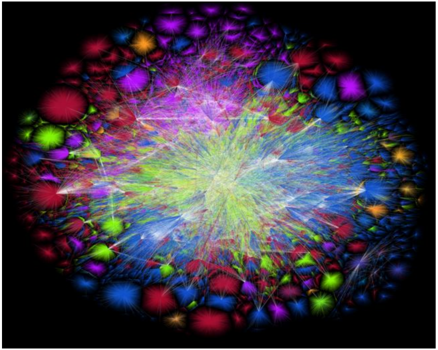
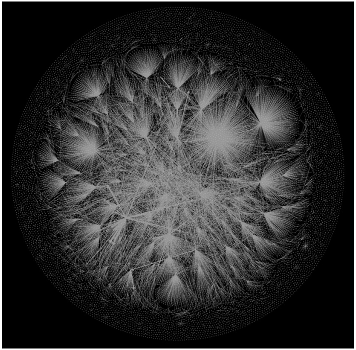
Complex Networks
Power laws

Nelly Litvak

Stochastic Operations Research & Data Science

Scale-free phenomenon





Left: Barrett Lyon / The Opte Project. Visualization of the routing paths of the Internet. 2015

Right: Tweets and retweets about Project X in Haren, 22-09-2012 5:00. Image by Marijn ten Thij

Scale-free graph sequences

- p_k is the fraction of nodes with degree k in a real-life network
- p_k is often roughly proportional to $k^{-\tau}$, $\tau \in (2, 4)$

Scale-free graph sequences

- p_k is the fraction of nodes with degree k in a real-life network
- p_k is often roughly proportional to $k^{-\tau}$, $\tau \in (2,4)$
- Plots in the log-log scale: $p_k \approx const \cdot k^{-\tau}$

$$\log(p_k) \approx -\tau \log(k) + const$$

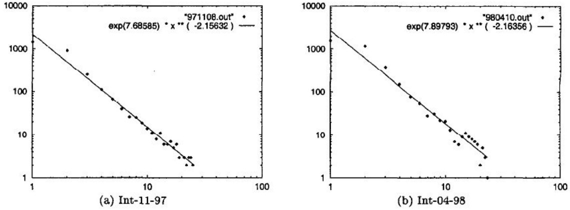
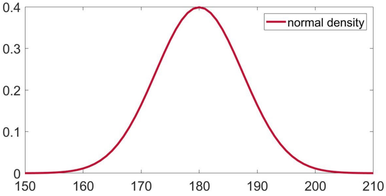
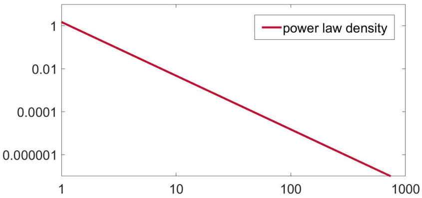


Figure 5: The outdegree plots: Log-log plot of frequency f_d versus the outdegree d.

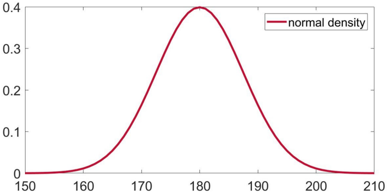
Faloutsos, Faloutsos, Faloutsos (1999): degrees of Internet routers and autonomous systems.

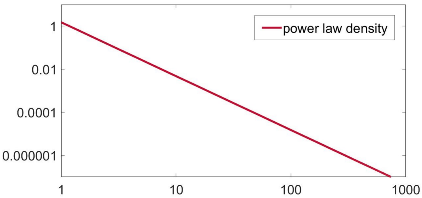
‘Non-normal’


normal density
0.4
0.3
0.2
0.1
0
150
160
170
180
190
200
210


power law density
1
0.01
0.0001
0.000001
1
10
100
1000

‘Non-normal’





Length of a man	EU Web crawl 2015 1 070 557 254 webpages
Average Dutch man: 180cm Tallest man on Earth: 250cm	Average in-degree: 85.743 Maximal in-degree: 20 252 239

P. Boldi, A. Marino, M. Santini, and S. Vigna. BUbiNG: Massive crawling for the masses. WWW 2014

Complementary cumulative distribution function

- X is a discrete random variable, $p_k = P(X = k)$, $k = 0, 1, \dots$
- $\bar{F}(k) = P(X \geq k) = \sum_{s=k}^{\infty} p_s$
- Power law (assume discrete version of Pareto distribution): $p_k = ck^{-\tau}$, $k \geq k_0$ (**)
- For any $x \geq k_0$

$$\int_k^{\infty} cx^{-\tau}\,dx \leq \sum_{s=k}^{\infty} cs^{-\tau} \leq \int_{k-1}^{\infty} cx^{-\tau}\,dx$$

$$\int_k^{\infty} cx^{-\tau}\,dx = -\frac{c}{\tau-1}x^{-\tau+1}\Big|_k^{\infty} = \frac{c}{\tau-1}k^{-\tau+1}$$

