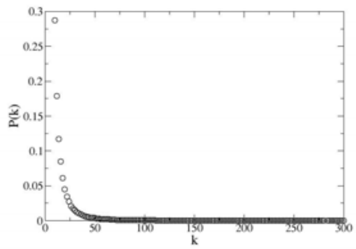
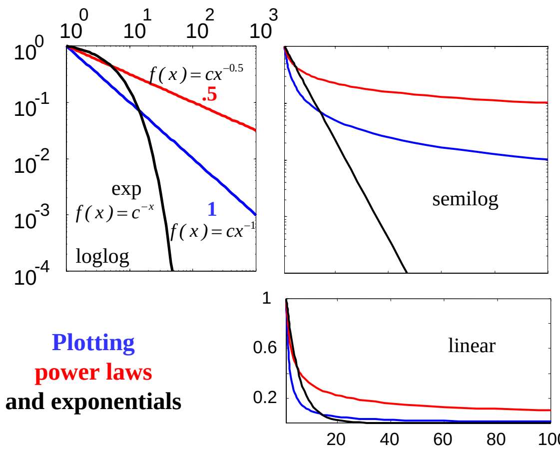
Properties of real networks: degree distribution

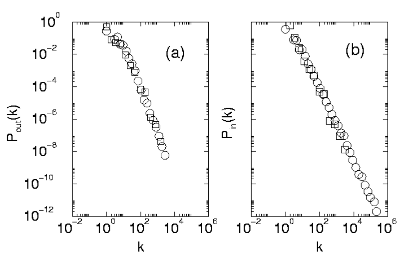


Nodes with small degrees are most frequent.
The fraction of highly connected nodes decreases, but is not zero.
Look closer: use a logarithmic plot.



Plotting power laws and exponentials

The in- and out-degree distribution of the WWW are power-laws



nodes: webpages
edges: hyperlinks

$P_{out}(k) \approx k^{-2.45}$

$P_{in}(k) \approx k^{-2.1}$

R. Albert, H. Jeong, A.-L. Barabási, Nature 401, 130 (1999)

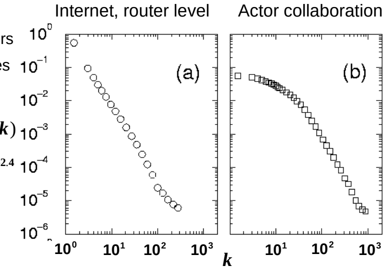
A. Broder *et al.*, Comput. Netw. 33, 309 (1999)

Power-law degree distributions were found in diverse networks

Internet, router level | Actor collaboration

nodes: routers
edges: cables

$P(k) \approx k^{-2.4}$

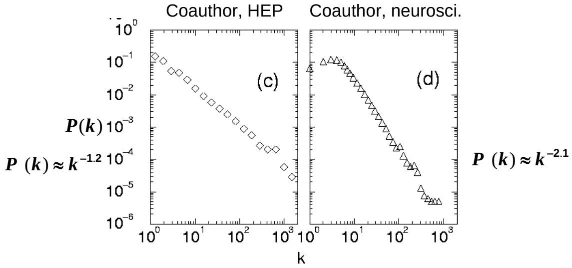


nodes: actors
edges: cast jointly

$P(k) \approx (31 + k)^{-3}$

R. Govindan, H. Tangmunarunkit, IEEE Infocom (2000)
A.-L. Barabási, R. Albert, Science 286, 509 (1999)

Networks of science collaborations also have power-law degree distributions



M. E. J. Newman, Phys. Rev. E 64, 016131 (2001)

A.-L. Barabási et al., cond-mat/0104162 (2001)

Metabolic networks have a power-law degree distribution

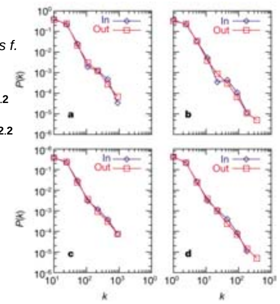
Archaeoglobus f.

E. coli

$P_{in}(k) \approx k^{-2.2}$

$P_{out}(k) \approx k^{-2.2}$

C. elegans



bipartite
nodes: metabolites, reactions
directed edges,
out: reactant (substrate)
in: product of reaction

H. Jeong et al., Nature 407, 651 (2000)

Cleaning up degree distributions

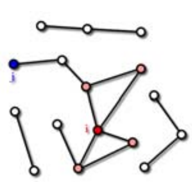
Often it is difficult to determine the best fit to the points that make up a degree distribution.

Methods of data cleanup:

1. logarithmic binning: bin the k range; use bins of exponentially increasing size
2. Display the cumulative degree distribution

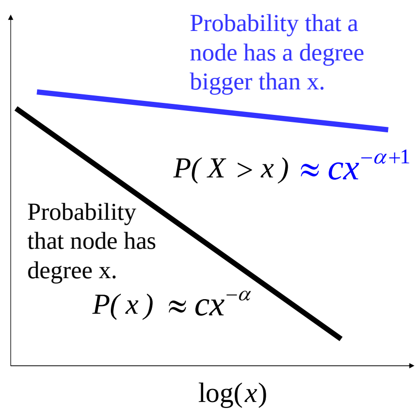
$$P(k \le K) = \sum_{k=k_{min}}^{K} P(k) \text{ or}$$

$$P(k > K) = 1 - P(k \le K)$$

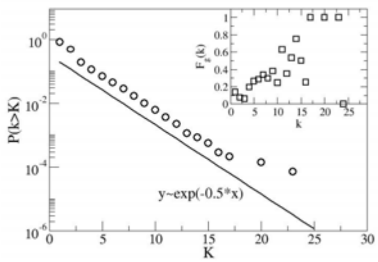


Ex. Determine the degree distribution and cumulative degree distribution of the graph on the right.

