

Information Networks

Networks and Measurements

Lecture 2





What is an information network?

- § Network: a collection of **entities** that are interconnected
- § A **link** (edge) between two entities (nodes) denotes an interaction between two entities
- § We view this interaction as **information exchange**, hence, Information Networks
- § The term encompasses more general networks



Why do we care?

- § Networks are everywhere
 - § more and more systems can be modeled as networks, and more data is collected
 - § traditional graph models no longer work
- § Large scale networks require new tools to study them

- § A fascinating “new” field (“new science”?)
 - § involves multiple disciplines: computer science, mathematics, physics, biology, sociology. economics



Types of networks

- § Social networks
- § Knowledge (Information) networks
- § Technology networks
- § Biological networks



Social Networks

§ Links denote a social interaction

§ Networks of acquaintances

§ actor networks

§ co-authorship networks

§ director networks

§ phone-call networks

§ e-mail networks

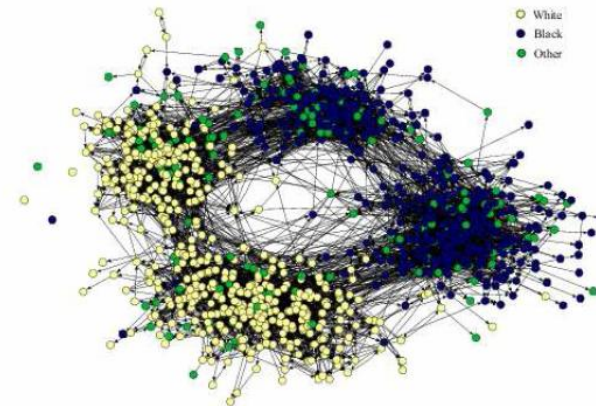
§ IM networks

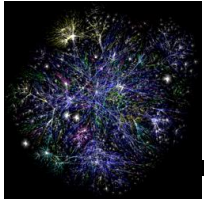
- Microsoft buddy network

§ Bluetooth networks

§ sexual networks

§ home page networks





Knowledge (Information) Networks

- § Nodes store information, links associate information
- § Citation network (directed acyclic)
- § The Web (directed)
- § Peer-to-Peer networks
- § Word networks
- § Networks of Trust
- § Bluetooth networks



