Network Theory and Dynamic Systems

08. Network Models

SOSE 2025

Dr. -Ing. Stefania Zourlidou

Institute for Web Science and Technologies
Universität Koblenz

Recap from Previous Lecture (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 😟

Recap from Previous Lecture (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 ☹

Objectives of this Lecture

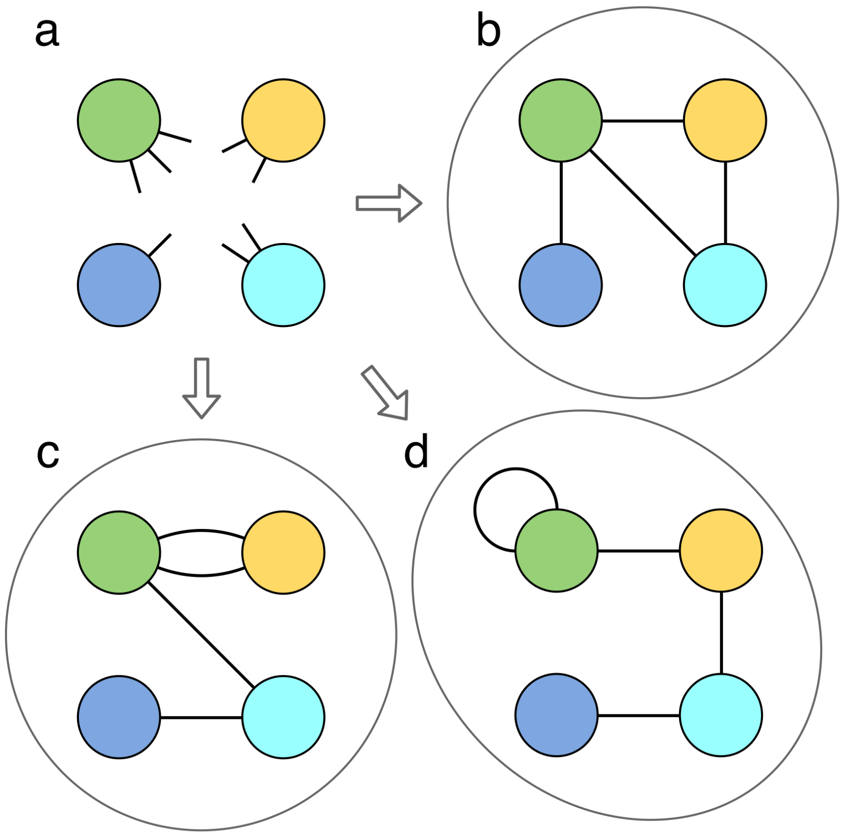
- The Configuration Model
- Preferential Attachment
- Other Preferential Models

universität koblenz
weiter:denken

1. The Configuration Model

The Configuration Model (1/2)

- **Problem**: is it possible to build networks with a predefined degree distribution?
- **Solution**: the configuration model
- **More specific focus:** build networks with a predefined degree sequence!
- **Degree sequence:** list of N numbers $(k_1, k_2, \ldots, k_{N-1}, k_N)$, where k_i is the degree of node i
- **Warning:** many degree sequences can be extracted from the same distribution!
- **Principle:** assign a degree to each node (*e.g.*, from the desired distribution or a real network), place as many stubs on each node as the degree of the node, and attach pairs of stubs at random



The Configuration Model (2/2)

- **Degree-preserving randomization:** generate randomized versions of a given network with the same degree sequence, using the configuration model
- **Why:** useful to see whether a specific property of the original network is determined by its degree distribution alone
  - If the property is maintained in the randomized configurations, then the degree distribution is the main driver
  - If the property is lost in the randomized configurations, other factors must be responsible for it

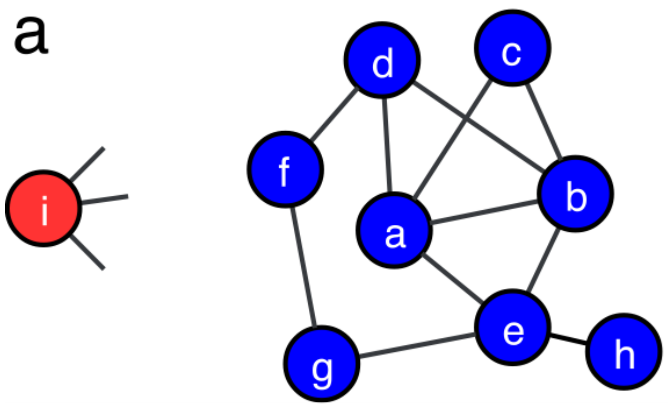
```
# network with degree sequence D
G = nx.configuration_model(D)
```

Network Growth (1/2)

- **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 (2/2)

- **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



universität koblenz
weiter:denken

2. Preferential Attachment

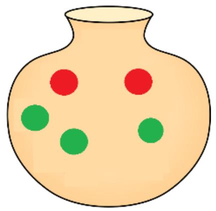
Preferential Attachment

- **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
    - The rich gets richer and the poor gets poorer!

