Power law and scale-free networks

Lecture 2

Natalia Semenova

Higher School of Economics

nasemenova_2@edu.hse.ru

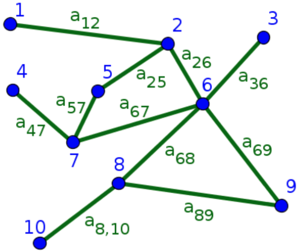
Structural Analysis and Visualization of Networks
23.01.2024

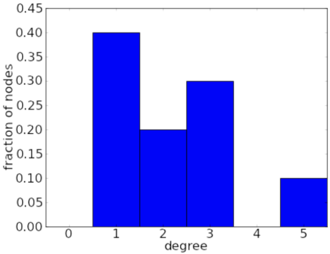
Lecture Overview

1. Power law
    - The degree distribution of a network
    - Hubs in networks

2. Scale-free networks
    - Small world: six degrees of separation
    - Hight clustering coefficient: transitivity

Undirected networks

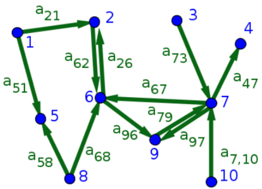


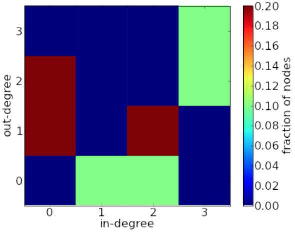


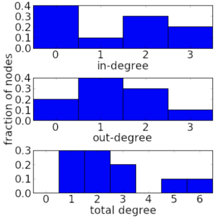
- k_i - node degree, i.e. number of nearest neighbors, $k_i = 1, 2, \ldots k_{\max}$
- n_k - number of nodes with degree k, $n_k = \sum_i \mathcal{I}(k_i == k)$
- total number of nodes $N = \sum_k n_k$
- Degree distribution is a fraction of the nodes with degree k

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

Directed networks

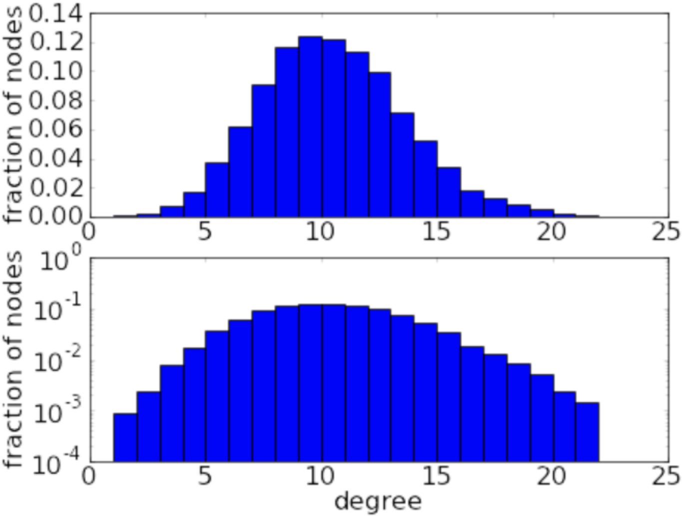




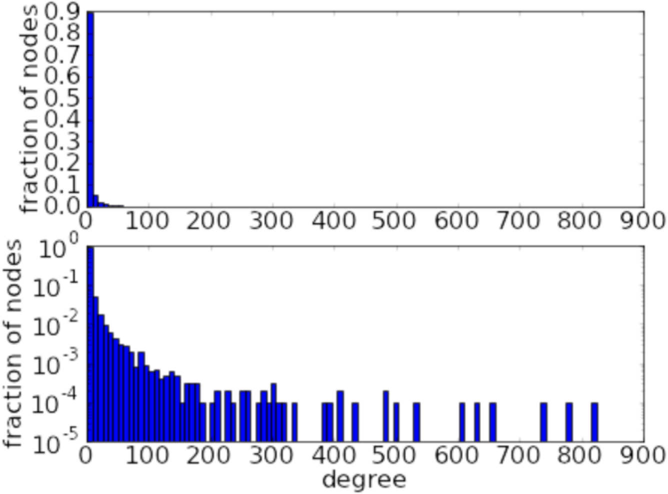


- $k_i^{tot} = k_i^{in} + k_i^{out}$; $k_i^{in} = \sum_{j=1} a_{ij}$; $k_i^{out} = \sum_{j=1} a_{ij}$
- $P_{deg}(k_{in}, k_{out})$ = the fraction of nodes in the graph with in-degree k_{in} and out-degree k_{out}

