Universitat Pompeu Fabra Barcelona

Lecture 6 · Centrality measures II

Networks, Crowds and Markets

Today's Lecture

1. Linear Algebra, Random walks, and PageRank
2. Why random graphs? Motivation and Erdős–Rényi models.
3. Probability recap for $G(N, p)$:
   3.1 Binomial distribution (edges, degrees).
   3.2 Poisson approximation in the sparse regime.

Basic spectral theory

Why Linear Algebra for Networks?

- Adjacency matrix A_G: encodes all links of G.
- Degree vector: $A_G\mathbf{1} = (\deg(v_1), \ldots, \deg(v_N))$.
- Laplacian $L = D - A_G$: central in diffusion, clustering, spanning trees.
- Many network measures (centrality, random walks, PageRank) reduce to eigenvalue/eigenvector problems.

Note

Eigenvalues of A_G reveal secrets of G.

- Google built its empire on one eigenvector (PageRank).
- Spotify/Youtube recommenders use eigenvector-like ideas.
- In social networks, eigenvector centrality captures being "friends with important people."

Recall: Eigenvalues and Eigenvectors

Definition

Let $A \in \mathbb{R}^{n \times n}$ then $\boldsymbol{v} \neq \mathbf{0}$ is called an eigenvector of A if

$$A\mathbf{v} = \lambda \mathbf{v}$$

for some λ, called eigenvalue. Assume $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}^\top \boldsymbol{v}} = 1$.

If A has only real eigenvalues then it can be diagonalized: $\exists$ invertible P s.t.

$$A \;=\; P \Lambda P^{-1} \quad \text{with } \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

The columns of P are the eigenvectors of A.

Note

If A is diagnosable then $A^k = P \Lambda^k P^{-1}$, $\Lambda^k = \mathrm{diag}(\lambda_1^k, \ldots, \lambda_n^k)$.

Spectral theorem

Theorem

If A is symmetric (i.e. $A = A^\top$), all eigenvalues are real, and eigenvectors form an orthogonal basis.

A is diagonalizable and for some orthogonal matrix U (i.e. $U^\top U = I_n$):

$$A = U \Lambda U^\top.$$

Spectral theorem

Theorem

If A is symmetric (i.e. $A = A^\top$), all eigenvalues are real, and eigenvectors form an orthogonal basis.

A is diagonalizable and for some orthogonal matrix U (i.e. $U^\top U = I_n$):

$$A = U \Lambda U^\top.$$

Note (Variational characterization of eigenvectors)

The eigenvectors are the saddle points of $\boldsymbol{x}^\top A \boldsymbol{x}$ subject to $\|\boldsymbol{x}\| = 1$:

- By KKT condition each optimum is a stationary point of

$$\text{Lagrangian} = \boldsymbol{x}^\top A \boldsymbol{x} - \lambda(\boldsymbol{x}^\top \boldsymbol{x} - 1).$$

- This gives $A\boldsymbol{x} = \lambda \boldsymbol{x}$. And for every such unit $\boldsymbol{x}$, $\boldsymbol{x}^\top A \boldsymbol{x} = \lambda$.

In particular, the maximal eigenvalue is $\lambda_{\max} = \max_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^\top A \boldsymbol{x}$.

Eigenvalue centrality

Motivation

In degree centrality all neighbours are treated equally.

Now: a node is important if connected to other important nodes.

Motivation

In degree centrality all neighbours are treated equally.

Now: a node is important if connected to other important nodes.

- We try to define an importance measure x_v for $v \in V$ s.t.

$$x_v \;\propto\; \sum_{u \sim v} x_u.$$

In matrix form: there exists $\lambda > 0$ and a positive $\boldsymbol{x}$ s.t.

$$A_G \boldsymbol{x} = \lambda \boldsymbol{x}.$$

Motivation

In degree centrality all neighbours are treated equally.

Now: a node is important if connected to other important nodes.

- We try to define an importance measure x_v for $v \in V$ s.t.

$$x_v \propto \sum_{u \sim v} x_u.$$

In matrix form: there exists $\lambda > 0$ and a positive $\boldsymbol{x}$ s.t.

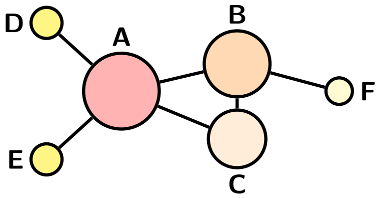
$$A_G \boldsymbol{x} = \lambda \boldsymbol{x}.$$

So centrality is given by an eigenvector of A_G with a positive eigenvalue.

