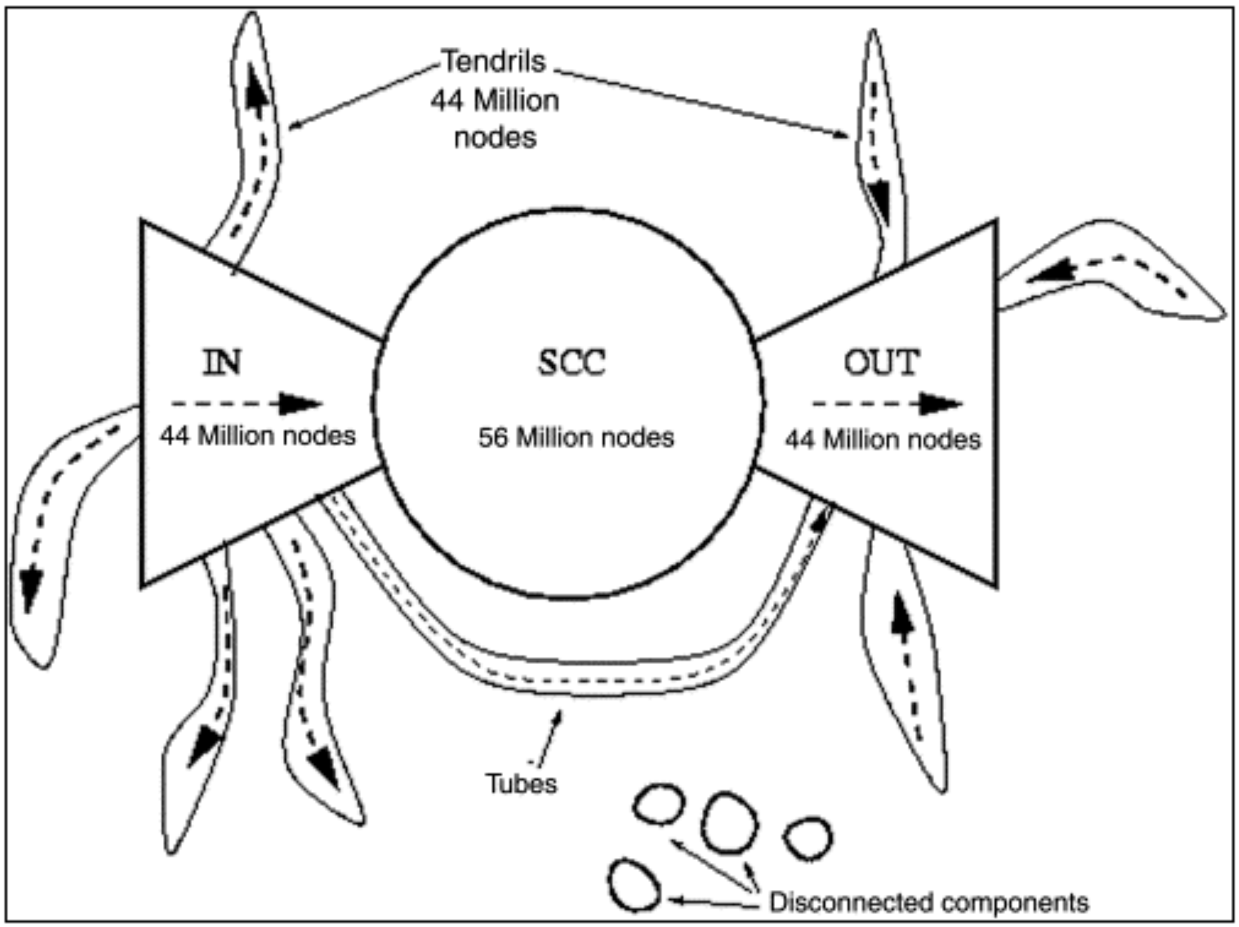
Graphs and Networks
Sep 7, 2010: Lecture 2
Daniel A. Spielman

