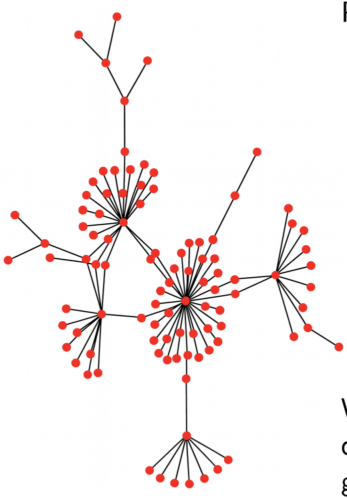
upf. Universitat Pompeu Fabra Barcelona

Seminar 3 · Networks, Crowds and Markets

Scale-Free Networks

Warm-up



Pick the best estimate for:

1. minimal degree:
   0, 1, 3, 30
2. maximal degree:
   5, 25, 100, 1000
3. the diameter:
   2,4,10,20
4. the clustering coefficient:
   0.05, 0.5, 0.75, 0.95.

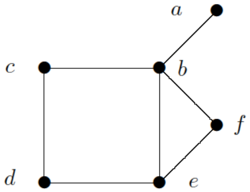
Would a social network be likely to have the diameter and clustering coefficient of this graph?

Exercise 1. Clustering Coefficient

Find the clustering coefficient for:

a) the central node in a Star Graph with N nodes.

b) node b, d and f of the following graph:



Exercise 2. Clustering in ER vs Real Data

Theory. For $G(N, p)$, show $\mathbb{E}[C_v] = p$ for any node v (condition on $\deg(v) = k$, then average over k).

Practice.

- Simulate $G(200, 0.05)$ and estimate the average clustering coefficient $\overline{C}$.
- Compute $\overline{C}$ on a real network (e.g., `karate_club_graph()`).
- Compare. Why does ER under-predict clustering in social networks?

Exercise 3: Estimating clustering coefficient

Suppose you want to compute your clustering coefficient on Facebook or any other social network.

We do not have access to the whole network so we can do things only manually: for any two of your friends check if they are friends.

Say you have 200 friends. The number of pairs – 19,900 – may be too large to explore by hand.

Propose a method to **estimate** the clustering coefficient that can be computed quicker.

Exercise 4: Power Law

Given a network with $N = 10^7$ nodes, $k_{min} = 1$, and with the following power-law distribution:

$$p_k = Ck^{-2}$$

a) Determine the probability of finding a node with 100 links attached.

b) What is the expected degree and variance?

c) Determine the value of C in the Continuum Formalism.

d) What is the probability that a node has between 1 and 10 edges?

e) How many hubs do we expect to find in this network? $(k \geq 10^5)$

Exercise 5: More on power laws

Given a SFN with $N = 25{,}000$ nodes and $\gamma = 2.12$, $k_{min} = 5$, determine:

a) Its degree distribution (in both formalisms).

b) The probability of having a node with exactly 10 links.

c) The expected degree.

d) The number of hubs that the network has (with degree $k \geq 5000$)

e) The expected number of links.

f) What is the probability of finding a node with the same amount or fewer links than the average.

Exercise 6: Cayley tree

A *Cayley tree* is a symmetric tree constructed starting from a central node of degree k. Each node at distance 1 from the central node has degree k. More generally, each node at distance ℓ from the central node has degree k until we reach the nodes at distance t, which have degree one and are called *leaves*.

1. Calculate the number of nodes reachable in s steps from the central node.
2. Calculate the degree distribution of the network.
3. Calculate the diameter d_{max}.
4. Find an expression for the diameter d_{max} in terms of the total number of nodes N.
5. Does the network display the small-world property?

Additional exercises

Exercise: Power law vs. Poisson

Consider the in-degree distribution of the World Wide Web, which is approximately a power law with exponent $\gamma_{\text{in}} = 2.1$ and minimum degree $k_{\text{min}} = 1$. There are about $N = 10^{12}$ pages.

For comparison, let us take a random Erdős–Rényi network of the same size and with the same average in-degree $\langle k_{\text{in}} \rangle = 4.6$.

1. Estimate the fraction of nodes with $1 \leq k \leq 5$ incoming links in both networks. Compare the results.
2. For the Erdős–Rényi network, find the approximate range of degrees that contains 68% of all nodes.
3. Estimate how many pages have more than 10^5 incoming links in each network, and discuss the qualitative difference.

Hint. For the power law, you may use the normalized form $p_k = k^{-\gamma}/\zeta(\gamma)$. For the ER network, assume a Poisson with $\lambda = 4.6$.