Theorem (special case of Perron-Frobenius)

As A_G has nonnegative entries, maximal eigenvalue is positive.
Since $\mathbf{1}^\top A_G \mathbf{1} = 2L > 0$ then $\lambda_{\max} > 0$.
The principal eigenvector has positive entries.

Eigenvector Centrality – Core–Periphery Example



Adjacency matrix (A):

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Setup. A small core (A,B,C) connected as a triangle; three peripheral nodes (D,E,F) each attach to the core.

Why sizes differ.

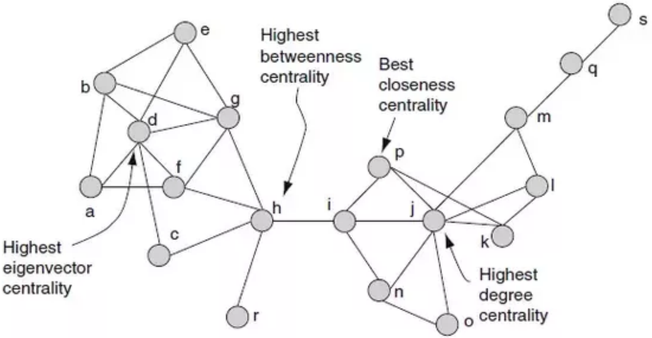
- $\boldsymbol{A}$ connects to two central nodes ($\boldsymbol{B}$,$\boldsymbol{C}$) and two peripherals ($\boldsymbol{D}$,$\boldsymbol{E}$) — very central.
- $\boldsymbol{B}$ beats $\boldsymbol{C}$ because it also connects to $\boldsymbol{F}$.
- $\boldsymbol{D}$, $\boldsymbol{E}$, $\boldsymbol{F}$ are peripheral and get low scores.

Note (Potential problems)

- What if G is disconnected?
- What if λ_{max} has multiplicity ≥ 2?

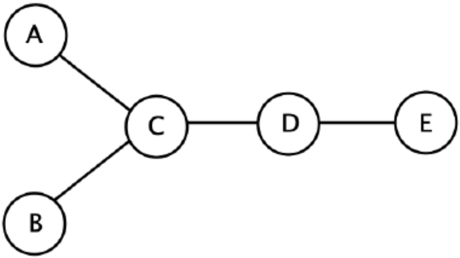
Normalized ratios:

$$x_A : x_B : x_C : x_D : x_E : x_F \approx 1.00 : 0.87 : 0.76 : 0.41 : 0.41 : 0.35.$$


e
b
g
d
f
a
c
r
h
i
j
p
m
q
s
l
k
n
o
Highest betweenness centrality
Best closeness centrality
Highest eigenvector centrality
Highest degree centrality

Exercise 1

Determine the eigenvector centrality for all the nodes in the graph:



You may use a software in order to find the eigenvalues and vectors.

Random Walks and PageRank

Random Walks on a Graph

Definition (Random Walk on a Graph $G = (V, E)$)

This is a stochastic process $(X_t)_{t=0}^{\infty}$ with each $X_t \in V$ s.t.:

- Start with a node $v_0 = X_0$ chosen uniformly at random.
- If $X_t = i$ then X_{t+1} is a neighbour of i chosen uniformly at random from all its neighbours:

$$P_{ij} \;:=\; \Pr(X_{t+1} = j | X_t = i) \;=\; \begin{cases} \frac{1}{\deg(i)}, & ij \text{ is a link} \\ 0, & \text{otherwise.} \end{cases}$$

Random Walks on a Graph

Definition (Random Walk on a Graph $G = (V, E)$)

This is a stochastic process $(X_t)_{t=0}^{\infty}$ with each $X_t \in V$ s.t.:

- Start with a node $v_0 = X_0$ chosen uniformly at random.
- If $X_t = i$ then X_{t+1} is a neighbour of i chosen uniformly at random from all its neighbours:

$$P_{ij} := \Pr(X_{t+1} = j | X_t = i) = \begin{cases} \frac{1}{\deg(i)}, & ij \text{ is a link} \\ 0, & \text{otherwise.} \end{cases}$$

The matrix $P = (P_{ij}) \in \mathbb{R}^{N \times N}$ is called the transition matrix.