Graph Structure in the Web

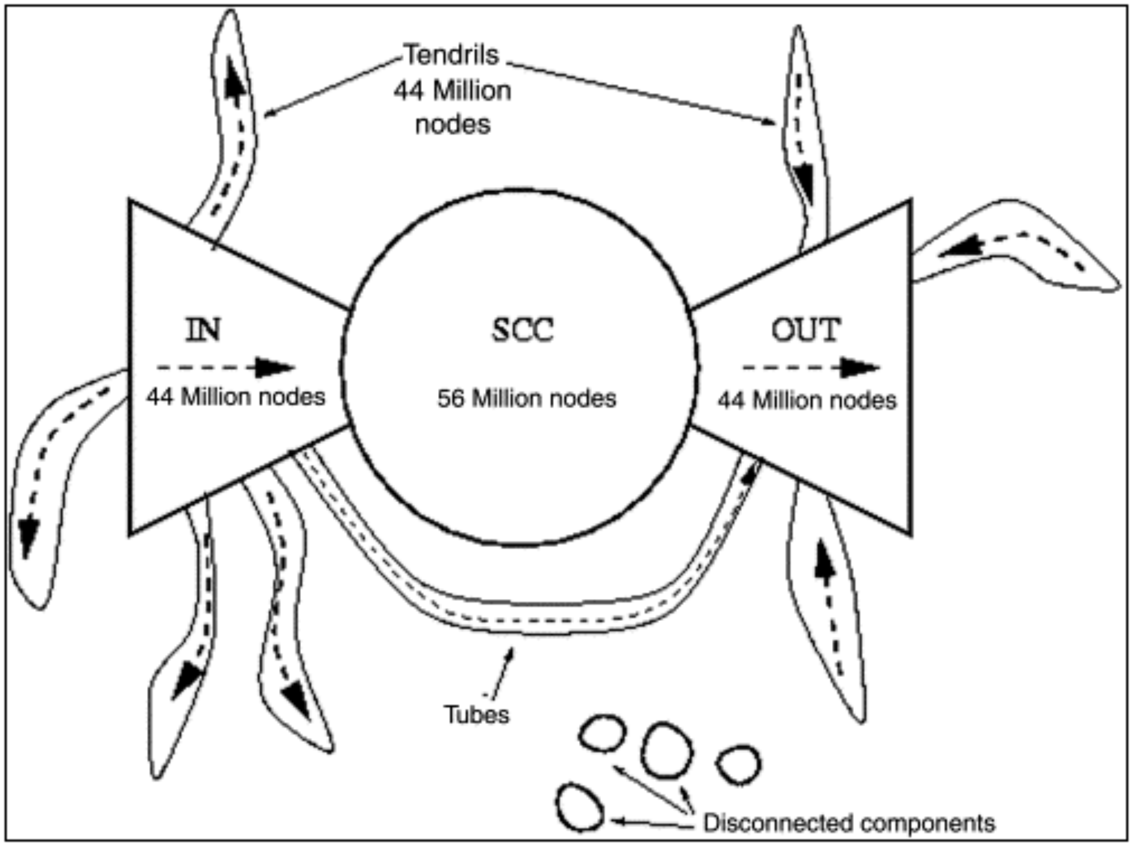
Broder et. al., Computer Networks 33, (2000) pp.309-320

Altavista Crawl of 200M pages, 1.5B links



Graph Structure in the Web

Broder et. al., Computer Networks 33, (2000) pp.309-320



For two random nodes, directed path exists with prob 25%

Graph Structure in the Web

Broder et. al., Computer Networks 33, (2000) pp.309-320

Distances when there is a path:

Edge type	In-links (directed)	Out-links (directed)	Undirected
Average connected distance	16.12	16.18	6.83

Breadth-first search from random nodes in SCC:

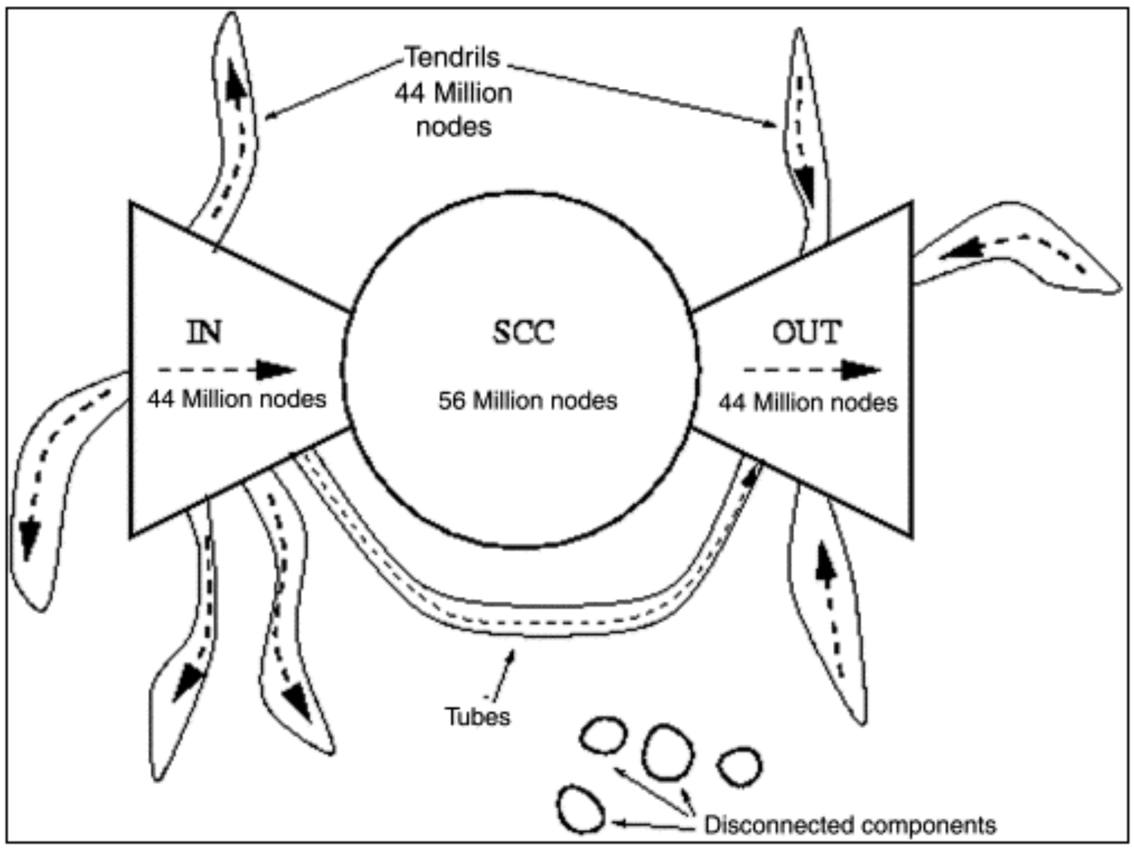
Measure	Minimum depth	Average depth	Maximum depth
In-links	475	482	503
Out-links	430	434	444

As the table shows, from some nodes in the SCC it is possible to complete the search at distance 475, while from other nodes distance 503 is required. This allows us to conclude that the directed diameter of SCC is at least 28.

Graph Structure in the Web

Broder et. al., Computer Networks 33, (2000) pp.309-320

Starting point	OUT	IN
Exploring outwards – all nodes	3093	171
Exploring inwards – unexpected nodes	3367	173



	MEDLINE	Los Alamos e-Print Archive				SPIRES	NCSTRL
		complete	astro-ph	cond-mat	hep-th		
total papers	2156769	98502	22029	22016	19085	66652	13169
total authors	1388989	52909	16706	16726	8361	56627	11994
first initial only	1006412	45685	14303	15451	7676	47445	10998
mean papers per author	5.5(4)	5.1(2)	4.8(2)	3.65(7)	4.8(1)	11.6(5)	2.55(5)
mean authors per paper	2.966(2)	2.530(7)	3.35(2)	2.66(1)	1.99(1)	8.96(18)	2.22(1)
collaborators per author	14.8(1.1)	9.7(2)	15.1(3)	5.86(9)	3.87(5)	173(6)	3.59(5)
size of giant component	1193488	44337	14845	13861	5835	49002	6396
first initial only	892193	39709	12874	13324	5593	43089	6706
as a percentage	87.3(7)%	85.4(8)%	89.4(3)	84.6(8)%	71.4(8)%	88.7(1.1)%	57.2(1.9)%
2nd largest component	56	18	19	16	24	69	42
mean distance	4.4(2)	5.9(2)	4.66(7)	6.4(1)	6.91(6)	4.0(1)	9.7(4)
maximum distance	21	20	14	18	19	19	31

The Structure of Scientific Collaboration Networks, M.E.J. Newman
http://arxiv.org/abs/cond-mat/0007214/

