Universitat Pompeu Fabra Barcelona

Lesson 9 · Clustering and small world

Networks, Crowds and Markets

Today's Lecture

We continue our discussion of various phenomena observed in real networks. We study to what extend a basic model as the Erdős-Renyi model could explain it.

1. The clustering coefficient: definition, motivation, formulas.

2. Small World Phenomenon

3. Milgram's experiment

4. Power laws and hubs

Clustering

Clustering

Real networks are not tree-like. Friends of friends often know each other (and so triangles are common).

Examples:

- **Social networks:** If Alice knows Bob and Carol, it's likely Bob and Carol also know each other. $\rightarrow$ Social circles, community structure.
- **Trade networks:** Countries trading with the same partner often trade with each other. $\rightarrow$ Formation of regional trade blocs.
- **Financial networks:** Two banks lending to the same counterparties are likely connected through risk exposures. $\rightarrow$ Triangles increase contagion channels.
- **Citation or collaboration networks:** If researcher A collaborates with both B and C, B–C collaboration becomes more probable. $\rightarrow$ Knowledge diffusion through closed triads.

Clustering coefficient: definition

Definition

For node v with degree $\deg(v) = k_v$:

$$C_v = \frac{\#\text{ links among neighbors of } v}{\binom{k_v}{2}} \in [0, 1].$$

- Measures "friend-of-friend closure."
- $C_v = 1$: neighbors form a clique; $C_v = 0$: none connected.
- Average clustering coefficient: $\overline{C} = \frac{1}{N} \sum_v C_v$.

Triangles in networks represent tightly knit groups and enable fast information spread or trust formation in social networks.

Clustering in Erdős–Rényi networks

Suppose $\deg(v) = k_v$ (fixed). Consider two neighbors u, w.

Each pair u, w gets connected (independently) with probability p.

The expected number of links among neighbors is $\mathbb{E}L_v = p\binom{k_v}{2}$.

Thus

$$\mathbb{E}[C_v] \;=\; \mathbb{E}\left[\frac{L_v}{\binom{k_v}{2}}\right] \;=\; \frac{\mathbb{E}[L_v]}{\binom{k_v}{2}} \;=\; p.$$

Clustering in Erdős–Rényi networks

Suppose $\deg(v) = k_v$ (fixed). Consider two neighbors u, w.

Each pair u, w gets connected (independently) with probability p.

The expected number of links among neighbors is $\mathbb{E}L_v = p\binom{k_v}{2}$.

Thus

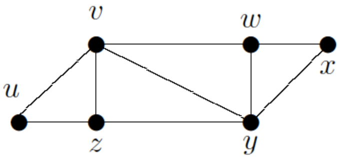
$$\mathbb{E}[C_v] = \mathbb{E}\left[\frac{L_v}{\binom{k_v}{2}}\right] = \frac{\mathbb{E}[L_v]}{\binom{k_v}{2}} = p.$$

Implications:

- In the sparse regime $p = c/N$: $\mathbb{E}[C_i] \approx c/N \to 0$.
- Conclusion: clustering vanishes as N grows.
- Real networks (social, financial, trade) exhibit far higher clustering. $\Rightarrow$ Mismatch: motivates richer models leading so sparse networks with nontrivial clustering coefficients.

Exercise

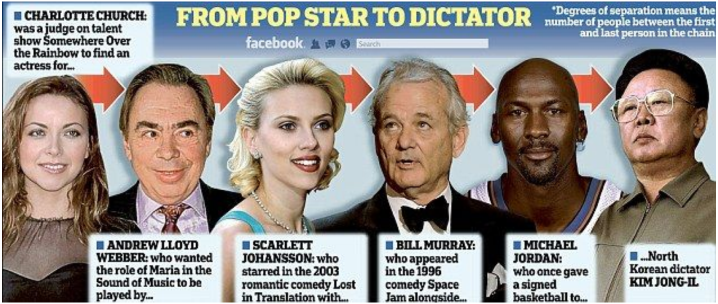
Determine the Clustering Coefficient for nodes w and y.



Small World Phenomenon

Six degrees of separation

- The idea: in social networks, any two people are separated by surprisingly few steps.
- Popularized as "six degrees of separation" in the 1960s.
- Relatively recent data (Facebook, 2016): average distance between users ≈ 3.5.
- LinkedIn advertises "3rd-degree connections" — the same principle in business networks.



