Power Laws in the Wild

Céline Comte and Fabien Mathieu

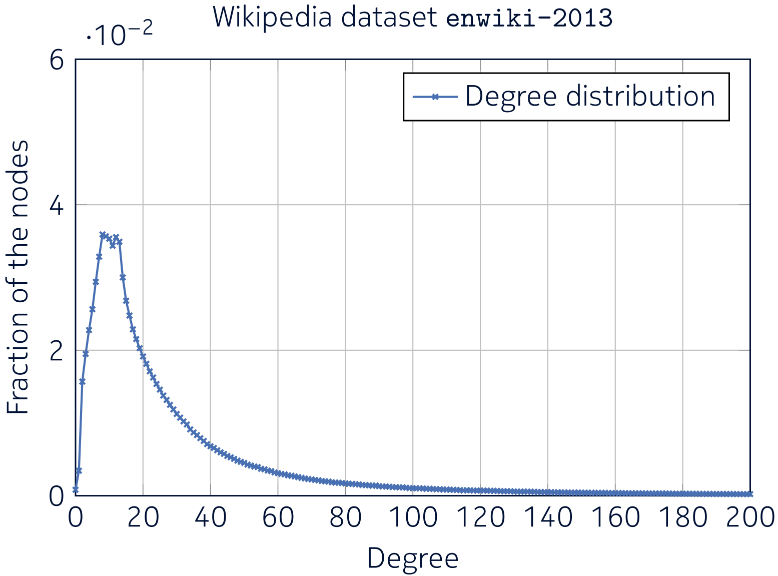
NOKIA Bell Labs

TELECOM ParisTech

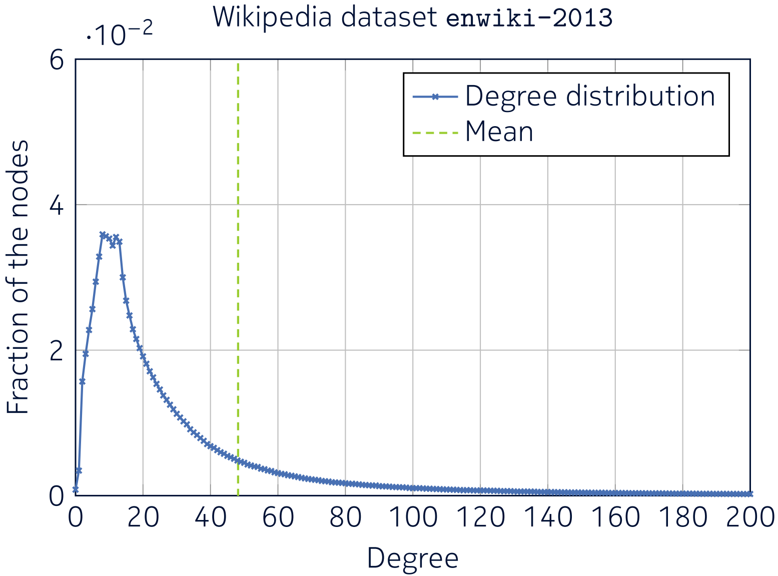
ACN – Propagation in Graphs

November 12, 2018

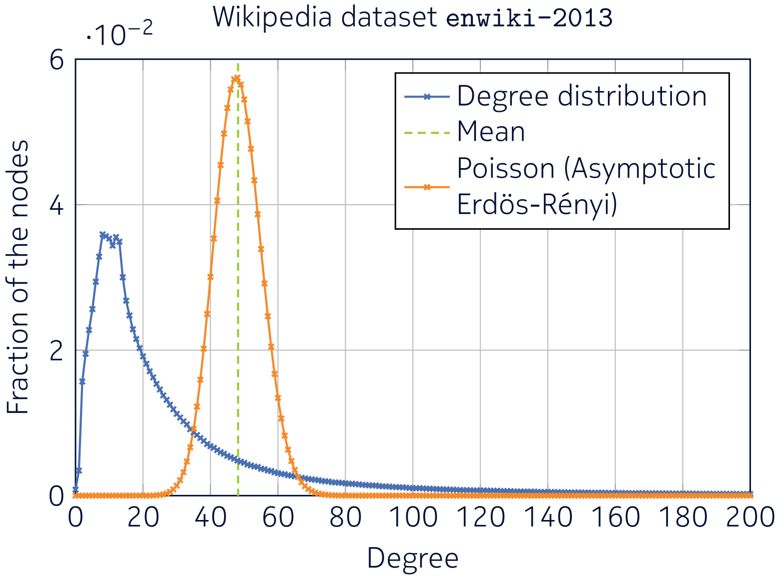
Introduction



Introduction



Introduction



  Public

Introduction

Observations

- The mean degree is surprisingly high compared to the region where the mass of the distribution seems to be
- The degree distribution is scattered (dispersée) over a large range of values. By comparison, the Poisson distribution is concentrated around its mean.

Objectives

- Characterize **heavy-tailed** degree distributions
- Understand how this distribution emerges