Note: $P = D^+ A_G$, where $D = \mathrm{diag}(\deg(1), \ldots, \deg(N))$.

$\rightarrow (D^+)_{ii} = 1/D_{ii}$ is $D_{ii} \neq 0$ and $(D^+)_{ii} = 0$ otherwise.

The resulting Markov chain

Let $\pi^{(t)} \in \mathbb{R}^N$ be the distribution of X_t, i.e., $\pi_i^{(t)} = \Pr(X_t = i)$. We have

$$\pi_i^{(t+1)} = \sum_{j=1}^{N} \Pr(X_t = j) \Pr(X_{t+1} = i | X_t = j) = \sum_{j=1}^{N} \pi_j^{(t)} P_{j,i}.$$

In other words, $\pi^{(t+1)} = P^\top \pi^{(t)}$.

The resulting Markov chain

Let $\pi^{(t)} \in \mathbb{R}^N$ be the distribution of X_t, i.e., $\pi_i^{(t)} = \Pr(X_t = i)$. We have

$$\pi_i^{(t+1)} = \sum_{j=1}^{N} \Pr(X_t = j) \Pr(X_{t+1} = i | X_t = j) = \sum_{j=1}^{N} \pi_j^{(t)} P_{j,i}.$$

In other words, $\pi^{(t+1)} = P^\top \pi^{(t)}$.

Note

- Define $\pi = \frac{1}{\mathrm{tr}(D)} D\mathbf{1}$ and recall $P = D^+ A_G$. So that

$$P^\top \pi = \frac{1}{\mathrm{tr}(D)} A_G D^+ D\mathbf{1} = \frac{1}{\mathrm{tr}(D)} A_G \mathbf{1} = \frac{1}{\mathrm{tr}(D)} D\mathbf{1} = \pi.$$

- We have $\pi_i = \frac{\deg(i)}{\sum_{j=1}^{N} \deg(j)}$ and so π is a probability distribution.

  (π defines the degree centrality!!)
- If $\pi^{(t)} = \pi$ then $\pi^{(s)} = \pi$ for all $s \geq t$; stationary distribution.

Eigenvalues of P

Note (Assume for simplicity all degrees positive; $D^{+} = D^{-1}$)

The transition matrix P is similar to a symmetric matrix:

$$P \;=\; D^{-1}A_G \;=\; D^{-1/2}D^{-1/2}A_G D^{-1/2}D^{1/2} = D^{-1/2}SD^{1/2}$$

and so it is diagonalizable. All eigenvalues lie in $[-1,1]$.

Theorem (About the eigenvalues of P)

If G has no bipartite component, eigenvalues lie in $(-1,1]$.

If G is connected, $\lambda = 1$ has multiplicity one.

Let $S = U\Lambda U^\top$ with U orthogonal. Let $\boldsymbol{u}_i$ be the i-th column of U. Then

$$S = \sum_{i=1}^{N} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \qquad \text{and so} \qquad S^k = \sum_{i=1}^{N} \lambda_i^k \boldsymbol{u}_i \boldsymbol{u}_i^\top \xrightarrow[k\to\infty]{} \boldsymbol{u}_1 \boldsymbol{u}_1^\top,$$

where $\boldsymbol{u}_1$ is s.t. $S\boldsymbol{u}_1 = \boldsymbol{u}_1$.

Eigenvalues of P

Note (Assume for simplicity all degrees positive; $D^{+} = D^{-1}$)

The transition matrix P is similar to a symmetric matrix:

$$P \;=\; D^{-1}A_G \;=\; D^{-1/2}D^{-1/2}A_G D^{-1/2}D^{1/2} = D^{-1/2}SD^{1/2}$$

and so it is diagonalizable. All eigenvalues lie in $[-1, 1]$.

Theorem (About the eigenvalues of P)

If G has no bipartite component, eigenvalues lie in $(-1, 1]$.

If G is connected, $\lambda = 1$ has multiplicity one.

Let $S = U\Lambda U^\top$ with U orthogonal. Let $\boldsymbol{u}_i$ be the i-th column of U. Then

$$S = \sum_{i=1}^{N} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \qquad \text{and so} \qquad S^k = \sum_{i=1}^{N} \lambda_i^k \boldsymbol{u}_i \boldsymbol{u}_i^\top \xrightarrow[k\to\infty]{} \boldsymbol{u}_1 \boldsymbol{u}_1^\top,$$

where $\boldsymbol{u}_1$ is s.t. $S\boldsymbol{u}_1 = \boldsymbol{u}_1$. It follows that $\boxed{P^k \longrightarrow \mathbf{1}\pi^\top}$.

