Universitat
Pompeu Fabra
Barcelona

Lecture 12 · Random Graph Models

Networks, Crowds and Markets

Summary

The goal of the lecture today is to give an overview of some approaches to model random networks.

Static random graph models

Graphs as random objects

Consider an undirected graph $G = (V, E)$.

Order all pairs of elements in V: $\{1, 2\}, \{1, 3\}, \ldots, \{N - 1, N\}$.

Each graph is uniquely identified by a vector $\boldsymbol{y} = (y_{ij}) \in \{0, 1\}^{\binom{N}{2}}$:

- $y_{ij} = 1$ if and only if $ij \in E$.

Graphs as random objects

Consider an undirected graph $G = (V, E)$.

Order all pairs of elements in V: $\{1, 2\}, \{1, 3\}, \dots, \{N-1, N\}$.

Each graph is uniquely identified by a vector $\boldsymbol{y} = (y_{ij}) \in \{0, 1\}^{\binom{N}{2}}$:

- $y_{ij} = 1$ if and only if $ij \in E$.

In this sense, every distribution for a random binary vector in $\{0, 1\}^{\binom{N}{2}}$ gives a distribution of a random graph with N nodes.

e.g. $(p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) = (\frac{1}{2}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14})$ gives a distribution over 3-node graphs.

Graphs as random objects

Consider an undirected graph $G = (V, E)$.

Order all pairs of elements in V: $\{1, 2\}, \{1, 3\}, \dots, \{N-1, N\}$.

Each graph is uniquely identified by a vector $\boldsymbol{y} = (y_{ij}) \in \{0, 1\}^{\binom{N}{2}}$:

- $y_{ij} = 1$ if and only if $ij \in E$.

In this sense, every distribution for a random binary vector in $\{0, 1\}^{\binom{N}{2}}$ gives a distribution of a random graph with N nodes.

e.g. $(p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}) = (\frac{1}{2}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14}, \frac{1}{14})$ gives a distribution over 3-node graphs.

Every family of distributions over $\{0, 1\}^{\binom{N}{2}}$ gives a statistical model for random graphs with N nodes.

Replacing with $\{0, 1\}^{N(N-1)}$ gives models for directed graphs.

Erdős–Rényi model as an example

Recall: Every family of distributions over $\{0,1\}^{\binom{N}{2}}$ gives a statistical model for random graphs with N nodes.

Erdős-Rényi model: for $\boldsymbol{y} = (y_{ij}) \in \{0,1\}^{\binom{N}{2}}$ consider distribution

$$p(\boldsymbol{y}) \;=\; \prod_{i<j} (1-p)^{1-y_{ij}} p^{y_{ij}}.$$

Denote $s = \sum_{i<j} y_{ij}$ (the number of edges) then

$$p(\boldsymbol{y}) \;=\; (1-p)^{\binom{N}{2}-s} p^s \;=\; (1-p)^{\binom{N}{2}} \left(\frac{p}{1-p}\right)^s.$$

Quick recall: exponential families

Let $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^n$, $T : \mathbb{R}^n \to \mathbb{R}^d$, $\theta \in \mathbb{R}^d$.

Definition (Eponential family)

A family of probability distributions on $\mathcal{X}$ is an *exponential family* if the probability mass functions (densities) take the form

$$p_\theta(\boldsymbol{x}) = h(\boldsymbol{x}) \exp\left(\theta^T T(\boldsymbol{x}) - \psi(\theta)\right).$$

- $T(\boldsymbol{x})$ = sufficient statistics (counts of edges, triangles, ...).
- θ = natural parameter.
- $\psi(\theta)$ = log-partition function (ensures normalization).

Bernoulli, binomial, Poisson, Ising models, multivariate Gaussian, and many other popular statistical models are exponential families.