Technological networks

§ Networks built for distribution of commodity

§ The Internet

- router level, AS level

§ Power Grids

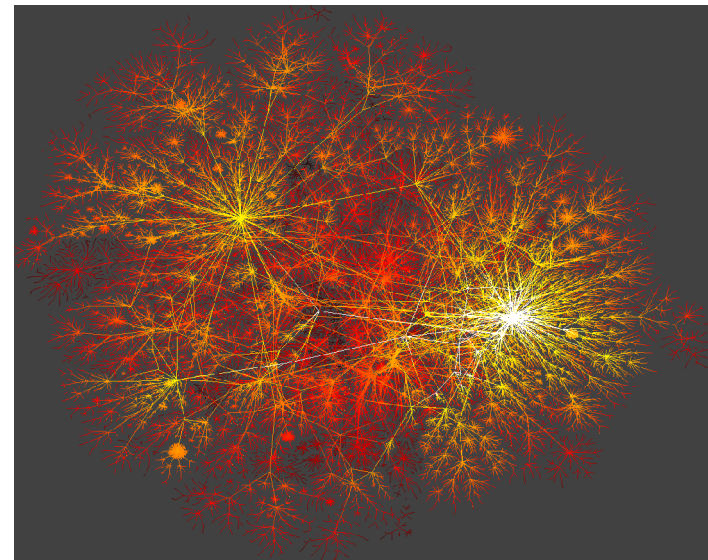
§ Airline networks

§ Telephone networks

§ Transportation Networks

- roads, railways, pedestrian traffic

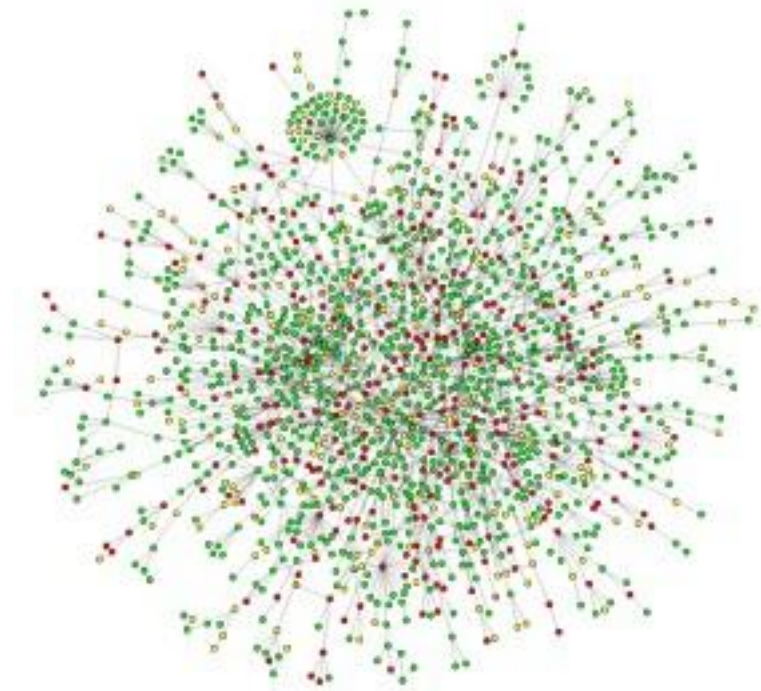
§ Software graphs

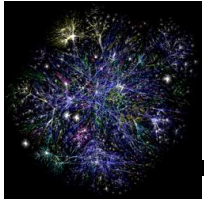




Biological networks

- § Biological systems represented as networks
 - § Protein-Protein Interaction Networks
 - § Gene regulation networks
 - § Metabolic pathways
 - § The Food Web
 - § Neural Networks



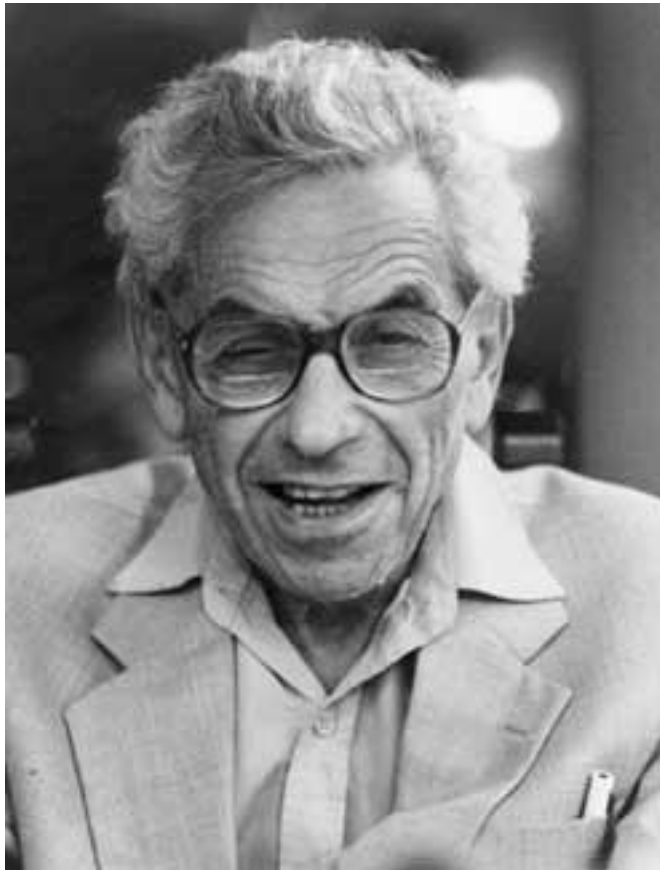


Now what?

- § The world is full with networks. What do we do with them?
- § understand their topology and measure their properties
- § study their evolution and dynamics
- § create realistic models
- § create algorithms that make use of the network structure



Erdős-Rényi Random graphs



Paul Erdős (1913-1996)



Erdős-Renyi Random Graphs

§ The $G_{n,p}$ model

§ n : the number of vertices

§ $0 \leq p \leq 1$

§ for each pair (i,j) , generate the edge (i,j) independently with probability p

§ Related, but not identical: The $G_{n,m}$ model



Graph properties

§ A property P holds **almost surely** (or for **almost every** graph), if

$$\lim_{n \rightarrow \infty} P[G \text{ has } P] = 1$$

§ Evolution of the graph: which properties hold as the probability p increases?

§ **Threshold phenomena**: Many properties appear suddenly. That is, there exist a probability p_c such that for $p < p_c$ the property does not hold a.s. and for $p > p_c$ the property holds a.s.



The giant component

- § Let $z=np$ be the average degree
- § If $z < 1$, then almost surely, the largest component has size at most $O(\ln n)$
- § if $z > 1$, then almost surely, the largest component has size $\Theta(n)$. The second largest component has size $O(\ln n)$
- § if $z = \omega(\ln n)$, then the graph is almost surely connected.



The phase transition

- § When $z=1$, there is a phase transition
 - § The largest component is $O(n^{2/3})$
 - § The sizes of the components follow a power-law distribution.



Random graphs degree distributions

§ The degree distribution follows a **binomial**

$$p(k) = B(n; k; p) = \binom{n}{k} p^k (1-p)^{n-k}$$

§ Assuming **$z=np$** is fixed, as $n \rightarrow \infty$ $B(n, k, p)$ is approximated by a **Poisson** distribution

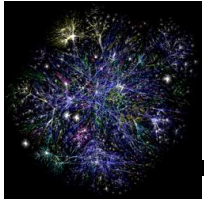
$$p(k) = P(k; z) = \frac{z^k}{k!} e^{-z}$$

§ Highly concentrated around the mean, with a tail that drops exponentially



Phase Transition

- § Starting from some vertex v perform a BFS walk
- § At each step of the BFS a Poisson process with mean z , gives birth to new nodes
- § When $z < 1$ this process will stop after $O(\log n)$ steps
- § When $z > 1$, this process will continue for $\Theta(n)$ steps



Random graphs and real life

- § A beautiful and elegant theory studied exhaustively
- § Random graphs had been used as idealized generative models
- § Unfortunately, they don't capture reality...