  → the **Albert-Barabási** generative model

Outline

Discrete power laws

Continuous power laws

Albert-Barabási generative graph model

Outline

Discrete power laws

Continuous power laws

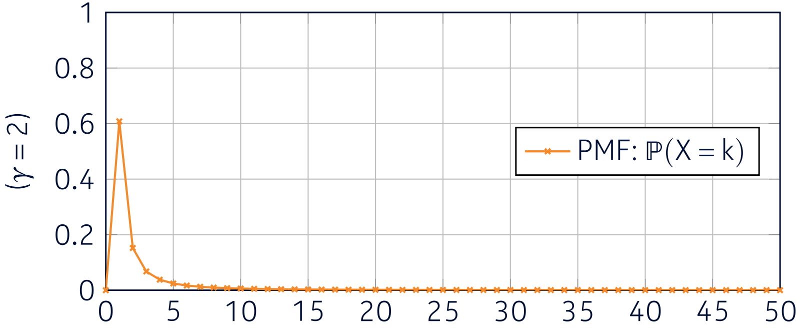
Albert-Barabási generative graph model

Power law distribution

A random variable X has a **power law** distribution if

$$\mathbb{P}(X = k) \sim \frac{P}{k^{\gamma}} \text{ as } k \to +\infty,$$

for some real $\gamma > 1$.

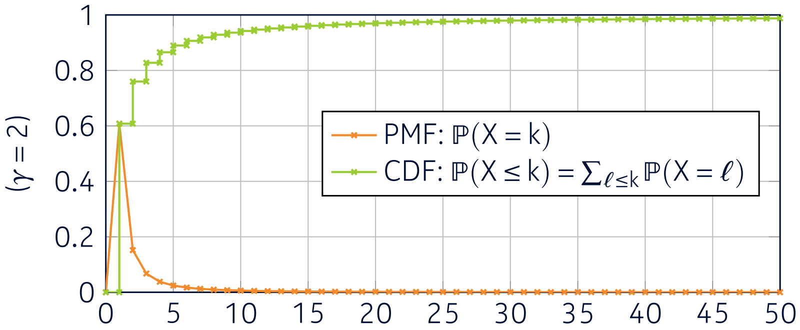


Power law distribution

A random variable X has a **power law** distribution if

$$\mathbb{P}(X = k) \sim \frac{P}{k^\gamma} \text{ as } k \to +\infty,$$

for some real $\gamma > 1$.