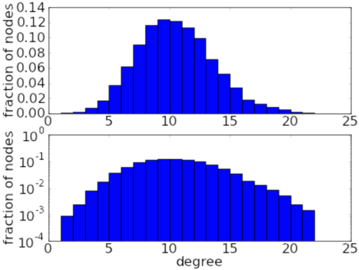
The degree distribution of a network: Expected

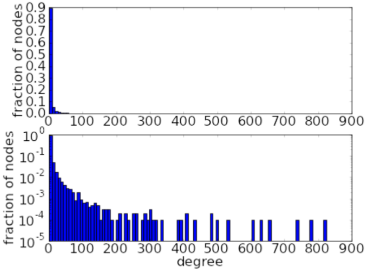


The degree distribution of a network: Real networks

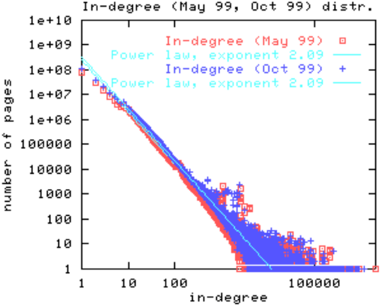


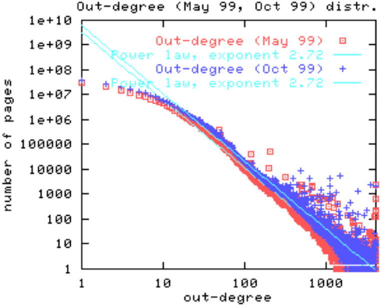
Expected distribution VS Real networks





Power law



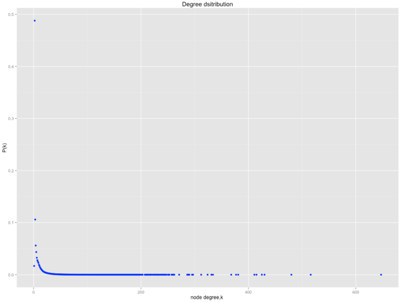


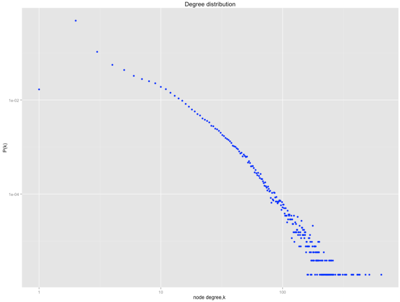
Scale-free networks In- and out- degrees of WWW crawl 1999
Broder et.al, 1999

Power law

- $P_{deg}(k) \propto k^{-\gamma}$, where γ is some exponent.
- Networks with power-law distributions a called **scale-free** because power laws have the same functional form at all scales. The power law $P_{deg}(k)$ remains unchanged (other than a multiplicative factor) when rescaling the independent variable k, as it satisfies $P_{deg}(ak) = a^{-\gamma} P_{deg}(k)$.

Power law





$P_{deg}(k) \propto k^{-\gamma}$

$\log P_{deg}(k) \propto -\gamma \log(k)$

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 $(\gamma > 1)$

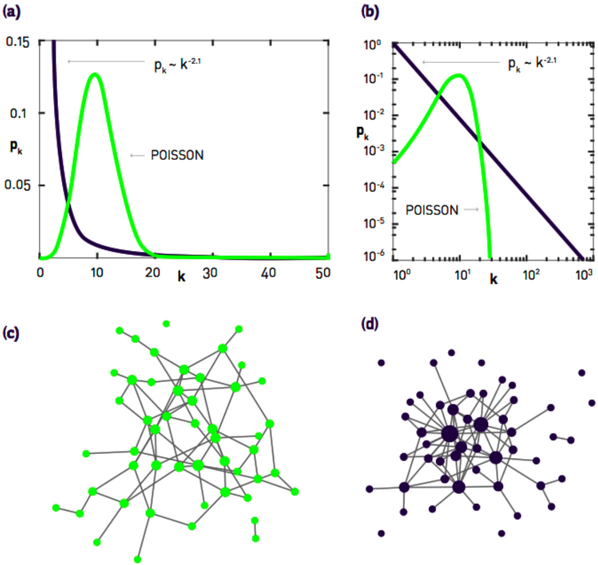
$$1 = \int_{k_{min}}^{\infty} p(k)dk = C \int_{k_{\text{min}}}^{\infty} \frac{dk}{k^{\gamma}} = \frac{C}{\gamma - 1} k_{\text{min}}^{-\gamma+1}$$