Polya’s Urn Model



- **Start**: an urn contains X white and Y black balls
- **Process**: a ball is drawn from the urn and put back in with another ball of the same color
- Example
  - If we first pick a white ball, there will be X+1 white and Y black balls in the urn; **white will become more likely to be picked than black in the future**
- Preferential attachment used to **explain heavy-tail distributions of many quantities:** the number of species per genus of flowering plants, the number of (distinct) words in a text, the populations of cities, individual wealth, scientific production, citation statistics, firm size, etc.

The Barabási-Albert Model (1/2)

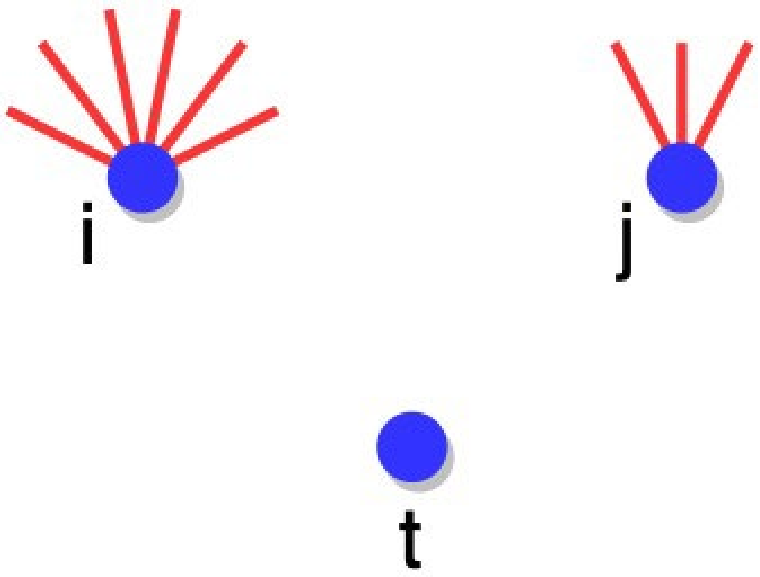
- **Procedure**
  - Start with a group of m_0 nodes, usually fully connected (clique)
  - At each step a new node i is added to the system, and sets m links with some of the older nodes ($m \leq m_0$)
  - The probability that the new node i chooses an older node j as neighbor is **proportional to the 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 (2/2)

- **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 (1/2)

- **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 (2/2)

