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.

We start with static graphs, directly generalizing Erdős-Rényi.

- In $ER(N, p)$ each of the $\binom{N}{2}$ edges is sampled independently from $\mathrm{Bern}(p)$.
- What if each edge gets a separate parameter p_{ij}.
- What if there is no independence?

Later we study models for a dynamic formation of 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: $\{1,2\},\{1,3\},\ldots,\{N-1,N\}$.

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 $ij \in E$.

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: $\{1, 2\}, \{1, 3\}, \ldots, \{N-1, N\}$.

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 $ij \in E$.

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.

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

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

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

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

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, ...).
- $\theta =$ 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).