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

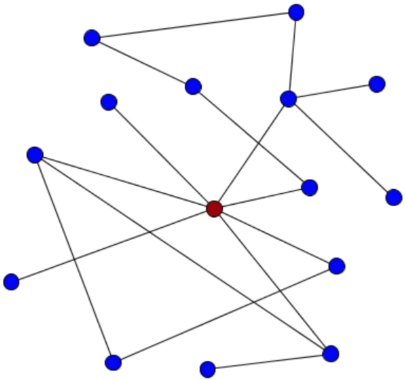
- Power law normalized PDF

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

Power law vs Poisson degree distribution



Hubs



The red node is an example of a hub.

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 $k \geq k_{\max}$:

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

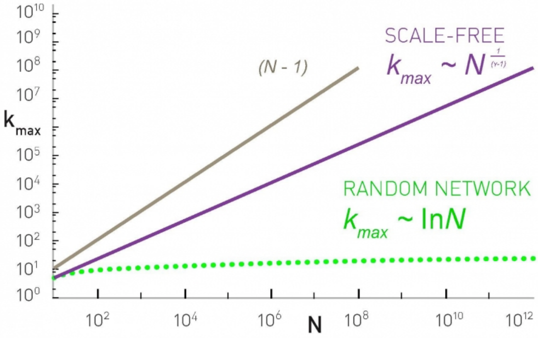
- Expected largest node degree in exponential network $p(k) = Ce^{-\lambda k}$

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

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

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

Hubs in networks



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), $\gamma > 2$:

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

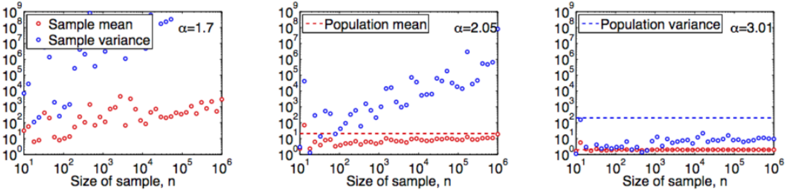
- Second moment, $\gamma > 3$:

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

- 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



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

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

Scale free network

Degree of a randomly chosen node:

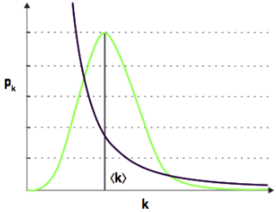
$$k = \langle k \rangle \pm \sigma_k, \quad \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$$

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:

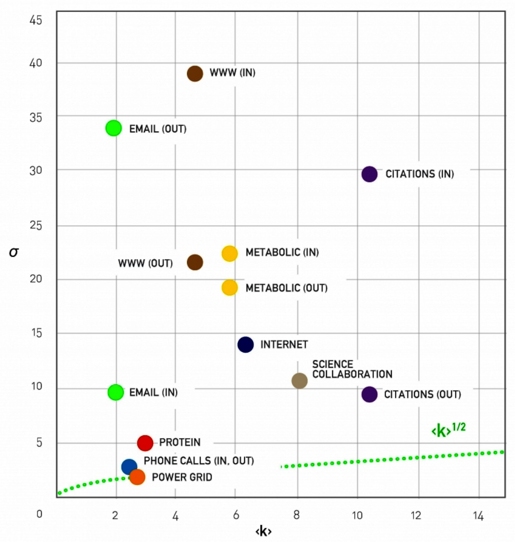
$$k = \langle k \rangle \pm \infty$$



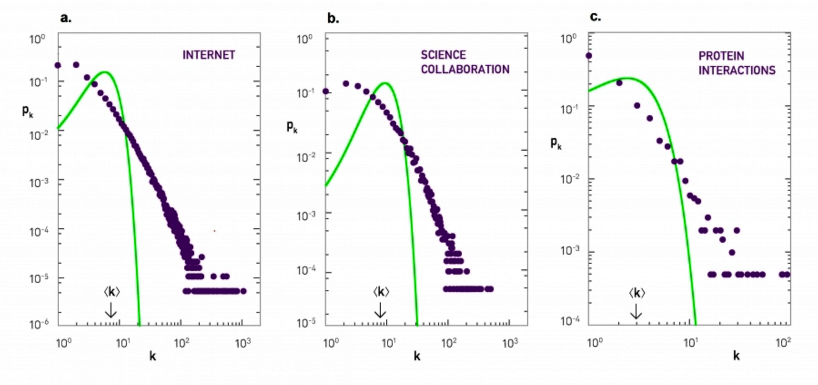
Degree fluctuation in real networks

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*-

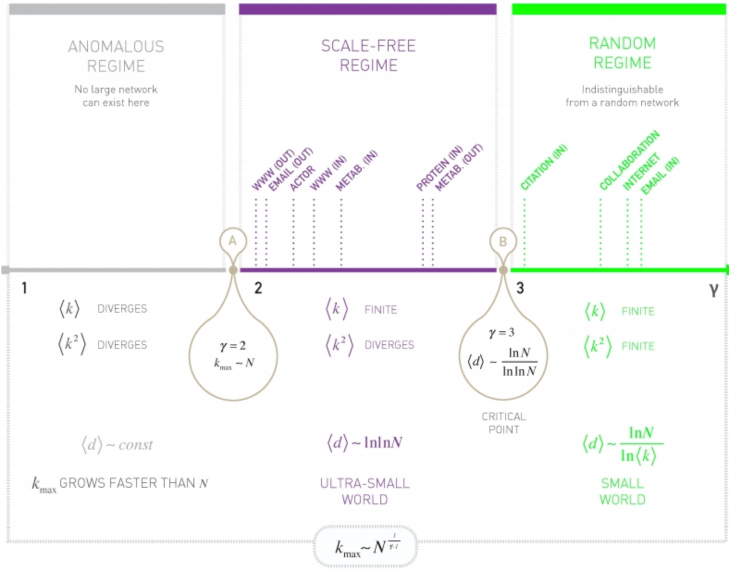
Degree fluctuation in real networks



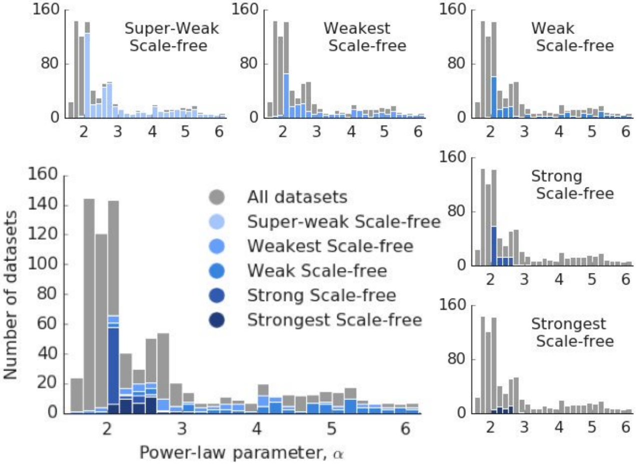
Degree fluctuation in real networks



Properties of scale free networks



Scale free networks in real world



Plotting power Law

- 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

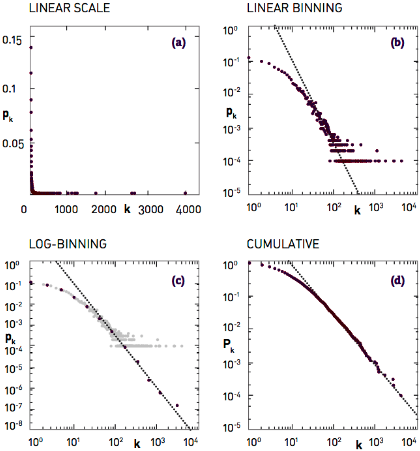
$$\bar{F}(k) = Pr(k_i > k) = 1 - F(k) = \int_k^\infty p(k)dk$$

- 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 Law



Parameter estimation: γ

Maximum likelihood estimation of parameter γ

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

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

- Probability of the sample

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

- Bayes' theorem

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

Maximum likelihood

- log-likelihood

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

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

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

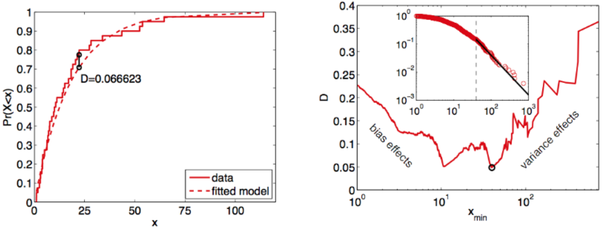
- error estimate

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

Parameter estimation: k_{min}

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

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



- find

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

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).

