NATIONAL RESEARCH UNIVERSITY

Power law and scale-free networks

Network Science
Lecture 2

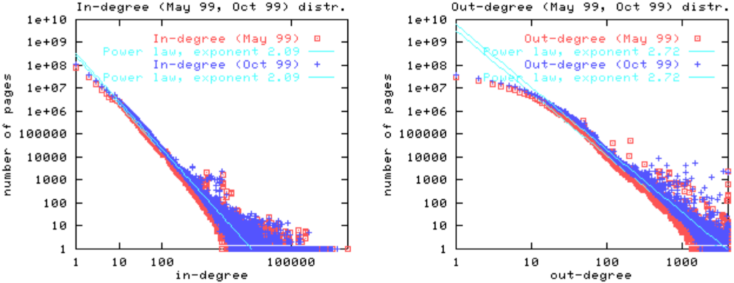
Leonid Zhukov

lzhukov@hse.ru

National Research University Higher School of Economics
School of Data Analysis and Artificial Intelligence, Department of Computer Science

January 17, 2022

Scale-free networks



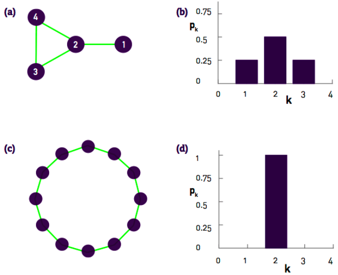
In- and out- degrees of WWW crawl 1999

Broder et.al, 1999

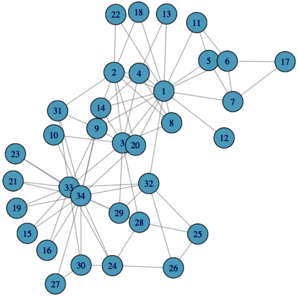
Node degree distribution

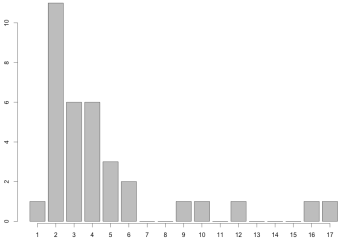
- k_i - node degree, i.e. number of nearest neighbors, $k_i = 1, 2, ...k_{\max}$
- n_k - number of nodes with degree k, $n_k = \sum_i \mathcal{I}(k_i == k)$
- total number of nodes $N = \sum_k n_k$
- Degree distribution is a fraction of the nodes with degree k