Diameter in scientific collaborations

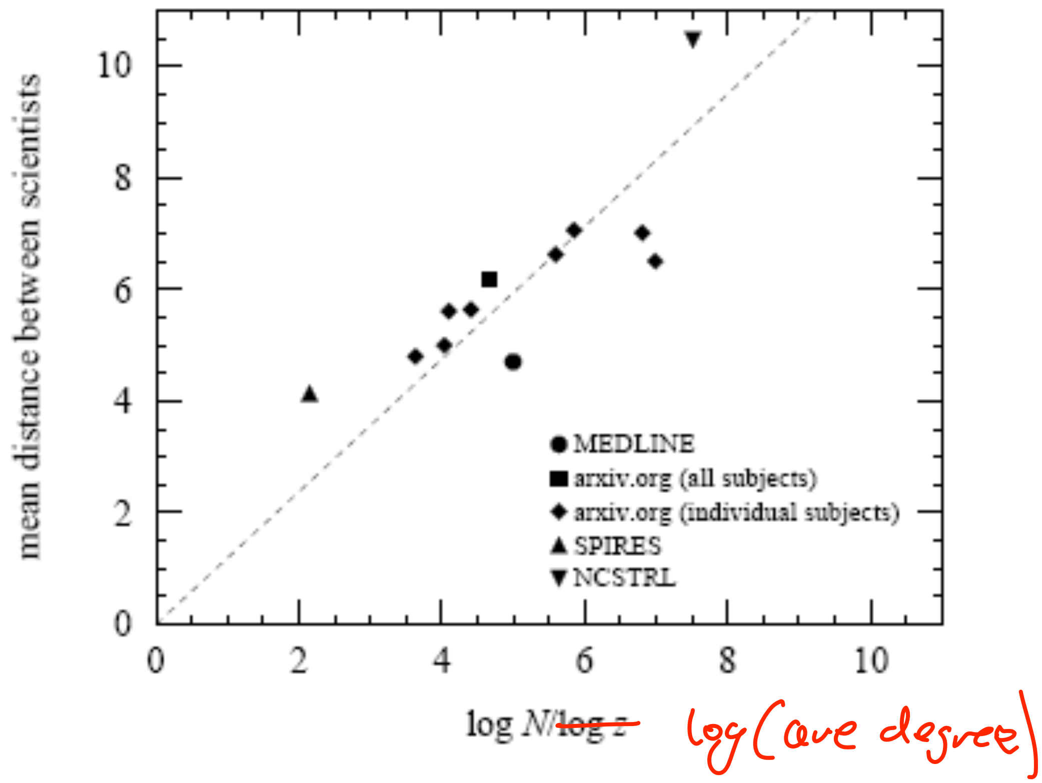


FIG. 3. Average distance between pairs of scientists in the various communities, plotted against the average distance on a random graph of the same size and average coordination number. The dotted line is the best fit to the data which also passes through the origin.

The Structure of Scientific Collaboration Networks, M.E.J. Newman
http://arxiv.org/abs/cond-mat/0007214/

	MEDLINE	Los Alamos e-Print Archive				SPIRES	NCSTRL
		complete	astro-ph	cond-mat	hep-th		
total papers	2156769	98502	22029	22016	19085	66652	13169
total authors	1388989	52909	16706	16726	8361	56627	11994
first initial only	1006412	45685	14303	15451	7676	47445	10998
mean papers per author	5.5(4)	5.1(2)	4.8(2)	3.65(7)	4.8(1)	11.6(5)	2.55(5)
mean authors per paper	2.966(2)	2.530(7)	3.35(2)	2.66(1)	1.99(1)	8.96(18)	2.22(1)
collaborators per author	14.8(1.1)	9.7(2)	15.1(3)	5.86(9)	3.87(5)	173(6)	3.59(5)
size of giant component	1193488	44337	14845	13861	5835	49002	6396
first initial only	892193	39709	12874	13324	5593	43089	6706
as a percentage	87.3(7)%	85.4(8)%	89.4(3)	84.6(8)%	71.4(8)%	88.7(1.1)%	57.2(1.9)%
2nd largest component	56	18	19	16	24	69	42
mean distance	4.4(2)	5.9(2)	4.66(7)	6.4(1)	6.91(6)	4.0(1)	9.7(4)
maximum distance	21	20	14	18	19	19	31
clustering coefficient C	0.072(8)	0.43(1)	0.414(6)	0.348(6)	0.327(2)	0.726(8)	0.496(6)

The Structure of Scientific Collaboration Networks, M.E.J. Newman
http://arxiv.org/abs/cond-mat/0007214/

	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

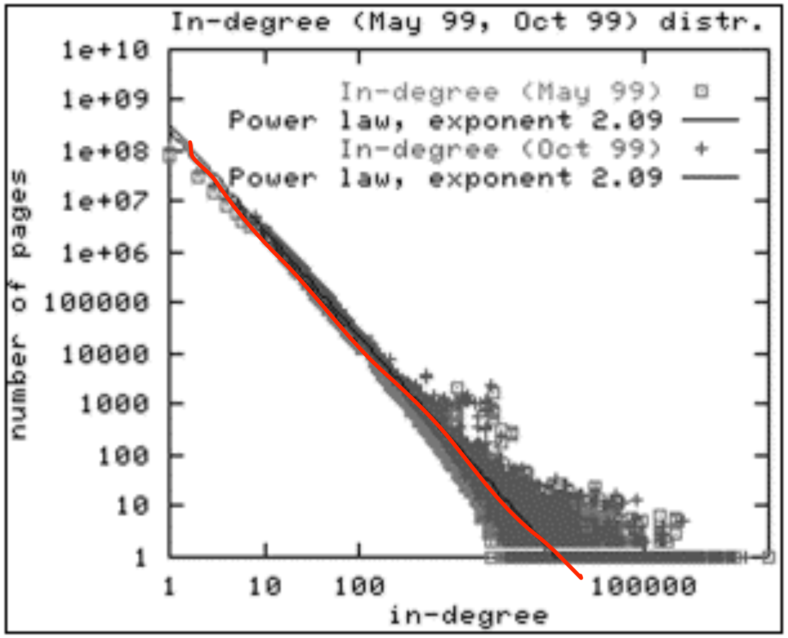
The Structure and Function of Complex Networks