Sandbox: Explore the Small World Yourself

The "six degrees" idea is easy to explore online and in data.

- **Wikipedia Game:** shortest path between two articles
  https://wiki.spaceface.dev/
- **Hollywood Game:** distance between any actor and Kevin Bacon
  https://oracleofbacon.org/
- **Mathematical Collaboration:** find distance between any two mathematicians (e.g., Paul Erdős and Piotr Zwiernik)
  https://mathscinet.ams.org/mathscinet/freetools/collab-dist
- **Social Graphs:** Facebook's *"Degrees of Separation"* study (avg. distance ≈ 3.5!) https://research.facebook.com/blog/2016/02/three-and-a-half-degrees-of-separation/

Even in enormous graphs, the average distance between two nodes often scales like $\log N$ — a hallmark of the Small World phenomenon.

Diameter in Erdős–Rényi Graphs

Let $c = \mathbb{E}(\overline{\deg}(G))$ in $G(N, p)$ with $p = c/N$ (sparse regime).

Branching-process heuristic:

- Running BFS, early layers are almost tree-like.
- Each node produces on average c new nodes.

Hence, the expected number of vertices within distance $\leq d$ from a node is approximately

$$N(d) \;\approx\; 1 + c + c^2 + \cdots + c^d \;=\; \frac{c^{d+1} - 1}{c - 1}.$$

Diameter in Erdős–Rényi Graphs

Let $c = \mathbb{E}(\overline{\deg}(G))$ in $G(N, p)$ with $p = c/N$ (sparse regime).

Branching-process heuristic:

- Running BFS, early layers are almost tree-like.
- Each node produces on average c new nodes.

Hence, the expected number of vertices within distance $\leq d$ from a node is approximately

$$N(d) \;\approx\; 1 + c + c^2 + \cdots + c^d \;=\; \frac{c^{d+1} - 1}{c - 1}.$$

Assume $c \gg 1$. We estimate the graph diameter (or typical distance) by solving $N(d_{\max}) \approx N$:

$$N \;\approx\; c^{\,d_{\max}} \quad \Rightarrow \quad d_{\max} \;\approx\; \log_c N \;=\; \frac{\ln N}{\ln c}.$$

The diameter formula

As we showed: If the average degree is $c \gg 1$, then

$$\mathrm{diam}(G) \approx \frac{\ln N}{\ln c}.$$

- Distances in random networks grow only logarithmically in N.
- This explains why even very large systems can feel “small”.

Example

Consider a network with $N = 10^6$ and $c = 10$. Then the diameter is approximately

$$\frac{\ln 10^6}{\ln 10} = \frac{6 \ln 10}{\ln 10} = 6$$

.

Diameter of Real Networks

Many real-world networks also have very short path lengths.

Network	N	L	‹k›	‹d›	d_{max}	$\ln N/\ln\langle k\rangle$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,437	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Note: The last column better approximated the average distance.

Milgram's experiment

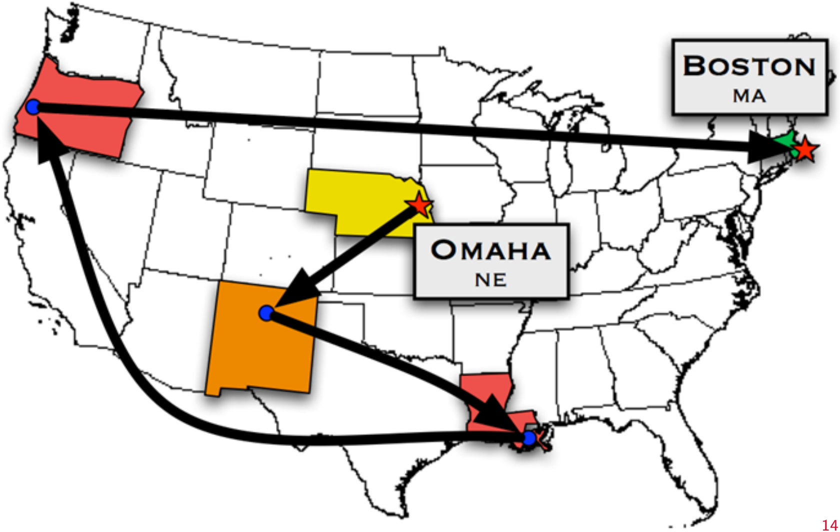
Subway

Milgram's Small-World Experiment (1967)

Psychologist Stanley Milgram wanted to measure the "social distance" between two random people in the US.

