Random Models

- Model $G(n,m)$ is a probability distribution $P(G)$ over all graphs with n nodes and m edges.
  - Properties of model = properties of ensemble
    - Examples
      - graph diameter $l(G)$ means

$$\langle l \rangle = \sum_G P(G)\, l(G) = \frac{1}{\Omega} \sum_G l(G)$$

Total number of graphs

- degree $\langle d(\cdot) \rangle = 2m/n$

Q: what is the total number of graphs with no loops and multi-edges?

$2^{\binom{n}{2}}$

Random Models

- Model $G(n,p)$ - graphs with n nodes and independent probability p for placing an edge between two vertices (aka Erdös-Rényi model).

- Properties of model = properties of ensemble where a particular graph G with m edges appears with probability

$$P(G) = p^m (1-p)^{\binom{n}{2}-m}$$

and probablity of drawing a graph with m edges from the ensemble is

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} \text{ and } \langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m P(m) = \binom{n}{2} p$$

- mean degree $\displaystyle\sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n}\binom{n}{2} p = (n-1)p = c$

mean degree in a graph with exactly *m* edges

- degree distribution
  - node is connected to a particular k others $q_k = p^k (1-p)^{n-1-k}$
  - node is connected to exacly k others $p_k = \binom{n-1}{k} q_k$

- mean degree $\sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n}\binom{n}{2} p = (n-1)p = c$

mean degree in a graph with exactly *m* edges

- degree distribution
  - node is connected to a particular k others $q_k = p^k (1-p)^{n-1-k}$
  - node is connected to exacly k others $p_k = \binom{n-1}{k} q_k$
  - in large-scale networks $p = c/(n-1)$ can be very small, i.e.,

$$\ln((1-p)^{n-1-k}) = (n-1-k)\ln(1-c/(n-1)) \approx -(n-1-k)\frac{c}{n-1} \approx -c$$