- $\bar{F}(k) \approx Ck^{-\tau+1}$ is less restrictive than (**). $\bar{F}(k)$ is used to represent the distribution of the data

Regular variation

$X \geq 0$ -regularly varying with exponent $\tau - 1$

$$P(X > x) = L(x)x^{-(\tau-1)}, \quad x \geq 0$$

$L(x)$ -Slowly varying function

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \quad \text{for all } t > 0$$

Regular variation

$X \geq 0$ -regularly varying with exponent $\tau - 1$

$$P(X > x) = L(x)x^{-(\tau-1)}, \quad x \geq 0$$

$L(x)$ -Slowly varying function

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \quad \text{for all } t > 0$$

Lemma. Let L be slowly varying function. Then for all $\epsilon > 0$ there exists $T > 0$ such that, for $x > T$

$$x^{-\epsilon} \leq L(x) \leq x^{\epsilon}.$$

Moments of scale-free distribution

- Let X has a discrete Pareto distribution:

$$P(X = k) = ck^{-\tau},\ k \geq k_0; \qquad c = \left[\sum_{k=k_0}^{\infty} k^{-\tau}\right]^{-1}$$

Moments of scale-free distribution

- Let X has a discrete Pareto distribution:

$$P(X=k)=ck^{-\tau},\ k\geq k_0; \qquad c=\left[\sum_{k=k_0}^{\infty} k^{-\tau}\right]^{-1}$$

$$E(X)=\sum_{k=k_0}^{\infty} k\cdot ck^{-\tau}=\sum_{k=k_0}^{\infty} ck^{-(\tau-1)}$$

Moments of scale-free distribution

- Let X has a discrete Pareto distribution:

$$P(X=k)=ck^{-\tau},\ k\geq k_0; \qquad c=\left[\sum_{k=k_0}^{\infty}k^{-\tau}\right]^{-1}$$

$$E(X)=\sum_{k=k_0}^{\infty}k\cdot ck^{-\tau}=\sum_{k=k_0}^{\infty}ck^{-(\tau-1)}$$

$E(X)<\infty$ only if $\tau>2$

Moments of scale-free distribution

- Let X has a discrete Pareto distribution:

$$P(X=k)=ck^{-\tau},\ k\geq k_0; \qquad c=\left[\sum_{k=k_0}^{\infty}k^{-\tau}\right]^{-1}$$

$$E(X)=\sum_{k=k_0}^{\infty}k\cdot ck^{-\tau}=\sum_{k=k_0}^{\infty}ck^{-(\tau-1)}$$

$E(X)<\infty$ only if $\tau>2$

$$E(X^2)=\sum_{k=k_0}^{\infty}k^2\cdot ck^{-\tau}=\sum_{k=k_0}^{\infty}ck^{-(\tau-2)}$$

$E(X^2)<\infty$ only if $\tau>3$

Moments of scale-free distribution

- Let X has a discrete Pareto distribution:

$$P(X=k)=ck^{-\tau},\ k\geq k_0; \qquad c=\left[\sum_{k=k_0}^{\infty}k^{-\tau}\right]^{-1}$$

$$E(X)=\sum_{k=k_0}^{\infty}k\cdot ck^{-\tau}=\sum_{k=k_0}^{\infty}ck^{-(\tau-1)}$$

$E(X)<\infty$ only if $\tau>2$

$$E(X^2)=\sum_{k=k_0}^{\infty}k^2\cdot ck^{-\tau}=\sum_{k=k_0}^{\infty}ck^{-(\tau-2)}$$

$E(X^2)<\infty$ only if $\tau>3$

Moments of scale-free distribution (continued)

- $E(X) < \infty$ only if $\tau > 2$
- $E(X^2) < \infty$ only if $\tau > 3$

Moments of scale-free distribution (continued)

- $E(X) < \infty$ only if $\tau > 2$
- $E(X^2) < \infty$ only if $\tau > 3$
- In particular,

$$Var(X) = E(X^2) - (E(X))^2 < \infty \text{ only if } \tau > 3$$

Moments of scale-free distribution (continued)

- $E(X) < \infty$ only if $\tau > 2$
- $E(X^2) < \infty$ only if $\tau > 3$
- In particular,

$$Var(X) = E(X^2) - (E(X))^2 < \infty \quad \text{only if } \tau > 3$$

- $\tau > 4$ is also an important 'nice' special case
- Many real networks have $\tau \in (2, 4)$

Maximal degree

- Let $X_1, X_2, \ldots, X_n$ be independent copies of X
- $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ – order statistics, see (2.6.1)

Maximal degree

- Let $X_1, X_2, \ldots, X_n$ be independent copies of X
- $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ – order statistics, see (2.6.1)
- $X_{(n)}$ - maximum of the n realizations. How large is $X_{(n)}$?