Appendix: More formal arguments for $\lambda = -1$

Statement: P has eigenvalue $\lambda = -1$ if and only if G is bipartite.

Proof. $\boxed{\Leftarrow}$ If G is bipartite with partition $V = A \cup B$ define e_A to be a 0/1-vector with 1s on coordinates corresponding to A and 0s otherwise. It is a direct check that $P(e_A - e_B) = -(e_A - e_B)$.
$\boxed{\Rightarrow}$ There exists $\boldsymbol{x}$ such that $P\boldsymbol{x} = -\boldsymbol{x}$. Assume that G is connected. Otherwise apply the same argument to each connected component. The condition implies that for all $i \in V$

$$\sum_{j=1}^{N} P_{ij} x_j \;=\; \frac{1}{\deg(i)} \sum_{j \sim i} x_j \;=\; -x_i. \tag{1}$$

If $x_i = 0$ then (1) implies that $x_j = 0$ for $j \sim i$. Since G is connected, we would have $\boldsymbol{x} = 0$, which is impossible. We conclude, that $x_i \neq 0$ for all i. By (1), $\deg(i)|x_i| = |\sum_{j \sim i} x_j| \leq \sum_{j \sim i} |x_j|$. Summing over all i we get $\sum_i \deg(i)|x_i| \leq \sum_i \deg(i)|x_i|$ and hence the inequality must be equality for each i. This is only possible if $\forall i$ the sign of all x_j for $j \sim i$ is the same. Since all x_i are non-zero, this is only possible if G is bipartite. $\square$

Appendix: More formal arguments for $\lambda = 1$

Statement: If G is connected then $\lambda = 1$ has multiplicity one or, in other words, if $P\boldsymbol{x} = \boldsymbol{x}$ then $\boldsymbol{x} = c\mathbf{1}$ for some $c \neq 0$.

Proof. For every i, we have

$$\sum_{j=1}^{N} P_{ij} x_j \;=\; \frac{1}{\deg(i)} \sum_{j \sim i} x_j \;=\; x_i. \tag{2}$$

Suppose that $x_k = \max_i x_i$. The equation $\frac{1}{\deg(k)} \sum_{j \sim k} x_j = x_k$ implies that $x_j = x_k$ for all $j \sim k$. Using the fact that G is connected, we propagate this equality across the whole graph and so all the entries of $\boldsymbol{x}$ must be equal (and non-zero). $\square$

PageRank

Note

We define random walk on a directed graph in analogous way.

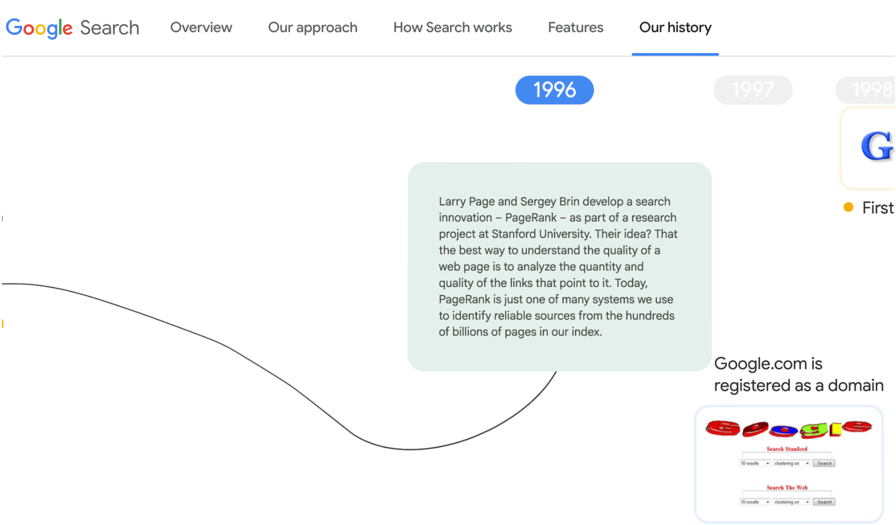
Algebraically more complicated as A_G is not symmetric and the eigenvalues are complex.

