Systems Biology Across Scales: A Personal View

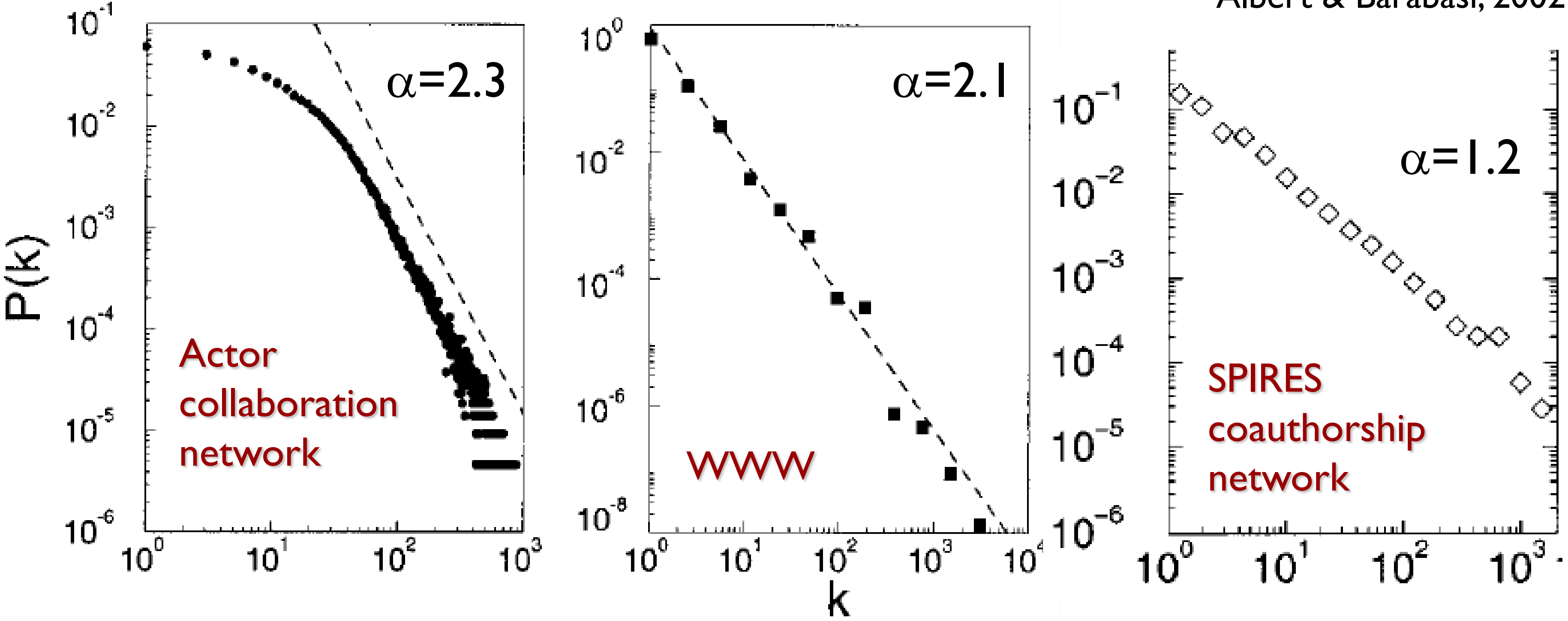
VIII. Networks: Models II

Sitabhra Sinha
IMSc Chennai

Example: *scale-free* networks

Barabasi and Albert (1999): In many large networks the vertex connectivities follow a scale-free distributions, i.e., the degree distrn has a power law tail: $P(k) \sim k^{-\alpha}$.

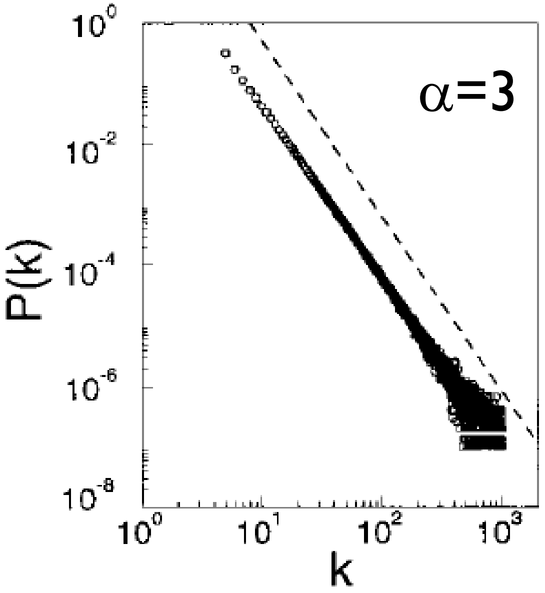
Albert & Barabasi, 2002



In contrast,

ER random networks (start with N nodes and connect each pair with probability p) have Poisson degree distrn: $P(k) = e^{-\lambda} (\lambda^k/k!)$

How can *scale-free* networks evolve ?