Small world experiment

An Experimental Study of the Small World Problem*

JEFFREY TRAVERS

Harvard University

AND

STANLEY MILGRAM

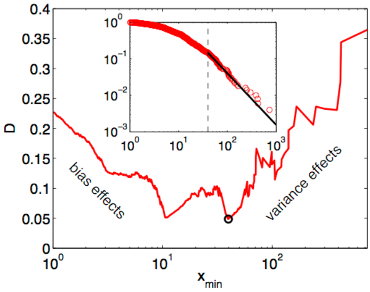
The City University of New York

Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.

© Al Satterwhite

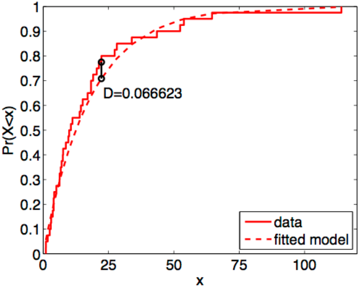
Stanley Milgram's 1967 experimen

- Starting persons:
  - 296 volunteers, 217 sent
  - 196 in Nebraska
  - 100 in Boston
- Target person - Boston stockbroker
- Information given: target name, address, occupation, place of employment, college, hometown



Stanley Milgram's 1967 experimen

- Reached the target N = 64(29%)
- Average chain length $\langle L \rangle$ = 5.2
- Channels:
  - hometown $\langle L \rangle$ = 6.1
  - business contacts $\langle L \rangle$ = 4.6
  - from Boston $\langle L \rangle$ = 4.4
  - from Nebraska $\langle L \rangle$ = 5.7



In popular culture

Connected: The Power of Six Degrees

Article Talk

From Wikipedia, the free encyclopedia

Connected: The Power of Six Degrees (alternate title: ***How Kevin Bacon Cured Cancer***[1]) is a 2008 documentary film by Annamaria Talas. It was first aired in 2009 on the Science Channel. The documentary introduces the audience to the main ideas of network science through the exploration of the concept of six degrees of separation.[2] It was awarded the 2009 AFI Award for Best Editing in a Documentary.[3]

Babel(2006), Six Degrees of Separation(1993), etc

Mean path length definition

- The mean path length l is the average of the shortest path length, averaged over all pairs of nodes. For an undirected graph of N nodes, the mean path length is

$$l = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$$

  where the sum is over all pairs of distinct nodes.
- If two nodes are disconnected, meaning there is no path between them, then the path length between them in **infinite**. As a consequence, if a network contains disconnected components (collections of nodes that have no paths between them), then the mean path length l also diverges to infinity. One way to avoid this problem is to calculate l only from nodes in **the largest connected component**.

Clustering coefficient definition

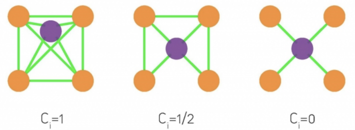
- The clustering coefficient of a graph is based on a local clustering coefficient for each node:

$$C_i = \frac{\text{number of triangles connected to node i}}{\text{number of triples centered around node i}}$$