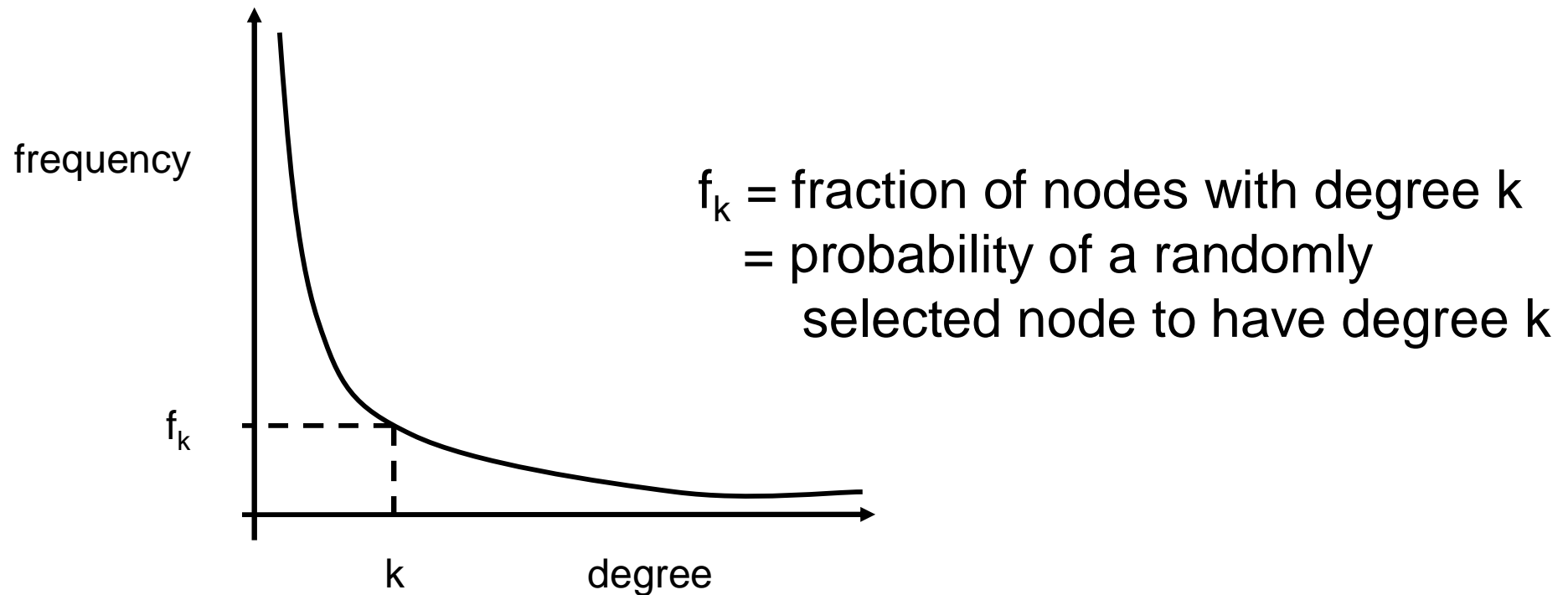


Measuring Networks

- § Degree distributions
- § Small world phenomena
- § Clustering Coefficient
- § Mixing patterns
- § Degree correlations
- § Communities and clusters