If the (noncumulative) degree distribution is a power law with exponent α>1, the cumulative degree distribution will be a power law with exponent α-1. Does not apply for α=1!



Power grid has exponential degree distribution



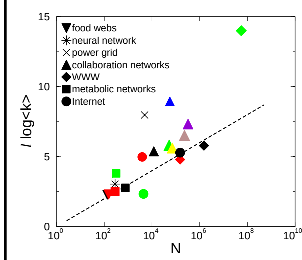
nodes: generators, power stations

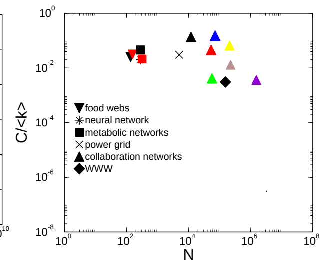
edges: power lines

$$P\ (k > K) \propto exp(0.5K)$$

R. Albert, I. Albert, G. L. Nakarado, Phys. Rev. E 69, 025103(R) (2004)

Path length and order in real networks





$$l \approx \frac{log\ N}{log\langle k\rangle} \qquad C \propto \langle k\rangle$$

Apparent scaling with the network size and average degree - as though these different networks were members of the same family.

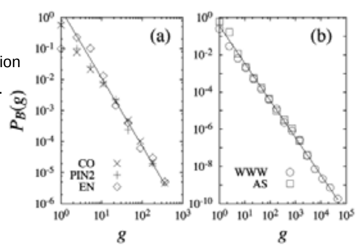
Distribution of betweenness centrality

Coauthorship

Protein interaction

Metabolic netw.

$$P_B(g) \approx g^{-2.2}$$



World-wide Web

Internet (AS level)

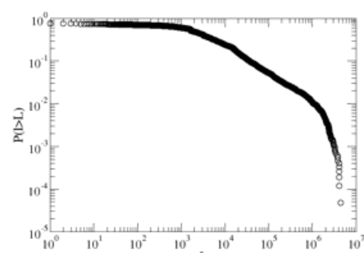
$$P_B(g) \approx g^{-2}$$

K. I. Goh et al., PNAS 99, 12583 (2002)

Betweenness centrality (load) distribution of the power grid

$$P(l > L) \approx (2500 + L)^{-0.7}$$

Q: How does the non-cumulative distribution look like in the region where the cumulative distribution is almost horizontal?



R. Albert, I. Albert, G. L. Nakarado, Phys. Rev. E 69, 025103(R) (2004)

Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

Mixing patterns in networks

Mixing in social networks
- assortative: people prefer to associate with others who are like them
- disassortative: people prefer to associate with others who are different

Mixing with respect of node degree:
- assortative: high degree nodes tend to be connected to high degree nodes
- disassortative: high degree nodes tend to be connected to low degree nodes

Focus on edge *i*, denote the excess in-degree of its starting point with j_i and the excess out-degree of its endpoint with k_i

Mixing is quantified by the correlation between j_i and k_i over all *i*

$$r = \frac{\sum_i j_i k_i - \sum_i j_i \sum_i k_i / N}{\left(\sum_i j_i^2 - (\sum_i j_i)^2 / N\right)^{0.5} \left(\sum_i k_i^2 - (\sum_i k_i)^2 / N\right)^{0.5}}$$

Positive correlation - assortative, Negative correlation - disassortative

	network	type	size n	assortativity r	error σ_r	ref.
social	physics coauthorship	undirected	52909	0.363	0.002	a
	biology coauthorship	undirected	1520251	0.127	0.0004	a
	mathematics coauthorship	undirected	253339	0.120	0.002	b
	film actor collaborations	undirected	449913	0.208	0.0002	c
	company directors	undirected	7673	0.276	0.004	d
	student relationships	undirected	573	−0.029	0.037	e
	email address books	directed	16881	0.092	0.004	f
technological	power grid	undirected	4941	−0.003	0.013	g
	Internet	undirected	10697	−0.189	0.002	h
	World-Wide Web	directed	269504	−0.067	0.0002	i
	software dependencies	directed	3162	−0.016	0.020	j
biological	protein interactions	undirected	2115	−0.156	0.010	k
	metabolic network	undirected	765	−0.240	0.007	l
	neural network	directed	307	−0.226	0.016	m
	marine food web	directed	134	−0.263	0.037	n
	freshwater food web	directed	92	−0.326	0.031	o

Social networks tend to be assortative, technological and biological networks tend to be disassortative.

Possible causes of assortativity: attraction of similars, group affiliation
Possible cause of disassortativity: service relationships (e.g. directories)

M. E. J. Newman, Phys. Rev. E (2003)

Universality in large-scale networks

The degree distribution is a decreasing function, usually a power-law
The betweenness centrality distribution is a power law as well
Both indicate heterogeneity and the existence of hubs.

The distances scale logarithmically with the network size

$$l \approx \frac{\log N}{\log\langle k \rangle}$$

The clustering coefficient does not seem to depend on the network size $C \propto \langle k \rangle$

Frequent subgraphs – not universal but common to several networks