- He selected a target person living in Sharon, Massachusetts (a stockbroker).
- Random participants across the US received a letter with instructions:
  - If they knew the stockbroker the chain terminates.
  - Otherwise, they had to forward the letter to one acquaintance whom they believed was closer to the target (geographically, professionally, or socially).
  - Each acquaintance repeated the same rule, until (hopefully) the letter reached the target.

Output of Milgram's Experiment

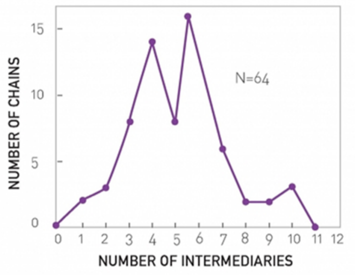


Results of Milgram's Experiment

- About 296 letters were originally sent out. Only 64 letters successfully reached the target.
- For those that did arrive, the average chain length was about

$$\text{Average path length} \approx 6.5.$$

- This gave rise to the popular phrase: **"six degrees of separation"**.



Why was this surprising?

- At the time, many expected paths to be much longer (tens or hundreds of steps).
- The experiment revealed that human social networks are extremely well connected.
- Even though only $\approx 20\%$ of letters arrived, the short paths were consistent.
- Interpretation: social networks have a **"small-world property"** — typical distances grow very slowly with population size.

Question: If distances are small, why do real networks still look so different from ER? Let's look at their degree distributions.

Power laws: motivating example

The World Wide Web

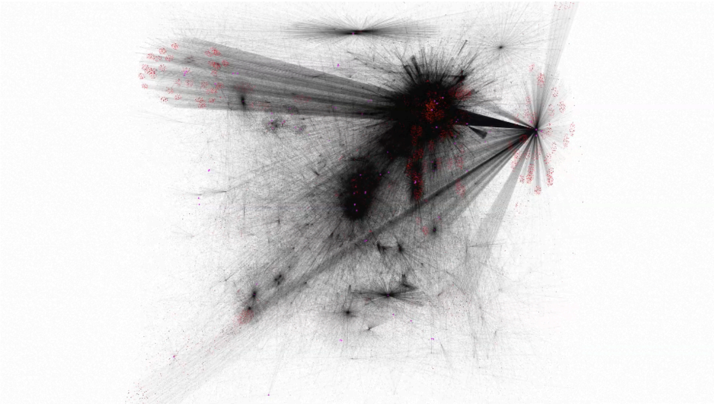
The visible Web today has roughly as many pages as the human brain has neurons — and far fewer links than the brain has synapses.

The World Wide Web

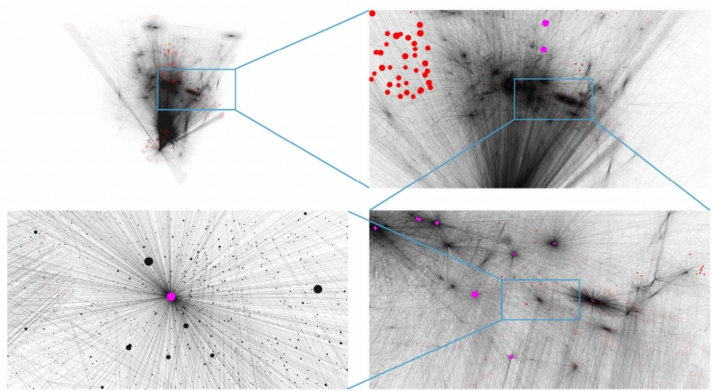
The visible Web today has roughly as many pages as the human brain has neurons — and far fewer links than the brain has synapses.

- The World Wide Web (WWW) is a network: nodes are web pages, edges are hyperlinks (URLs).
- The site WorldWideWebSize.com estimates that the indexed web (the portion that major search engines have indexed) contains at least 4 billion pages (as of Jan 15, 2025).
- The human brain has roughly 86 billion neurons.
- The structure can be mapped using a crawler that follows hyperlinks.
- The first large-scale map of the WWW for scientific purposes was created by Hawoong Jeong (Notre Dame, 1998).

