Lecture 13 · Dynamic Random Models

Networks, Crowds and Markets

Universitat Pompeu Fabra Barcelona

Quick recap

Erdős-Renyi model is a simple baseline model but it has problems:

- Not a good generative model for realistic networks.
- Degree distribution highly contrated around the mean.
- No community structure.

One problem with this model is that each node/edge is treated equally.

We introduced some static models giving heterogeneity in edges.

- Exponential Random Graph Models: e.g. p_2-model.
- Latent space model.

Today we study the preferential attachment and configuration models.

- Generating networks with arbitrary degree distribution

Configuration model

The Configuration Model

Goal: Generate graph with a given degree sequence $\{k_1, \ldots, k_n\}$.

Algorithm:

1. Give each node i exactly k_i *stubs* (half–edges).
2. Randomly pair all $2L = \sum_i k_i$ stubs to form L edges.
3. Optionally discard self–loops or multi–edges for a simple graph.

Key property:

- Every network with the same degrees has equal probability.

Expected adjacency:

$$E[A_{ij}] = \frac{k_i k_j}{2L - 1} \approx \frac{k_i k_j}{2L}.$$

Configuration Model: intuition and limitations

Intuition:

- Each node keeps the degree, but partners are chosen uniformly at random.
- This gives a uniform distribution on pairings with given degree sequence.
- $E[A_{ij}] \propto k_i k_j$: nodes with many stubs have higher expected connectivity, even without any preference or dynamics.

Limitations:

- Can create self–loops or parallel edges (rare for large n).
- Produces no community structure or clustering.

A static, structureless baseline for networks with given degree sequence.

Preferential attachment

From Static to Growing Models

All previous models assumed a fixed number of nodes and edges.

But real networks *grow* over time: new users, new webpages, new firms.

Preferential attachment: New node attaches to existing node v with probability proportional to $\deg(v)$.

- "Rich get richer" $\rightarrow$ hubs emerge.

From Static to Growing Models

All previous models assumed a fixed number of nodes and edges.

But real networks *grow* over time: new users, new webpages, new firms.

Preferential attachment: New node attaches to existing node v with probability proportional to $\deg(v)$.

- "Rich get richer" $\rightarrow$ hubs emerge.

Result: degree distribution follows a *power law*.

- Few very large hubs.
- Many low-degree nodes.
- Matches data: web, citation networks, finance.

Preferential Attachment: Formal Definition

We construct a growing sequence of graphs $G_m, G_{m+1}, G_{m+2}, \ldots$.

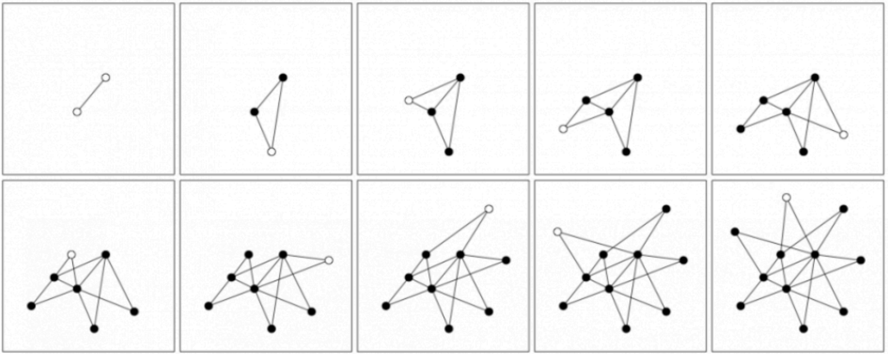
1. **Initialization:** Start from a complete graph G_m on m nodes (so each node initially has degree $m-1$).
2. **Growth rule:** For each step $t = m+1, m+2, \ldots$:
   - Add a new node v_t and m edges sticking out of it.
   - Connect each edge to a node u with probability

$$\mathbb{P}(v_t \to u) \;=\; \frac{\deg(u, t-1)}{\sum_w \deg(w, t-1)}.$$

Thus, high-degree nodes are more likely to receive new links.

This process defines the **Barabási–Albert (BA) model**.

