Power law and scale-free networks

Leonid E. Zhukov

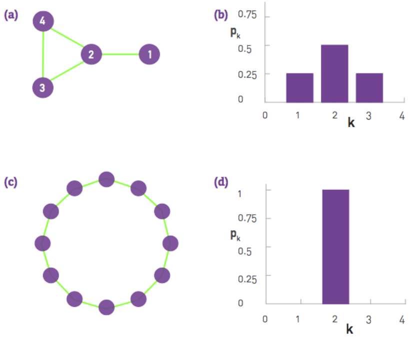
School of Data Analysis and Artificial Intelligence
Department of Computer Science
National Research University Higher School of Economics

Network Science

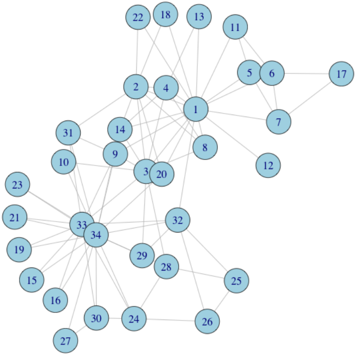
Node degree distribution

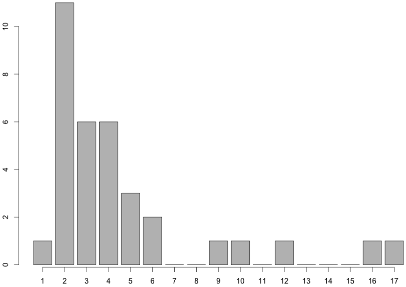
- k_i - node degree, i.e. number of nearest neighbors, $k_i = 1, 2, ...k_{\max}$
- n_k - number of nodes with degree k, $n_k = \sum_i \mathcal{I}(k_i == k)$
- total number of nodes $N = \sum_k n_k$
- Degree distribution is a fraction of the nodes with degree k