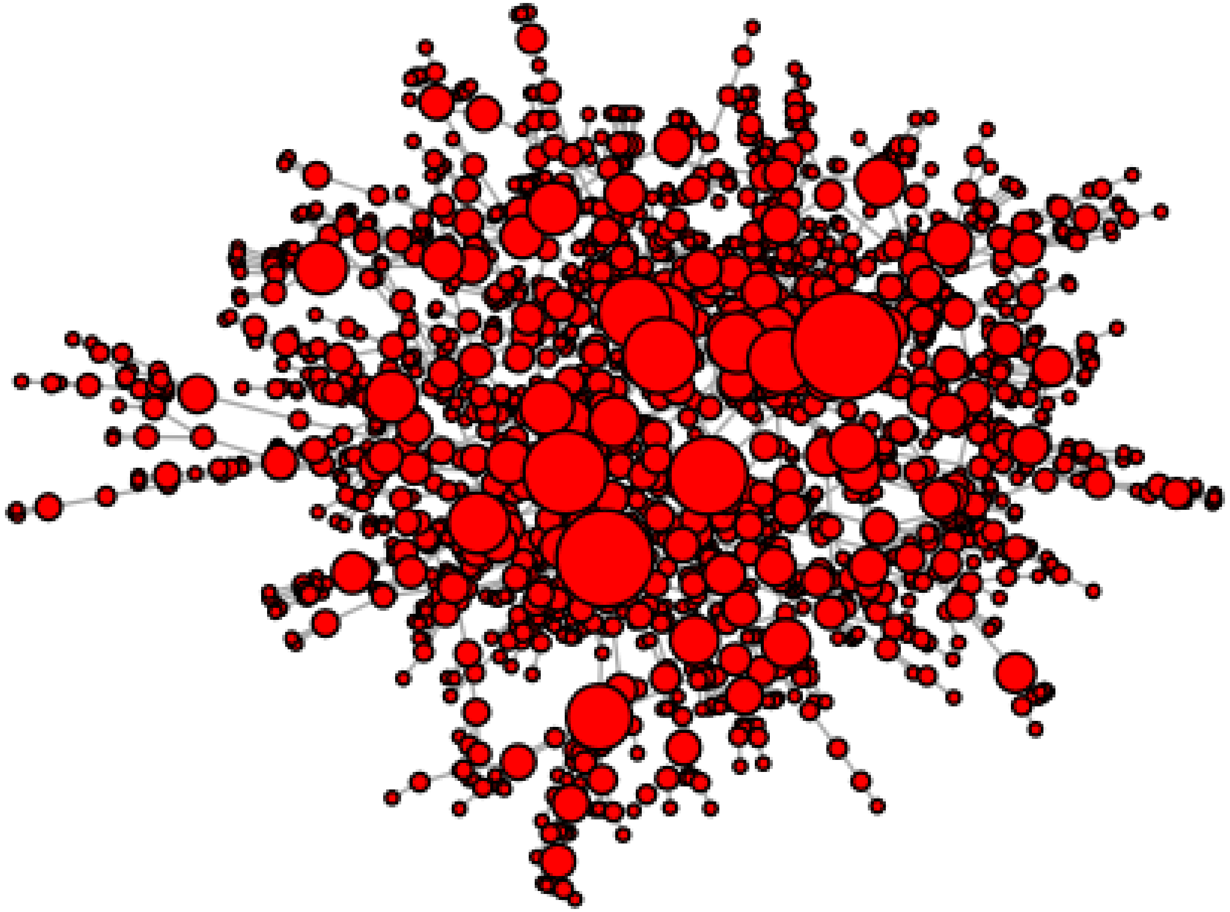
- **Question:** what if nodes have to be picked with different probabilities?
- **Answer:** we need to provide a second list, whose elements are the weights associated with the nodes
- **Note:** weights are used to calculate probabilities, but do not have to be integers or add up to one
- **Example:**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

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

The Barabási-Albert Model (1/3)

- **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

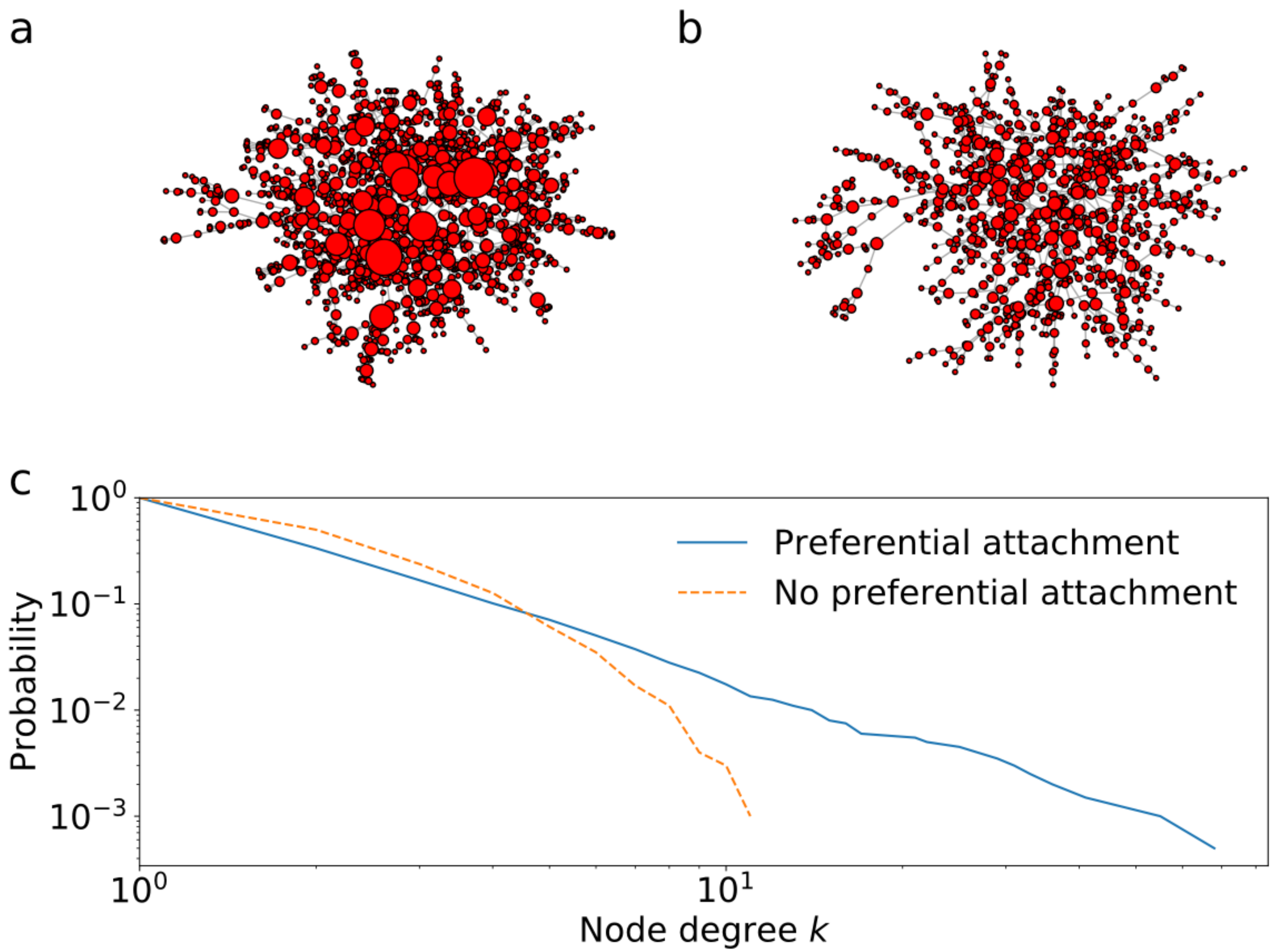
```
# BA model network
G = nx.barabasi_albert_graph(N,m)
```



