3)

- Equivalence with the tossing of a biased coin, which yields heads with probability p
- Number of independent trials (tosses): t
- Number of heads after t trials: h
- **Special cases**
  - $p = 0 \Rightarrow$ the coin never yields heads $\Rightarrow h = 0$
  - $p = 1 \Rightarrow$ the coin always yields heads $\Rightarrow h = \mathrm{t}$
  - $p = 1/2 \Rightarrow$ the coin yields heads (about) half of the times $\Rightarrow h \approx t/2$
- **General rule:** $h \approx pt$

Random Networks: Number of Links, Density, Average Degree (2/3)

- **(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:

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

- **(Expected) density of links d of a random network with N nodes:**

$$d = \frac{\langle L \rangle}{N(N-1)/2} = \frac{pN(N-1)/2}{N(N-1)/2} = p$$

- Real-world networks are **sparse:** for random networks to be good models of real nets ***p* must be very small!**

Random Networks: Number of Links, Density, Average Degree (3/3)

- **(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:

$$t = N - 1 \quad \rightarrow \quad \langle k \rangle = p(N-1)$$

Random Networks: Degree Distribution (1/2)

universität koblenz
weiter:denken

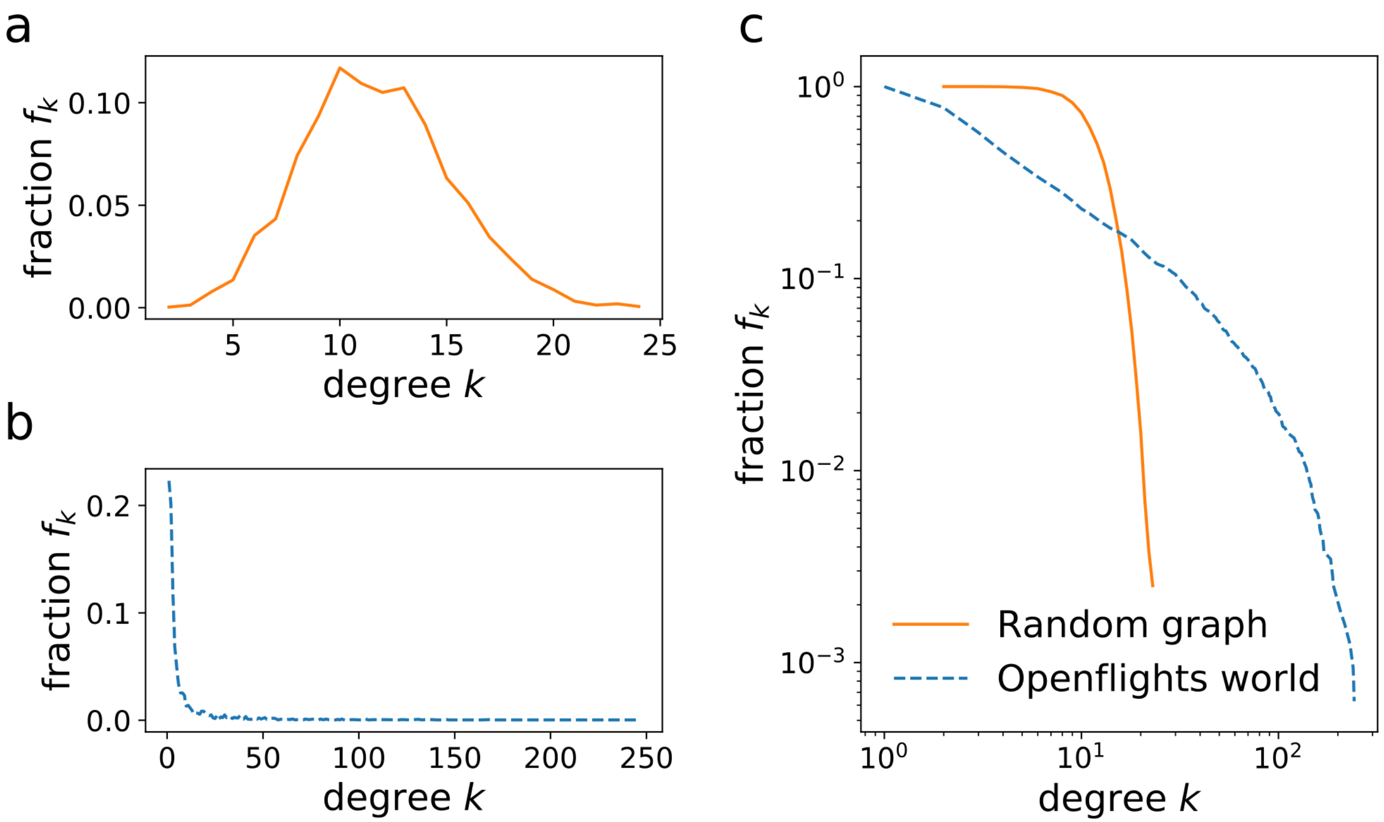
Now, suppose that we have constructed a random network.....