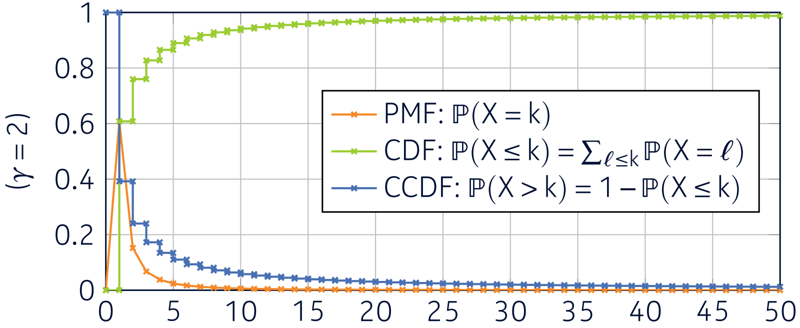


Power law distribution

A random variable X has a **power law** distribution if

$$\mathbb{P}(X = k) \sim \frac{P}{k^{\gamma}} \text{ as } k \to +\infty,$$

for some real $\gamma > 1$.



Public

NOKIA Bell Labs TELECOM ParisTech

Interpretation

Typically, X ~ **measure of popularity** of individuals

- Node degree in a graph
- Frequency of occurrence of a word in a text
- Number of file downloads

$\mathbb{P}(X = k)$ ~ fraction of the individuals with popularity k

Heavy-tailed property

- Most of the individuals have a very small popularity
- A few individuals have an outstanding popularity

In practice: Zipf law

- A random variable X has a **Zipf law** if

$$\mathbb{P}(X = k) = \frac{\frac{1}{k^\gamma}}{\sum_{\ell=1}^{N} \frac{1}{\ell^\gamma}}, \quad \forall k = 1, \ldots, N,$$

  γ is a non-negative real and N is a positive integer.

- It is not a problem if the power-law behavior is not observed for the smallest values. The most important is the **tail** of the distribution.

- Unbounded version: Zeta distribution.

Examples (Newman, 2005)

- Node degree in some real-life graphs
- Town sizes (in number of individuals)
- Word frequency
- Citations of scientific papers
- Web hits
- Number of emails received per user and per day
- Magnitude of earthquakes
- Diameter of moon craters
- Intensity of solar flares
- Wealth of the richest people
- ...

Outline

Discrete power laws

Continuous power laws

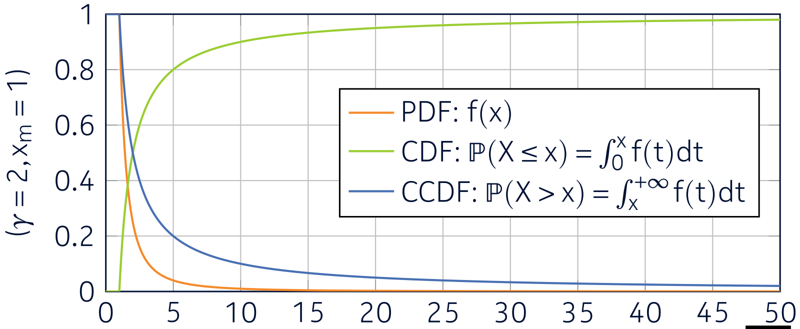
Albert-Barabási generative graph model

Pareto distribution

A continuous random variable X with values in $[x_m, +\infty[$ has a Pareto distribution if its PDF is

$$f(x) = \frac{\gamma - 1}{x_m}\left(\frac{x_m}{x}\right)^{\gamma} \propto \frac{1}{x^{\gamma}}, \quad \forall x \geq x_m,$$

$x_m > 0$ is the **scale** and $\gamma > 1$ is the **exponent**



Pareto distribution

A continuous random variable X with values in $[x_m, +\infty[$ has a Pareto distribution if its PDF is

