



UNIVERSITÀ
DI TORINO

Analisi e Visualizzazione delle Reti Complesse

NS09 - Preferential Attachment Models

Prof. Rossano Schifanella





Outline

- Preferential attachment
- Barabási-Albert model
- Other preferential models
 - Attractiveness model
 - Fitness model
 - Random walk model
 - Copy model

Network growth

Note: real-world networks are **dynamic**!

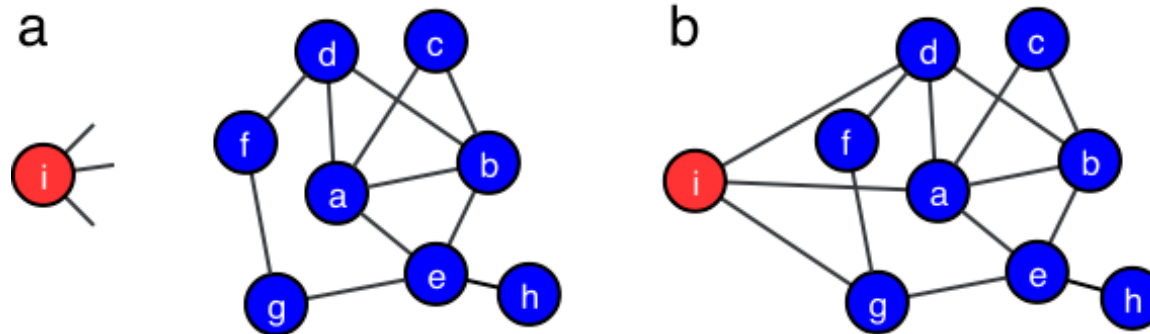
Examples:

- 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 for over 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 after adding 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

"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 get richer, and the poor get poorer

[\[Example in NetLogo\]](#)

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, and many others!

The Barabási-Albert model

Procedure:

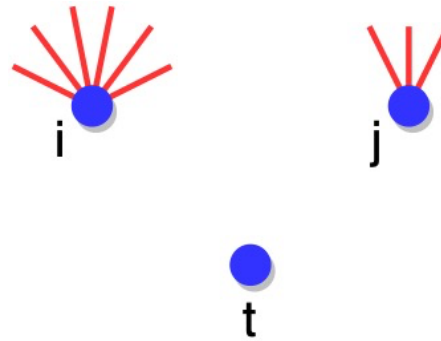
- Start with a group of m_0 nodes, usually fully connected (clique)
- At each step a new node i is added to the system, and sets m links with older nodes ($m \leq m_0$)
- The probability that the new node i chooses an older node j as a neighbor is **proportional to the degree k_j of j** :

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

- 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

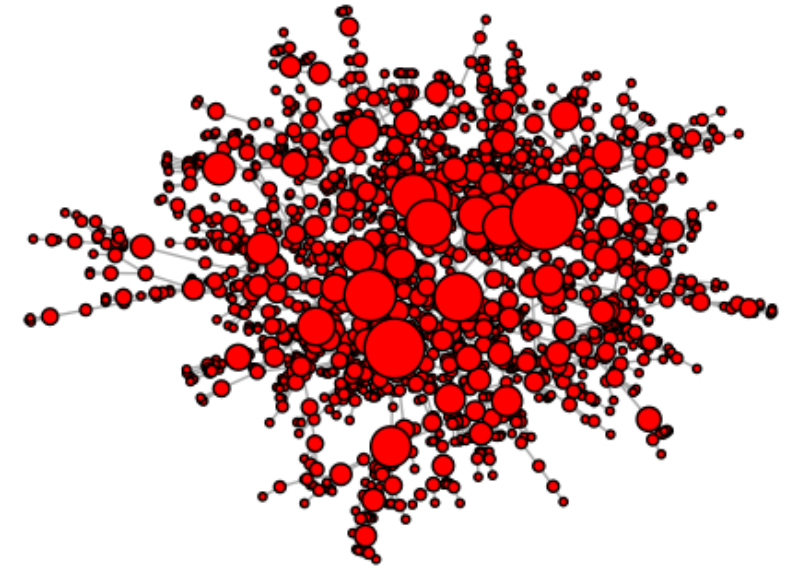


The Barabási-Albert model

- **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**

In NetworkX:

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



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

Key properties of the Barabási-Albert model

The Barabási-Albert model generates networks with three key characteristics:

1. Scale-free (power-law) degree distribution

- Follows $P(k) \sim k^{-3}$
- Creates hub nodes with very high degree
- Heavy-tailed distribution unlike random or Watts-Strogatz networks

2. Small-world property

- Average path length scales logarithmically with network size: $\ell \sim \ln N / \ln \ln N$
- Even shorter paths than random networks of same size and density
- Hubs act as "super-connectors" creating shortcuts through the network
- Efficient information transfer despite preferential growth

Key properties of the Barabási-Albert model

3. Low clustering

- Clustering coefficient $C \sim N^{-0.75}$ (decreases with network size)
- Much lower than real social networks and Watts-Strogatz model
- Approaches zero for large networks

Low clustering in the Barabási-Albert model

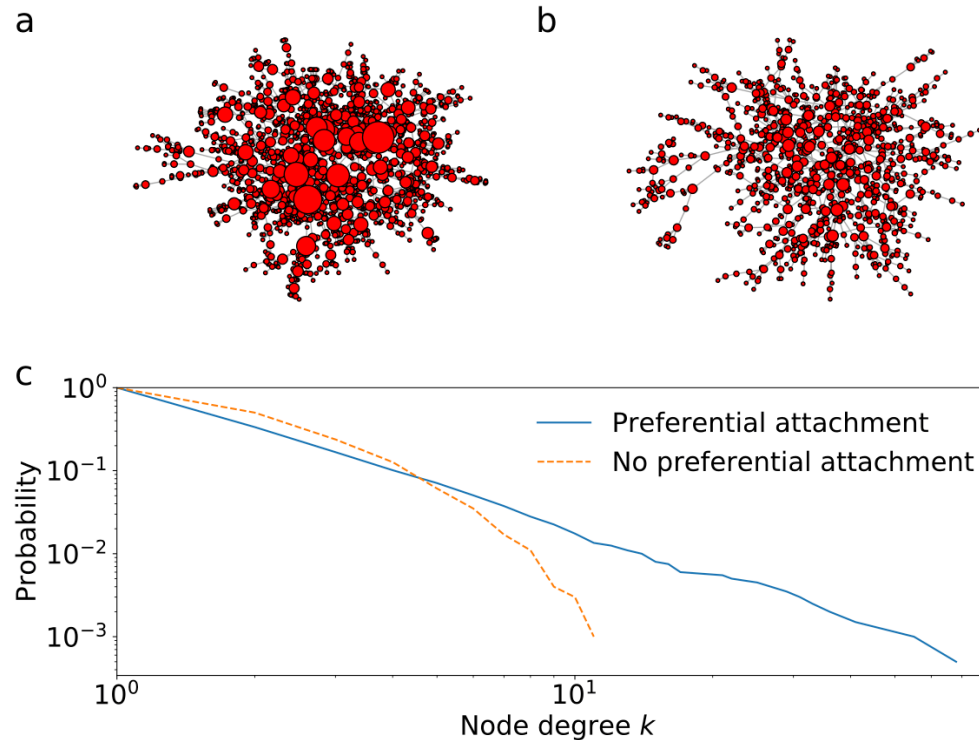
Why does the BA model generate networks with low clustering?

- **Tree-like growth:** New nodes typically connect to existing nodes that aren't connected to each other
 - This creates a tree-like structure without many triangles
- **No mechanism for triangle formation:** The model has no explicit mechanism to close triangles
 - Preferential attachment only considers degree, not network proximity
- **Hub-centric connections:** When new nodes connect to hubs, these connections rarely form triangles
 - The neighbors of a hub are typically not connected to each other

The Barabási-Albert model

Conclusion: growth + random attachment does not generate hubs.

Preferential attachment is necessary!



Final Remarks

Network models: summary comparison

High clustering	✗ Low	✓ High	✗ Low
Small-world	✓ Yes	✓ Yes	✓ Yes
Scale-free	✗ No	✗ No	✓ Yes

Comparison of models:

- **Random networks:** Generate short paths but lack clustering and hubs
- **Watts-Strogatz:** Captures both high clustering and short paths, but lacks hubs
- **Barabási-Albert:** Creates scale-free networks with hubs and short paths, but low clustering



Other preferential models

Non-linear preferential attachment

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 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** :

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

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

Non-linear preferential attachment

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

For $\alpha = 1$ we recover the linear preferential attachment (BA model)

Question: what happens when $\alpha \neq 1$?