Monitoring the map of WWW



Hubs in the WWW

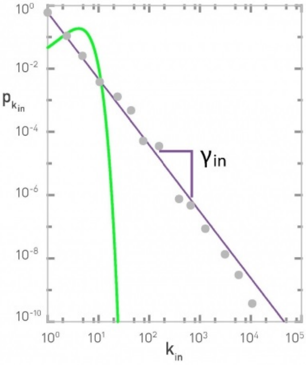


Close-ups reveal the presence of a few very highly connected nodes, often called *hubs*.

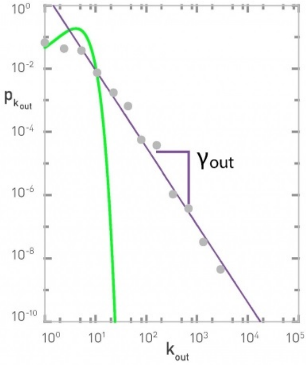
Degree distribution of the WWW

- Most pages have few links, but a few have extremely many.
- This is visible on a **log-log plot** of the degree distribution.

In-degree



Out-degree



$$\log P(k) \sim -\gamma \log k \quad \Longrightarrow \quad P(k) \sim k^{-\gamma}.$$

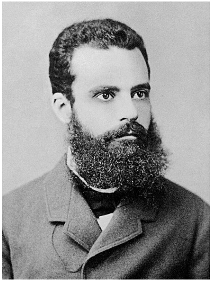
Power laws and hubs

Why "scale-free"?

- In most familiar distributions (e.g., Gaussian, Poisson) there is a *typical scale*: most values cluster around the mean.
- In a power law, $p_k \propto k^{-\gamma}$, there is no characteristic degree.
  - Small degrees are common, but very large ones also occur with non-negligible probability.
  - The ratio
  $$\frac{p_{ak}}{p_k} = \frac{(ak)^{-\gamma}}{k^{-\gamma}} = a^{-\gamma}$$
  is **independent of** k!
- Hence the term **scale-free**: the distribution looks the same no matter how we "zoom" in on the scale of k.

Historical roots: Pareto and the 80/20 law

Vilfredo Pareto (1848–1923), Italian economist, observed that income distribution in society is very uneven.



- Incomes followed a distribution with a heavy tail: a small fraction of people held most of the wealth.
- This became the well-known "80/20 rule": e.g. 20% of people control 80% of wealth.
- Similar patterns appear in many domains:
  - 80% of web links point to about 20% of webpages.
  - A small number of firms or banks control a large share of markets.
  - A few researchers or papers receive most citations.

Connection: Pareto's law is an early example of a *power law* in economics, closely related to what we now see in network degree distributions.

Power law: Discrete formalism

We model the degree distribution of a scale–free network as

$$p_k \;=\; \frac{1}{\zeta(\gamma)} k^{-\gamma}, \quad k \geq 1,$$

where $\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$ is the normalizing constant.

The function $\zeta(\gamma)$ is called the Riemann zeta function.

- The series converges only if $\gamma > 1$.
- In real networks, empirical exponents typically satisfy $2 < \gamma \leq 3$.

First two moments

If $Z \sim (p_k)$ with $p_k = \frac{1}{\zeta(\gamma)} k^{-\gamma}$ for $k \geq 1$, then

$$\mathbb{E}Z \;=\; \sum_{k\geq 1} k p_k \;=\; \frac{1}{\zeta(\gamma)} \sum_{k\geq 1} k k^{-\gamma} \;=\; \frac{1}{\zeta(\gamma)} \sum_{k\geq 1} k^{-(\gamma-1)} \;=\; \frac{\zeta(\gamma-1)}{\zeta(\gamma)}.$$

$$\mathbb{E}Z^2 \;=\; \frac{1}{\zeta(\gamma)} \sum_{k\geq 1} k^2 k^{-\gamma} \;=\; \frac{\zeta(\gamma-2)}{\zeta(\gamma)}$$

First two moments

If $Z \sim (p_k)$ with $p_k = \frac{1}{\zeta(\gamma)} k^{-\gamma}$ for $k \geq 1$, then

$$\mathbb{E}Z \; = \; \sum_{k \geq 1} k p_k \; = \; \frac{1}{\zeta(\gamma)} \sum_{k \geq 1} k k^{-\gamma} \; = \; \frac{1}{\zeta(\gamma)} \sum_{k \geq 1} k^{-(\gamma-1)} \; = \; \frac{\zeta(\gamma-1)}{\zeta(\gamma)}.$$

$$\mathbb{E}Z^2 \; = \; \frac{1}{\zeta(\gamma)} \sum_{k \geq 1} k^2 k^{-\gamma} \; = \; \frac{\zeta(\gamma-2)}{\zeta(\gamma)}$$

Recall, $\zeta(\gamma)$ only defined for $\gamma > 1$. The regime $2 < \gamma \leq 3$ is special:

- Since $\gamma - 1 > 1$, the mean exists.
- Since $\gamma - 2 \leq 1$, the variance does not!