Evolution of the Barabási-Albert model



Expected degree growth in the BA model

If t_u is the time when u appears, we can show

$$\mathbb{E}[d_t] \;\approx\; m\sqrt{\frac{t}{t_u}}.$$

Derivation: At time $t \geq m$, the network has $L_t = \binom{m}{2} + m(t-m)$ edges. Up to constants depending only on m, we may write $L_t \;\approx\; mt$ (think large t).

Expected degree growth in the BA model

If t_u is the time when u appears, we can show

$$\mathbb{E}[d_t] \approx m\sqrt{\frac{t}{t_u}}.$$

Derivation: At time $t \geq m$, the network has $L_t = \binom{m}{2} + m(t-m)$ edges. Up to constants depending only on m, we may write $L_t \approx mt$ (think large t).

Fix a node u, $d_t := \deg(u, t)$. When a new node arrives, it creates m new edges, each connecting to an existing node with probability proportional to its degree:

$$\mathbb{P}(\text{edge connects to } u) = \frac{d_t}{\sum_v \deg(v,t)} = \frac{d_t}{2L_t} \approx \frac{d_t}{2mt}.$$

Expected degree growth in the BA model

If t_u is the time when u appears, we can show

$$\mathbb{E}[d_t] \;\approx\; m\sqrt{\frac{t}{t_u}}.$$

Derivation: At time $t \geq m$, the network has $L_t = \binom{m}{2} + m(t-m)$ edges. Up to constants depending only on m, we may write $L_t \;\approx\; mt$ (think large t).

Fix a node u, $d_t := \deg(u, t)$. When a new node arrives, it creates m new edges, each connecting to an existing node with probability proportional to its degree:

$$\mathbb{P}(\text{edge connects to } u) \;=\; \frac{d_t}{\sum_v \deg(v,t)} \;=\; \frac{d_t}{2L_t} \;\approx\; \frac{d_t}{2mt}.$$

Expected increment:

$$\mathbb{E}[d_{t+1} - d_t \mid d_t] \;\approx\; m \cdot \frac{d_t}{2mt} \;=\; \frac{d_t}{2t}.$$

Expected degree growth in the BA model

If t_u is the time when u appears, we can show

$$\mathbb{E}[d_t] \;\approx\; m\sqrt{\frac{t}{t_u}}.$$

Derivation: At time $t \geq m$, the network has $L_t = \binom{m}{2} + m(t-m)$ edges. Up to constants depending only on m, we may write $L_t \;\approx\; mt$ (think large t).

Fix a node u, $d_t := \deg(u,t)$. When a new node arrives, it creates m new edges, each connecting to an existing node with probability proportional to its degree:

$$\mathbb{P}(\text{edge connects to } u) \;=\; \frac{d_t}{\sum_v \deg(v,t)} \;=\; \frac{d_t}{2L_t} \;\approx\; \frac{d_t}{2mt}.$$

Expected increment:

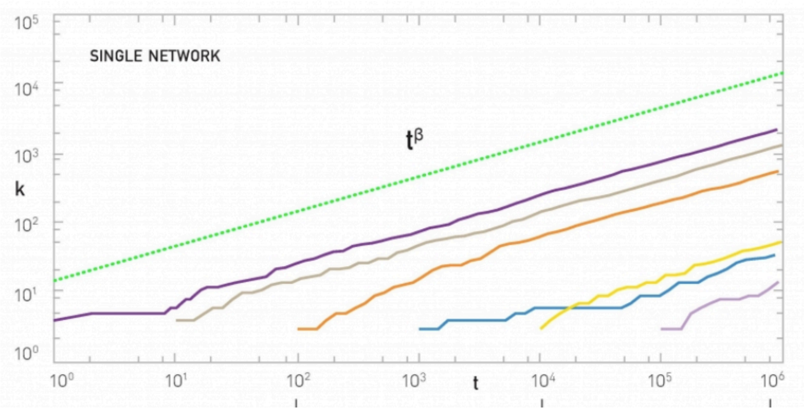
$$\mathbb{E}[d_{t+1} - d_t \mid d_t] \;\approx\; m \cdot \frac{d_t}{2mt} \;=\; \frac{d_t}{2t}.$$

This gives a recursion that gives the expected degree growth:

$$\mathbb{E}[d_{t+1}] \;\approx\; \mathbb{E}[d_t]\Big(1 + \frac{1}{2t}\Big), \qquad t \geq m.$$

Evolution of the degree

Simulated degrees of a few nodes in the log-log scale:



Quick comparison (who becomes a hub?)