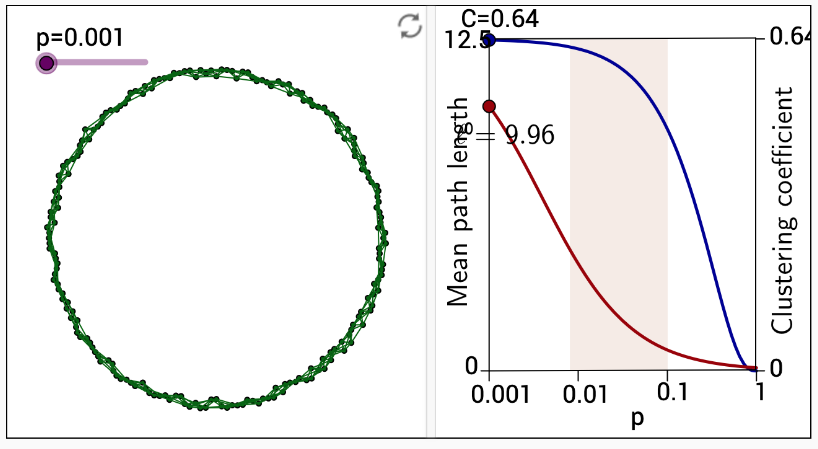
where a triple centered around node *i* is a set of two edges connected to node *i*. (If the degree of node *i* is 0 or 1, we which gives us $Ci = \frac{0}{0}$, we can set C_i=0.)

- The **clustering coefficient** for the whole graph is the average of the local values C

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i$$

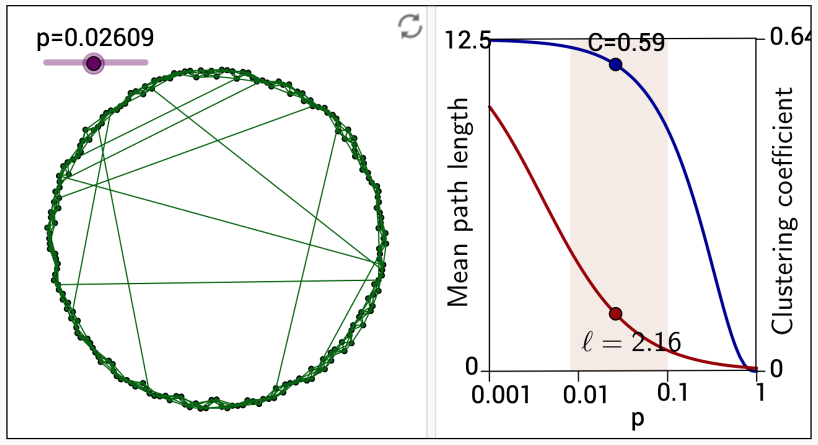


Properties of the small world network(0)



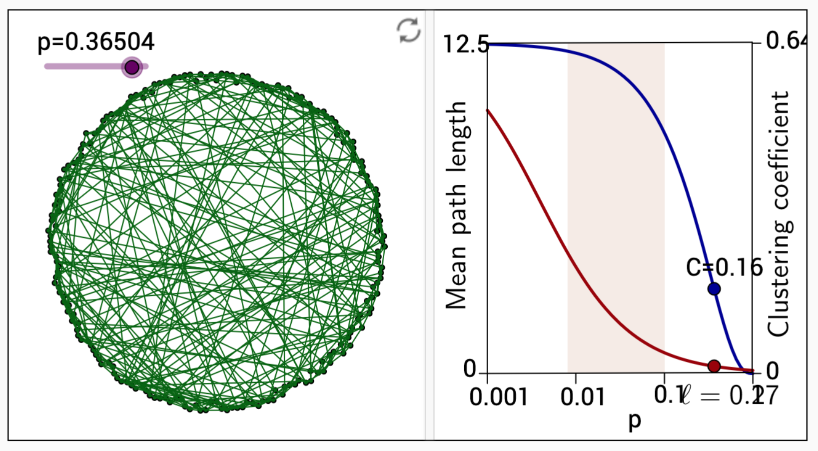
$N = 200$ $l = 9.96$ $C = 0.64$

Properties of the small world network(1)



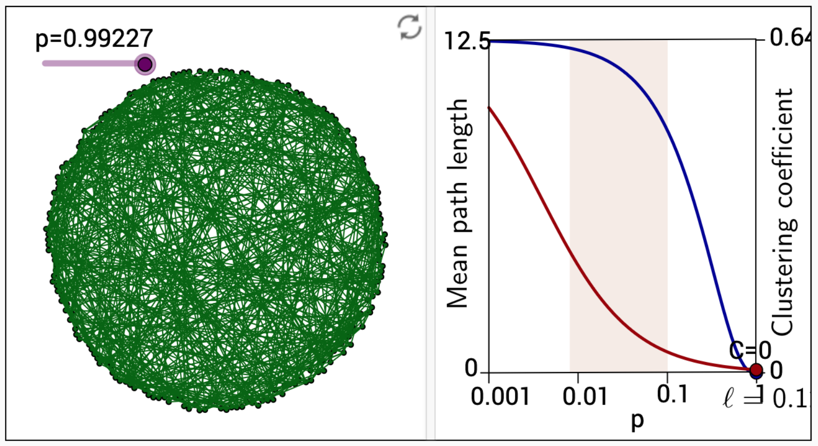
$N = 200$ $l = 2.16$ $C = 0.59$

Properties of the small world network(2)