(a very heavy-tailed distribution)

Recall that the degrees in $\mathrm{ER}(N, p)$ are concentrated around the mean.

Power law: Continuum formalism

Sums like $\sum_k k^{-\gamma}$ are hard to work with. For large networks, it is convenient to approximate degrees k by a continuous variable.

If $k_{\min}$ large, $\int_{k_{\min}}^{\infty} k^{-\gamma} \mathrm{d}k \approx \sum_{k=k_{\min}}^{\infty} k^{-\gamma}$.

Define $p(k) = Ck^{-\gamma}$ for for $k \geq k_{\min} \geq 1$, where C is such that

$$\int_{k_{\min}}^{\infty} p(k)\, dk = 1.$$

Power law: Continuum formalism

Sums like $\sum_k k^{-\gamma}$ are hard to work with. For large networks, it is convenient to approximate degrees k by a continuous variable.

If $k_{\min}$ large, $\int_{k_{\min}}^{\infty} k^{-\gamma}\mathrm{d}k \approx \sum_{k=k_{\min}}^{\infty} k^{-\gamma}$.

Define $p(k) = Ck^{-\gamma}$ for for $k \geq k_{\min} \geq 1$, where C is such that

$$\int_{k_{\min}}^{\infty} p(k)\, dk = 1.$$

Normalization gives

$$C \int_{k_{\min}}^{\infty} k^{-\gamma}\, dk = 1 \quad \Rightarrow \quad C = (\gamma - 1) k_{\min}^{\gamma-1}.$$

Power law: Continuum formalism

Sums like $\sum_k k^{-\gamma}$ are hard to work with. For large networks, it is convenient to approximate degrees k by a continuous variable.

If $k_{\min}$ large, $\int_{k_{\min}}^{\infty} k^{-\gamma} \mathrm{d}k \approx \sum_{k=k_{\min}}^{\infty} k^{-\gamma}$.

Define $p(k) = Ck^{-\gamma}$ for for $k \geq k_{\min} \geq 1$, where C is such that

$$\int_{k_{\min}}^{\infty} p(k)\, dk = 1.$$

Normalization gives

$$C \int_{k_{\min}}^{\infty} k^{-\gamma}\, dk = 1 \quad \Rightarrow \quad C = (\gamma - 1) k_{\min}^{\gamma-1}.$$

Thus

$$\boxed{p(k) = (\gamma - 1)\, k_{\min}^{\gamma-1}\, k^{-\gamma}}, \quad k \geq k_{\min}.$$

Extreme value of a power law: scaling of $k_{\max}$

For $p(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}$ $(k \geq k_{\min})$, the survival function is

$$\Pr\{K \geq k\} = \int_k^{\infty} p(x)\,dx = \left(\frac{k_{\min}}{k}\right)^{\gamma-1}.$$

In a network with N nodes, the *typical* maximum degree $k_{\max}$ is found by the extreme-value heuristic

$$N\,\Pr\{K \geq k_{\max}\} \;\approx\; 1 \quad\Longrightarrow\quad N\left(\frac{k_{\min}}{k_{\max}}\right)^{\gamma-1} \approx 1.$$

$$\boxed{k_{\max} \;\approx\; k_{\min}\, N^{1/(\gamma-1)}}$$

Notes.

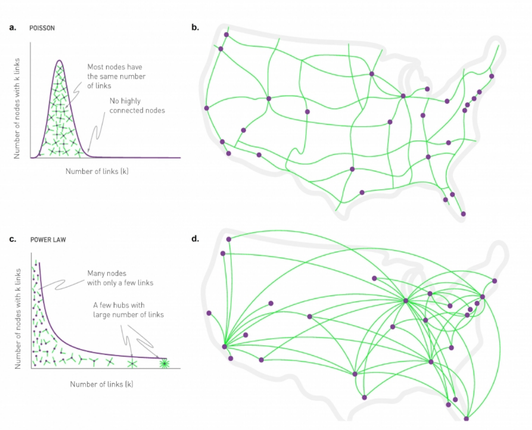
- This gives the correct order for the *typical* (median) maximum; fluctuations are smaller-order.
- Works the same way for the discrete model (replace integrals by sums). Just harder calculations.