$$P(k_i = k) = P_k = \frac{n_k}{\sum_k n_k} = \frac{n_k}{N}$$

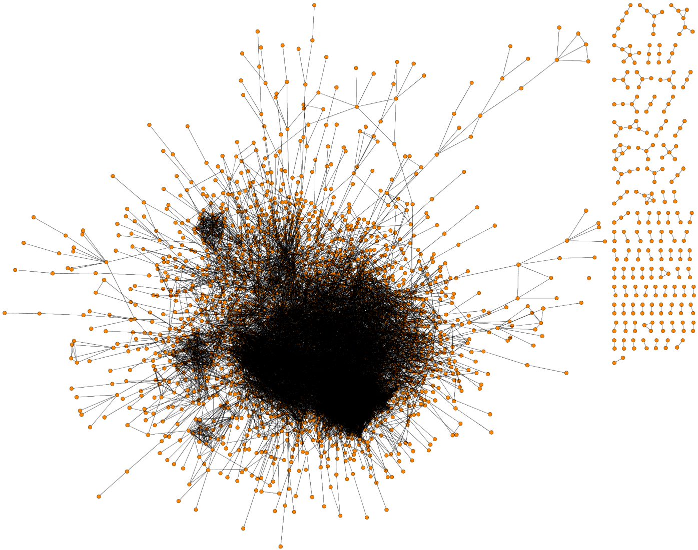


Node degree distribution

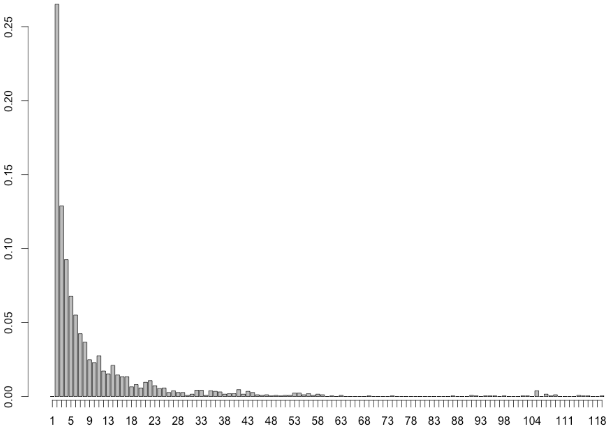




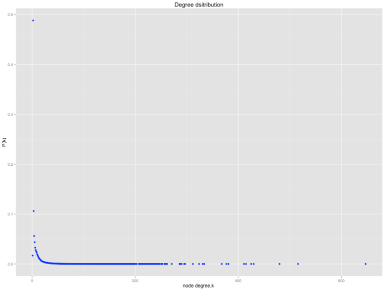
Degree distribution

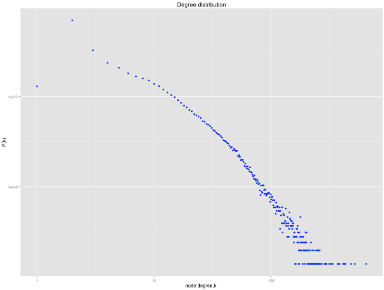


Degree distribution



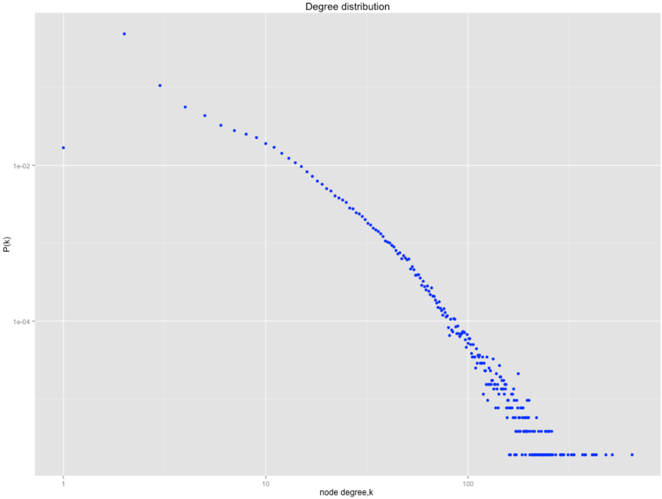
Power law degree distribution





$$P_k \sim k^{-\gamma} \qquad \log P_k \sim -\gamma \log(k)$$

Power law degree distribution



Discrete power law distribution

- Power law distribution, $k \in \mathbb{N}$, $\gamma \in \mathbb{R} > 0$

$$P_k = Ck^{-\gamma} = \frac{C}{k^\gamma}$$

- Log-log coordinates

$$\log P_k = -\gamma \log k + \log C$$

- Normalization

$$\sum_{k=1}^{\infty} P_k = C \sum_{k=1}^{\infty} k^{-\gamma} = C\zeta(\gamma) = 1; \;\; C = \frac{1}{\zeta(\gamma)}$$

- Riemann zeta function, $\gamma > 1$

$$P_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$$

Power law continuous approximation

- Power law, $k \in \mathbb{R}$, $\gamma \in \mathbb{R} > 0$

$$p(k) = Ck^{-\gamma} = \frac{C}{k^{\gamma}}, \quad \text{for } k \geq k_{min}$$

- Normalization ($\gamma > 1$)

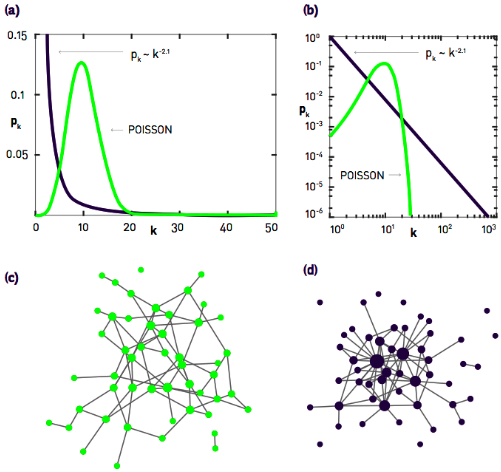
$$1 = \int_{k_{min}}^{\infty} p(k)dk = C \int_{k_{\min}}^{\infty} \frac{dk}{k^{\gamma}} = \frac{C}{\gamma - 1} k_{\min}^{-\gamma+1}$$

$$C = (\gamma - 1) k_{\min}^{\gamma - 1}$$

- Power law normalized PDF

$$p(k) = (\gamma - 1) k_{\min}^{\gamma - 1} k^{-\gamma} = \frac{\gamma - 1}{k_{min}} \left(\frac{k}{k_{min}} \right)^{-\gamma}$$