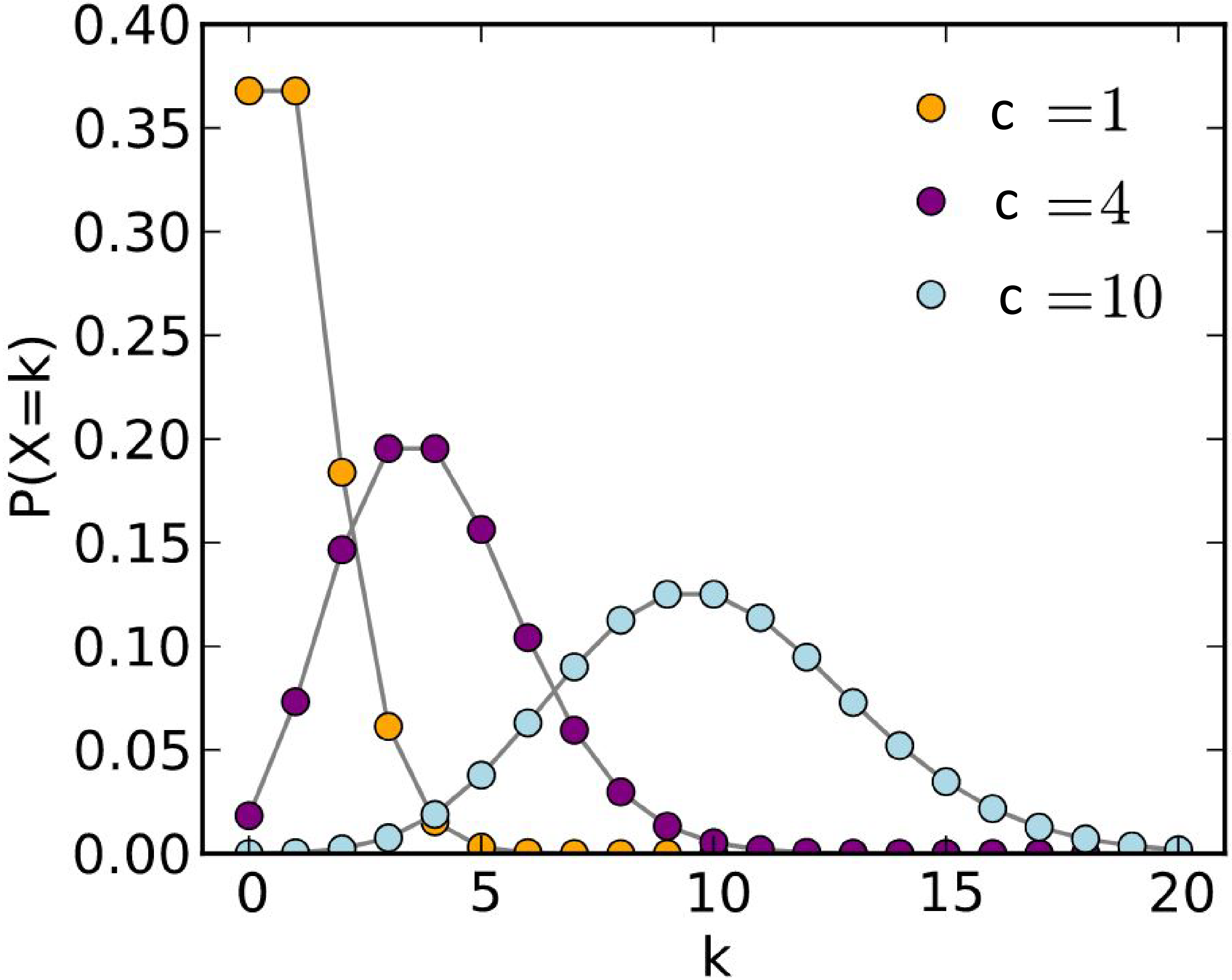
∞

Taylor series reminder: $\ln(1+\frac{1}{x}) = 2\left(A + \frac{1}{3}A^3 + \frac{1}{5}A^5 + \ldots\right)$, where $A = \frac{1}{2x+1}$

also if $\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)!k!} \approx \frac{(n-1)^k}{k!}$ then

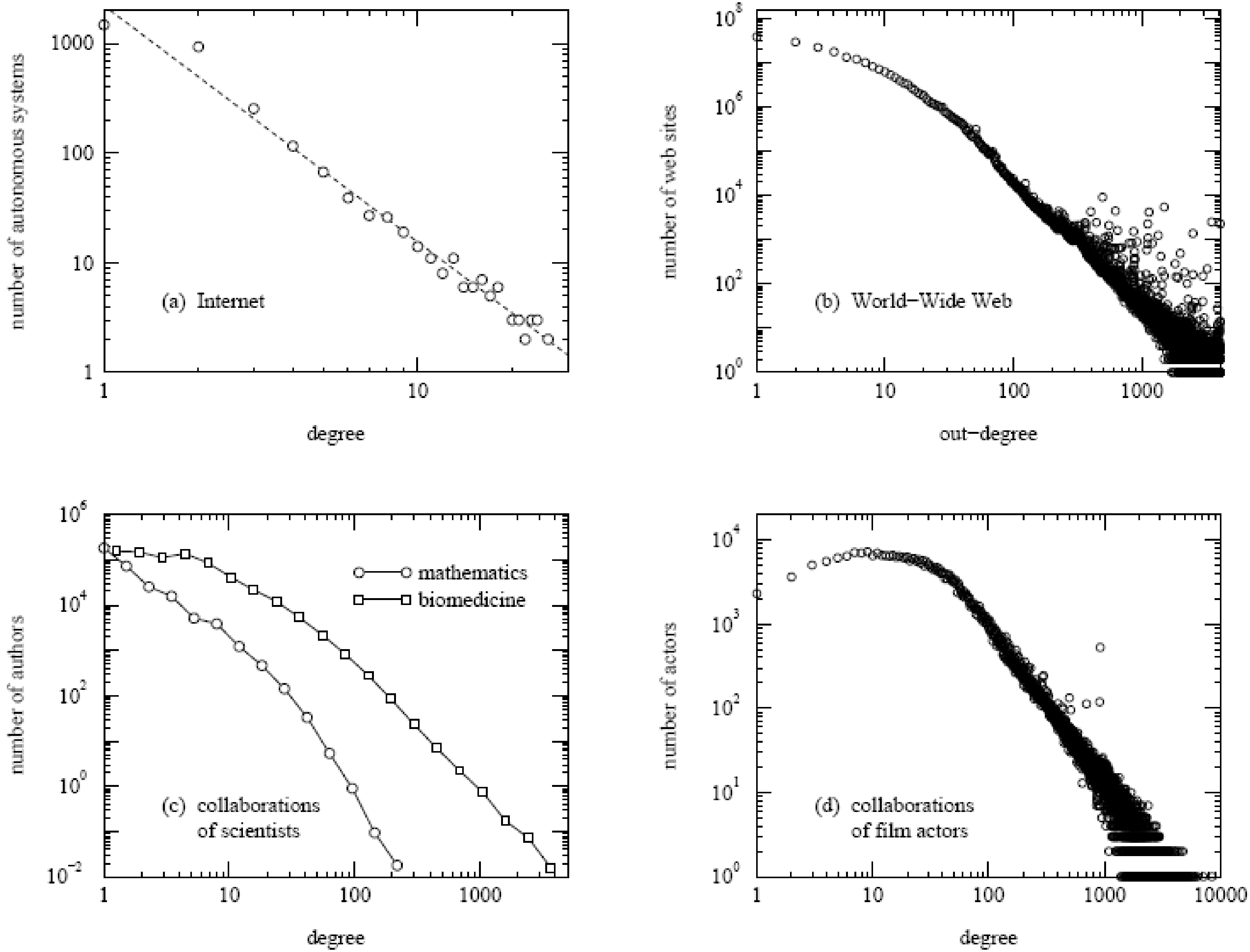
$$\boxed{p_k = \frac{(n-1)^k}{k!} p^k e^{-c} = \frac{(n-1)^k}{k!}\left(\frac{c}{n-1}\right)^k e^{-c} = e^{-c}\frac{c^k}{k!}}$$