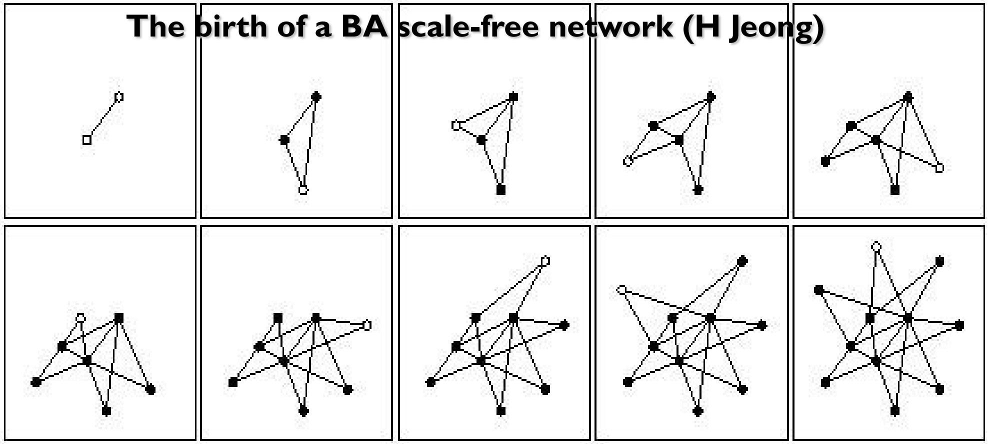
Albert, Jeong & Barabasi, 2000

The Price-Barabasi-Albert preferential attachment scheme:

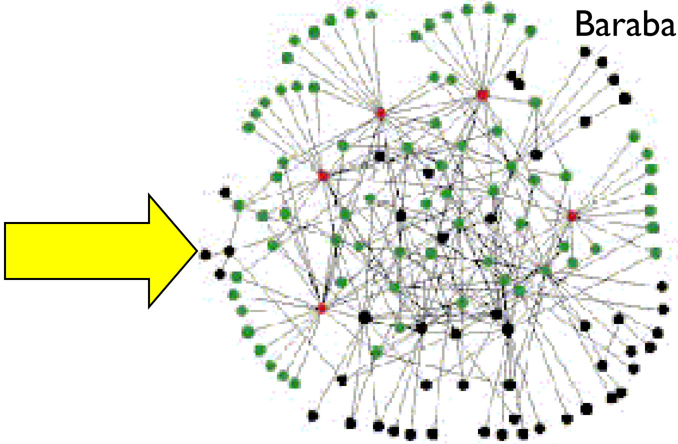
(A) Networks expand continuously by addition of new nodes

(B) New nodes attach preferentially to nodes already well-connected, i.e., probability that a new node is connected to a node of degree k_i is $\Pi(k_i)=k_i/\Sigma_j k_j$ ("*linear*" scheme)

Resulting network degree distribution: $P(k) \sim k^{-3}$



The birth of a BA scale-free network (H Jeong)



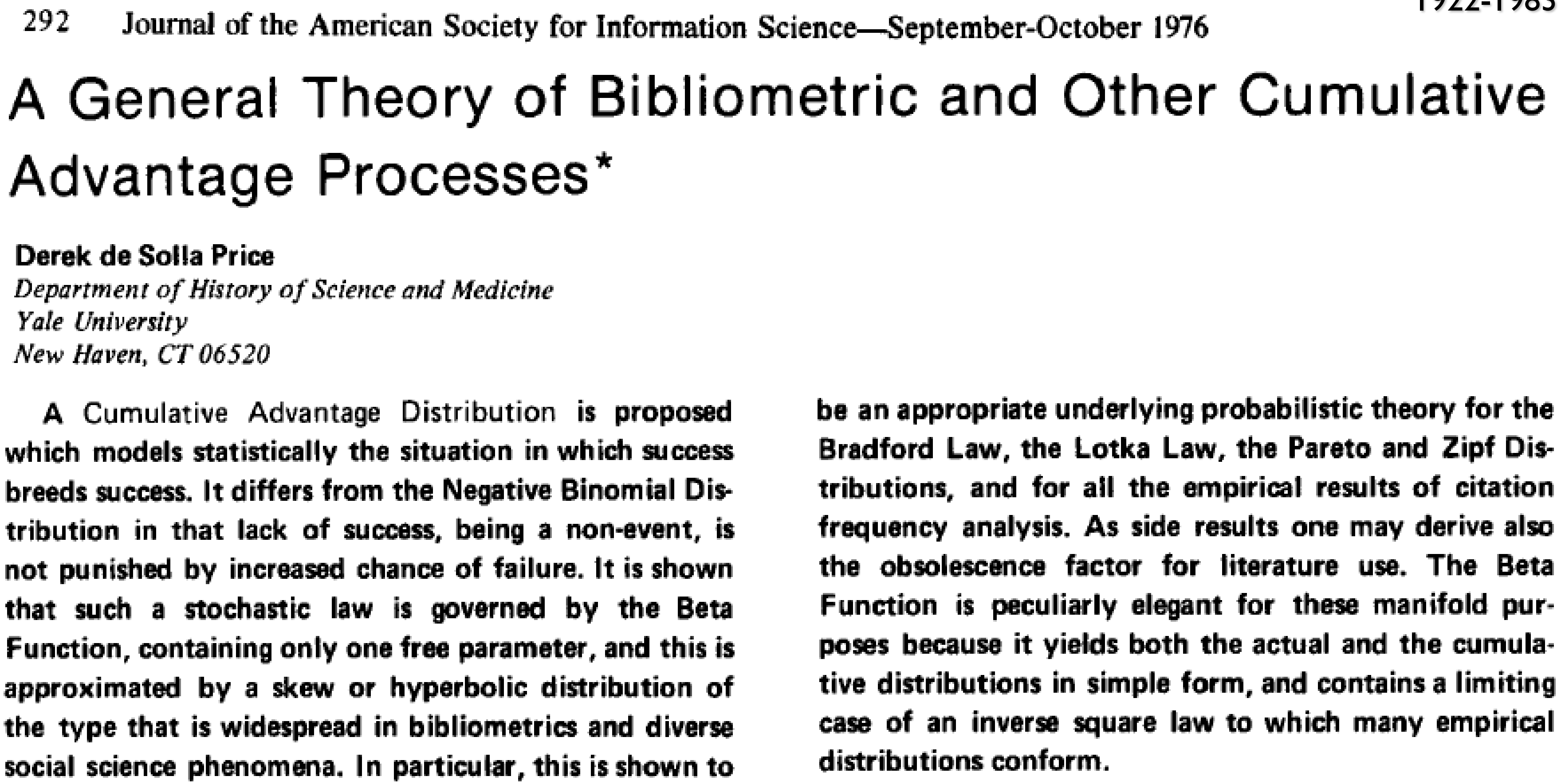
Characterised by highly connected "hubs", which hold the network together