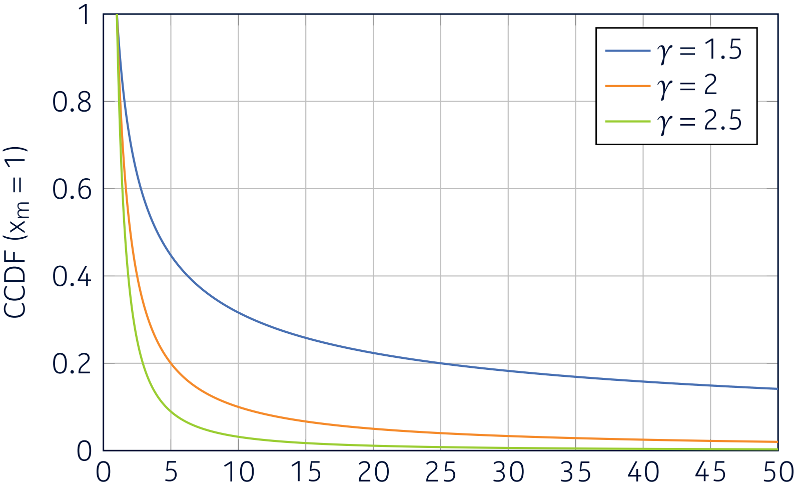
$$f(x) = \frac{\gamma - 1}{x_m}\left(\frac{x_m}{x}\right)^{\gamma} \propto \frac{1}{x^{\gamma}}, \quad \forall x \geq x_m,$$

$x_m > 0$ is the **scale** and $\gamma > 1$ is the **exponent**

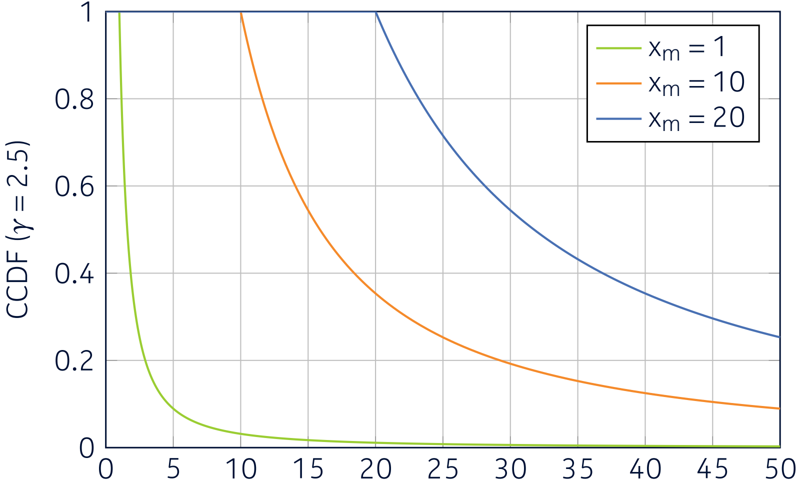
Its CCDF is given by:

$$\begin{aligned}\mathbb{P}(X > x) &= \int_x^{+\infty} f(t)dt = (\gamma - 1){x_m}^{\gamma-1} \int_x^{+\infty} \frac{1}{t^{\gamma}}dt \\ &= (\gamma - 1){x_m}^{\gamma-1}\left[\frac{t^{1-\gamma}}{1-\gamma}\right]_x^{+\infty} \\ &= \left(\frac{x_m}{x}\right)^{\gamma-1} \propto \frac{1}{x^{\gamma-1}}\end{aligned}$$