- **Question:** what is the probability that a node has k neighbors? (what is the degree distribution)
- **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, $PN \approx \langle k \rangle$, is well approximated by a bell-shaped curve $\Rightarrow$ **most degree values are concentrated around the peak, so the average degree is a good descriptor of the distribution**

Random Networks: Degree Distribution (2/2)



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

Random Networks: Small-world Property (1/4)

universität koblenz
weiter:denken

- **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 (steps) d = 1 there are k nodes
  - At distance d = 2 there are k(k – 1) nodes
  - …
  - 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 (2/4)

universität koblenz
weiter:denken

- **Question:** how many steps does it take to cover the whole network? (how far from a node do we have to go to reach all other nodes?)

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

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

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

- 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: Small-world Property (3/4)

universität koblenz
weiter:denken

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

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

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

- The fact that the maximum distance to go from any node to any other node (the diameter) in a random network is small compared to the size of the network means that Erdős– Rényi networks indeed have short paths

Random Networks: Small-world Property (4/4)

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

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

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

- **Example**: let us consider the world's network of social contacts and imagine that it is a random network
- If we take k=150, which is the average number of regular contacts that humans can maintain (*Dunbar's number*), at distance five, the number of reachable people is $150^5 \approx$ 75 billion, a factor of 10 larger than the world population
- In principle we could reach any individual in five steps or less, which is compatible with the result of Milgram's small-world experiment

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 $\langle k \rangle$ is usually a small number, the average clustering coefficient of random networks with realistic values for $\langle k \rangle$ and N is **much smaller** than the ones observed in real-world networks (e.g. real social networks)

Random Networks: Summary

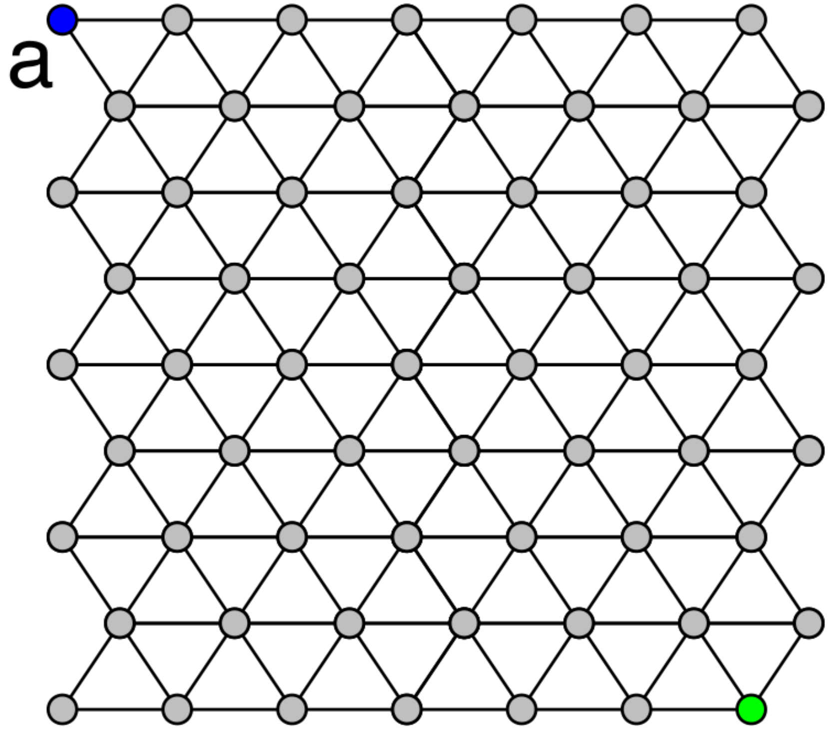
- Links are placed at random, independently of each other
  - Distances between pairs of nodes are short (small-world property): 😊
- The average clustering coefficient is much lower than on real networks of the same size and average degree: ☹
- The nodes have approximately the same degree, there are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

```
G = nx.gnm_random_graph(N,L) # Erdos-Renyi random graph
G = nx.gnp_random_graph(N,p) # Gilbert random graph
```