- Web graph = directed network of pages and hyperlinks.
- Eigenvector centrality does not work directly in directed graphs with sinks or disconnected components.
- PageRank modifies the random walk with teleportation:

$$P_\alpha = \alpha P + (1-\alpha)\frac{1}{N}\mathbf{1}\mathbf{1}^T,$$

  where P is the transition matrix of the web, $\alpha \in (0,1)$.
- Stationary distribution of P_α = PageRank vector.

https://www.google.com/search/howsearchworks/our-history/



- Solving for π = solving a huge eigenvector problem ($\sim 10^{10}$ nodes).
- Power iteration with $\alpha = 0.85$ converges in ~ 50 steps.

Computing Centrality in Python (NetworkX)

```
import networkx as nx

G = nx.karate_club_graph()

# Eigenvector centrality
eig = nx.eigenvector_centrality(G)
print(max(eig, key=eig.get))

# PageRank
pr = nx.pagerank(G, alpha=0.85)
print(max(pr, key=pr.get))
```

Karate club example: - Eigenvector centrality highlights the main hub (node 33). - PageRank is similar but also adapts to directed networks.

Conclusions

- Eigenvector centrality: nodes are important if linked to other important nodes.
- Perron–Frobenius ensures uniqueness and positivity of the principal eigenvector.
- PageRank extends the same idea to the Web via teleportation.
- Linear algebra (largest eigenvalue, eigenvector) is the foundation of centrality measures.

Random graphs and Erdős–Rényi model

Why random graphs?

Real networks (social, economic, financial) are noisy and constantly evolving. We need a simple *baseline model* to compare against.

Definition (Erdős–Rényi (ER) model)

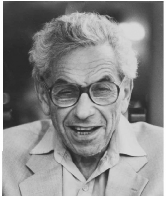
$G(N, p)$: a random graph on N nodes where each of the $\binom{N}{2}$ possible edges appears independently with prob. p.

Why random graphs?

Real networks (social, economic, financial) are noisy and constantly evolving. We need a simple *baseline model* to compare against.

Definition (Erdős–Rényi (ER) model)

$G(N, p)$: a random graph on N nodes where each of the $\binom{N}{2}$ possible edges appears independently with prob. p.

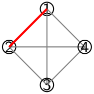


Paul Erdős (1913 - 1996) Alfréd Rényi (1921-1970)

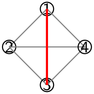
Erdős and Rényi (1959–60) launched the probabilistic study of graphs. Their program connected combinatorics and probability, leading to modern random graph theory.

$G(N, p)$ Model

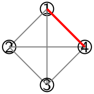
Take $N = 4$ then the graph can have up to six edges. Each with distribution $\mathrm{Bern}(p)$:



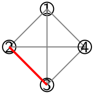
12



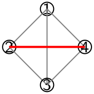
13



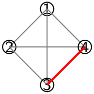
14



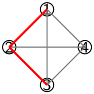
23



24



34

Pr(

$) = p^2(1-p)^4$

If $p = \frac{1}{2}$, each graph appears with the same probability $\frac{1}{2^6} = \frac{1}{64}$.

Probability recap: Binomial

Definition

If $X \sim \mathrm{Bin}(n, p)$ then

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mathbb{E}[X] = np, \quad \mathrm{Var}(X) = np(1-p).$$

Useful characterization: $X = \sum_{i=1}^{n} Z_i$ with independent $Z_i \sim \mathrm{Bern}(p)$.

In $G(N, p)$**:**

- Number of edges:

$$L \;\sim\; \mathrm{Bin}\left(\binom{N}{2}, p\right).$$

- Degree of a fixed vertex v:

$$\deg(v) \;\sim\; \mathrm{Bin}(N-1, p).$$

Probability recap: Poisson (as Binomial limit)

Theorem

If $X_n \sim \mathrm{Bin}(n, p_n)$ with $n \to \infty$ and $np_n \to \lambda > 0$, then

$$X_n \longrightarrow X \sim \mathrm{Pois}(\lambda), \qquad \Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

The approximation $\mathrm{Bin}(n, p) \approx \mathrm{Poiss}(\lambda)$ for $\lambda = pn$ is particularly good if p is small.

Example (Quick check)

For $n = 2000$, $p = 0.003$, $\lambda = np = 6$. Compare $\Pr(X = 0)$: Binomial $\approx (1 - p)^{2000}$ vs. Poisson e^{-6} (very close).