Achilles' heel: Network fragile to directed attack on "hubs"

The Mechanism:
A Cumulative Advantage Process

Derek J de Solla Price proposed in 1976 the earliest mathematically detailed mechanism by which scale-free degree distribution can arise in the context of networks of citations between scientific papers

Derek J de Solla Price
1922-1983



292 Journal of the American Society for Information Science—September-October 1976

A General Theory of Bibliometric and Other Cumulative Advantage Processes*

Derek de Solla Price
Department of History of Science and Medicine
Yale University
New Haven, CT 06520

A Cumulative Advantage Distribution is proposed which models statistically the situation in which success breeds success. It differs from the Negative Binomial Distribution in that lack of success, being a non-event, is not punished by increased chance of failure. It is shown that such a stochastic law is governed by the Beta Function, containing only one free parameter, and this is approximated by a skew or hyperbolic distribution of the type that is widespread in bibliometrics and diverse social science phenomena. In particular, this is shown to be an appropriate underlying probabilistic theory for the Bradford Law, the Lotka Law, the Pareto and Zipf Distributions, and for all the empirical results of citation frequency analysis. As side results one may derive also the obsolescence factor for literature use. The Beta Function is peculiarly elegant for these manifold purposes because it yields both the actual and the cumulative distributions in simple form, and contains a limiting case of an inverse square law to which many empirical distributions conform.