Consequences of the $k_{\max}$ scaling

From $k_{\max} \approx k_{\min} N^{1/(\gamma-1)}$:

$$\begin{aligned}
\gamma = 2 \quad &\Rightarrow \quad k_{\max} \sim k_{\min}\, N \quad \text{(one hub touches a linear fraction of nod} \\
2 < \gamma < 3 \quad &\Rightarrow \quad k_{\max} \sim k_{\min}\, N^{1/(\gamma-1)} \text{ sublinear but large} \\
\gamma = 3 \quad &\Rightarrow \quad k_{\max} \sim k_{\min}\, N^{1/2} \\
\gamma > 3 \quad &\Rightarrow \quad k_{\max} \text{ grows slowly; tails are lighter}
\end{aligned}$$

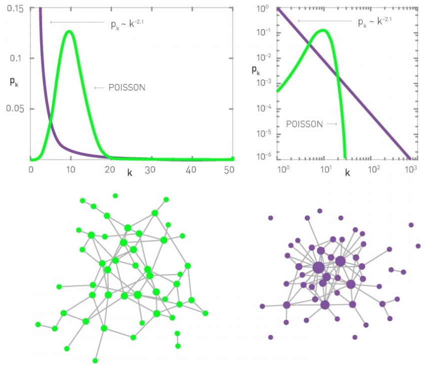
Scale-free vs Erdős–Rényi networks



Main differences in degree distribution

- The key difference lies in the **tail behavior**.
- For small k, power law predicts many low-degree nodes — ER does not.
- Around the mean degree $\mathbb{E}[\text{deg}]$, ER has a sharp peak (Poisson), while scale-free is much flatter.
- For large k, scale-free has a heavy tail: a non-negligible number of very high degree nodes.

Implication: scale-free networks generate hubs, while ER almost never does.



ER Network

Scale-Free Network

Universality of power law

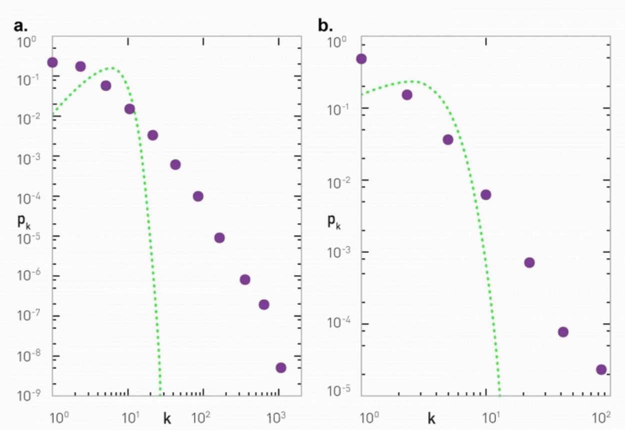
Universality

Power laws appear across very different networks:

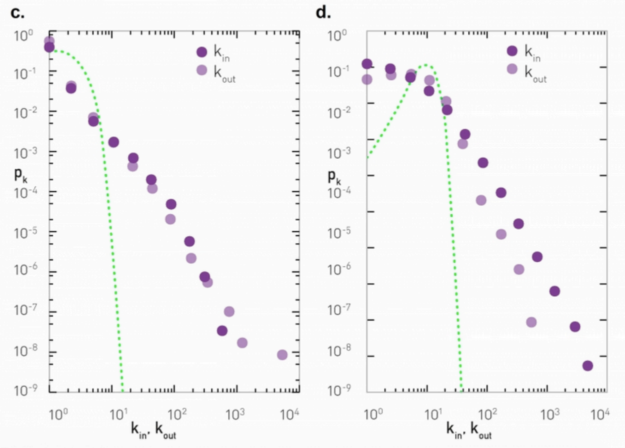
- Air transportation networks.
- The WWW.
- The internet (routers, AS graph).
- Protein–protein interaction networks.
- Email and communication networks.
- Citation networks.

Remark: these systems are very different in origin, but share the same statistical signature — power-law degree distributions.