Answer: it depends on whether $\alpha > 1$ or $\alpha < 1$

Non-linear preferential attachment

- For $\alpha < 1$, 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 $\alpha > 1$, 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 $\alpha > 2$, 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 (Dorogovtsev-Mendes-Samukhin)

- **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

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 ($m < m_0$)
- 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 , indicating the intrinsic appeal of j :

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

Key features:

- **The attractiveness parameter A :**
 - Works as a "degree offset" or initial attractiveness
 - Ensures every node has a non-zero chance of receiving links, solving the "zero-degree problem"
 - Makes degree differences less important when A is large (more uniform attachment)
- **Mathematical behavior:**
 - For very large $A \gg \langle k \rangle$: Approaches uniform random attachment
 - For $A = 0$: Recovers the original BA model
 - For $0 < A < \infty$: Generates power-law degree distributions with exponent $\gamma = 3 + A/m$
- **Applications:**
 - Models citation networks where new papers have intrinsic value but no citations
 - Web page networks where new pages have visibility before gaining links
 - Social networks where individuals have baseline popularity

Consequences and insights:

- **Flexibility in fitting real data:**

- Can fit networks with power-law exponents $\gamma > 3$ (unlike standard BA where $\gamma = 3$)
- The attractiveness parameter provides a tuning knob for the degree distribution

- **Network evolution dynamics:**

- Reduces "first-mover advantage" - newer nodes can still gain significant connectivity
- Creates more balanced competition between nodes of different ages
- Emergence of hubs is less dramatic than in the standard BA model

- **Variations:**

- Class-based attractiveness: Different node types have different baseline attractiveness
- Time-dependent attractiveness: A node's intrinsic appeal can decay or grow over time

Extensions of the BA model: Fitness model (Bianconi-Barabási)

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

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 :

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

Key features:

- **Node-specific fitness values η_j :**
 - Represents intrinsic quality or attractiveness of each node
 - Usually drawn from some probability distribution $\rho(\eta)$
 - Allows young nodes with high fitness to overcome age disadvantage
 - Creates a "meritocratic" component to network growth
- **Mathematical formulation:**
 - Multiplicative effect (unlike additive effect in Attractiveness model)
 - Results in "preferential attachment with fitness" where both connectivity and quality matter
- **Analytical results:**
 - The model can produce power-law degree distributions with a fitness-dependent exponent
 - Can lead to "winner-takes-all" effects when fitness distribution has certain properties
 - Introduces the concept of dynamic monopoly: a high-fitness node can eventually capture a finite fraction of all links

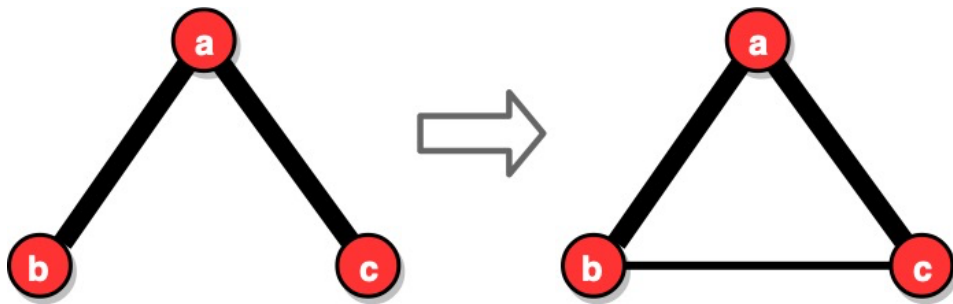
Consequences and real-world applications:

- **Breaking the age barrier:**
 - Allows young nodes with high fitness to become hubs ("Google effect")
 - Creates more realistic scenarios where quality matters alongside age
- **Fitness distribution effects:**
 - Uniform fitness: More equal competition, but still power-law distribution
 - Heavy-tailed fitness: Extreme domination by a few super-fit nodes
 - Bounded fitness: More balanced competition with power-law degree distribution
- **Real-world interpretations of fitness:**
 - Scientific papers: Intrinsic quality/importance
 - Websites: Usefulness/relevance to users
 - Social networks: Inherent sociability/popularity of individuals
 - Business networks: Innovation capability of companies