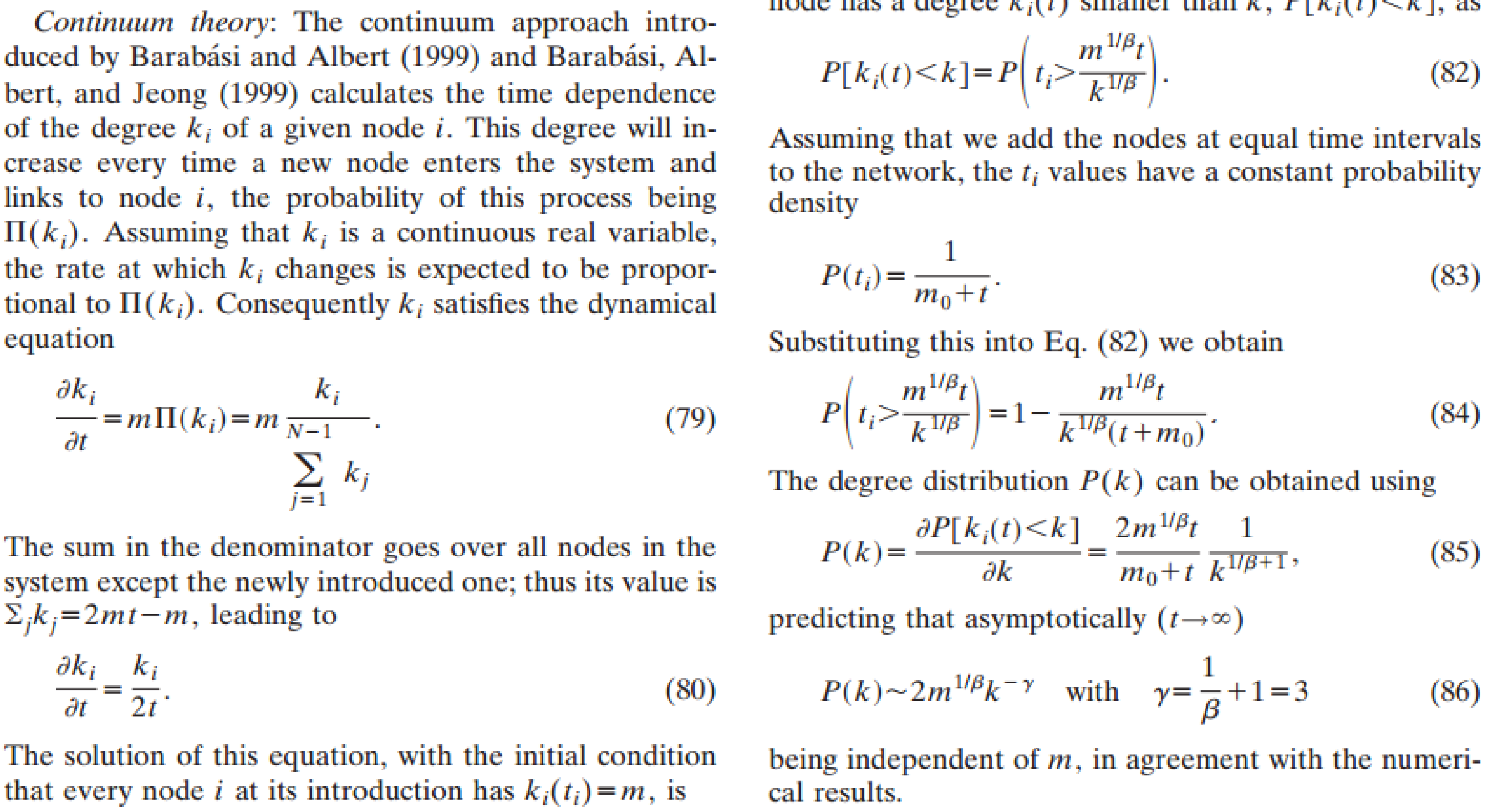
Why the degree distribution is *scale-free*

From

Statistical mechanics of complex networks

Reka Albert and Albert-Laszlo´Barabasi

Rev Mod Phys 74 (2002) 47-97



Continuum theory: The continuum approach introduced by Barabási and Albert (1999) and Barabási, Albert, and Jeong (1999) calculates the time dependence of the degree k_i of a given node i. This degree will increase every time a new node enters the system and links to node i, the probability of this process being $\Pi(k_i)$. Assuming that k_i is a continuous real variable, the rate at which k_i changes is expected to be proportional to $\Pi(k_i)$. Consequently k_i satisfies the dynamical equation

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}. \tag{79}$$

The sum in the denominator goes over all nodes in the system except the newly introduced one; thus its value is $\Sigma_j k_j = 2mt - m$, leading to

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}. \tag{80}$$

The solution of this equation, with the initial condition that every node i at its introduction has $k_i(t_i) = m$, is

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \quad \text{with} \quad \beta = \frac{1}{2}. \tag{81}$$

Equation (81) indicates that the degree of all nodes evolves the same way, following a power law, the only difference being the intercept of the power law.

Using Eq. (81), one can write the probability that a node has a degree $k_i(t)$ smaller than k, $P[k_i(t) < k]$, as

$$P[k_i(t) < k] = P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right). \tag{82}$$

Assuming that we add the nodes at equal time intervals to the network, the t_i values have a constant probability density

$$P(t_i) = \frac{1}{m_0 + t}. \tag{83}$$

Substituting this into Eq. (82) we obtain

$$P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta}t}{k^{1/\beta}(t + m_0)}. \tag{84}$$

The degree distribution $P(k)$ can be obtained using

$$P(k) = \frac{\partial P[k_i(t) < k]}{\partial k} = \frac{2m^{1/\beta}t}{m_0 + t}\frac{1}{k^{1/\beta+1}}, \tag{85}$$

predicting that asymptotically ($t \to \infty$)

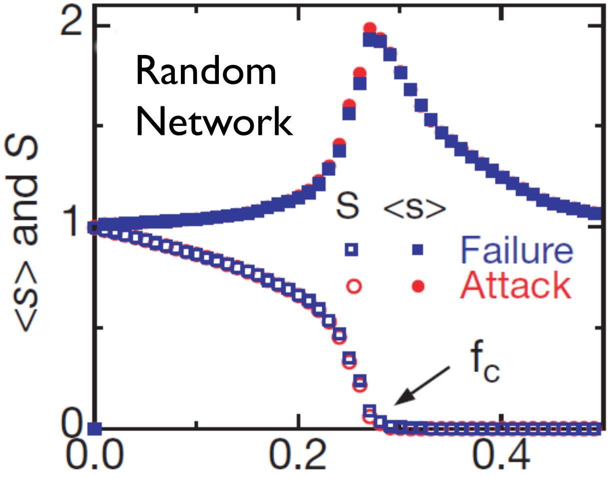
$$P(k) \sim 2m^{1/\beta}k^{-\gamma} \quad \text{with} \quad \gamma = \frac{1}{\beta} + 1 = 3 \tag{86}$$