Power-law vs Poisson degree distribution



from A.-L., Barabasi, 2016

Hubs in networks

- How does the network size affect the size of its hubs(natural cutoff)?
- Probability of network having a node with degree $k > k_{\max}$:

$$Pr(k \geq k_{\max}) = \int_{k_{\max}}^{\infty} p(k)dk$$

- Expected number of nodes with degree $k \geq k_{\max}$:

$$N \cdot Pr(k \geq k_{\max}) = 1$$

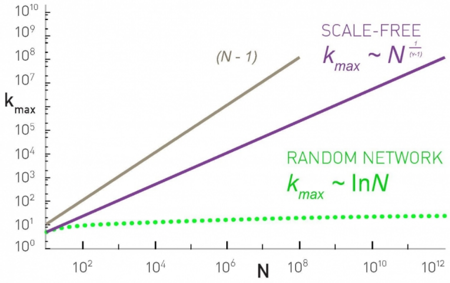
- Expected largest node degree in exponential network $p(k) = Ce^{-\lambda k}$

$$k_{max} = k_{min} + \frac{\ln N}{\lambda}$$

- Expected largest node degree in power law network $p(k) = Ck^{-\gamma}$

$$k_{max} = k_{min} N^{\frac{1}{\gamma - 1}}$$

Hubs in networks



from A.-L., Barabasi, 2016

Moments

- Power law PDF, $\gamma > 1$:

$$p(k) = \frac{C}{k^{\gamma}},\ k \geq k_{min}; C = (\gamma - 1)k_{\text{min}}^{\gamma-1}$$

- First moment (mean value), $\gamma > 2$:

$$\langle k \rangle = \int_{k_{\text{min}}}^{\infty} kp(k)dk = C\int_{k_{\text{min}}}^{\infty} \frac{dk}{k^{\gamma-1}} = \frac{\gamma-1}{\gamma-2}k_{\text{min}}$$

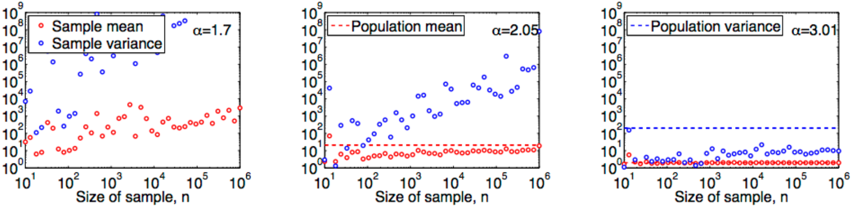
- Second moment, $\gamma > 3$:

$$\langle k^2 \rangle = \int_{k_{\text{min}}}^{\infty} k^2 p(k)dk = C\int_{k_{\text{min}}}^{\infty} \frac{dk}{k^{\gamma-2}} = \frac{\gamma-1}{\gamma-3}k_{\text{min}}^2$$

- m-th moment, $\gamma > m + 1$:

$$\langle k^m \rangle = \int_{k_{\text{min}}}^{k_{\text{max}}} k^m p(k)dk = C\frac{k_{\text{max}}^{m+1-\gamma} - k_{\text{min}}^{m+1-\gamma}}{m+1-\gamma}$$

Moments



$$\langle k \rangle = C\frac{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}{2-\gamma}, \quad \langle k^2 \rangle = C\frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{3-\gamma}$$

$$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2, \quad k_{max} = k_{min}\, N^{\frac{1}{\gamma-1}}$$

Clauset et.al, 2009

Scale free network

Degree of a randomly chosen node:

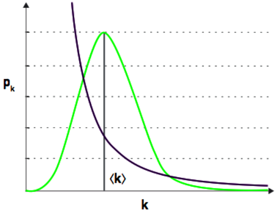
$$k = \langle k \rangle \pm \sigma_k, \quad \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

Poisson degree distribution (random network) has a scale $\langle k \rangle$:

$$k = \langle k \rangle \pm \sqrt{\langle k \rangle}$$

Power law network with $2 < \gamma < 3$ is scale free:

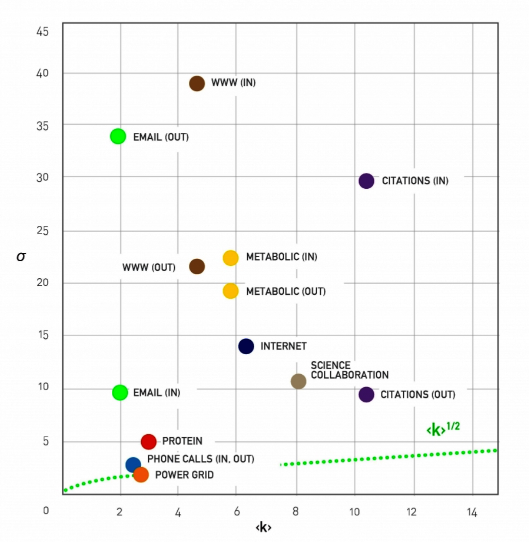
$$k = \langle k \rangle \pm \infty$$



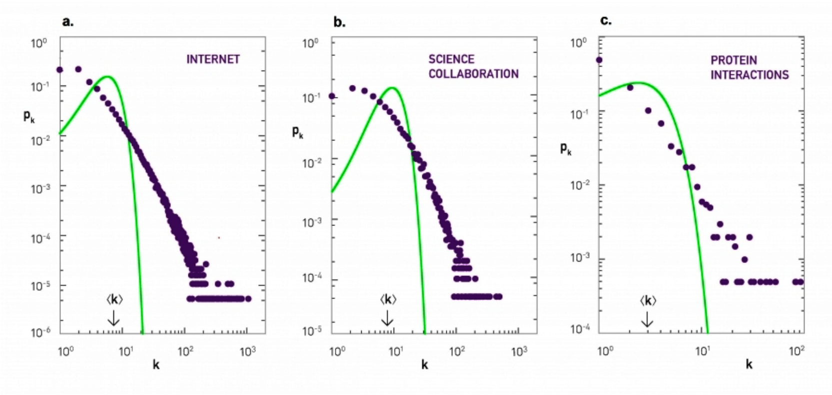
Degree fluctuation in real networks

Network	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	240.1	-	-	3.42*
WWW	325,729	1,497,134	4.60	1546.0	482.4	-	2.00	2.31	-
Power Grid	4,941	6,594	2.67	-	-	10.3	-	-	Exp.
Mobile-Phone Calls	36,595	91,826	2.51	12.0	11.7	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	94.7	1163.9	-	3.43*	2.03*	-
Science Collaboration	23,133	93,437	8.08	-	-	178.2	-	-	3.35*
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7	-	-	2.12*
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-	3.03*	4.00*	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-	2.43*	2.90*	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3	-	-	2.89*-

Degree fluctuation in real networks

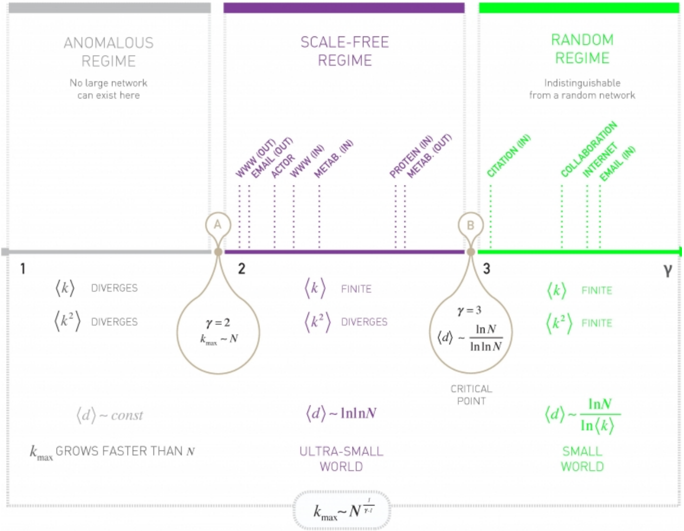


Scale-free networks

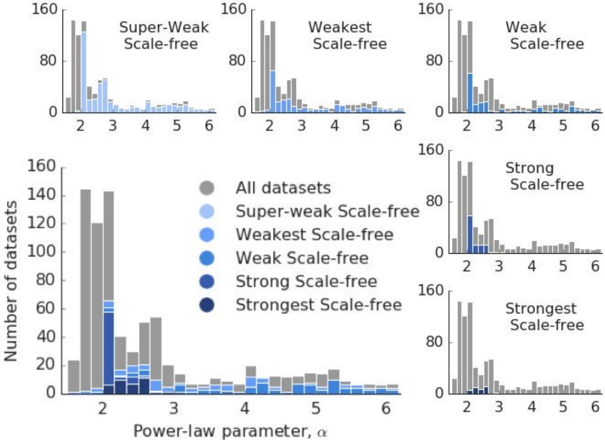


from A.-L. Barabasi, 2016

Properties of scale free networks



Scale free networks in real world



from A. Clauset, 2018

Plotting power Laws

- Power law PDF

$$p(k) = Ck^{-\gamma}; \quad \log p(k) = \log C - \gamma \log k$$

- Cumulative distribution function (CDF)