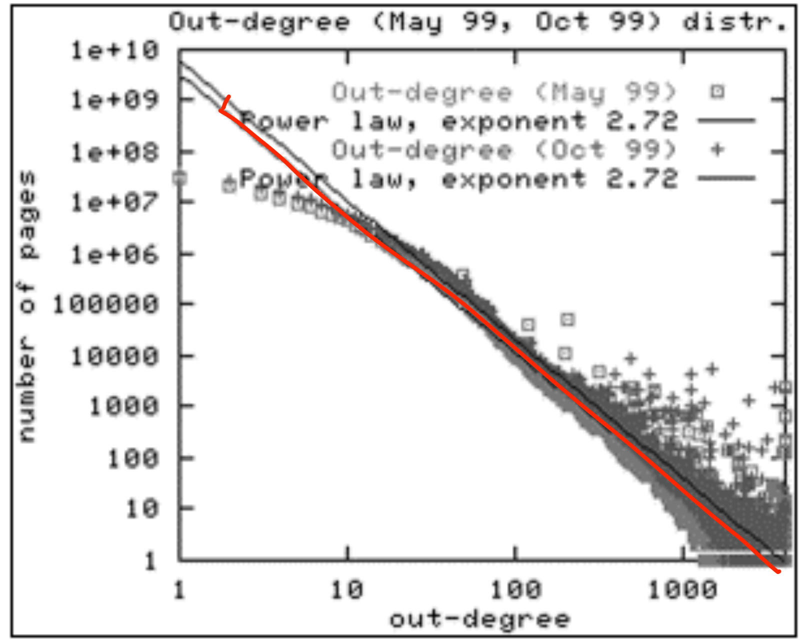
M.E.J. Newman, cond-mat/0303516 v1 15 Mar 2003

	Group	Network	Type	Size n	Assortativity r	Error σ_r
Social	a	Physics coauthorship	undirected	52 909	0.363	0.002
	a	Biology coauthorship	undirected	1 520 251	0.127	0.0004
	b	Mathematics coauthorship	undirected	253 339	0.120	0.002
	c	Film actor collaborations	undirected	449 913	0.208	0.0002
	d	Company directors	undirected	7 673	0.276	0.004
	e	Student relationships	undirected	573	−0.029	0.037
	f	Email address books	directed	16 881	0.092	0.004
Technological	g	Power grid	undirected	4 941	−0.003	0.013
	h	Internet	undirected	10 697	−0.189	0.002
	i	World Wide Web	directed	269 504	−0.067	0.0002
	j	Software dependencies	directed	3 162	−0.016	0.020
Biological	k	Protein interactions	undirected	2 115	−0.156	0.010
	l	Metabolic network	undirected	765	−0.240	0.007
	m	Neural network	directed	307	−0.226	0.016
	n	Marine food web	directed	134	−0.263	0.037
	o	Freshwater food web	directed	92	−0.326	0.031

Mixing Patterns In Networks, M.E.J. Newman
Phys. Rev. E 67, 026126 (2003)

Power-law degree distributions





Graph structure in the Web, Broder et. al.,

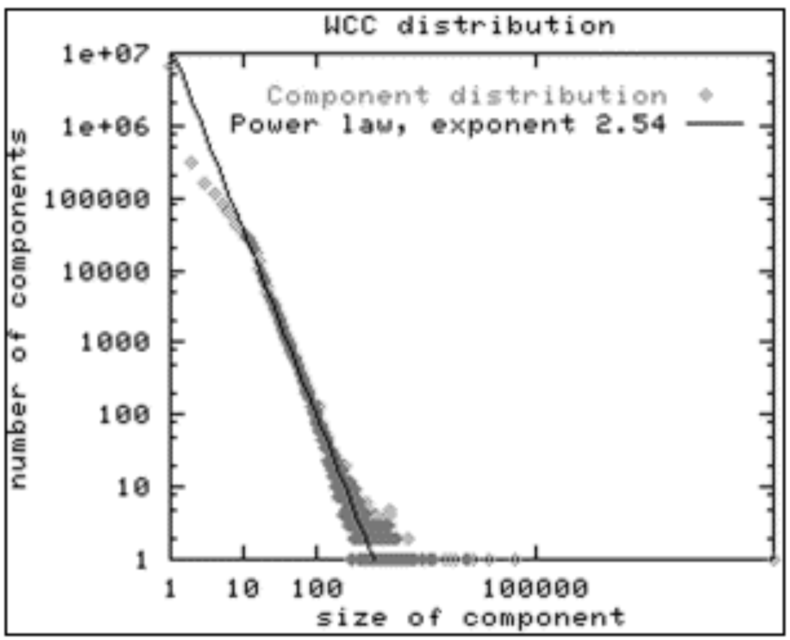
Computer Networks Vol 33, No 1-6 , June 2000, pp. 309-320