2. Small-World Networks

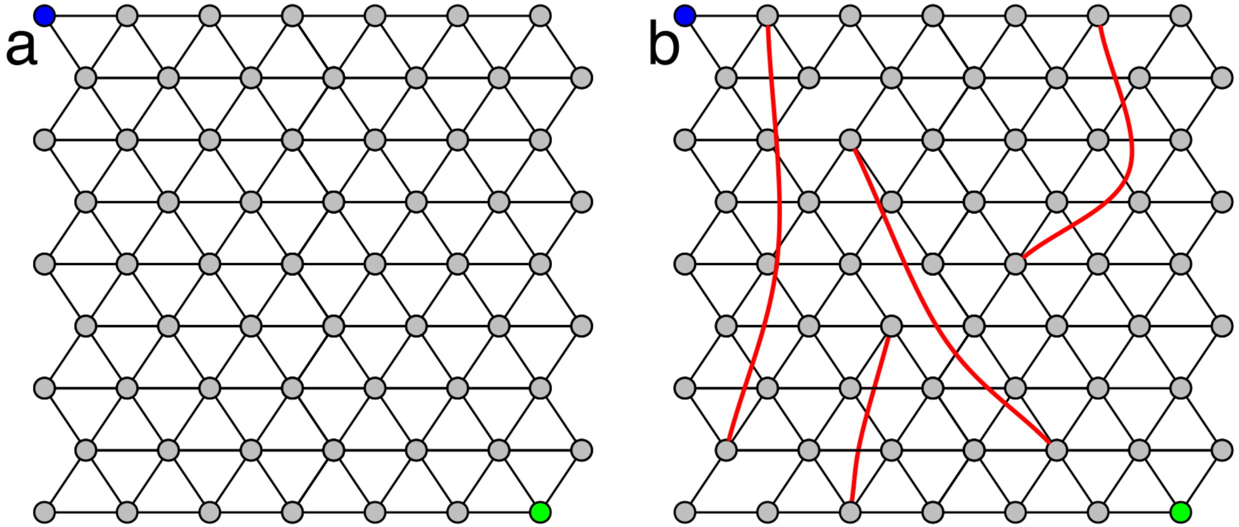
Small-World Networks (1/3)

- **Goal:** building networks with the small-world property and high clustering coefficient
- **Solution:** interpolating between a regular lattice (high clustering) and a random network (small-world property)
- **Clustering coefficient of lattice is high:**
  - The internal nodes have $k = 6$ neighbors, 6 pairs of which are connected
  - $C = 6/[(6*5)/2] = 6/15 = 2/5 = 0.4$
  - Most nodes are internal, so the average clustering coefficient of the network is close to 0.4!