Exponent parameter γ



Scale parameter x_m

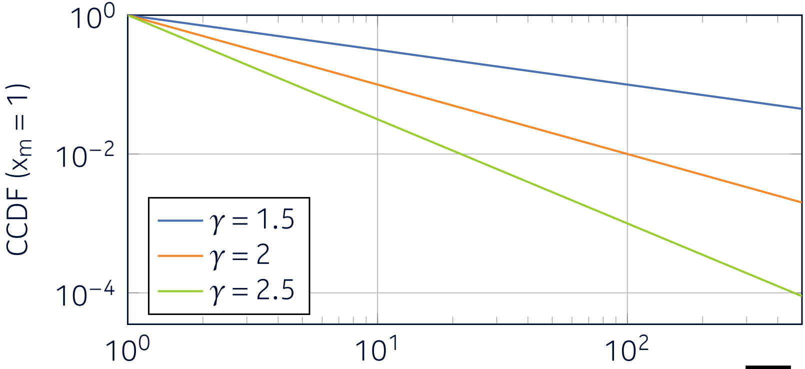


Plot in a log-log scale

$\mathbb{P}(X > x) = \left(\frac{x_m}{x}\right)^{\gamma-1}$ means that

$$\log(\mathbb{P}(X > x)) = \log\left(x_m{}^{\gamma-1}\right) - (\gamma - 1)\log(x)$$

→ Line of slope $-(\gamma - 1)$ in a log-log scale

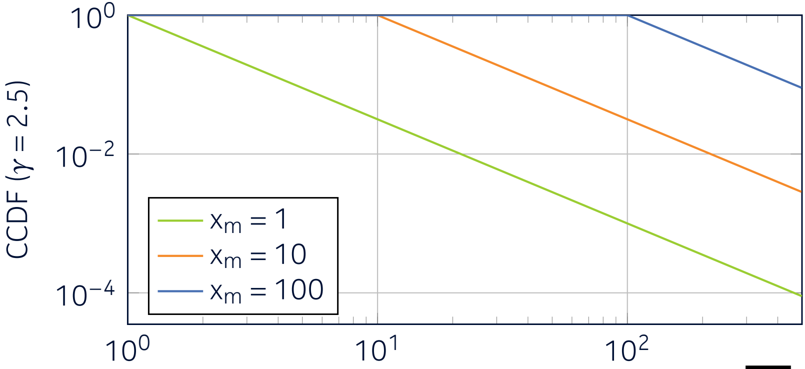


Plot in a log-log scale

$\mathbb{P}(X > x) = \left(\frac{x_m}{x}\right)^{\gamma-1}$ means that

$$\log(\mathbb{P}(X > x)) = \log\left(x_m{}^{\gamma-1}\right) - (\gamma - 1)\log(x)$$

→ Line of slope $-(\gamma - 1)$ in a log-log scale



Other quantities

- PDF: $f(x) = \frac{\gamma - 1}{x_m} \left(\frac{x_m}{x}\right)^{\gamma}, \quad \forall x \geq x_m$

  $\rightarrow$ Line of slope $-\gamma$ in a log-log scale

- CCDF: $\mathbb{P}(X > x) = \left(\frac{x_m}{x}\right)^{\gamma - 1}, \quad \forall x \geq x_m$

  $\rightarrow$ Line of slope $-(\gamma - 1)$ in a log-log scale

- CDF: $\mathbb{P}(X \leq x) = 1 - \left(\frac{x_m}{x}\right)^{\gamma - 1}, \quad \forall x \geq x_m$

- Mean: $\mathbb{E}(X) = \begin{cases} \frac{(\gamma - 1){x_m}^{\gamma - 1}}{\alpha - 2} & \text{if } \gamma > 2 \\ +\infty & \text{otherwise} \end{cases}$

Scale-free properties

- **Scale-free property**

$$\mathbb{P}(\theta X > x) = \mathbb{P}\left(X > \frac{x}{\theta}\right) = \left(\frac{\theta x_m}{x}\right)^{\gamma-1}$$

→ Pareto distribution with scale θx_m and exponent γ

- **Conditional distribution**

$$\mathbb{P}(X > x \mid X > t) = \frac{\mathbb{P}(X > x)}{\mathbb{P}(X > t)} = \frac{\left(\frac{x_m}{x}\right)^{\gamma-1}}{\left(\frac{x_m}{t}\right)^{\gamma-1}} = \left(\frac{t}{x}\right)^{\gamma-1}$$

→ Pareto distribution with scale t and exponent γ

Outline

Discrete power laws

Continuous power laws

Albert-Barabási generative graph model

NOKIA Bell Labs

TELECOM ParisTech

Undirected graph model (1^{st} version)

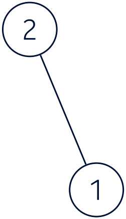
Generative model

- We start with an arbitrary undirected graph that contains at least one edge.
- **Expansion**: We add one new node at each step.
- **Preferential attachment**: The new node is attached to an existing node chosen at random with a probability that is proportional to its degree.