Analysis of Web Graph

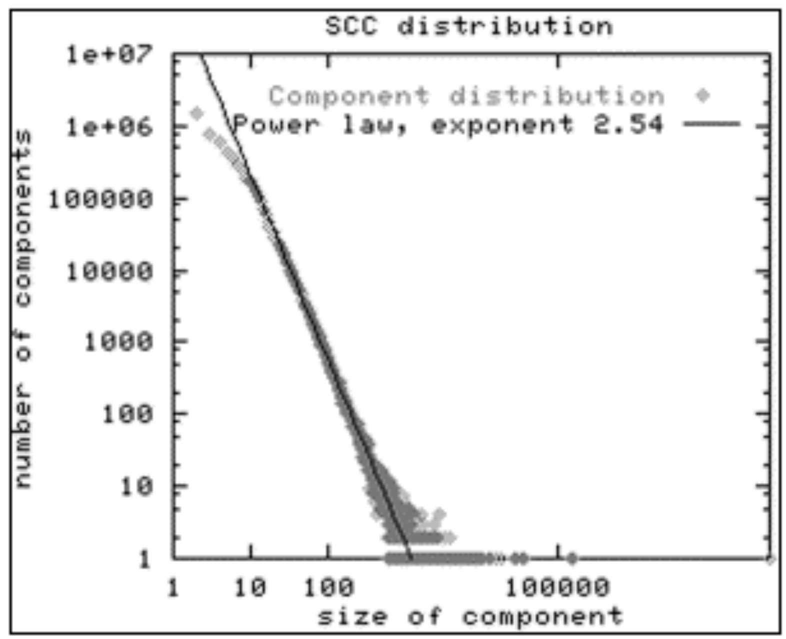
Graph structure in the Web, Broder et. al.,

Computer Networks Vol 33, No 1-6 , June 2000, pp. 309-320

Weakly-connected components
(traverse edge either way)
largest had 186m pages = 91%