Degree distributions



§ Problem: find the probability distribution that best fits the observed data



Power-law distributions

- § The degree distributions of most real-life networks follow a **power law**

$$p(k) = Ck^{-\alpha}$$

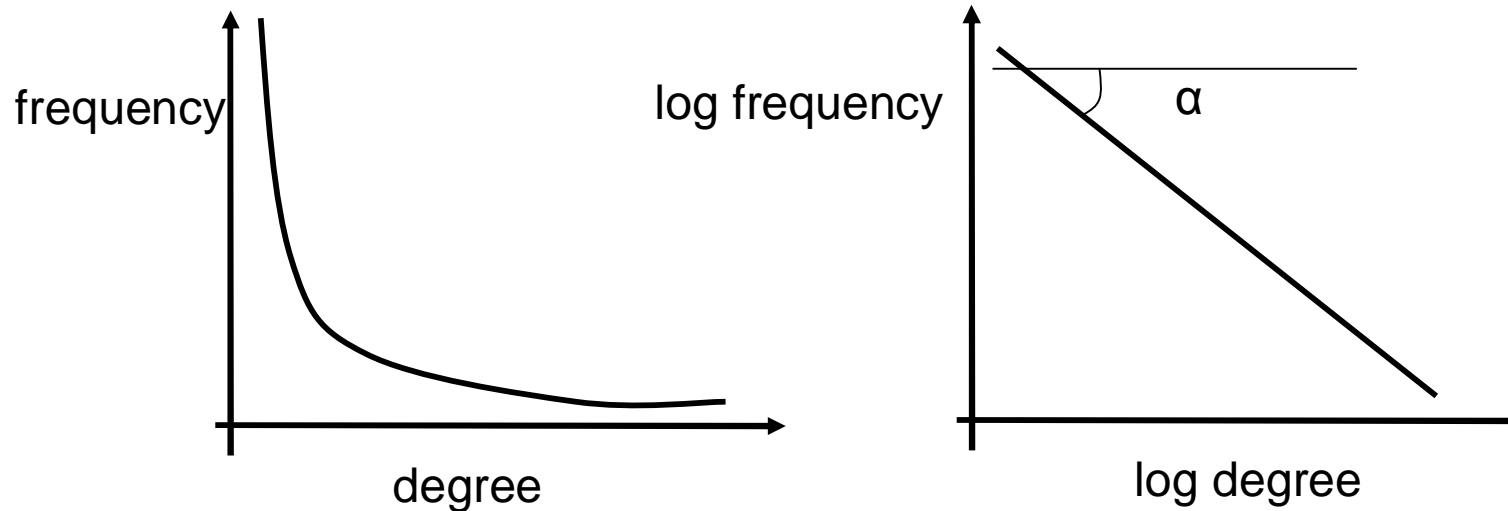
- § Right-skewed/Heavy-tail distribution
 - § there is a non-negligible fraction of nodes that has very high degree (hubs)
 - § **scale-free**: no characteristic scale, average is not informative
- § In stark contrast with the random graph model!
 - § highly concentrated around the mean
 - § the probability of very high degree nodes is exponentially small



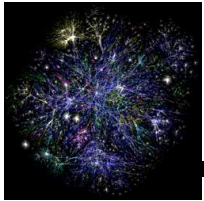
Power-law signature

§ Power-law distribution gives a line in the log-log plot

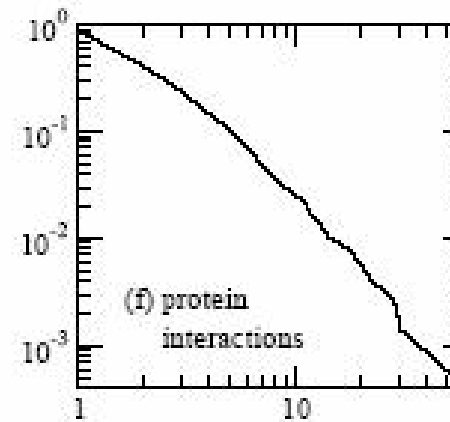
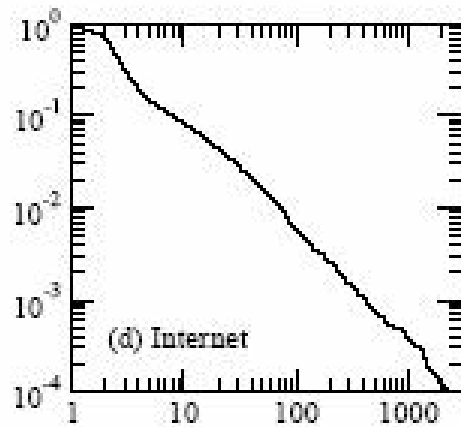
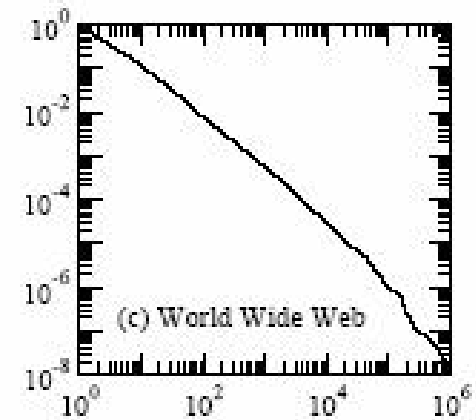
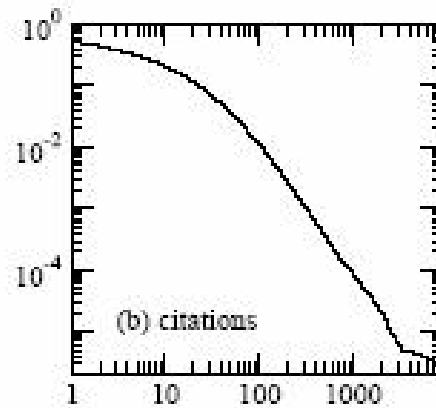
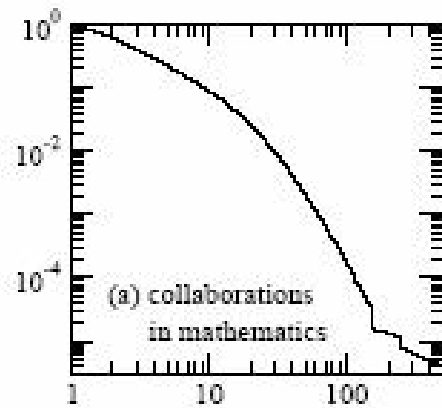
$$\log p(k) = -\alpha \log k + \log C$$



