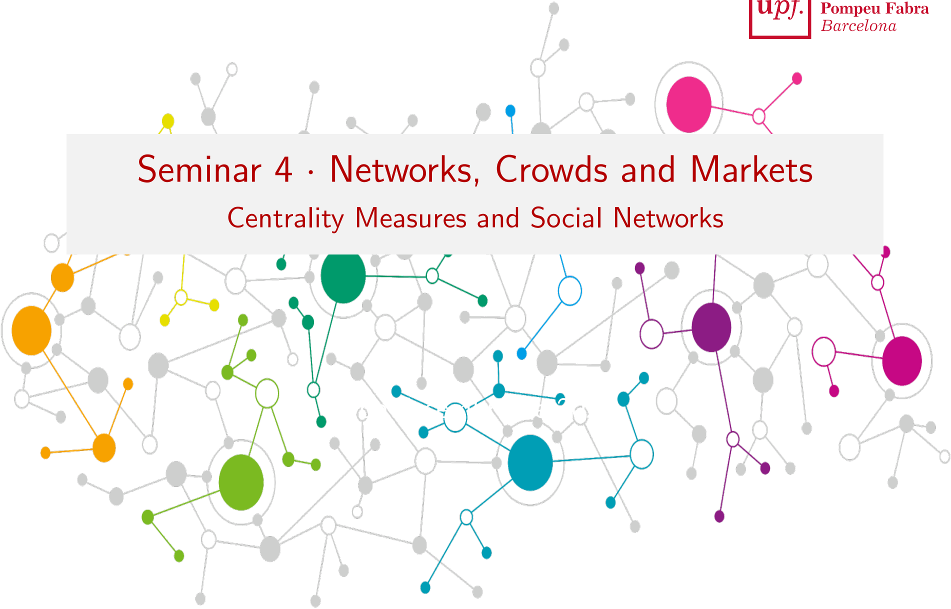
Universitat Pompeu Fabra Barcelona

# Seminar 4 · Networks, Crowds and Markets

## Centrality Measures and Social Networks



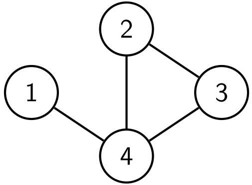
# Warm-up · Building a small $G(4, p)$ graph

We consider an Erdős–Rényi model $G(4, p)$ with $p = \frac{2}{3}$. There are $\binom{4}{2} = 6$ possible edges.

1. List all possible pairs of nodes: 12, 13, 14, 23, 24, 34.
2. Roll a fair die once per potential edge. Connect if the roll is 1–4.
3. Draw the resulting graph.

Example:

| pair | 12 | 13 | 14 | 23 | 24 | 34 |
|---|---|---|---|---|---|---|
| die | 5 | 6 | 2 | 3 | 3 | 1 |
| edge | 0 | 0 | 1 | 1 | 1 | 1 |




$$\mathbb{E}[L] = p\binom{4}{2} = 4, \qquad \mathbb{E}[\deg(v)] = (N-1)p = 2.$$

# Midterm review

I went through the solutions of the midterm.

# Additional exercises

# Exercise · Independence in Erdős–Rényi

Show that in $G(N, p)$:

- Any two different edges are independent.
- Degrees of two nodes are not independent.

Compute:

$$\mathrm{Cov}(\deg(i), \deg(j)) \quad (i \neq j).$$

# Solution · Independence in Erdős–Rényi

In $G(N, p)$, each possible edge is an independent Bernoulli($p$).

**1. Distinct edges.** If $\{i, j\} \neq \{k, \ell\}$, then:

$$P(Y_{ij} = a, Y_{k\ell} = b) = P(Y_{ij} = a)P(Y_{k\ell} = b).$$

**2. Degrees are dependent.**

$$\deg(i) = \sum_{k \neq i} Y_{ik}, \qquad \deg(j) = \sum_{\ell \neq j} Y_{j\ell}.$$

They share the term $Y_{ij}$.

**Covariance.** Only the shared edge contributes:

$$\mathrm{cov}(\deg(i), \deg(j)) = \mathrm{var}(Y_{ij}) = p(1 - p).$$

# Exercise · ER as an Exponential Family

Show that the Erdős–Rényi model

$$P(Y = y) \propto \exp(\theta\, s(y)), \qquad s(y) = \#\text{edges}(y), \quad \theta = \log \frac{p}{1-p}.$$

Identify:

- the sufficient statistic,
- the natural parameter,
- the log-partition function.

# Solution · ER as an Exponential Family

$$P(Y = y) = p^{s(y)}(1-p)^{\binom{N}{2}-s(y)}.$$

Write it as:

$$P(Y = y) \propto \exp\big(\theta\, s(y)\big), \qquad \theta = \log \frac{p}{1-p}.$$

**Sufficient statistic:**

$$s(y) = \#\text{edges}(y).$$

**Natural parameter:**

$$\theta = \log \frac{p}{1-p}.$$

**Log-partition function:** $Z$ is the normalizing constant and $A$ is its logarithm

$$Z(\theta) = (1+e^{\theta})^{\binom{N}{2}}, \qquad A(\theta) = \binom{N}{2} \log(1+e^{\theta}).$$

# Exercise · The $p_2$ Model (undirected)

$$\text{logit } \Pr(Y_{ij} = 1 \mid \alpha_i, \alpha_j) = \theta + \alpha_i + \alpha_j.$$

1. Interpret $\theta$ and $\alpha_i$.
2. What if all $\alpha_i = 0$?
3. Link to mixed-effects logistic regression.

# Solution · The $p_2$ Model

**(1) Interpretation.**

- $\theta$: baseline connectivity.
- $\alpha_i$: sociality of node $i$; higher implies more edges.

**(2) If all $\alpha_i = 0$:** Reduces to ER with

$$p = \text{logit}^{-1}(\theta).$$

**(3) Mixed-effects link.** It is a logistic regression with random node effects:

$$Y_{ij} \sim \text{logit}^{-1}(\theta + \alpha_i + \alpha_j).$$

# Exercise · Latent Space Model

Each node $i$ has position $z_i \in \mathbb{R}^2$, and

$$\operatorname{logit} P(Y_{ij} = 1) = \alpha - \|z_i - z_j\|.$$

1. Explain the effect of distance.
2. What can this model capture that ER or $p_2$ cannot?
3. Give a real-world system where such a geometry is natural.

# Solution · Latent Space Model

**(1) Distance effect.** Closer nodes so higher probability of an edge.
**(2) Captures:**

- clustering and communities,
- transitivity (triangle formation),
- geometric effects,
- heterogeneity induced by spatial layout.

**(3) Real-world examples.**

- Social networks with ideological or spatial proximity,
- Ecological interaction networks,
- Communication or mobility networks.