Small-World Networks (2/3)

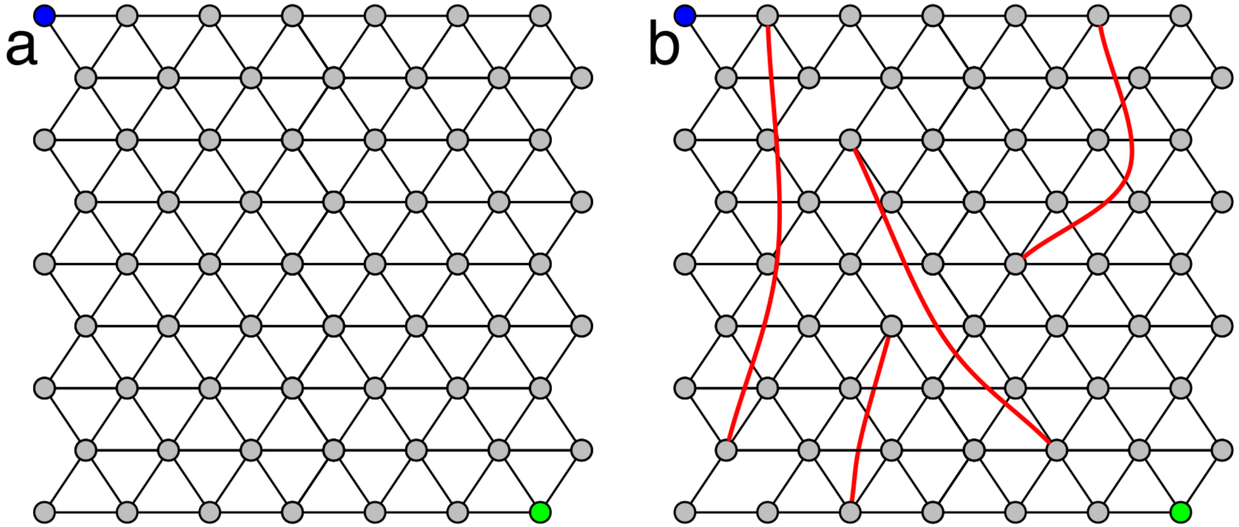
- **Large average shortest path length:** Going from a node to another can take a large number of steps, which grows rapidly with the size of the network/grid



- **Solution?**

Small-World Networks (3/3)

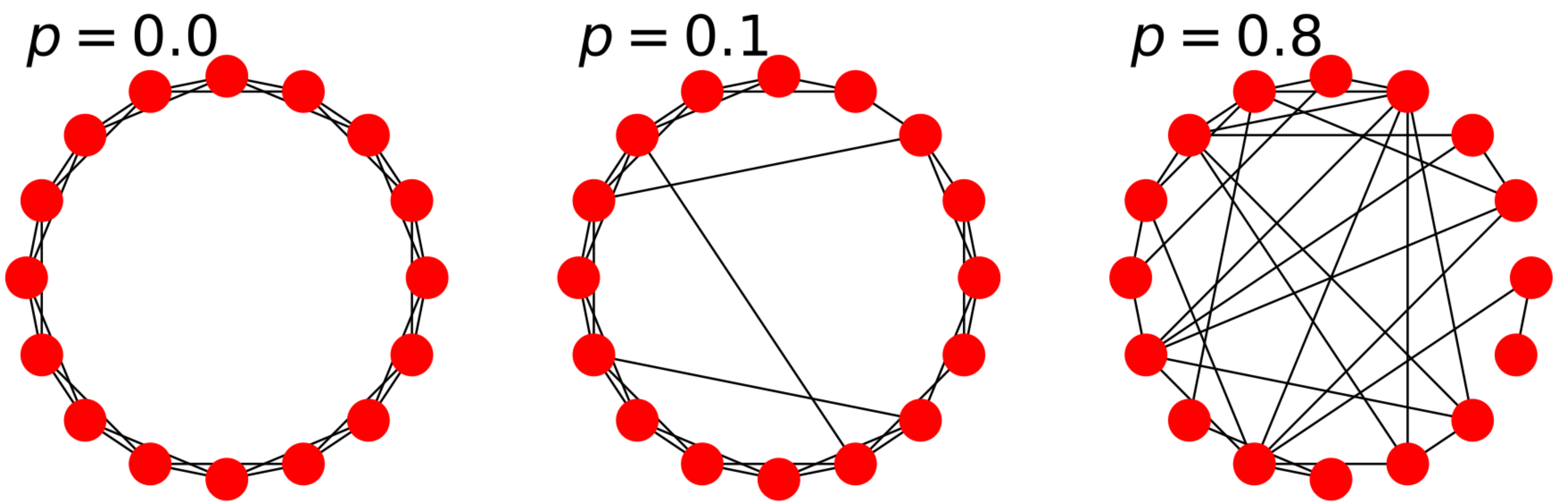
- **Large average shortest path length:** Going from a node to another can take a large number of steps, which grows rapidly with the size of the network/grid