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

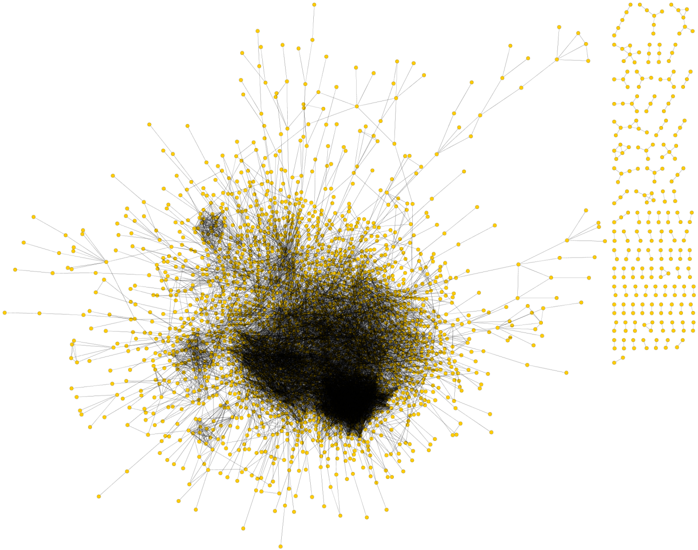


Node degree distribution

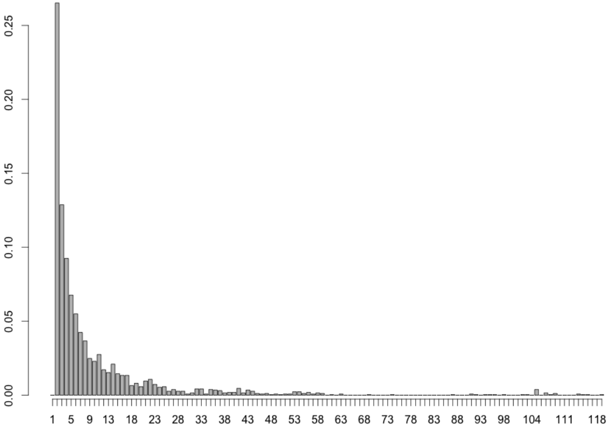




Degree distribution

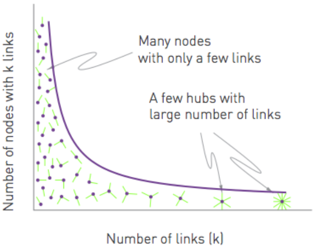


Degree distribution

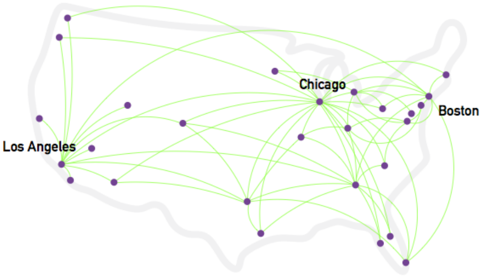


Power law degree distribution

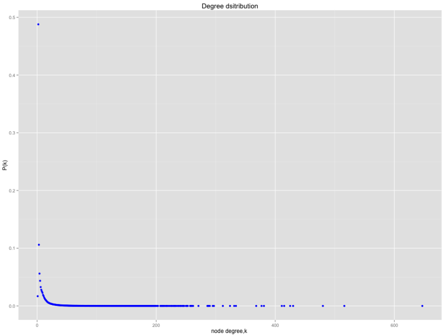
(c) POWER LAW

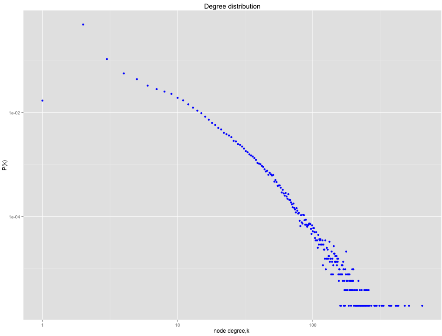


(d)



Power law degree distribution





Discrete power law distribution

- Power law distribution

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

- Log-log coordinates

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

- Normalization

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

- Riemann zeta function, $\gamma > 1$

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

Power law continuous approximation

- Power law

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

- Normalization ($\gamma > 1$)

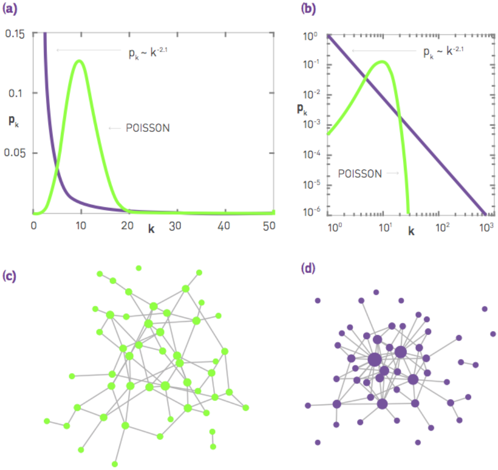
$$1 = \int_{k_{min}}^{\infty} p(k)dk = C\int_{k_{\text{min}}}^{\infty} \frac{dk}{k^{\gamma}} = \frac{C}{\gamma - 1} k_{\text{min}}^{-\gamma+1}$$

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

- Power law normalized PDF

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

Hubs in networks



from A.-L., Barabasi, 2016

Hubs in networks

- Expected number of nodes with degree $k > k_{\max}$: $N\ Pr(k > k_{\max})$
- Probability of observing a single node with degree $k > k_{\max}$:

$$Pr(k > k_{\max}) = \int_{k_{\max}}^{\infty} p(k) dk = \frac{1}{N}$$

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

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

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

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

Moments

- Power law PDF

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

- 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, $\gamma > m + 1$:

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

Scale free network

Degree of a randomly chosen node:

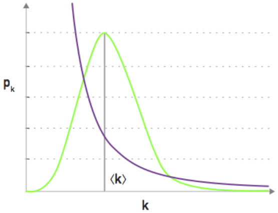
$$k = \langle k \rangle \pm \sigma_k, \quad \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

Poisson degree distribution:

$$k = \langle k \rangle \pm \langle k \rangle^2$$

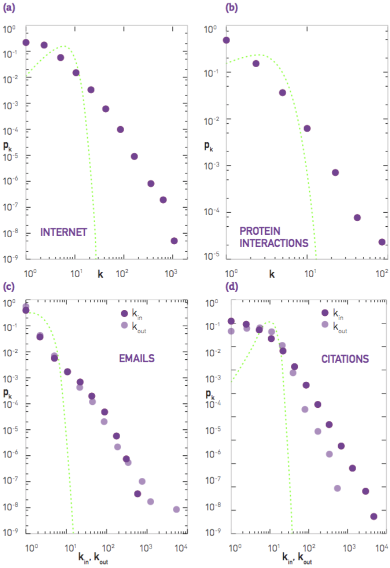
Power law network with $2 < \gamma < 3$

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

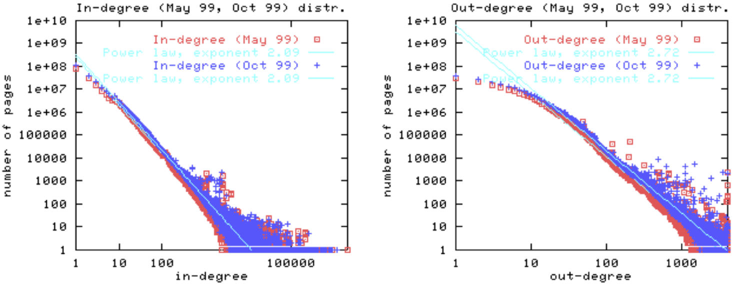


from A.-L., Barabasi, 2016

Scale-free networks



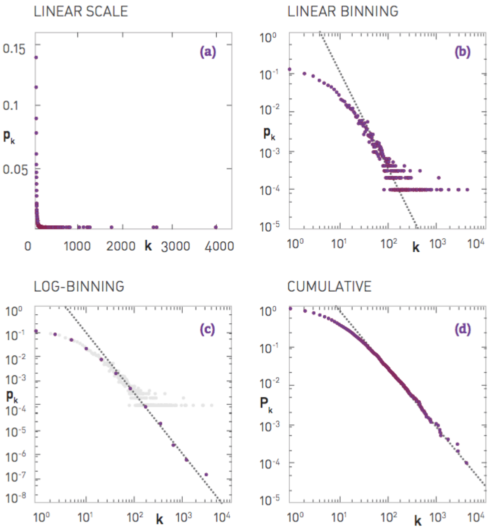
Scale-free networks



In- and out- degrees of WWW crawl 1999

Broder et.al, 1999

Plotting power laws



from A.-L., Barabasi, 2016

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

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

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

Rank-frequency plot

- Discrete complementary CDF

$$\bar{F}_k = \sum_{k' \geq k}^{N} P_{k'} = \frac{1}{N} \sum_{k' \geq k}^{N} n_{k'}$$

This is the number of vertices with degree greater or equal to k

- Sort the degrees of vertices in descending order and number them from 1 to N, these are ranks r_i
- Plot vertices ranks r_i/N as a function of degree k_i

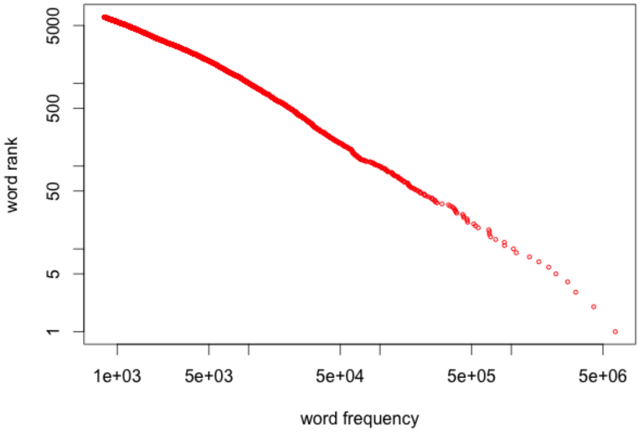
Word counting

Word frequency table (6318 unique words, min freq 800, corpus size $> 85 mln$):

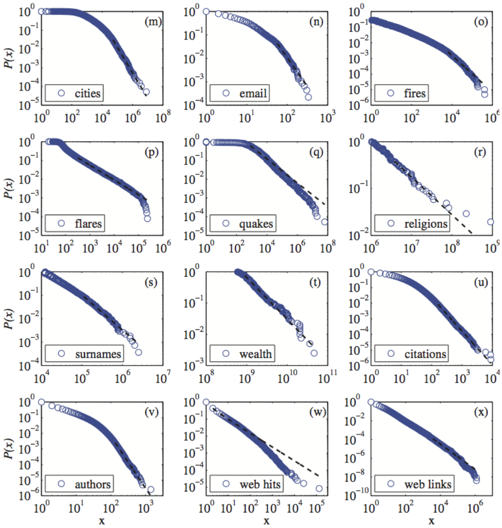
```
6187267 the
4239632 be
3093444 of
2687863 and
2186369 a
1924315 in
1620850 to
........
801 incredibly
801 historically
801 decision-making
800 wildly
800 reformer
800 quantum
```

Zips'f law

Zipf's law - the frequency of a word in an natural language corpus is inversely proportional to its rank in the frequency table $f(k) \sim 1/k$.



More empirical data



Clauset et.al, 2009

Parameter estimation: γ

Maximum likelihood estimation of parameter γ

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

$$P(x_i) = \frac{\gamma - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\gamma}$$

- Probability of the sample

$$P(\{x_i\}|\gamma) = \prod_i^n \frac{\gamma - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\gamma}$$

- Bayes' theorem

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

Maximum likelihood

- log-likelihood

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

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

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

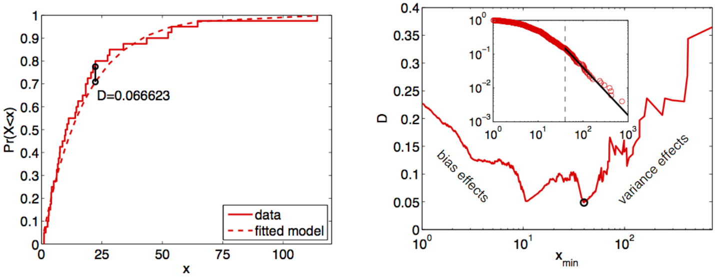
- error estimate

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

Parameter estimation: k_{min}

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

$$D = \max_x |F(x|\gamma, x_{min}) - F_{exp}(x)|$$



- find

$$x^*_{min} = argmin_{x_{min}} D$$

Clauset et.al, 2009

References

- Power laws, Pareto distributions and Zipfs law, M. E. J. Newman, Contemporary Physics, pages 323351, 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.