Poisson distribution in random models

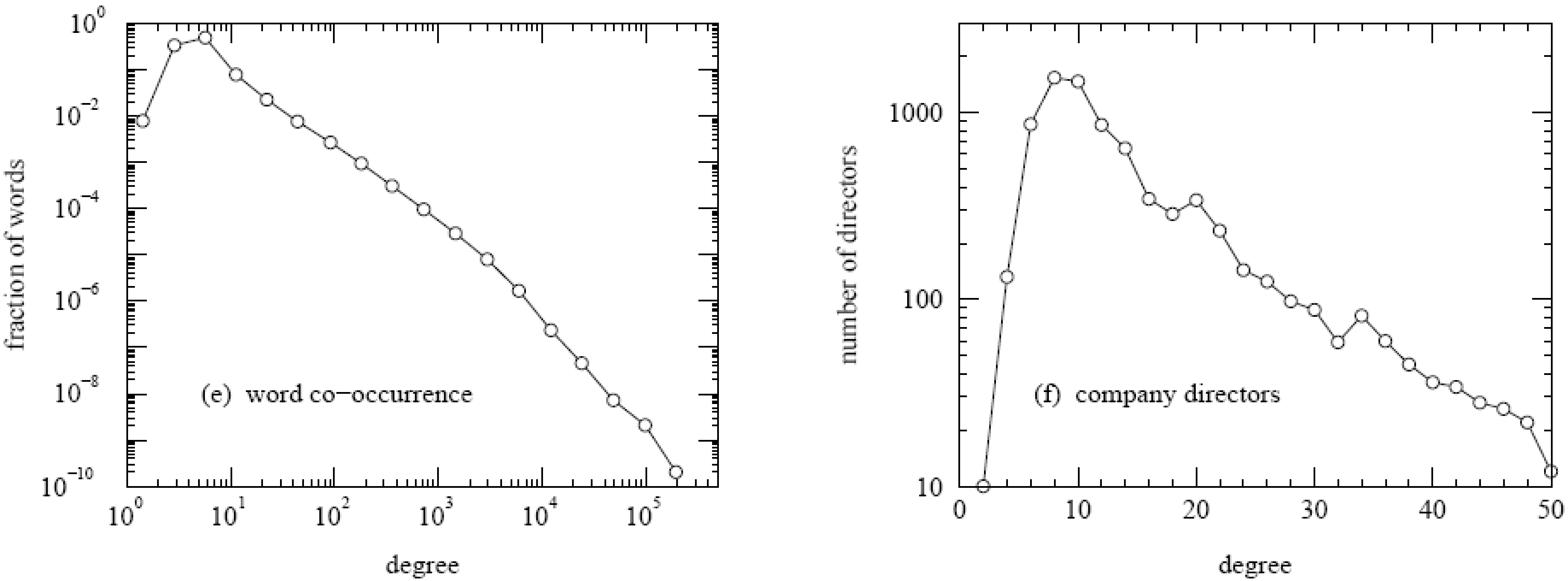


$$p_k = e^{-c}\frac{c^k}{k!}$$

In contrast to the degree distribution in random model, many real network degree distributions are different.



In contrast to the degree distribution in random model, many real network degree distributions are different.