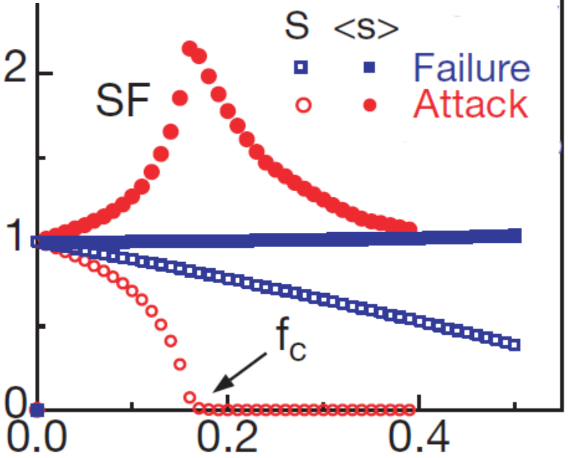
being independent of m, in agreement with the numerical results.

Importance of "hubs"

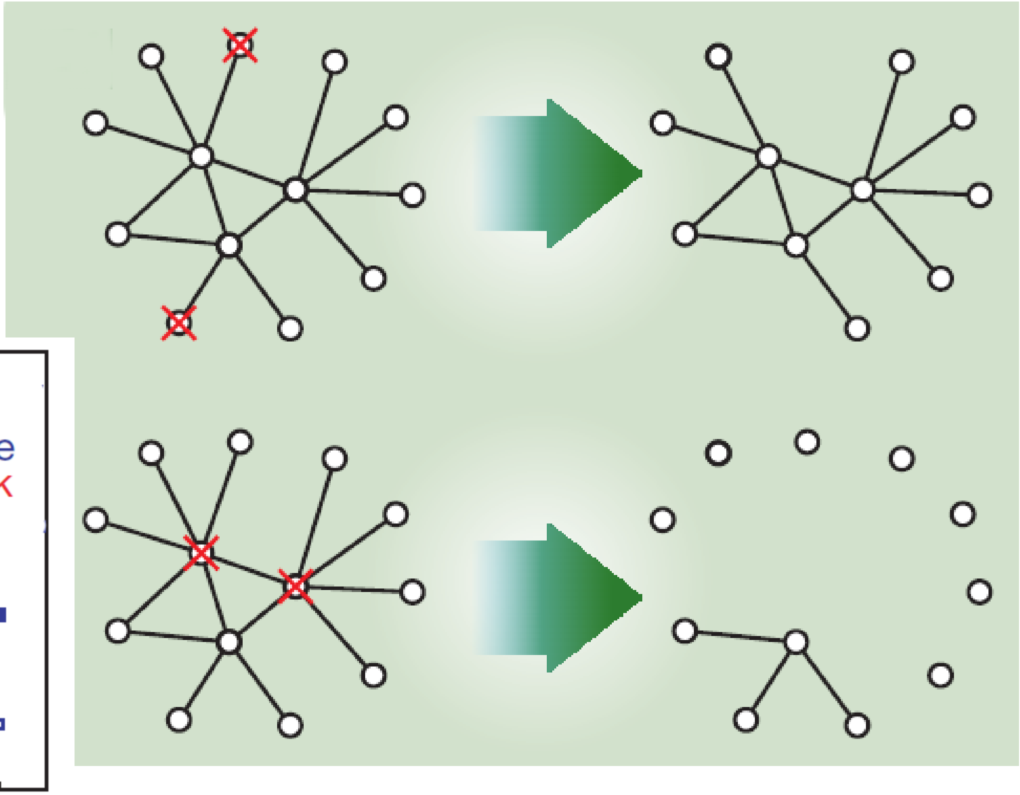
Random failure of nodes typically has little effect on scale-free network as most nodes connect only to a few other nodes: *Robustness to random node removal*

Newman 2008





S: relative size of largest cluster, and <s> : average size of the isolated clusters, shown as a function of the fraction of removed nodes



However targeting the highest-degree nodes (hubs) has devastating effect on the network – most nodes become isolated on removing a few hubs: *Vulnerability to targeted removal of hubs*

No threshold for epidemics in *scale-free networks*

Networks of sexual relations have been claimed to be scale-free !
A few highly promiscuous individuals act as "hub" nodes
May play a crucial role in spreading sexually transmitted diseases !

If the contact structure of a disease is network with inhomogeneous degree distribution, the condition for occurrence of an epidemic is:

$$R = bN/g \quad > \quad \langle k\rangle / \langle k^2 \rangle$$

b: rate of infection spreading, g: recovery rate (=1/period of infection)
Initial popn of susceptibles, $S(t = 0) = N$, the total population

For a scale-free network having degree exponent $2<\alpha \leq 3$, $\langle k^2\rangle \rightarrow \infty$
$\Rightarrow$There is no epidemic threshold !