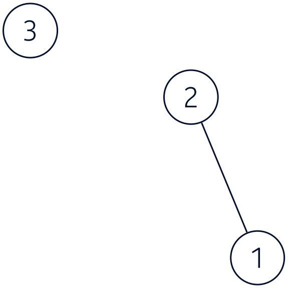
Notations

- $X_k(n)$ = number of degree-k nodes in the n-node graph
- $P_k(n) = \frac{X_k(n)}{n}$ = fraction of the nodes that have degree k

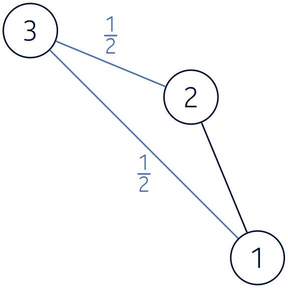
An example



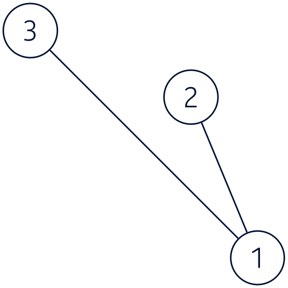
An example



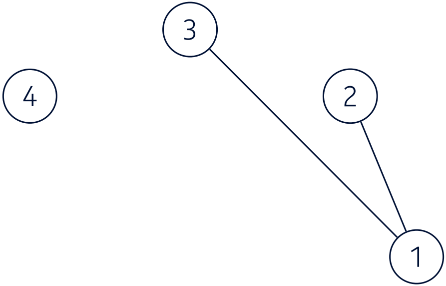
An example



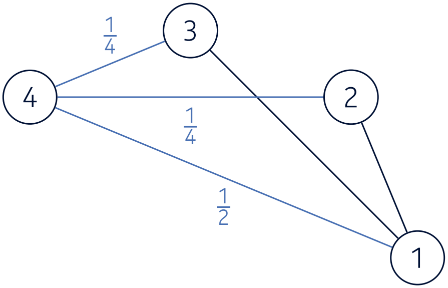
An example



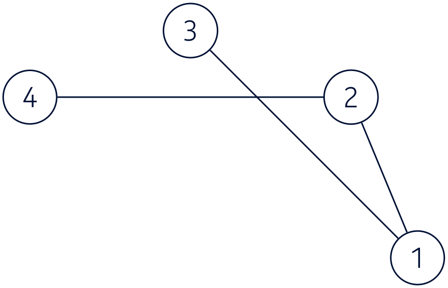
An example



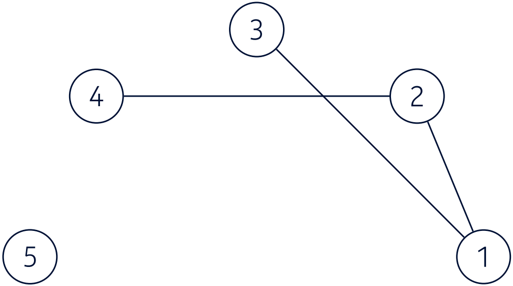
An example



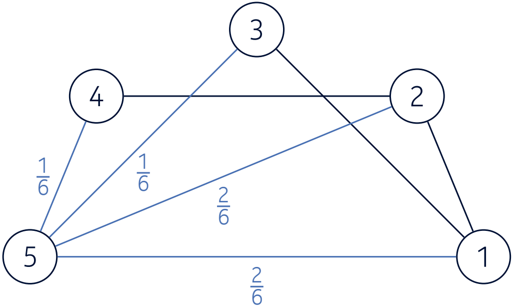
An example



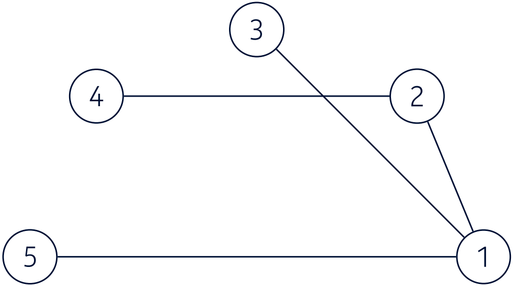
An example



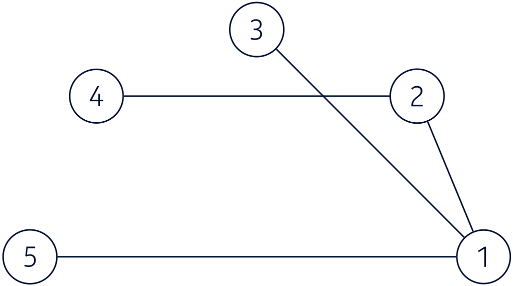
An example



An example

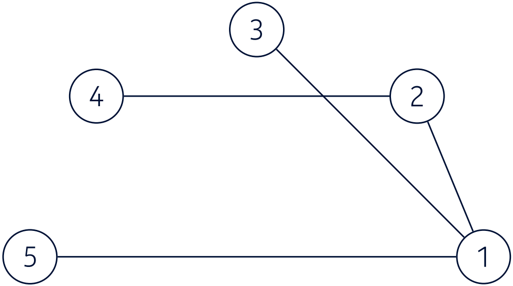


An example



…

An example



...

We have $n = 5$ nodes

$X_1(5) = 3,\ X_2(5) = 1,\ X_3(5) = 1$

$P_1(5) = \frac{3}{5},\ P_2(5) = \frac{1}{5},\ P_3(5) = \frac{1}{5}$

Remarks

Rich-get-richer phenomenon

- “Because of the preferential attachment, a vertex that acquires more connections than another will increase its connectivity at a higher rate”. (Barabási and Albert, 1999)