Degree distribution in Erdős–Rényi model

Degree distribution in $G(N, p)$

For a fixed v, if $p = \lambda/(N-1)$,

$$\deg(v) \;\sim\; \mathrm{Bin}(N-1, p) \;\approx\; \mathrm{Pois}(\lambda)$$

- Mean degree: $\mathbb{E}[\deg(v)] = (N-1)p$.
- Sparse regime $p = \lambda/(N-1)$: $\Pr\{\deg(v) = k\} \approx \dfrac{\lambda^k}{k!} e^{-\lambda}$.
- Why useful: closed forms for expectations; Poisson is a great approximation when N is large and p small.

Concentration: Chebyshev (simple but general)

Theorem (Chebyshev inequality)

For any r.v. X with mean μ and variance σ^2,

$$\Pr(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

For degree: $\deg(v) \sim \mathrm{Bin}(N-1, p)$, so

$$\Pr\big(|\deg(v) - (N-1)p| \geq t\big) \leq \frac{(N-1)p(1-p)}{t^2}.$$

Chebyshev is loose but distribution-free; good first control of deviations.

Sharper concentration: Hoeffding for Binomial

Theorem (Hoeffding inequality)

If $X = \sum_{i=1}^{n} Y_i$ with independent $Y_i \in [0,1]$ and $\mathbb{E}X = \mu$, then for $t > 0$,

$$\Pr(|X - \mu| \geq t) \leq 2\exp\left(-\frac{2t^2}{n}\right).$$

Applied to degree: $\deg(v)$ has $N-1$ independent Bernoulli summands,

$$\Pr(|\deg(v) - (N-1)p| \geq t) \leq 2\exp\left(-\frac{2t^2}{N-1}\right).$$

Taking $t_0 = \sqrt{(N-1)\log N}$ gives

$$\Pr\left(|\deg(v) - (N-1)p| \geq t_0\right) \leq \frac{2}{N^2}.$$

A union bound over all v shows all degrees concentrate near $(N-1)p$ with high probability.

Asymptotic Thinking in Random Graphs

Why asymptotics?

- We study $G(N, p)$ as $N \to \infty$ to reveal general patterns.
- Precise constants matter less than the scaling behavior of p with N.

Notation recap:

- $f(N) = o(g(N))$ means $f(N)/g(N) \to 0$.
- $f(N) = O(g(N))$ means $|f(N)| \leq C|g(N)|$ for large N.
- $f(N) \sim g(N)$ means $f(N)/g(N) \to 1$.

Probabilistic language:

- "With high probability" (w.h.p.) means $\Pr(\text{event}) \to 1$ as $N \to \infty$.
- Example: in $G(N, p)$ with $p = \frac{\log N}{N}$, the graph is connected *w.h.p.*

Mindset: We think of N as huge and ask: "At what scale of p does a qualitative change occur?" This scale is called a threshold.

Maximum degree in $G(N, p)$

Let $\Delta = \max_v \deg(v)$ be the **maximum degree**.

1. **Dense regime (p constant, not tiny):**
   - Each $\deg(v) \sim \mathrm{Bin}(N-1, p)$ with mean $\mathbb{E}\deg(v) \approx Np$.
   - With high probability:

$$\Delta \; = \; Np \, + \, O\big(\sqrt{N \log N}\big).$$

2. **Sparse regime ($p = \lambda/N$):**
   - Each $\deg(v) \approx \mathrm{Pois}(\lambda)$ — mean λ.
   - By extreme–value theory for Poisson tails:

$$\Delta \; \approx \; \frac{\log N}{\log \log N}.$$

Takeaway: Even in purely random graphs, a few nodes will look like *"hubs"* simply due to chance.

Notation: average degree vs expected degree

For a graph G with N vertices and L edges:

- The **empirical average degree** is (a random variable)

$$\overline{\deg}(G) \;=\; \frac{1}{N}\sum_{v\in V}\deg(v) \;=\; \frac{2L}{N}.$$

- The **expected degree** under a random graph model is

$$\mathbb{E}[\deg] \;:=\; \mathbb{E}[\overline{\deg}(G)].$$

Example (Erdős–Rényi $G(N,p)$):

$$\overline{\deg}(G) \approx (N-1)p, \qquad \mathbb{E}[\deg] = (N-1)p.$$

We saw that for large N, $\overline{\deg}(G)$ is tightly concentrated around $\mathbb{E}[\deg]$.