Even diseases with extremely low transmission probabilities are likely to cause a major outbreak involving a signficant fraction of population

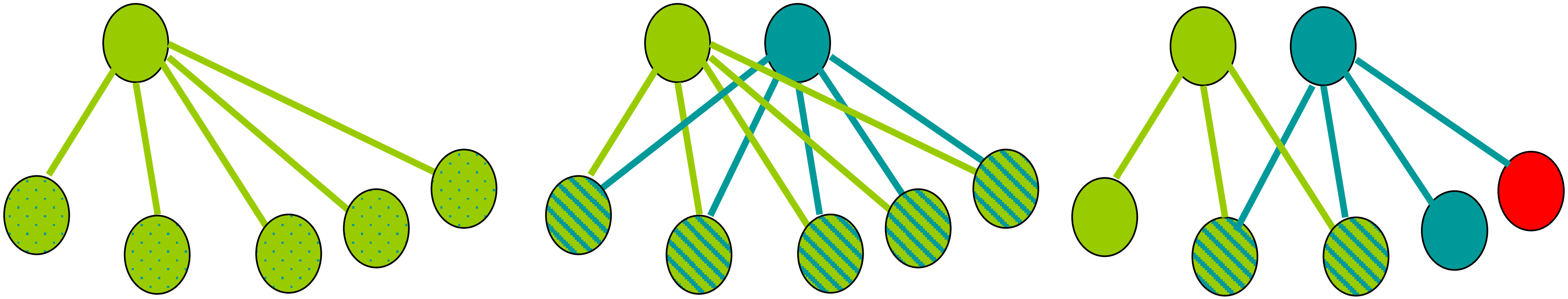
Alternative to preferential attachment *Duplication and Divergence*

While preferential attachment may be appropriate for explaining scale-free character of World-Wide Web, it is less clear how it might play a role in biological systems, e.g., the protein-protein interaction network that has been claimed to be scale-free

As most biological systems have emerged through a long history of evolution, can evolutionary processes give rise to a network with scale-free property ?

Vazquez et al, ComPlexUs (2003)

In the Duplication-Divergence mechanism, a node along with all its interactions are duplicated with some probability p and then some of the interactions are mutated with probability q ⇒ leads to a networks with scale-free degree distribution



Target Protein

Duplication

Divergence

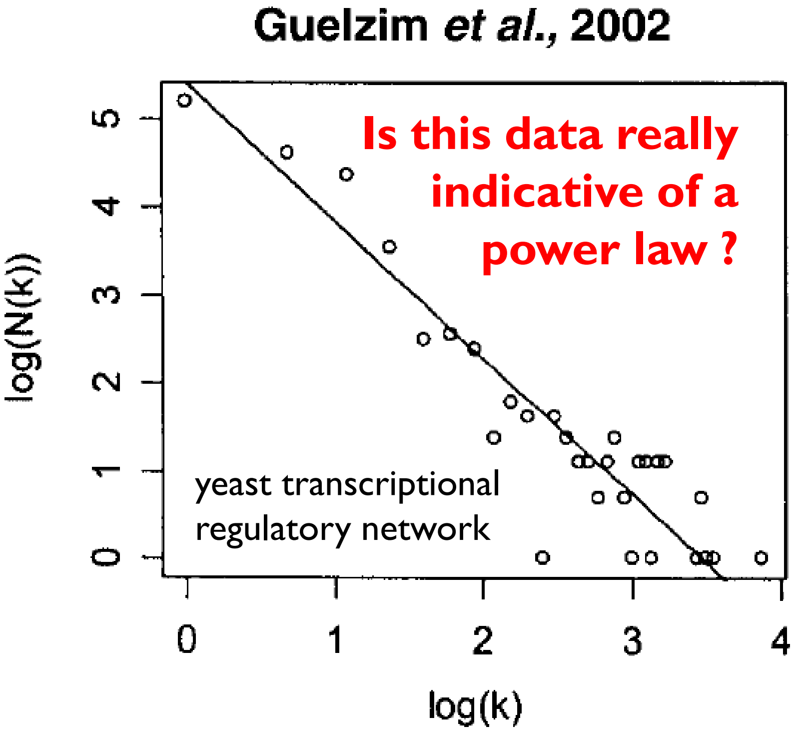
Interactions

Shared interactions

Loss of some interactions

Are biological networks really scale-free ?