The Barabási-Albert Model (2/3)

- 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 (3/3)



- **Conclusion:** growth + random attachment **does not** generate hubs

Preferential attachment is necessary!

3. Other Preferential Models

Other Preferential Models (1/2)

- The BA model uses **linear preferential attachment:** the linking probability is **proportional** to the degree
- **Question:** what happens if the linking probability is **proportional to a power of the degree?**

Other Preferential Models (2/2)

- The BA model uses **linear preferential attachment:** the linking probability is **proportional** to the degree
- **Question:** what happens if the linking probability is **proportional to a power of the degree?**
- **Non-linear preferential attachment!**

Non-linear Preferential Attachment (1/4)

universität koblenz
weiter:denken

- 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 power α of the degree k_j of j:**

$$\Pi_\alpha(i \leftrightarrow j) = \frac{k_j^\alpha}{\sum_l k_l^\alpha}$$

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

Non-linear Preferential Attachment (2/4)

$$\Pi_\alpha(i \leftrightarrow j) = \frac{k_j^\alpha}{\sum_l k_l^\alpha}$$

- For $\boldsymbol{\alpha} = 1$ we recover the **linear preferential attachment (BA model)**
- **Question**: what happens when $\boldsymbol{\alpha} \neq 1$?

Non-linear Preferential Attachment (3/4)

$$\Pi_\alpha(i \leftrightarrow j) = \frac{k_j^\alpha}{\sum_l k_l^\alpha}$$

- For $\boldsymbol{\alpha} = 1$ we recover the **linear preferential attachment (BA model)**
- **Question**: what happens when $\boldsymbol{\alpha} \neq 1$?
- **Answer**: it depends on whether $\boldsymbol{\alpha} > 1$ or $\boldsymbol{\alpha} < 1$

Non-linear Preferential Attachment (4/4)

universität koblenz
weiter:denken

- For $\boldsymbol{\alpha}$ <1, the link probability does not grow fast enough with degree, so the advantage of high-degree nodes over the others is not as big. As a result, **the degree distribution does not have a heavy tail: the hubs disappear!**
- If $\boldsymbol{\alpha}$ >1, high-degree nodes accumulate new links much faster than low-degree nodes. As a consequence, one of the nodes will end up being connected to a fraction of all other nodes. For $\boldsymbol{\alpha}$ >2, a single node may be connected to all other nodes (**winner-takes-all effect**), all other nodes having low degree
- **Conclusion:** Non-linear preferential attachment fails to generate hubs. Linear preferential attachment is the only way to go
- **Problem:** Strict proportionality of linking probability to degree appears unrealistic!