Internet & Protein interaction networks



Email & Citation networks



Path lenghts in Scale-Free Networks

Average path length

Let $d(u, v)$ be the distance between two vertices and $\mathbb{E}[d]$ the average distance across all pairs.

- In Erdős–Rényi:

$$\mathbb{E}[d] \sim \ln N.$$

- In scale-free networks:

$$\mathbb{E}[d] \sim \begin{cases} \text{constant}, & \gamma = 2, \\ \ln \ln N, & 2 < \gamma < 3, \\ \frac{\ln N}{\ln \ln N}, & \gamma = 3, \\ \ln N, & \gamma > 3. \end{cases}$$

Case $\gamma = 2$: hub-and-spoke

Recall: $\mathbb{E}[d] = \mathrm{const}$

- Heuristic: $k_{\text{max}} \sim N$, i.e. the largest hub connects to a linear fraction of all nodes.
- The network becomes star-like: most nodes connect to the same hub.
- The average distance $\mathbb{E}[d]$ stays bounded (independent of N).

Case $2 < \gamma < 3$: ultra-small

- The average distance grows as $\mathbb{E}[d] \sim \ln \ln N$, much smaller than the $\ln N$ typical of Erdős–Rényi graphs.
- Hubs act as *shortcuts*, drastically reducing path lengths.
- Most real-world scale-free networks are believed to lie in this regime.

Case $\gamma = 3$: critical point

- Distances grow as $\mathbb{E}[d] \sim \frac{\ln N}{\ln \ln N}$.
- The second moment $\mathbb{E}[D^2]$ of the degree distribution stops diverging.
- A logarithmic correction shortens paths compared to Erdős–Rényi.

Case $\gamma > 3$: small world

- Variance $\mathrm{Var}(D)$ is finite.
- The average distance recovers the Erdős–Rényi scaling: $\mathbb{E}[d] \sim \ln N$.
- Hubs exist but are not large enough to alter the global scaling.

When $\gamma < 2$

- Then $1/(\gamma - 1) > 1$, so the estimate

$$k_{\max} \sim N^{1/(\gamma-1)}$$

  grows faster than N.
- This would imply a hub degree larger than the total number of nodes — impossible without multiple edges.
- Moreover, $\mathbb{E}[D]$ diverges.
- $\Rightarrow$ infinite simple networks with $\gamma < 2$ cannot exist.

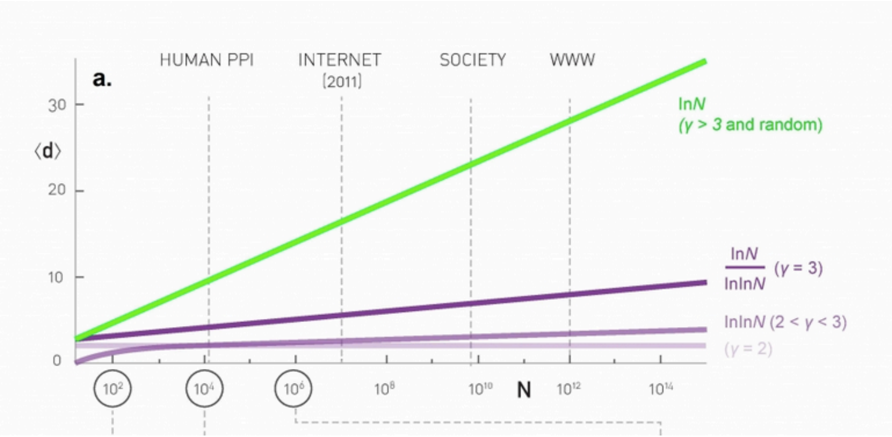
Conclusions

In summary, the effects on the distances in SFN are:

- Shrinks the average path lengths. Most scale-free networks of practical interest are "ultra-small". This is a consequence of the hubs, that act as bridges between many small degree nodes.

- Changes the dependence of $\langle d \rangle$ on the system size. The smaller is γ, the shorter are the distances between the nodes.

- Only for $\gamma > 3$ we recover the $\ln N$ dependence, the signature of the small-world property characterizing random networks.



Note that for large networks the difference in average degrees between the four regimes is much larger than for small networks.

Summary and Transition

- Erdős–Rényi explains short path lengths, but fails to capture:
  - high clustering (triangles),
  - heavy-tailed degree distributions.
- Real networks combine both properties — they are clustered *and* scale-free.
- Next: we explore richer models that explain how such networks emerge.