- Scale-free networks characterized by long-tailed degree distribution (power laws) have been proposed as unifying concept for biological complex systems – have been reported in metabolic, protein, and gene interaction networks.
- But many of these reports of scale-free networks are possibly just a result of bad statistics (a combination of extremely limited data and faulty analysis) !
- Almost any distribution seen over a small enough range in a double logarithmic scale would appear linear – and wrongly interpreted as power law

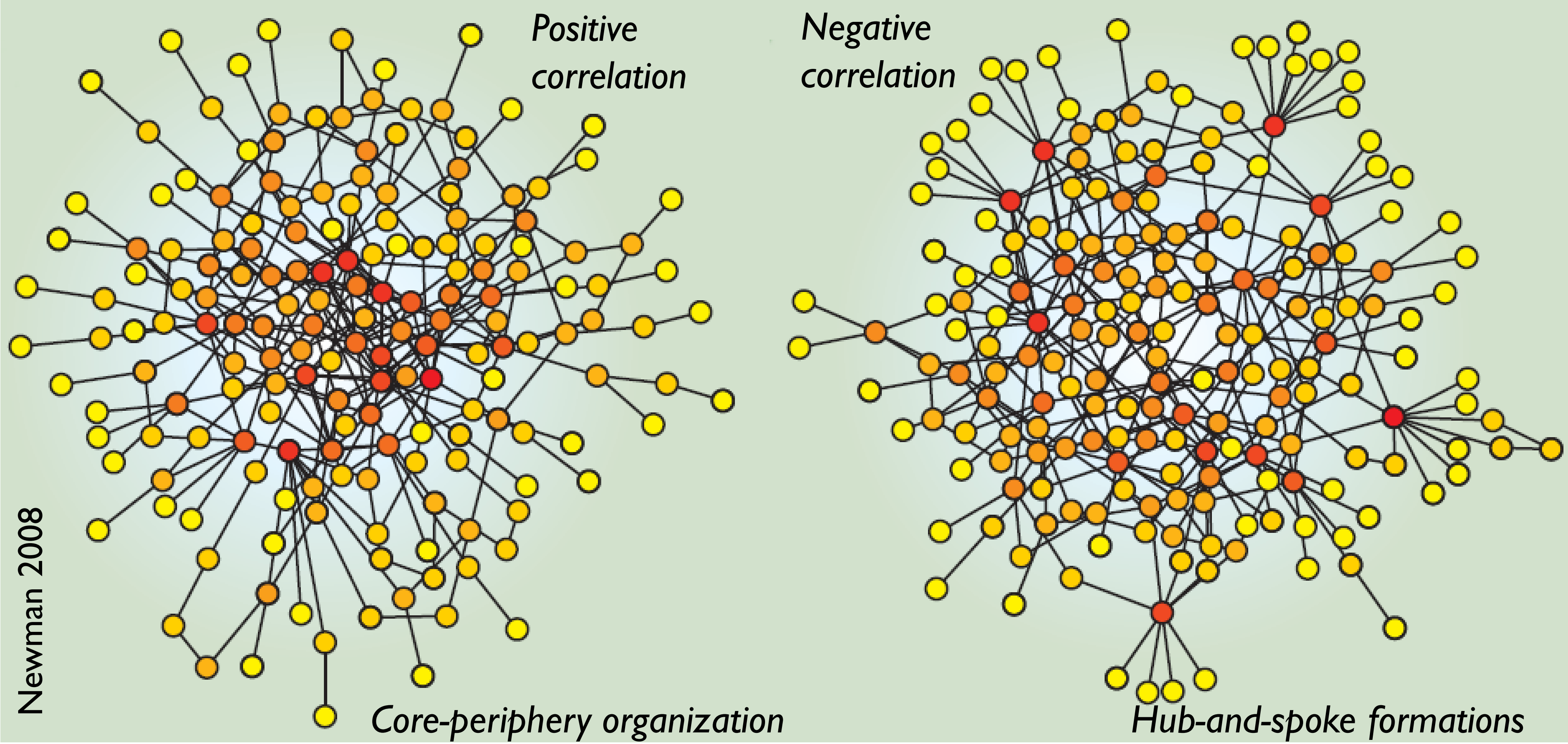

Guelzim *et al.*, 2002

Is this data really indicative of a power law ?

yeast transcriptional regulatory network

- To establish power laws from finite data one has to use unbiased techniques such as *maximum likelihood estimation*.
- Recent rigorous re-analysis of many of the data sets used by earlier studies that claimed power-law degree distributions for protein and gene interaction networks have shown little evidence for scale-free nature! (E.g., R Khanin & E Wit, *J Comp Biol* **13** (2006) 810)

Nodes may prefer to connect to nodes with similar or dissimilar connectivity



Newman 2008

Two networks may have the same degree distribution but different connectivity patterns overall because high-degree nodes may prefer to connect to other high-degree nodes (positive degree correlation) or may want to avoid them (negative degree correlation)

Measuring degree correlations

Random networks (e.g., random Price-Barabasi-Albert networks) do not exhibit any correlations between the degrees of connected nodes, i.e., The probability a link connects nodes of degrees k & k' is

$P(k,k') = k\,P(k)\,k'\,P(k')/\langle k\rangle^2$ (degree-uncorrelated network)

Most real-life networks exhibit degree correlations measured by

Assortativity, r

$$= \frac{\frac{1}{L}\left(\sum_{i=1}^{L} j_i k_i\right) - \left(\frac{1}{L}\sum_{i=1}^{L} \frac{1}{2}(j_i + k_i)\right)^2}{\frac{1}{L}\left(\sum_{i=1}^{L} \frac{1}{2}(j_i^2 + k_i^2\right) - \left(\frac{1}{L}\left(\sum_{i=1}^{L} \frac{1}{2}(j_i + k_i)\right)\right)^2}$$