Limits of Preferential Attachment

- It yields a **fixed pattern for the degree distribution:** the slope is the same for any choice of the model parameters. Degree distributions in real-world networks could decay faster or more slowly
- The **hubs are the oldest nodes**: new nodes cannot overcome their degree
- It **does not create many triangles**: the average clustering coefficient is much lower than in many real-world networks
- **Nodes and links are only added:** in real networks they can also be deleted
- Since each node is attached to older nodes, the **network consists of a single connected component**. Many real-world networks have multiple components

Extensions of the BA Model: Attractiveness Model (1/3)

- **Pitfall of preferential attachment:** What happens if a node has no neighbors (degree zero)? It will never get connections from other nodes!
- **No problem for standard initial condition:** the initial subgraph is complete (clique), so every node has nonzero degree
- **What if the network is directed and the linking probability is proportional to the in-degree? Bad,** as each new node has in-degree zero, so it will never be linked by future nodes!

Extensions of the BA Model: Attractiveness Model (2/3)

- **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 sum of the degree k_j of j and an attractiveness A:

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

Extensions of the BA Model: Attractiveness Model (3/3)

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

- For $A = 0$ we recover the **BA model**
- For every value of A we get networks with heavy-tailed degree distributions
- The pattern of the distribution **changes with *A***, so it is possible to match distributions of real-world networks, **unlike the BA model**

Extensions of the BA Model: Fitness Model (1/4)

universität koblenz
weiter:denken

- **Pitfall of preferential attachment:** the hubs are the oldest nodes. **Unrealistic!**
- **Examples:**
  - In the Web, new pages can overrun old pages (*e.g.,* Google!)
  - In science, new papers can be more successful than (many) old papers
- **Reason:** each node has its own individual appeal!

Extensions of the BA Model: Fitness Model (2/4)

universität koblenz
weiter:denken

- **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 product of the degree k_j of j with a fitness η_j**, indicating the intrinsic appeal of j:

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

Extensions of the BA Model: Fitness Model (3/4)

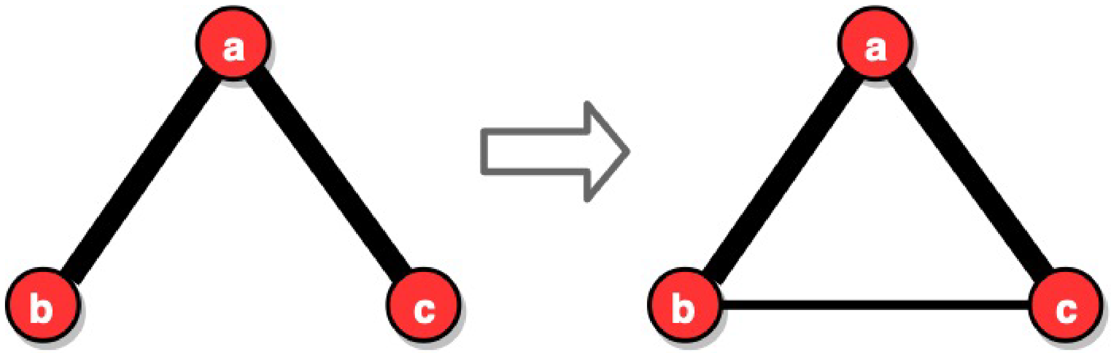
- The fitness values are extracted from a distribution $\rho(\eta)$ and assigned to each new node
- **Difference with attractiveness model**
  - The fitness enters as a **factor** in the link probability, not as a summand
  - The fitness is characteristic of each node, it is not a constant

Extensions of the BA Model: Fitness Model (4/4)

- **Results**
  - If the fitness distribution $\rho(\eta)$ has finite support, *i.e.*, the fitness is distributed over a finite range of values, the degree distribution of the network is heavy-tailed
  - If the fitness distribution $\rho(\eta)$ has infinite support, *i.e.*, the fitness is distributed over an infinite range of values, the node with largest fitness attracts a fraction of all links (**monopoly**)
  - Nodes with large fitness can acquire a large degree even if they are introduced late in the system (**good!**)

Extensions of the BA Model: Random Walk Model (1/6)