Solution: Power law vs. Poisson (1/4)

We compare two in-degree models on $N = 10^{12}$ pages:

$$\textbf{Power law:} \quad p_k^{\mathrm{PL}} = \frac{k^{-\gamma}}{\zeta(\gamma)}, \quad \gamma = 2.1,\ k \geq 1,\ \zeta(2.1) \approx 1.58.$$

$$\textbf{ER/Poisson:} \quad K \sim \mathrm{Poisson}(\lambda), \quad \lambda = 4.6.$$

Truncation for $k_{\mathrm{min}} = 1$: If we enforce $k \geq 1$ for the Poisson model, use the truncated distribution

$$p_k^{\mathrm{Pois}|k \geq 1} = \frac{\Pr(K = k)}{\Pr(K \geq 1)} = \frac{e^{-\lambda}\lambda^k/k!}{1 - e^{-\lambda}}, \quad k \geq 1,$$

where $1 - e^{-\lambda} \approx 0.9900$. This reweights probabilities by a factor $\approx 1/0.99 \approx 1.0101$ (a $\sim$1% effect).

Solution: Power law vs. Poisson (2/4)

(a) Fraction with $1 \le k \le 5$

Power law:

$$\Pr_{\mathrm{PL}}(1 \le k \le 5) = \frac{\sum_{k=1}^{5} k^{-2.1}}{\zeta(2.1)} \approx \frac{1.422}{1.58} \approx 0.90 \quad \Rightarrow \quad \text{about } 9.0{\times}10^{11} \text{ nodes.}$$

Poisson ($k \ge 1$ truncated):

$$\Pr(1 \le K \le 5 \mid K \ge 1) = \frac{\Pr(1 \le K \le 5)}{1 - e^{-\lambda}} \approx \frac{0.669}{0.990} \approx 0.676.$$

(b) "68%" confidence intervals

For $K \sim \mathrm{Poisson}(\lambda)$, the *normal approximation* gives

$$\mu = \lambda = 4.6, \quad \sigma = \sqrt{\lambda} \approx 2.144.$$

About 68% of a normal law lies in $[\mu - \sigma, \mu + \sigma]$. With a continuity correction:

$$[\,\mu - \sigma,\ \mu + \sigma\,] \approx [\,2.46,\ 6.74\,].$$

The integer degrees in this band are $k \in \{3, 4, 5, 6\}$ (or one may report $\{2, \ldots, 7\}$ for $\approx 68\%$ by symmetry).

Solution: Power law vs. Poisson (3/4)

(c) Nodes with $k > 10^5$

Power law (continuous tail):

$$\Pr(K > k_0) \approx \frac{1}{\zeta(\gamma)} \int_{k_0}^{\infty} x^{-\gamma}\, dx = \frac{1}{\zeta(\gamma)} \cdot \frac{k_0^{1-\gamma}}{\gamma - 1}.$$

For $\gamma = 2.1$, $k_0 = 10^5$:

$$\Pr(K > 10^5) \approx \frac{1}{1.58 \cdot 1.1} (10^5)^{-1.1} \approx 1.8 \times 10^{-6},$$

so the expected count is

$$N \Pr(K > 10^5) \approx 10^{12} \times 1.8 \times 10^{-6} \approx 1.8 \times 10^6 \text{ pages}.$$

Poisson: Note that using Hoeffding for $\mathrm{Bin}(N-1, p)$ is useless at this scale. We can use the Hoeffding bound to conclude that $N\mathbb{P}(K \geq k_0) \approx 0$.

Conclusion: The heavy tail of the power law yields millions of hubs with $k > 10^5$, whereas the ER/Poisson model yields essentially none.

Solution: Sharper intuition (beyond Hoeffding) (4/4)

Even though Hoeffding fails numerically, the true probability is *astronomically small.*

- A more refined approach (not covered in this course) uses **Chernoff bounds** or the **large-deviation principle** for binomial variables.
- Let $Z \sim \mathrm{Bin}(N-1, \lambda/N)$. Again, by Markov's inequality, for any $s > 0$, $\mathbb{P}(Z \geq z) \leq e^{-sz}\mathbb{E}e^{sZ} = e^{-sz}(1 - \frac{\lambda}{N} + \frac{\lambda}{N}e^{s})^{N-1} \approx e^{\lambda e^{s} - sz}$. Optimize the bound with respect to s (taking s such that $e^{s} = \frac{z}{c}$) to get the bound $e^{z - z\log(\frac{z}{\lambda})}$, which is extremely small for $z = 10^5$.

$$\Pr(S \geq 10^5) \lesssim e^{-9\times 10^5}.$$

So even across $N = 10^{12}$ nodes,

$$N \Pr(S \geq 10^5) \approx 0.$$

Conclusion: Erdős–Rényi networks cannot produce hubs, while power-law networks do so naturally.

Exercise: Snobbish network

Consider a network of N blue and N red nodes. Any two nodes of the same color are connected independently with probability p, and any two nodes of different colors with probability $q \leq p$.

(a) Compute the expected degree of a blue node:
- within its own color class (blue–blue links),
- and in the full network (blue–blue plus blue–red links).

(b) For large N, what happens to the network when $q = 0$? Describe qualitatively how the picture changes as q increases from 0 to values comparable to p.

(c) Suppose p and q are both of order $1/N$ so that the average degree stays around a constant. Argue (heuristically, not rigorously) that in this regime the typical distance between two nodes still grows roughly like $\log N$, even when $p \gg q$.

(d) Extra: What is the smallest value of q that is likely to make the two color groups part of the same connected component? (Hint: think about how many cross-color edges there are on average.)