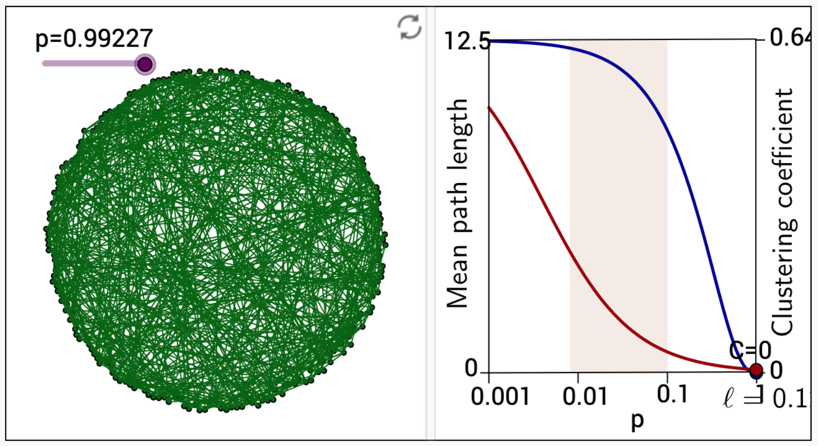
$N = 200\ l = 0.27\ C = 0.16$

Properties of the small world network(3)



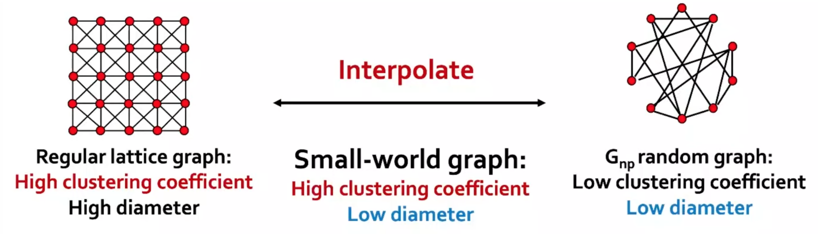
$N = 200 \; l = 0.1 \; C = 0$

Properties of the small world network(3)



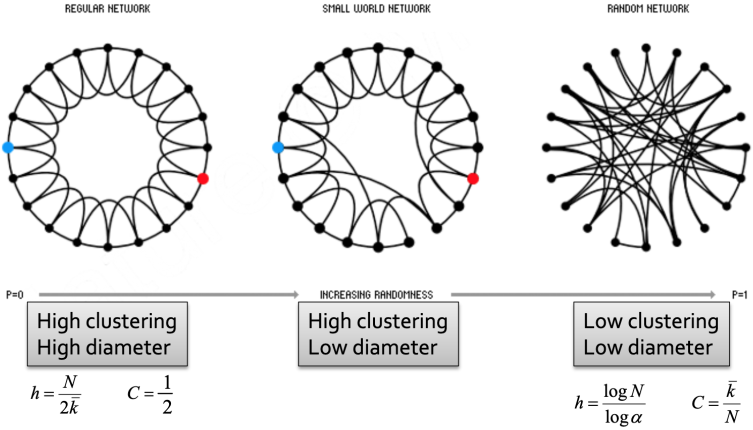
$N = 200$ $l = 0.1$ $C = 0$

From Lattice to Random Graph



$N = 200$ $l = 0.1$ $C = 0$

From Lattice to Random Graph



$$N = 200 \; l = 0.1 \; C = 0$$

Next time: Network Properties for comparison

- Degree distribution
- Clustering coefficient
- Connectivity
- Path length

References

- Collective dynamics of small-world networks. Duncan J. Watts and Steven H. Strogatz. Nature 393 (6684): 440-442, 1998
- Emergence of Scaling in Random Networks, A.L. Barabasi and R. Albert, Science 286, 509-512, 1999