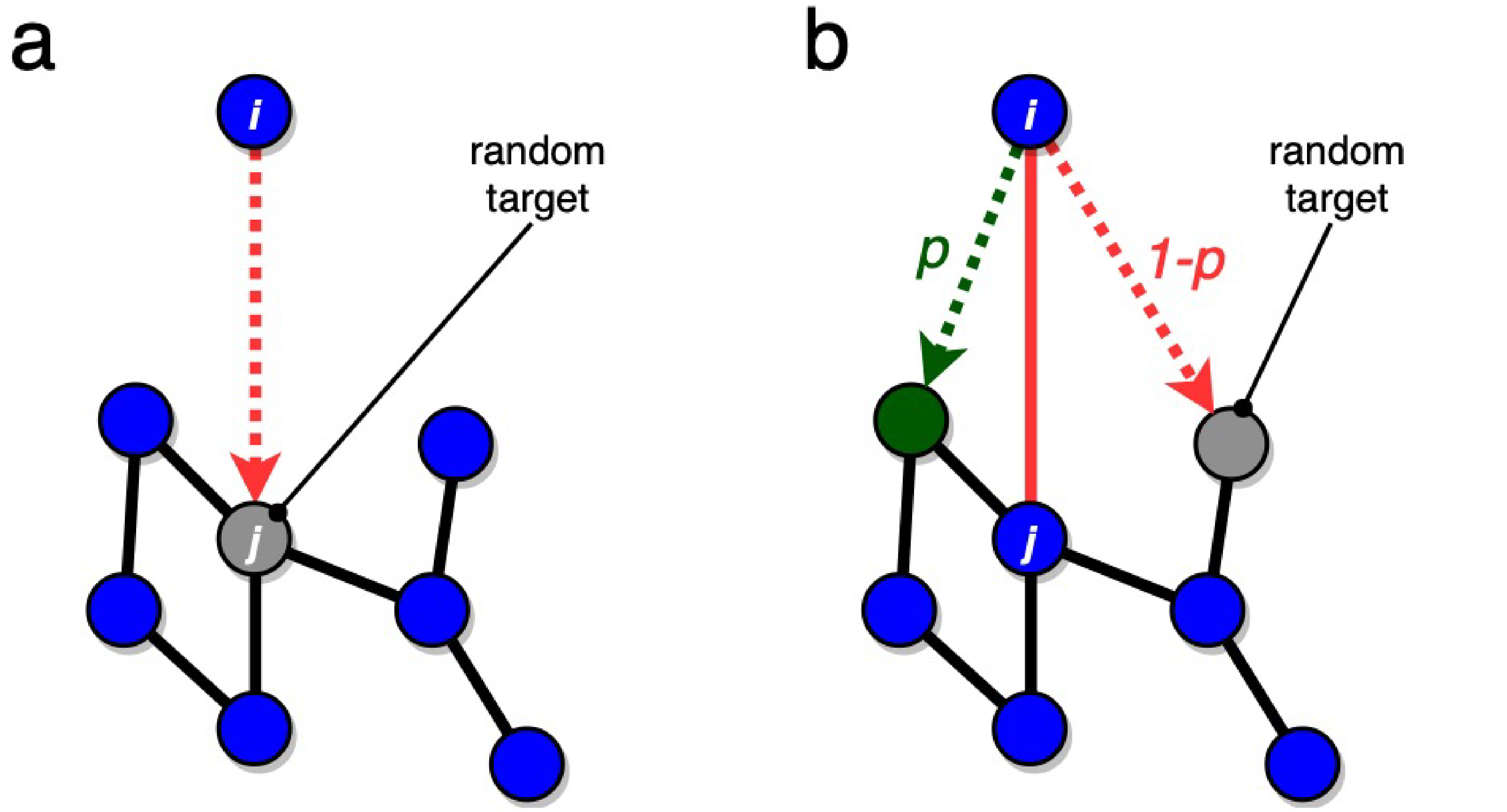
- **Pitfall of preferential attachment:** the BA model does not generate many triangles. **Why?**
- To close a triangle we need to set a link between two neighboring nodes, whereas in the BA model links are set based on degree, regardless of whether the future neighbors have common neighbors
- **Solution:** introduce a mechanism for triadic closure in the model!