The obtained graph is a tree. There exist extensions of the Albert-Barabási model in which it is not the case.

Asymptotic results

- **“Law of large numbers”**: For each $k \geq 1$, we have

$$P_k(n) \to P_k \text{ almost surely as } n \to +\infty,$$

  where the sequence P_k is defined recursively by

$$\begin{cases} P_1 = \frac{2}{3}, \\ P_k = \frac{k-1}{k+2} P_{k-1}, \quad \forall k \geq 2. \end{cases}$$

- **Heavy-tailed distribution**:

$$P_k \sim \frac{4}{k^3} \text{ as } k \to +\infty.$$

Intuition of the proof

- Probability of choosing a given degree-k node $\approx \frac{k}{2n}$.

- We first consider $k \geq 2$.
- Variation when we add the $n+1^{th}$ node:

$$X_k(n+1) - X_k(n) \approx \frac{k-1}{2n} X_{k-1}(n) - \frac{k}{2n} X_k(n).$$

Since $X_k(n) = nP_k(n)$, we obtain

$$(n+1)P_k(n+1) - nP_k(n) \approx \frac{k-1}{2} P_{k-1}(n) - \frac{k}{2} P_k(n).$$

Intuition of the proof

- Variation when we add the $n+1^{th}$ node:

$$n\left(P_k(n+1) - P_k(n)\right) \approx \frac{(k-1)}{2} P_{k-1}(n) - \frac{k}{2} P_k(n) - P_k(n).$$