Strongly-connected components
(only following links)
largest had 56m pages = 28%



Power-law degree distributions

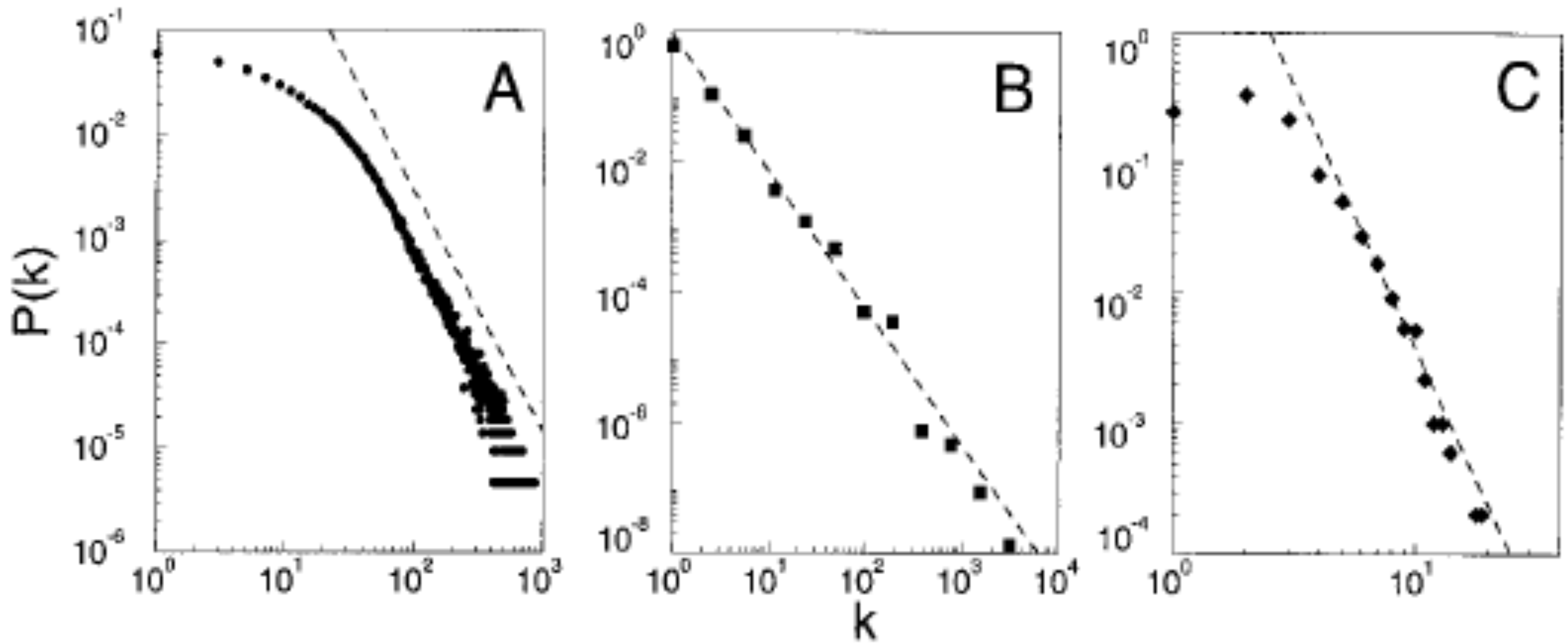


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Emergence of Scaling in Random Networks, Barabasi and Albert, Science, vol 286, 15 Oct 1999

Power-Law Degree Distributions?

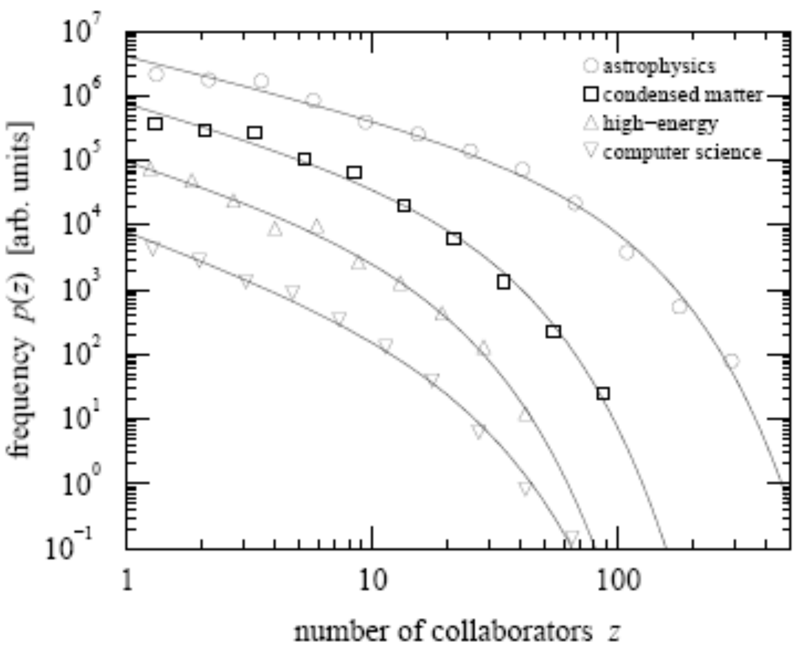


FIG. 1. Histograms of the number of collaborators of scientists in four of the databases studied here. The solid lines are least-squares fits to Eq. (1).