- **Solution?** **Shortcuts!**

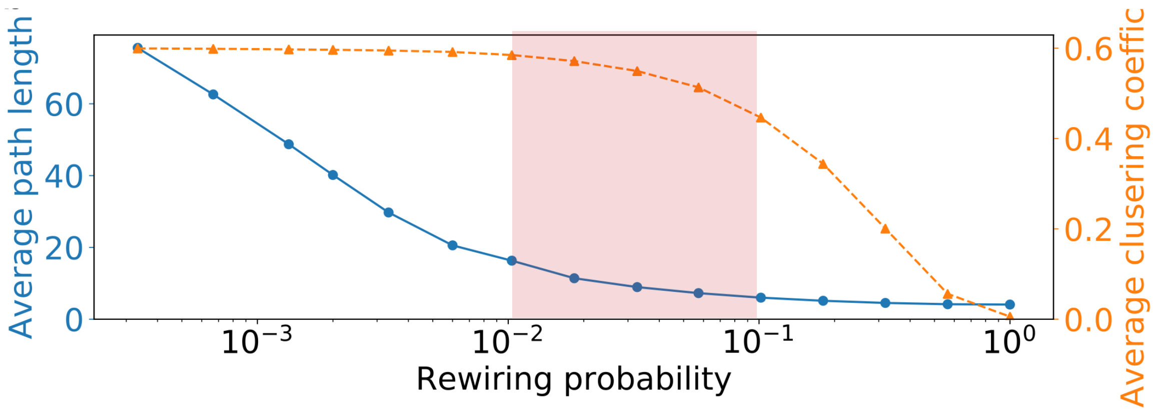
The Watts-Strogatz Model (1/2)

- N nodes form a regular ring lattice, with even degree k. With probability p, each link is rewired randomly



The Watts-Strogatz Model (2/2)