Exponential Random Graph Models

Definition (Exponential Random Graph Models (ERGMs):)

$$\mathbb{P}(Y = \boldsymbol{y}) \propto \exp\{\theta_1 \cdot \#\text{edges}(\boldsymbol{y}) + \theta_2 \cdot \#\text{triangles}(\boldsymbol{y}) + \cdots\}.$$

- The parameters: θ_1 tunes density, θ_2 tunes clustering, etc.

Erdős-Rényi model is a special case of ERGM:

$$\mathbb{P}(Y = \boldsymbol{y}) \; = \; (1-p)^{\binom{N}{2}} \left(\frac{p}{1-p}\right)^s \propto \exp(\theta \cdot s),$$

where $s = \sum_{i<j} y_{ij}$ and $\theta = \log\left(\frac{p}{1-p}\right)$

Example: the p_2 model for undirected networks

Extension of Erdős–Rényi that introduces *node-specific propensities* to form ties.

Model: All edges are independent with

$$\Pr(Y_{ij} = 1 \mid \alpha_i, \alpha_j) = \frac{\exp(\mu + \alpha_i + \alpha_j)}{1 + \exp(\mu + \alpha_i + \alpha_j)}.$$

Interpretation:

- μ — overall network density.
- α_i — sociability of node i (a random effect).

Example: the p_2 model for undirected networks

Extension of Erdős–Rényi that introduces *node-specific propensities* to form ties.

Model: All edges are independent with

$$\Pr(Y_{ij} = 1 \mid \alpha_i, \alpha_j) = \frac{\exp(\mu + \alpha_i + \alpha_j)}{1 + \exp(\mu + \alpha_i + \alpha_j)}.$$

Interpretation:

- μ — overall network density.
- α_i — sociability of node i (a random effect).

Remarks:

- Reduces to Erdős–Rényi when $\alpha_i \equiv 0$.
- Adds degree heterogeneity while preserving tractability.
- Foundation for later hierarchical and latent-space models.

Why is it an ERGM? What are the sufficient statistics?

Latent space random graphs

Definition

Each node i has a position z_i in a latent feature space (e.g. $\mathbb{R}^d$). The probability of an edge depends on distance:

$$\mathbb{P}(i \sim j) = f\left(\|z_i - z_j\|\right), \quad f \text{ decreasing.}$$

As an example, imagine an interaction network in a big company. Apart from the usual topology that follows the company's structure, unexpected links may occur (e.g. among smokers etc).

Example: latent space in economic networks

- Think of banks, firms, or households as nodes.
- Each actor has a position in a **latent space**:
  - Geography (local vs. international).
  - Sector (energy, tech, manufacturing).
  - Risk profile or credit rating.
- Links (e.g. loans, partnerships, trade) are more likely between nearby nodes in this space.
- A few "long-distance" links (large international banks, global supply chains) can connect distant clusters and reduce path lengths.

Takeaway

Latent space models explain why real networks show both clustering (local ties) and small-world shortcuts (rare global ties).

Dynamic random graph models

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.

Let's study a simple dynamic rule that explains why hubs emerge:
preferential attachment.

Recursive growth: preferential attachment

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

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

Recursive growth: preferential attachment

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 $G_m, G_{m+1}, G_{m+2}, \ldots$.

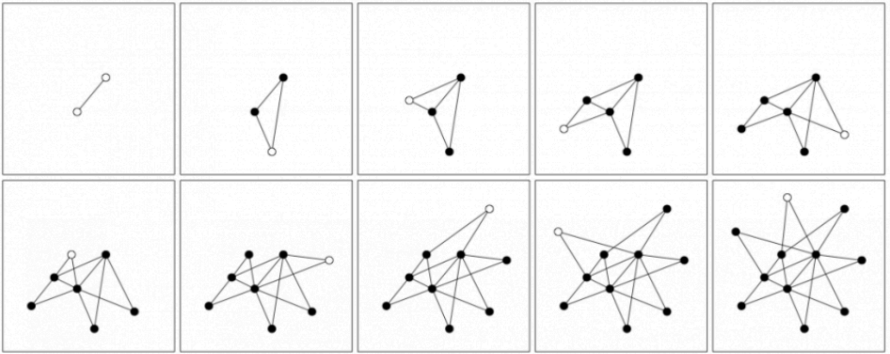
1. **Initialization:** Start from a complete graph G_m on m nodes (so each node initially has degree $m-1$).
2. **Growth rule:** For each step $t = m+1, m+2, \ldots$:
   - Add a new node v_t and m edges sticking out of it.
   - Connect each edge to a node u with probability

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

   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

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 $2L_t \;\approx\; 2mt$.