Extensions of the BA model: Random walk model (Holme-Kim)

- **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

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

Key features:

- **Triadic closure mechanism:**
 - Creates triangles in the network, unlike standard BA model
 - Parameter p controls the strength of triadic closure tendency
 - Mimics real-world connection formation through mutual friends/references
- **Mathematical behavior:**
 - Effective preferential attachment emerges naturally
 - Higher p values lead to higher clustering coefficients
 - Still maintains scale-free degree distribution despite different mechanism
- **Network properties:**
 - High clustering coefficient (increases with p)
 - Community structure emerges for high p values
 - Power-law degree distribution (implicit preferential attachment)
 - Small-world characteristics

Mechanism and implications:

- **Implicit preferential attachment:**

- High-degree nodes have more neighbors, making them more likely endpoints in a random walk
- This creates an effective preferential attachment without explicit degree calculations
- Demonstrates how simple local rules can lead to complex global structures

- **Community formation:**

- As p approaches 1, dense local clusters form naturally
- These communities are connected by fewer inter-community links
- Models how social networks develop community structure

- **Applications:**

- Social networks: People meet through existing friends
- Academic networks: Researchers find papers through references of papers they've read
- Information networks: Content discovery through related content
- Geographic networks: Spatial proximity influences connection patterns

Extensions of the BA model: Copy model

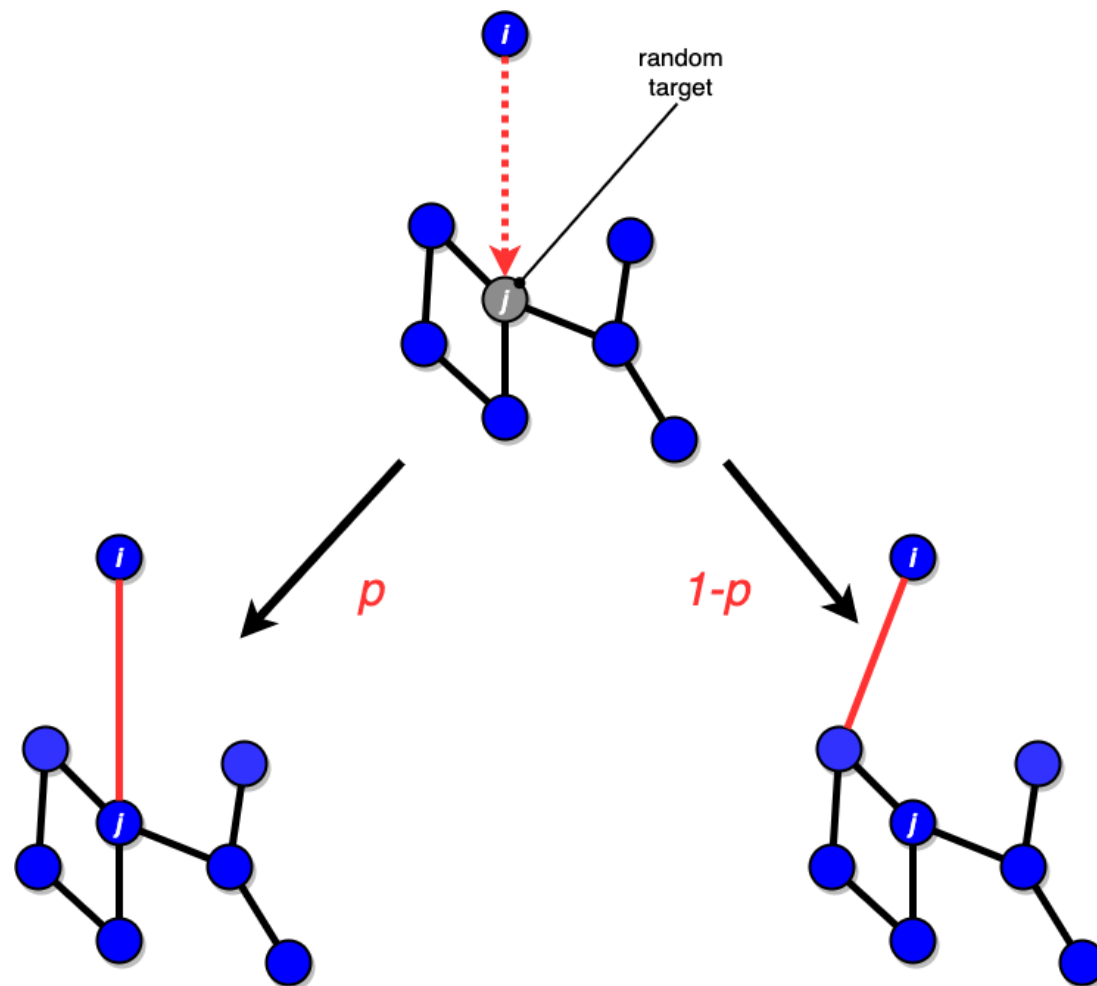
Motivation: the authors of a new web page tend to copy the hyperlinks of other pages on the same topic