(Newman, 2002)

j_i, k_i: degrees of vertices at ends of the i-th edge

L: total number of links

r < 0: disassortative mixing

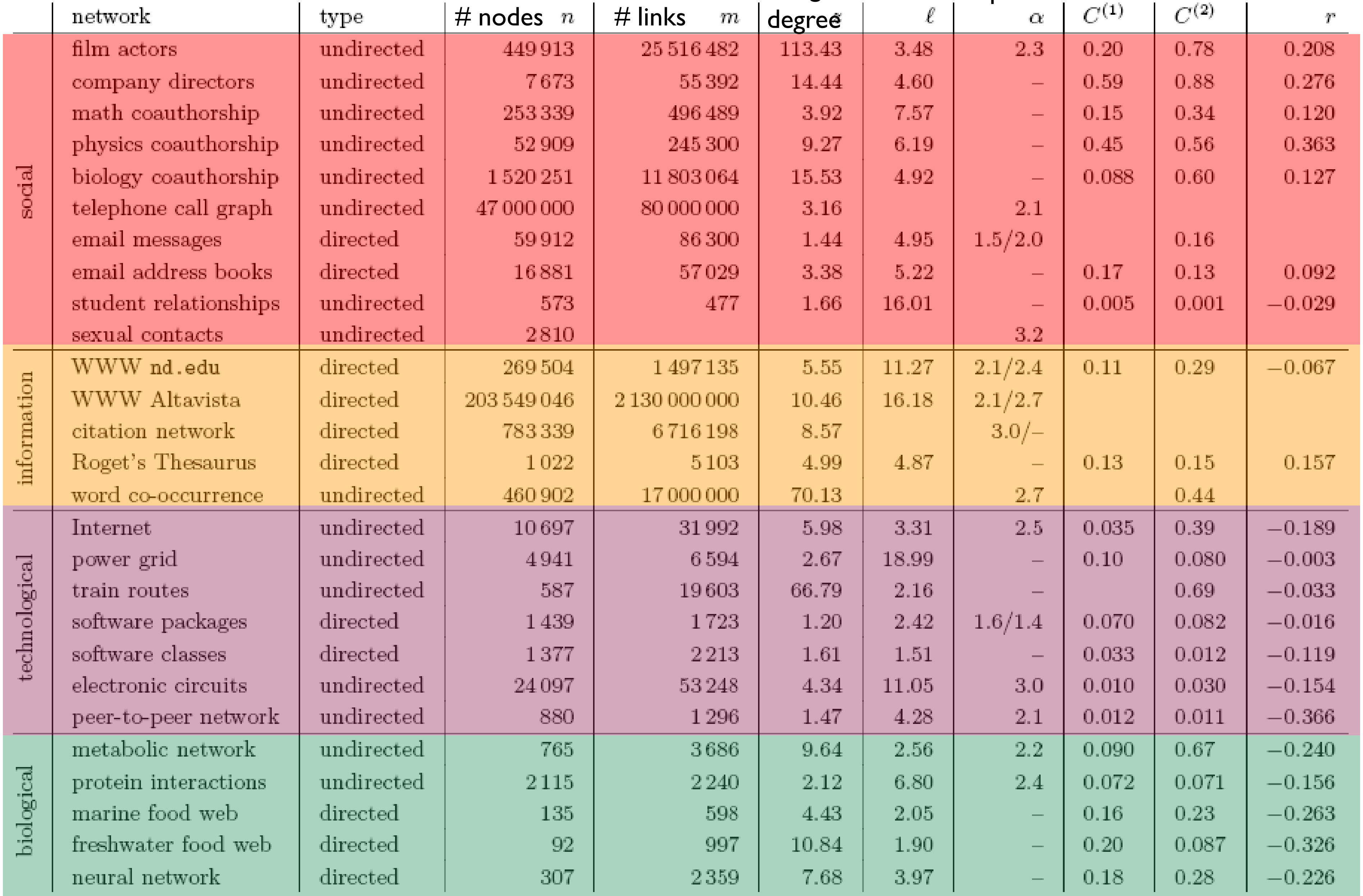
Nodes of high degree mostly have nearest nbrs of low degree
E.g.,most biological & technolgical networks

r > 0: assortative mixing

Nodes of high degree mostly have nearest nbrs of high degree
E.g., social networks

Macroscopic properties of networks

Newman, *SIAM Review*, 2003



	network	type	# nodes n	# links m	avg degree	avg path length ℓ	degree distrn exponent α	clustering coefficient $C^{(1)}$	clustering coefficient $C^{(2)}$	degree correln r
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1			
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16	
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029
	sexual contacts	undirected	2 810				3.2			
information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7			
	citation network	directed	783 339	6 716 198	8.57		3.0/–			
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44	
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226