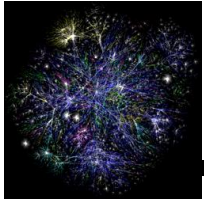
§ α : power-law exponent (typically $2 \leq \alpha \leq 3$)



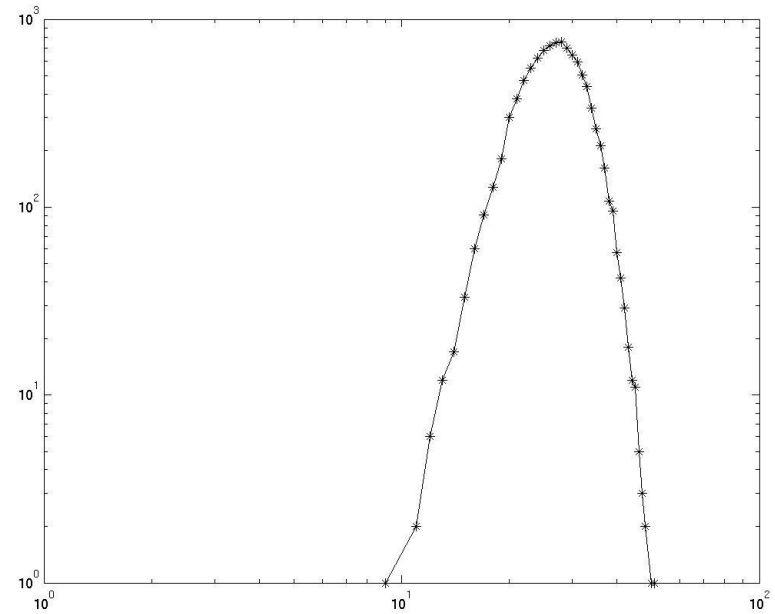
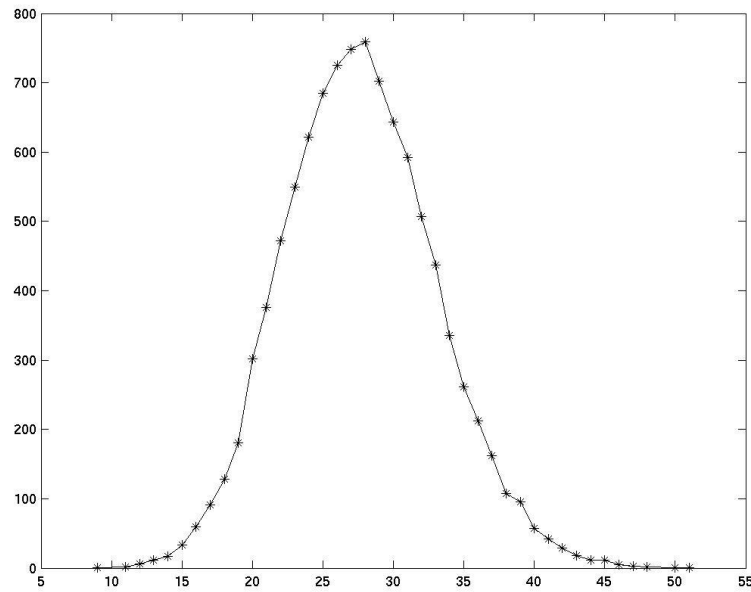
Examples



Taken from [Newman 2003]



A random graph example





Maximum degree

§ For random graphs, the maximum degree is highly concentrated around the average degree z