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

Copy model

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



Key features:

- **Link copying mechanism:**

- Models how new entities often imitate existing ones
- Parameter p controls balance between innovation (new links) and imitation (copied links)
- Creates implicit preferential attachment through copying

- **Mathematical properties:**

- Generates power-law degree distribution with exponent $\gamma = (2 - p)/(1 - p)$
- Tunable exponent based on copying probability p
- Creates higher clustering than BA model but lower than Random Walk model

- **Network characteristics:**

- Scale-free degree distribution
- Moderate clustering coefficient
- Community structure emerges based on copying patterns
- Short average path lengths (small-world property)

Applications and variations:

- **Web page creation:**
 - New websites often copy links from existing sites with similar content
 - Explains why web graph has both scale-free structure and community organization
- **Biological networks:**
 - Gene duplication: When genes duplicate, the new gene initially interacts with the same proteins
- **Content networks:**
 - Social media content sharing with partial copying of formats/references
 - Academic citation patterns where authors reference similar sources as existing papers
- **Variations:**
 - Partial copy model: Only a fraction of links are eligible for copying
 - Weighted copy model: Copying probability depends on link strength
 - Aging copy model: Older nodes are less likely to be selected as targets

Comparison of Preferential Attachment Extensions

Model	Key Mechanism	Degree Distribution	Clustering
BA (original)	Linear PA	Power-law ($\gamma = 3$)	Low
Attractiveness	Degree offset	Power-law ($\gamma > 3$)	Low
Fitness	Node quality	Power-law (variable γ)	Low
Random Walk	Triadic closure	Power-law	High
Copy	Link imitation	Power-law (tunable γ)	Medium

Extensions of BA Model: Comparative Analysis

Which model best replicates real-world networks?

- **Random Walk Model:**
 - **Best for social networks** where triangle formation is crucial
 - High clustering coefficient matches observed patterns in human networks
 - Emergent community structure aligns with social group formation
 - Naturally explains how preferential attachment emerges from simple processes
- **Fitness Model:**
 - **Best for meritocratic networks** like scientific citations or technological adoption
 - Captures the phenomenon of late-arriving high-quality nodes becoming hubs
 - Explains competitions between entities of different intrinsic quality
 - More realistic than pure age-dependent attachment models

Extensions of BA Model: Comparative Analysis

- **Copy Model:**
 - **Best for information networks** like the Web or citation networks
 - Effectively models how new content builds on existing content
 - Good balance between clustering and scale-free structure
 - Well-suited for biological networks formed through duplication processes
- **Attractiveness Model:**
 - **Best for networks with baseline visibility** requirements
 - Provides more realistic power-law exponents for many empirical networks
 - Helps explain how new nodes gain initial connections
 - Simpler mathematically than other extensions

Extensions of BA Model: Conclusion

Real-world network modeling requires:

1. **Hybrid approaches** combining multiple mechanisms
 - For example: Random Walk + Fitness provides both clustering and quality-based growth
2. **Context-specific models** tailored to specific domains
 - No single model fits all real-world network types
3. **Additional mechanisms** beyond these basic extensions
 - Node deletion, edge rewiring, aging effects, and geographical constraints
4. **Empirical validation** against multiple network metrics
 - Degree distribution alone is insufficient; clustering, path lengths, assortativity, and community structure must be considered



The most successful models incorporate both **structural mechanisms** (how nodes connect) and **node-specific attributes** (what makes some nodes special).



Reading material

References

[ns1] **Chapter 5 (Network Models)**

Q&A