Extensions of the BA Model: Random Walk Model (2/6)

- 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 ($1 < m \leq m_0$)
  - The first link targets a randomly chosen node j
  - From the second link onwards:
    - With probability p the link is set with a neighbor of j, chosen at random
    - With probability $1-p$ the link is set with a randomly chosen node

Extensions of the BA Model: Random Walk Model (3/6)



Extensions of the BA Model: Random Walk Model (4/6)

- Results
  - The degree distribution is heavy-tailed
  - The average clustering coefficient is much higher than in BA networks (the larger, the greater the probability p of triadic closure)
  - When the triadic closure probability p is sufficiently high that many triangles are formed ($p \sim 1$) the network has **community structure**, *i.e.*, it is made of cohesive groups of nodes, loosely connected to each other

Extensions of the BA Model: Random Walk Model (5/6)

- **Question:** if links are set at random, as it seems, how can the model generate hubs?
- **Answer:**
  - Choosing a random node and a random neighbor of the node amounts to choosing a link at random
  - The probability that the endpoint(s) of a randomly selected link have a given degree **is proportional to the degree**

Extensions of the BA Model: Random Walk Model (6/6)

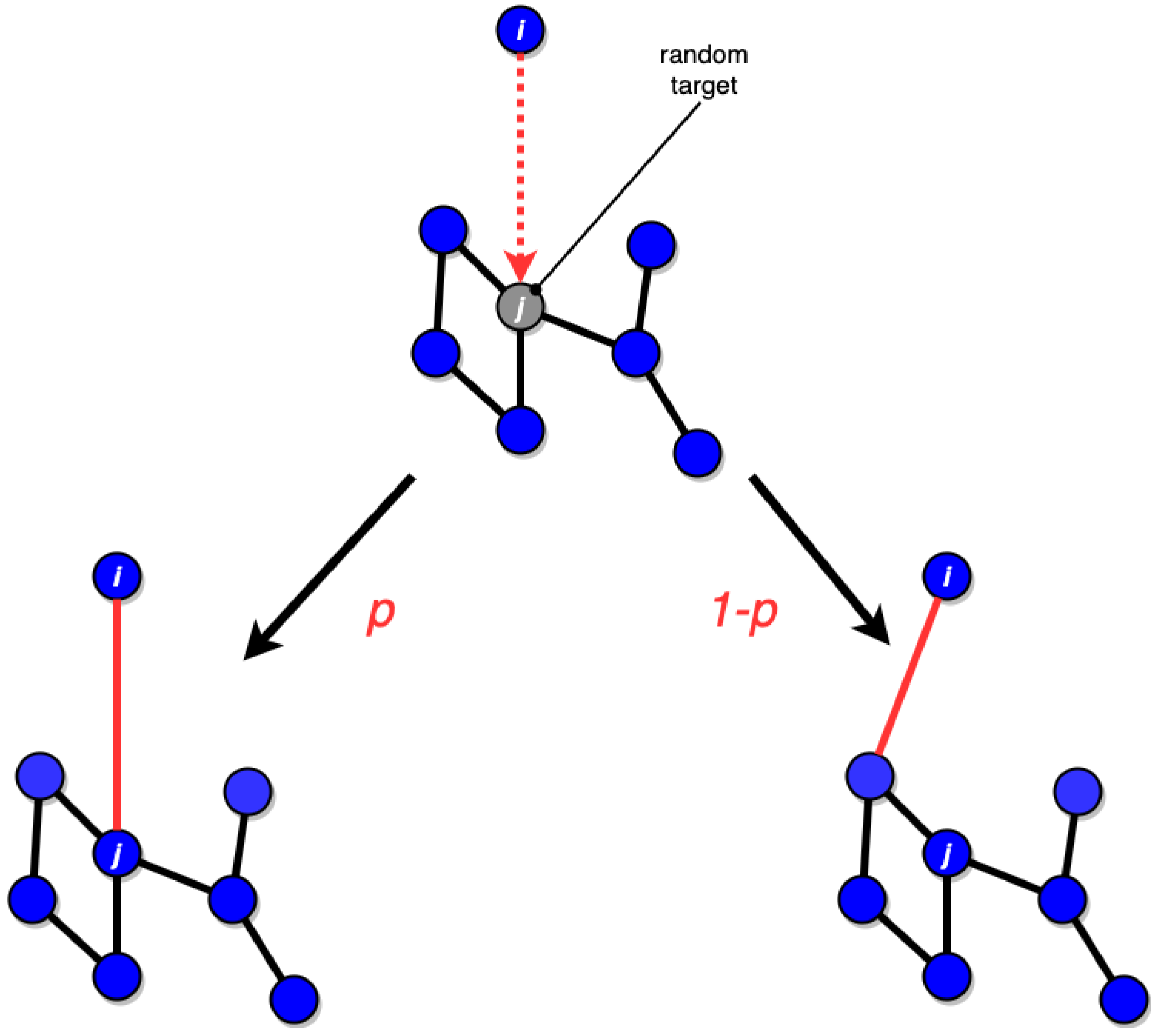
- **Conclusion:** the triadic closure mechanism of the random walk model **induces effective preferential attachment!**
- **Take-home message:** preferential attachment can be induced by simple mechanisms based on random choices; it is not necessary to require the knowledge of the degree of the nodes, nor a strict expression of the link probability!