§ For power law graphs

$$k_{\max} \approx n^{1/(a-1)}$$

§ Rough argument: solve $nP[X \geq k] = 1$



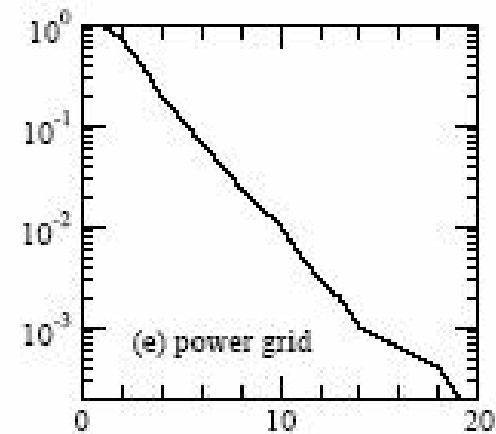
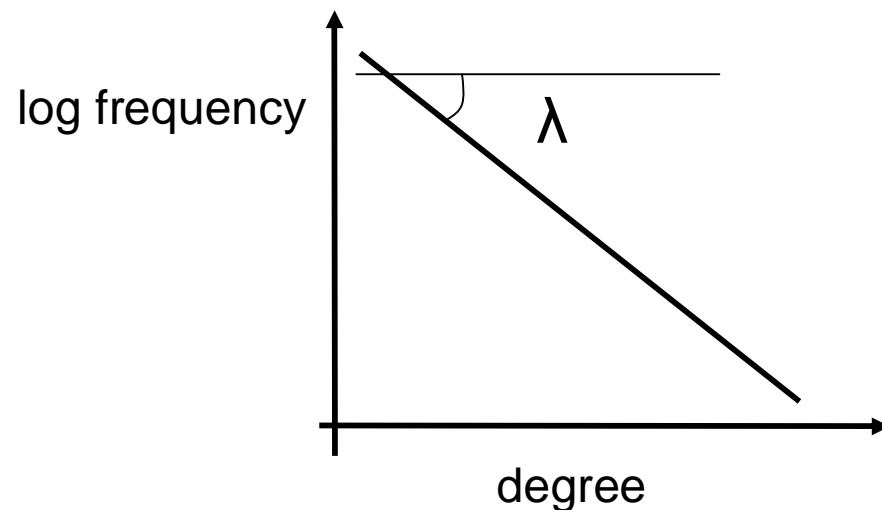
Exponential distribution

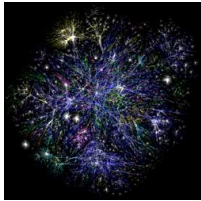
§ Observed in some technological or collaboration networks

$$p(k) = \lambda e^{-\lambda k}$$

§ Identified by a line in the log-linear plot

$$\log p(k) = -\lambda k + \log \lambda$$





Collective Statistics (M. Newman 2003)

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
sexual contacts	undirected	2 810				3.2				265, 266	
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex-vertex distance ℓ ; exponent α of degree distribution if the distribution follows a power law (or “–” if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.



Clustering (Transitivity) coefficient

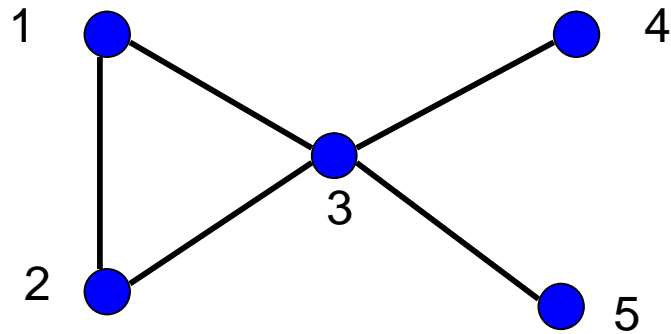
- § Measures the density of triangles (local clusters) in the graph
- § Two different ways to measure it:

$$C^{(1)} = \frac{\sum_i \text{triangles centered at node } i}{\sum_i \text{triples centered at node } i}$$

- § The ratio of the means



Example



$$C^{(1)} = \frac{3}{1+1+6} = \frac{3}{8}$$



Clustering (Transitivity) coefficient

§ Clustering coefficient for node i

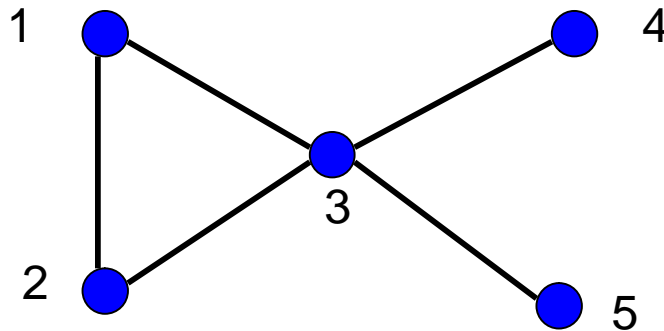
$$C_i = \frac{\text{triangles centered at node } i}{\text{triples centered at node } i}$$

$$C^{(2)} = \frac{1}{n} C_i$$

§ The mean of the ratios



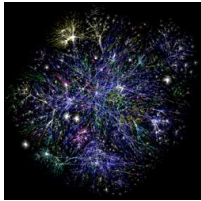
Example



$$C^{(2)} = \frac{1}{5} (1 + 1 + 1/6) = \frac{13}{30}$$

$$C^{(1)} = \frac{3}{8}$$

- § The two clustering coefficients give different measures
- § $C^{(2)}$ increases with nodes with low degree



Collective Statistics (M. Newman 2003)

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
sexual contacts	undirected	2 810				3.2				265, 266	
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex-vertex distance ℓ ; exponent α of degree distribution if the distribution follows a power law (or “–” if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

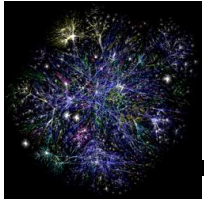


Clustering coefficient for random graphs

- § The probability of two of your neighbors also being neighbors is p , independent of local structure
 - § clustering coefficient $C = p$
 - § when z is fixed $C = z/n = O(1/n)$

Table 1: Clustering coefficients, C , for a number of different networks; n is the number of node, z is the mean degree. Taken from [146].