- 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**

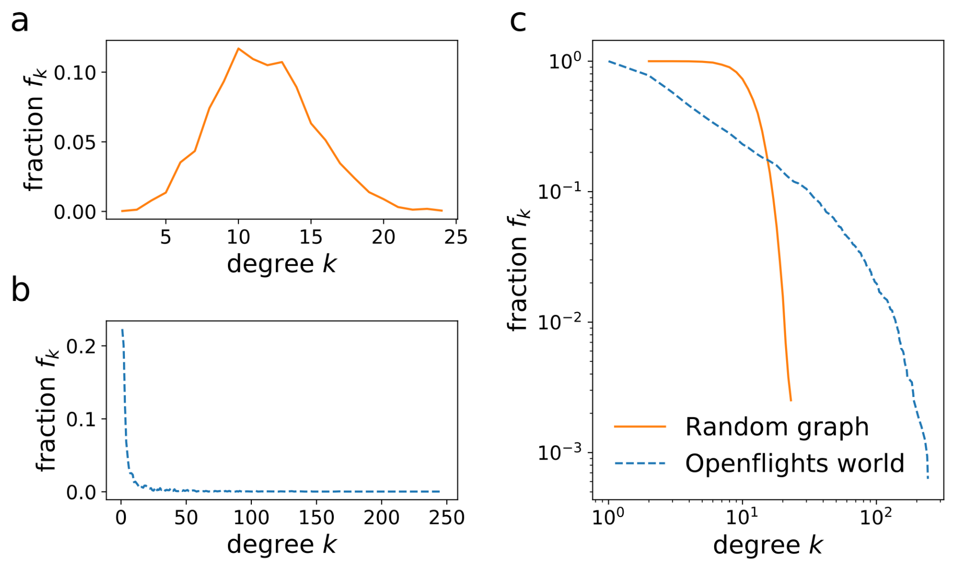


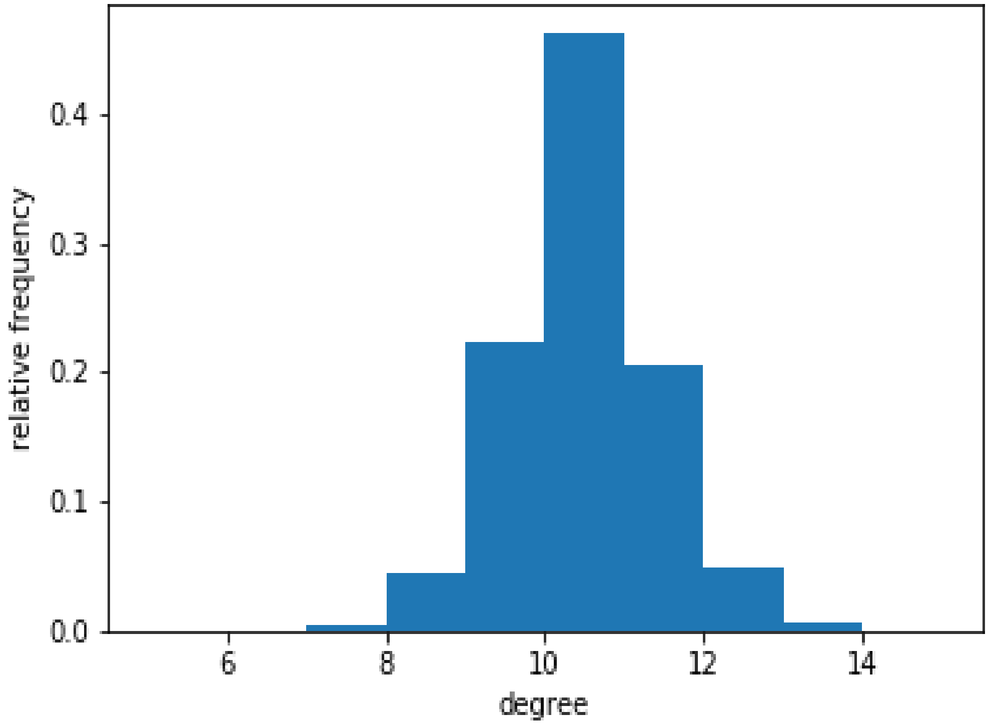
Distances become short already for low values of p; the average clustering coefficient stays high up to large values of p. **There is a range of values of p where the average path length is short and the clustering coefficient is high!**

The Watts-Strogatz Model: Degree Distribution

universität koblenz
weiter:denken

- 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





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 😊
- The nodes have approximately the same degree, there are no hubs 😟

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

3. Summary

Summary

- **Purpose of Network Models**
  - Help us understand the basic mechanisms responsible for characteristic structural features in real networks
- **Core Elements**
  - Rules that determine how nodes get connected to each other

Summary (1/2)

- **Random Networks (Erdős–Rényi Model)**
  - **Uniform Probability**: Every node can become a neighbor of any other node with the same likelihood
  - **Short Path Lengths**: The average distance between nodes is small, leading to efficient communication across the network 😊
  - **Low Clustering**: There are very few triangles, indicating that the network lacks tightly knit groups 😟
  - **Absence of Hubs**: No nodes with significantly higher degrees than others, leading to a homogeneous degree distribution 😟

Summary (2/2)

- **Small-World Model (Watts-Strogatz Model)**
  - Begins with a regular lattice where each node is connected to its nearest neighbors, creating a high average clustering coefficient
  - Random shortcuts are introduced by rewiring some edges, creating long-range connections between nodes
  - **Small-World Property**: A few shortcuts drastically reduce the average distance between nodes, making the network highly navigable 😊
  - **High Clustering Coefficient**: Despite the shortcuts, the network retains a high level of clustering, maintaining local interconnectedness 😊
  - **No Hubs**: The model does not naturally create hubs, which are nodes with significantly higher connections than average 😟

References

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). A First Course in Network Science Cambridge: Cambridge University Press.

- Chapter 5.1-5.2

[2] OLAT course page: https://olat.vcrp.de/url/RepositoryEntry/4669112833