Extensions of the BA Model: Copy Model (1/3)

- **Motivation:** the authors of a new Webpage tend to "copy" the hyperlinks of other pages on the same topic
- **Steps**
  - **Growth:** at each time step a new node i is added to the network
  - **Target selection:** a node j is selected at random
  - **Random connection:** with probability p the new node is connected to j
  - **Link copying:** with probability $1{-}p$ the new node is connected to a neighbor of j, chosen at random

Extensions of the BA Model: Copy Model (2/3)

- Difference from random walk model: here we do not link to the target and to its neighbor (no triadic closure)!

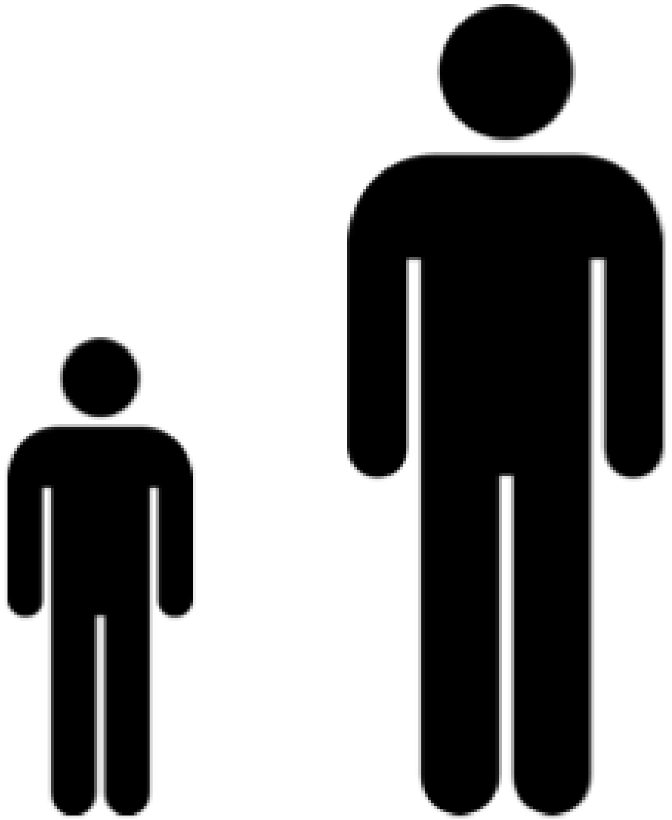


Extensions of the BA Model: Copy Model (3/3)

- **Relevance to real systems**
  - **Social Networks:** The more acquaintances an individual has, the higher is the chance that she will be introduced to new individuals by her existing acquaintances (we "copy" the friends of our friends)
  - **Citation Networks:** Authors decide what to read and cite by "copying" references from the papers they have read. Consequently papers with more citations are more likely to be studied and cited again
  - **Gene duplication:** Responsible for the emergence of new genes in a cell, can be mapped into the copying model

Extensions of the BA Model: Rank Model (1/3)

- **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 (2/3)

- **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*:**

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

Extensions of the BA Model: Rank Model (3/3)

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

- **Remark:** highly-ranked nodes (those with low values of *R*) have high probabilities of being linked, much higher than poorly-ranked nodes
- **Result:** the model generates networks with hubs, for any value of the exponent $\boldsymbol{\alpha}$ and any property used to rank the nodes!

4. Summary

Summary (1/2)

- The configuration model creates networks with any given degree sequence, meaning the structure is manually imposed rather than generated by the model itself
  - It is often used as a baseline to determine whether a network's properties are solely due to its degree distribution or influenced by other factors
    - ✓ This is done by comparing the original network's properties to those of randomized networks with the same degree sequence created by the model
- Realistic network models account for network growth, where nodes and links are added over time, reflecting the development of many real-world networks such as the Internet and the Web

Summary (2/2)

- The Barabási–Albert model combines network growth and preferential attachment, resulting in networks with heavy-tailed degree distributions, which explains the emergence of hubs
- Preferential attachment can also be induced implicitly through simple processes involving random choices, such as triadic closure and link selection
- Various models have been proposed to address the limitations of the Barabási–Albert model by incorporating factors like attractiveness, fitness, triadic closure, and ranking

References

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

- Chapter 5.3-5.5

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