$$F(k) = Pr(k_i \leq k) = \int_0^k p(k)dk$$

- Complimentary cumulative distribution function cCDF

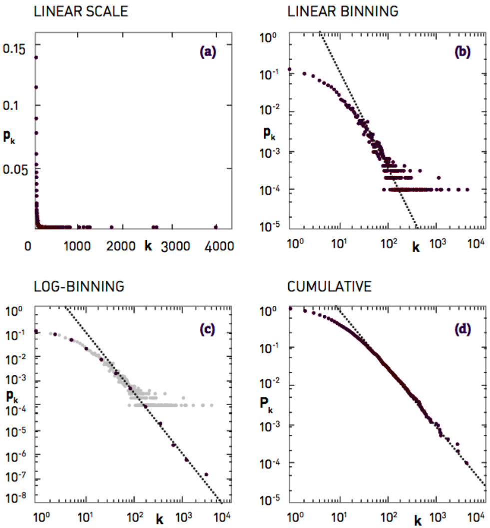
$$\bar{F}(k) = Pr(k_i > k) = 1 - F(k) = \int_k^\infty p(k)dk$$

- Power law cCDF

$$\bar{F}(k) = \frac{C}{\gamma - 1} k^{-(\gamma-1)}$$

$$\log \bar{F}(k) = \log \frac{C}{\gamma - 1} - (\gamma - 1) \log k$$

Plotting power laws



Parameter estimation: γ

Maximum likelihood estimation of parameter γ

- Let $\{k_i\}$ be a set of n observations (points) independently sampled from the distribution

$$P(k_i) = \frac{\gamma - 1}{k_{\text{min}}} \left(\frac{k_i}{k_{\text{min}}} \right)^{-\gamma}$$

- Probability of the sample

$$P(\{k_i\}|\gamma) = \prod_i^n \frac{\gamma - 1}{k_{\text{min}}} \left(\frac{k_i}{k_{\text{min}}} \right)^{-\gamma}$$

- Bayes' theorem

$$P(\gamma|\{k_i\}) = P(\{k_i\}|\gamma) \frac{P(\gamma)}{P(\{k_i\})}$$

Maximum likelihood

- log-likelihood

$$\mathcal{L} = \ln P(\gamma|\{k_i\}) = n \ln(\gamma - 1) - n \ln k_{\text{min}} - \gamma \sum_{i=1}^{n} \ln \frac{k_i}{k_{\text{min}}}$$

- maximization $\frac{\partial \mathcal{L}}{\partial \gamma} = 0$

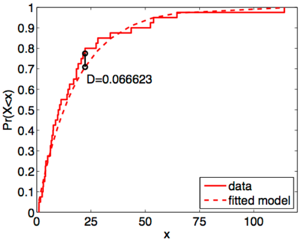
$$\gamma = 1 + n \left[\sum_{i=1}^{n} \ln \frac{k_i}{k_{\text{min}}} \right]^{-1}$$

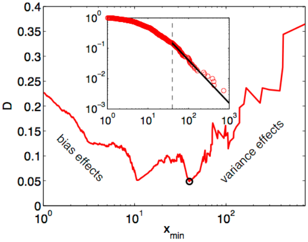
- error estimate

$$\sigma = \sqrt{n} \left[\sum_{i=1}^{n} \ln \frac{k_i}{k_{\text{min}}} \right]^{-1} = \frac{\gamma - 1}{\sqrt{n}}$$

- Kolmogorov-Smirnov test (compare model and experimental CDF)

$$D = \max_k |F(k|\gamma, k_{min}) - F_{exp}(k)|$$





- find

$$k^*_{min} = argmin_{k_{min}} D$$

Clauset et.al, 2009

References

- Power laws, Pareto distributions and Zipf's law, M. E. J. Newman, Contemporary Physics, pages 323–351, 2005.
- Power-Law Distribution in Empirical Data, A. Clauset, C.R. Shalizi, M.E.J. Newman, SIAM Review, Vol 51, No 4, pp. 661-703, 2009.
- A Brief History of Generative Models for Power Law and Lognormal Distributions, M. Mitzenmacher, Internet Mathematics Vol 1, No 2, pp 226-251.
- Scale-free networks are rare, Anna D. Broido and Aaron Clauset, Nature Communications 10, 1017 (2019).