Fix a node u with current degree $d_t = \deg(u, t)$. When a new node arrives at step $t{+}1$, it creates m new edges, each connecting to an existing node v with probability proportional to its degree:

$$\mathbb{P}(\text{edge connects to } u) \;\approx\; \frac{d_t}{2mt}.$$

Expected degree growth in the BA model

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 $2L_t \;\approx\; 2mt$.

Fix a node u with current degree $d_t = \deg(u, t)$. When a new node arrives at step $t{+}1$, it creates m new edges, each connecting to an existing node v with probability proportional to its degree:

$$\mathbb{P}(\text{edge connects to } u) \;\approx\; \frac{d_t}{2mt}.$$

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

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 $2L_t \approx 2mt$.

Fix a node u with current degree $d_t = \deg(u,t)$. When a new node arrives at step $t+1$, it creates m new edges, each connecting to an existing node v with probability proportional to its degree:

$$\mathbb{P}(\text{edge connects to } u) \approx \frac{d_t}{2mt}.$$

Expected increment:

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

Recursion for the expected degree:

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

Expected degree growth in the BA model

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 $2L_t \;\approx\; 2mt$.

Fix a node u with current degree $d_t = \deg(u, t)$. When a new node arrives at step $t+1$, it creates m new edges, each connecting to an existing node v with probability proportional to its degree:

$$\mathbb{P}(\text{edge connects to } u) \;\approx\; \frac{d_t}{2mt}.$$

Expected increment:

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

Recursion for the expected degree:

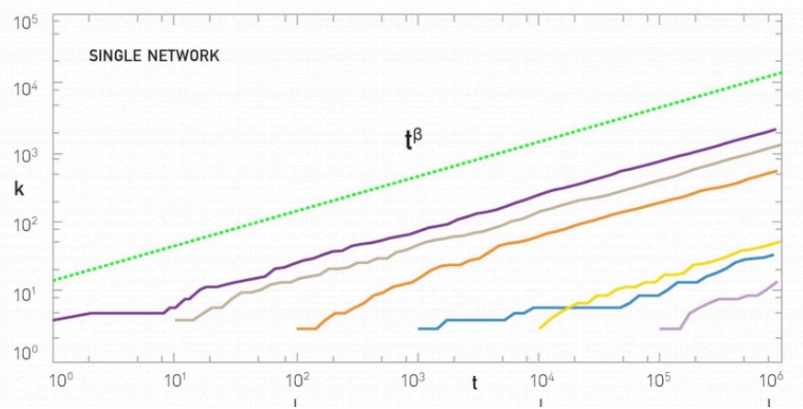
$$\mathbb{E}[d_{t+1}] \;\approx\; \mathbb{E}[d_t]\Big(1 + \frac{1}{2t}\Big), \qquad t \geq m.$$

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

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

Evolution of the degree

In the log − log scale:



Quick comparison (who becomes a hub?)

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.

Heuristic derivation of the degree distribution

From the recursion we found:

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

Hence, older nodes (small t_u) have larger expected degree.

To find the degree distribution at time t, note that

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

Heuristic derivation of the degree distribution

From the recursion we found:

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

Hence, older nodes (small t_u) have larger expected degree.

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

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

Discrete exact formula and 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

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

Discrete exact formula and 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).
- The discrete 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 $1/38$: $4 \times \frac{1}{38} = \frac{4}{38}$.

(Here $2L = \sum \deg = 38$.)