- Assuming that $P_k(n)$ has a limit P_k as $n \to +\infty$,

$$0 \approx \frac{(k-1)}{2} P_{k-1} - \frac{k}{2} P_k - P_k,$$

that is,

$$\left(\frac{k}{2} + 1\right) P_k \approx \frac{k-1}{2} P_{k-1},$$

that is,

$$P_k \approx \frac{k-1}{k+2} P_{k-1}.$$

Intuition of the proof

- We now focus on $k = 1$.
- Variation when we add the $(n+1)^{th}$ node:

$$X_1(n+1) - X_1(n) \approx 1 - \frac{1}{2n} X_1(n).$$

Since $X_1(n) = nP_1(n)$, we obtain

$$(n+1)P_1(n+1) - nP_1(n) \approx 1 - \frac{1}{2} P_1(n),$$

that is,

$$n(P_1(n+1) - P_1(n)) \approx 1 - \frac{1}{2} P_1(n) - P_1(n).$$

Intuition of the proof

- Variation when we add the $(n+1)^{th}$ node:

$$n\left(P_1(n+1) - P_1(n)\right) \approx 1 - \frac{1}{2}P_1(n) - P_1(n).$$

- Assuming that $P_1(n)$ has a limit P_1 as $n \to +\infty$,

$$0 \approx 1 - \frac{3}{2}P_1,$$

that is,

$$P_1 \approx \frac{2}{3}.$$

Intuition of the proof

- We have shown that

$$\begin{cases} P_1 = \frac{2}{3}, \\ P_k = \left(1 - \frac{3}{k+2}\right) P_{k-1}, \quad \forall k \geq 2. \end{cases}$$

- By expanding the recursion, we obtain, for each $k \geq 2$,

$$P_k = \frac{2}{3} \prod_{\ell=2}^{k} \frac{\ell - 1}{\ell + 2} = \frac{2}{3} \frac{\prod_{\ell=2}^{k} (\ell - 1)}{\prod_{\ell=2}^{k} (\ell + 2)},$$

$$= \frac{2}{3} \frac{\prod_{\ell=1}^{k-1} \ell}{\prod_{\ell=4}^{k+2} \ell} = \frac{2}{3} \frac{1 \times 2 \times 3}{k \times (k+1) \times (k+2)},$$

$$\sim \frac{4}{k^3} \text{ as } k \to +\infty. \quad \square$$

Undirected graph model (2nd version)

Generative model

- We start with an arbitrary undirected graph that contains at least one edge.
- **Expansion**: We add one new node at each step.
- The connecting node is chosen as follows:
  - With probability α, **uniform attachment**: All nodes are chosen with the same probability
  - With probability $1-\alpha$, **preferential attachment**: A node is chosen with a probability that is proportional to its degree

The 1st version of the model: $\alpha = 0$

Asymptotic results

- **"Law of large numbers"**: For each $k \geq 1$, we have

$$P_k(n) \to P_k \text{ almost surely as } n \to +\infty,$$

where the sequence P_k is defined recursively by

$$\begin{cases} P_1 = \frac{2}{3+\alpha}, \\ P_k = \frac{\alpha + \frac{1-\alpha}{2}(k-1)}{1+\alpha+\frac{1-\alpha}{2}k} P_{k-1}, \quad \forall k \geq 2. \end{cases}$$

- **Heavy-tailed distribution**:
$P_k \sim \frac{P}{k^\gamma}$ as $k \to +\infty$, where $P > 0$ is a constant and

$$\gamma = \frac{3-\alpha}{1-\alpha} = 1 + \frac{2}{1-\alpha} > 1.$$

References

- M. E. J. Newman (2005). "Power laws, Pareto distributions and Zipf's law". In: Contemporary Physics
  → An excellent overview
- Albert-László Barabási and Réka Albert (1999). "Emergence of Scaling in Random Networks". In: Science
  → Albert-Barabási model
- Moez Draief and Laurent Massoulié (2009). Epidemics and Rumours in Complex Networks. Cambridge University Press
  → Chapter 7 "Power Laws via preferential attachment"