Maximal degree

- Let $X_1, X_2, \ldots, X_n$ be independent copies of X
- $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ – order statistics, see (2.6.1)
- $X_{(n)}$ - maximum of the n realizations. How large is $X_{(n)}$?
- Heuristic argument:

$$P(X > x_{max}) \approx \frac{1}{n}$$

Maximal degree

- Let $X_1, X_2, \ldots, X_n$ be independent copies of X
- $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ – order statistics, see (2.6.1)
- $X_{(n)}$ - maximum of the n realizations. How large is $X_{(n)}$?
- Heuristic argument:

$$P(X > x_{max}) \approx \frac{1}{n}$$

- Take Pareto distribution

Maximal degree

- Let $X_1, X_2, \ldots, X_n$ be independent copies of X
- $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ – order statistics, see (2.6.1)
- $X_{(n)}$ - maximum of the n realizations. How large is $X_{(n)}$?
- Heuristic argument:

$$P(X > x_{max}) \approx \frac{1}{n}$$

- Take Pareto distribution

$$C\, x_{max}^{-(\tau-1)} \approx \frac{1}{n}$$

Maximal degree

- Let $X_1, X_2, \ldots, X_n$ be independent copies of X
- $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ – order statistics, see (2.6.1)
- $X_{(n)}$ - maximum of the n realizations. How large is $X_{(n)}$?
- Heuristic argument:

$$P(X > x_{max}) \approx \frac{1}{n}$$

- Take Pareto distribution

$$C\, x_{max}^{-(\tau-1)} \approx \frac{1}{n}$$

$$x_{max} \approx C^{1/(\tau-1)}\, n^{1/(\tau-1)}$$

- The maximum is of the 'order of magnitude' $n^{1/(\tau-1)}$

Maximal degree

- Let $X_1, X_2, \ldots, X_n$ be independent copies of X
- $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ – order statistics, see (2.6.1)
- $X_{(n)}$ - maximum of the n realizations. How large is $X_{(n)}$?
- Heuristic argument:

$$P(X > x_{max}) \approx \frac{1}{n}$$

- Take Pareto distribution

$$C\, x_{max}^{-(\tau-1)} \approx \frac{1}{n}$$

$$x_{max} \approx C^{1/(\tau-1)}\, n^{1/(\tau-1)}$$

- The maximum is of the 'order of magnitude' $n^{1/(\tau-1)}$
- Compare to the maximum degree in the web crawl on slide 4

Extreme value theory

$$u_n = \sup\{u : 1 - F_X(u) \geq 1/n\}$$

Note: u_n is what we called 'x_{max}'

$$u_n = \sup\{u : 1 - F_X(u) \geq 1/n\}$$

Note: u_n is what we called 'x_{max}'

Theorem (alternative version of Theorem 2.33)

Let $(X_n))n \geq 0$ be a sequence of i.i.d. regularly varying random variables for some $\tau > 1$. Then for any fixed k

$$\frac{1}{u_n}(X_{(n+1-i)})_{i=1}^{k} \xrightarrow{d} (\xi_i)_{i=1}^{k},$$

where $\xi_i = \Gamma_i^{-1/(\tau-1)}, \quad \Gamma_i = E_1 + \cdots + E_i$, and $E_1, E_2, \ldots$ are independent exponential random variables with mean 1. In particular (Theorem 2.30), ξ_1 has a Fréchet distribution with parameter $\alpha = \tau - 1$

Hill's estimator

- How to estimate τ?

- How to estimate τ?
- Extreme Value Theory

Hill's estimator

- How to estimate τ?
- Extreme Value Theory
- Hill's estimator for $1/(\tau - 1)$

Hill's estimator

- How to estimate τ?
- Extreme Value Theory
- Hill's estimator for $1/(\tau - 1)$

$$H_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^{n} \log(X_{(i)}/X_{(n-k)})$$

Hill's estimator

- How to estimate τ?
- Extreme Value Theory
- Hill's estimator for $1/(\tau-1)$

$$H_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^{n} \log(X_{(i)}/X_{(n-k)})$$

- $k/n \to 0,\ k \to \infty$

Hill's estimator

- How to estimate τ?
- Extreme Value Theory
- Hill's estimator for $1/(\tau - 1)$

$$H_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^{n} \log(X_{(i)}/X_{(n-k)})$$

- $k/n \to 0, \; k \to \infty$
- How to choose k?

Exponential cut-off

- Sometimes the spread of the degrees is very large but it is hard to have extremely high degrees.

Exponential cut-off

- Sometimes the spread of the degrees is very large but it is hard to have extremely high degrees.
- Examples?

Exponential cut-off

- Sometimes the spread of the degrees is very large but it is hard to have extremely high degrees.
- Examples?
- Social networks, collaboration graphs

Exponential cut-off

- Sometimes the spread of the degrees is very large but it is hard to have extremely high degrees.
- Examples?
- Social networks, collaboration graphs
- Power law with exponential cut-off

$$p_k = ck^{-\tau} e^{-k/A}, \quad k \geq 1.$$

When A is very large then the exponent influences only the probabiltiy of the high values of k.