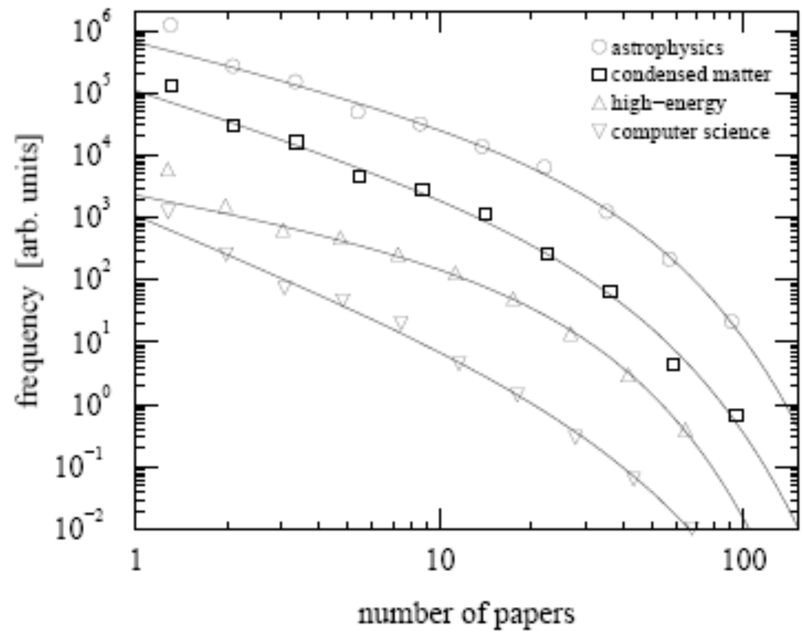


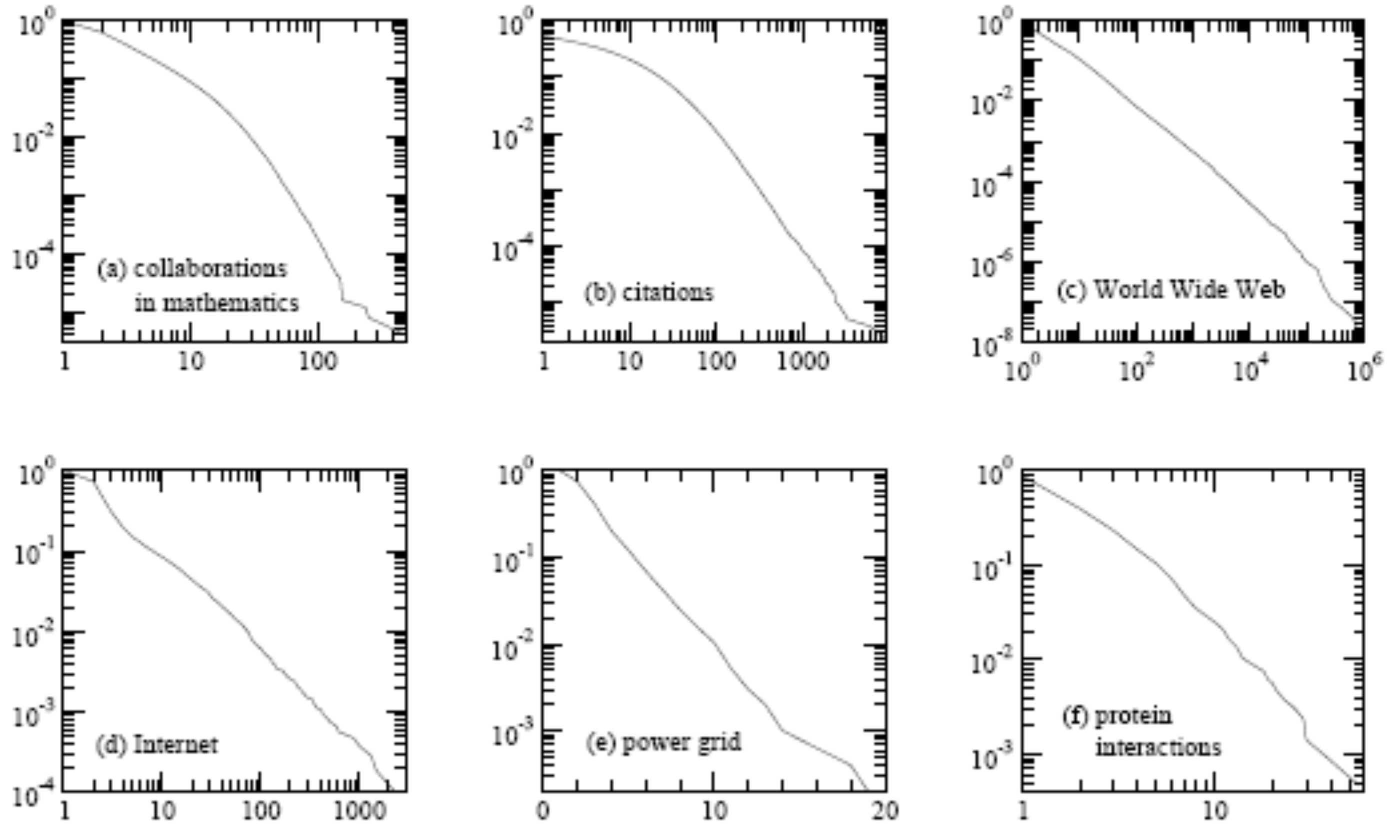
FIG. 2. Histograms of the number of papers written by scientists in four of the databases. As with Fig. 1, the solid lines are least-squares fits to Eq. (1).

used. However, our data are well fitted by a power-law form with an exponential cutoff:

$$P(z) \sim p^{-\tau} \mathrm{e}^{-z/z_c}, \qquad (1)$$

The Structure of Scientific Collaboration Networks, M.E.J. Newman
http://arxiv.org/abs/cond-mat/0007214/

Power-Law Degree Distributions?



The Structure and Function of Complex Networks
M.E.J. Newman, cond-mat/0303516 v1 15 Mar 2003