Network	n	z	C measured	C for random graph
Internet [153]	6,374	3.8	0.24	0.00060
World Wide Web (sites) [2]	153,127	35.2	0.11	0.00023
power grid [192]	4,941	2.7	0.080	0.00054
biology collaborations [140]	1,520,251	15.5	0.081	0.000010
mathematics collaborations [141]	253,339	3.9	0.15	0.000015
film actor collaborations [149]	449,913	113.4	0.20	0.00025
company directors [149]	7,673	14.4	0.59	0.0019
word co-occurrence [90]	460,902	70.1	0.44	0.00015
neural network [192]	282	14.0	0.28	0.049
metabolic network [69]	315	28.3	0.59	0.090
food web [138]	134	8.7	0.22	0.065



Small world phenomena

§ Small worlds: networks with short paths



Stanley Milgram (1933-1984):
“The man who shocked the world”

Obedience to authority (1963)

Small world experiment (1967)



Small world experiment

- § Letters were handed out to people in Nebraska to be sent to a target in Boston
- § People were instructed to pass on the letters to someone they knew on first-name basis
- § The letters that reached the destination followed paths of length around 6
- § **Six degrees of separation:** (play of John Guare)

- § Also:
 - § The Kevin Bacon game
 - § The Erdős number
- § Small world project:
<http://smallworld.columbia.edu/index.html>



Measuring the small world phenomenon

§ d_{ij} = shortest path between i and j

§ Diameter:

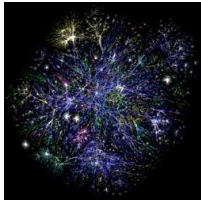
$$d = \max_{i,j} d_{ij}$$

§ Characteristic path length:

$$\bar{d} = \frac{1}{n(n-1)/2} \sum_{i>j} d_{ij}$$

§ Harmonic mean

$$\bar{d}^{-1} = \frac{1}{n(n-1)/2} \sum_{i>j} d_{ij}^{-1}$$



Collective Statistics (M. Newman 2003)

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
sexual contacts	undirected	2 810				3.2				265, 266	
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
biological	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex-vertex distance ℓ ; exponent α of degree distribution if the distribution follows a power law (or “–” if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.



Is the path length enough?

§ Random graphs have diameter

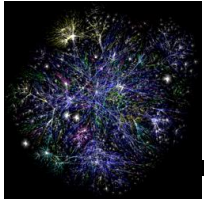
$$d = \frac{\log n}{\log z}$$

§ $d = \log n / \log \log n$ when $z = \omega(\log n)$

§ Short paths should be combined with other properties

§ ease of navigation

§ high clustering coefficient



Mixing patterns

§ Assume that we have various **types** of nodes.
What is the probability that two nodes of different type are linked?

§ assortative mixing (homophily)

E : mixing matrix

$p(i,j)$ = mixing probability

$$p(i, j) = \frac{E(i, j)}{\sum_{ij} E(i, j)}$$

$p(j | i)$ = conditional mixing probability

$$p(j | i) = \frac{E(i, j)}{\sum_j E(i, j)}$$

		women			
		black	hispanic	white	other
men	black	506	32	69	26
	hispanic	23	308	114	38
	white	26	46	599	68
	other	10	14	47	32

TABLE III Couples in the study of Catania *et al.* [85] tabulated by race of either partner. After Morris [302].



Mixing coefficient

§ Gupta, Anderson, May 1989

$$Q = \frac{\sum_i p(i|i) - 1}{N - 1}$$

§ Advantages:

§ Q=1 if the matrix is diagonal

§ Q=0 if the matrix is uniform

§ Disadvantages

§ sensitive to transposition

§ does not weight the entries



Mixing coefficient

§ Newman 2003

$$a(i) = \sum_j p(i, j) \quad (\text{row marginal})$$

$$b(i) = \sum_j p(j, i) \quad (\text{column marginal})$$

$$r = \frac{\sum_i p(i|i) - \sum_i a(i)b(i)}{N-1}$$

		women				a_i
		black	hispanic	white	other	
men	black	0.258	0.016	0.035	0.013	0.323
	hispanic	0.012	0.157	0.058	0.019	0.247
	white	0.013	0.023	0.306	0.035	0.377
	other	0.005	0.007	0.024	0.016	0.053
b_i		0.289	0.204	0.423	0.084	

TABLE I: The mixing matrix e_{ij} and the values of a_i and b_i for sexual partnerships in the study of Catania *et al.* [23]. After Morris [24].

$$r=0.621$$

$$Q=0.528$$

§ Advantages:

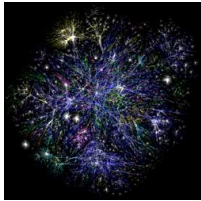
§ $r = 1$ for diagonal matrix , $r = 0$ for uniform matrix

§ not sensitive to transposition, accounts for weighting



Degree correlations

- § Do high degree nodes tend to link to high degree nodes?
- § Pastor Satoras et al.
 - § plot the mean degree of the neighbors as a function of the degree
- § Newman
 - § compute the correlation coefficient of the degrees of the two endpoints of an edge
 - § assortative/disassortative



Collective Statistics (M. Newman 2003)

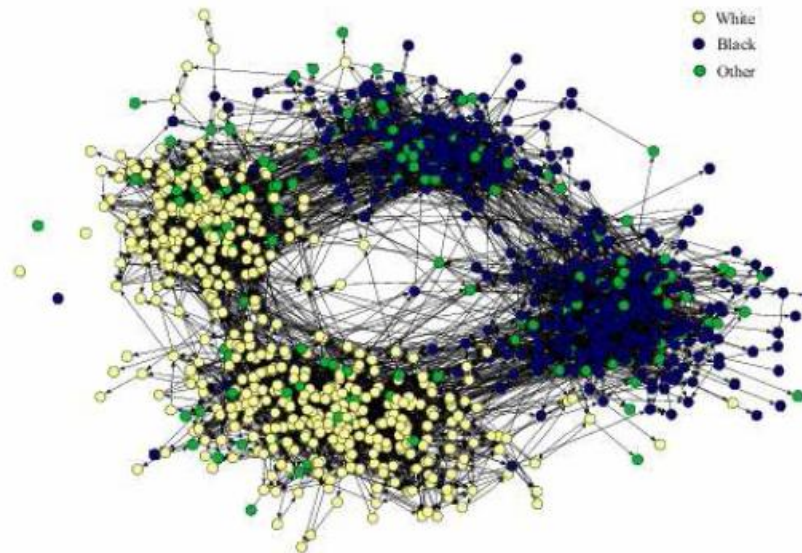
	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
sexual contacts	undirected	2 810				3.2				265, 266	
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
biological	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n ; total number of edges m ; mean degree z ; mean vertex-vertex distance ℓ ; exponent α of degree distribution if the distribution follows a power law (or “–” if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.



Communities and Clusters

- § Use the graph structure to discover communities of nodes
 - § essentially clustering and classification on graphs





Other measures

- § Frequent (or interesting) motifs
 - § bipartite cliques in the web graph
 - § patterns in biological and software graphs

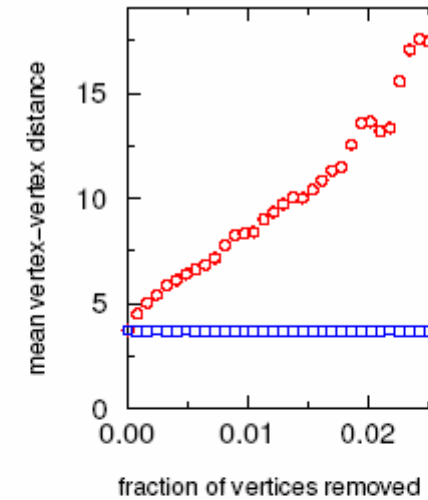
- § Use graphlets to compare models
[Przulj, Corneil, Jurisica 2004]



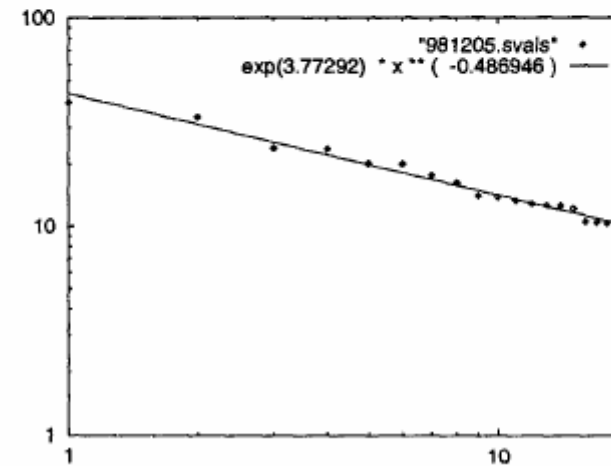
Other measures

§ Network resilience

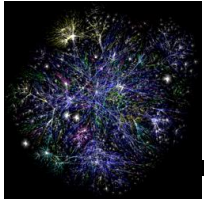
§ against random or targeted node deletions



§ Graph eigenvalues



(a) Int-12-98



Other measures

§ The giant component

§ Other?



References

- § M. E. J. Newman, **The structure and function of complex networks**, SIAM Reviews, 45(2): 167-256, 2003
- § M. E. J. Newman, **Random graphs as models of networks** in *Handbook of Graphs and Networks*, S. Bornholdt and H. G. Schuster (eds.), Wiley-VCH, Berlin (2003).
- § N. Alon J. Spencer, **The Probabilistic Method**