Recall: $\mathbb{E}[d_t] \approx m\sqrt{\frac{t}{t_u}}$.

Two nodes joined at $t_u = 10$ and $t_v = 100$. After $t = 1000$ with $m = 3$:

$$\frac{\deg(u, 1000)}{\deg(v, 1000)} = \sqrt{\frac{1000/10}{1000/100}} = \sqrt{10} \approx 3.16,$$

$$\deg(u, 1000) = 3\sqrt{100} = 30, \qquad \deg(v, 1000) = 3\sqrt{10} \approx 9.48.$$

Earlier arrival systematically advantages degree.

The ratio of expected degrees depends on t_u, t_v but not on m.

Heuristic derivation of the degree distribution

From the recursion we found:

$$\mathbb{E}[\deg(u,t)] \;\approx\; m\left(\frac{t}{t_u}\right)^{1/2}.$$

To find the degree distribution at time t, note that

$$\deg(u,t) \approx m\left(\frac{t}{t_u}\right)^{1/2} \quad\Longleftrightarrow\quad t_u \approx t\,\frac{m^2}{\deg(u,t)^2}.$$

Heuristic derivation of the degree distribution

From the recursion we found:

$$\mathbb{E}[\deg(u,t)] \;\approx\; m\left(\frac{t}{t_u}\right)^{1/2}.$$

To find the degree distribution at time t, note that

$$\deg(u,t) \approx m\left(\frac{t}{t_u}\right)^{1/2} \quad\Longleftrightarrow\quad t_u \approx t\,\frac{m^2}{\deg(u,t)^2}.$$

Since arrival times t_u are roughly *uniform* on $\{1, 2, \ldots, t\}$, we can compute

$$\mathbb{P}(\deg(u,t) \geq k) \;\approx\; \mathbb{P}\left(t_u \leq t\,\frac{m^2}{k^2}\right) \;\approx\; \frac{m^2}{k^2}.$$

Asymptotic tail

The prob. that a node has degree $\geq k$ decreases quadratically in k:

$$\mathbb{P}(\deg \geq k) \;\propto\; k^{-2} \quad\Longrightarrow\quad p_k = \mathbb{P}(\deg = k) \;\propto\; k^{-3}.$$

(at least for large k)

Hence: the Barabási–Albert model produces a *power-law* degree distribution with

$$\boxed{\gamma = 3.}$$

Asymptotic tail

The prob. that a node has degree $\geq k$ decreases quadratically in k:

$$\mathbb{P}(\deg \geq k) \;\propto\; k^{-2} \quad\Longrightarrow\quad p_k = \mathbb{P}(\deg = k) \;\propto\; k^{-3}.$$

(at least for large k)

Hence: the Barabási–Albert model produces a *power-law* degree distribution with

$$\boxed{\gamma = 3.}$$

Takeaways

- The tail exponent $\gamma = 3$ is universal for the BA model (all m).
- This formula matches simulations closely.

Exercise (preferential attachment probabilities)

Consider the preferential attachment model with $m = 1$. Given the degree multiset $\{1, 1, 1, 1, 2, 3, 3, 4, 4, 5, 5, 8\}$, let a new node add *one* link using BA attachment $\Pr\{u\} = \deg(u)/(2L)$.

a) Probability it attaches to the highest–degree node:

$$\Pr\{\text{choose } k = 8\} \; = \; \frac{8}{\sum \deg} \; = \; \frac{8}{38}.$$

b) Probability it attaches to a node of degree 1: there are four such nodes, each with probability $1/38$: $4 \times \frac{1}{38} = \frac{4}{38}$.