- clustering coefficient $C = c/(n-1)$ = prob that any two nodes are neighbors

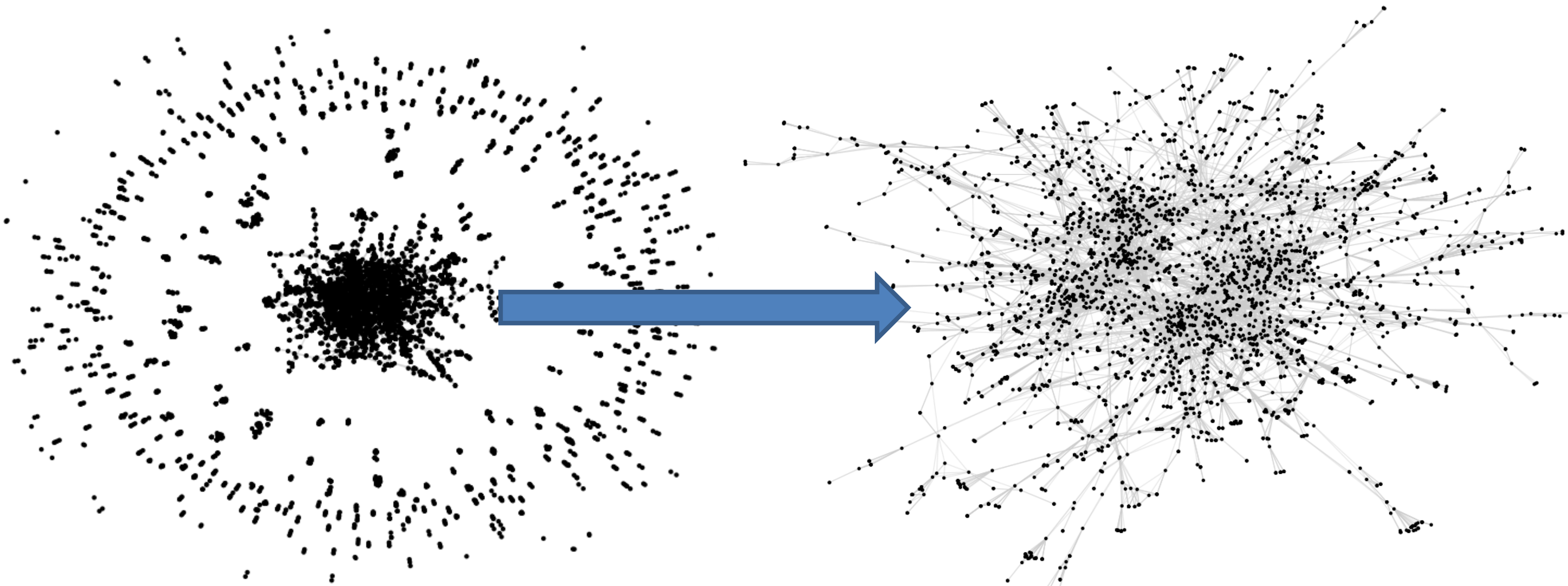
network	n	z	clustering coefficient C measured	clustering coefficient C random graph
Internet (autonomous systems)[a]	6 374	3.8	0.24	0.00060
World-Wide Web (sites)[b]	153 127	35.2	0.11	0.00023
power grid[c]	4 941	2.7	0.080	0.00054
biology collaborations[d]	1 520 251	15.5	0.081	0.000010
mathematics collaborations[e]	253 339	3.9	0.15	0.000015
film actor collaborations[f]	449 913	113.4	0.20	0.00025
company directors[f]	7 673	14.4	0.59	0.0019
word co-occurrence[g]	460 902	70.1	0.44	0.00015
neural network[c]	282	14.0	0.28	0.049
metabolic network[h]	315	28.3	0.59	0.090
food web[i]	134	8.7	0.22	0.065

Newman, "Random graphs as models of networks"

- giant component in $G(n,p)$

  Giant component is a network component whose size grows in proportion to n.

Q: When p=0 then |gc|=1; when p=1 then |gc|=n. What is the difference between them?



Co-authorship network — its largest connected component

- giant component in $G(n,p)$

Giant component is a network component whose size grows in proportion to n.
u = avg fraction of vertices that do not belong to the giant component.

Q: When p=0 then |gc|=1; when p=1 then |gc|=n. Is this transition smooth? Is there a point of transition?

i does not belong to gc if for every node *j* either

a) there is no edge *ij*, prob = *1-p*

b) there is edge *ij* but *j* is not in giant component, prob = *pu*

$\Pr[i \text{ does not belong to gc via } j] = 1 - p + pu$, i.e.,
total probability of not being connected to gc via any of $n-1$ other vertices is

$$u = (1-p+pu)^{n-1} = \left(1 - \frac{c}{n-1}(1-u)\right)^{n-1}$$

$$\ln u \overset{n\to\infty}{\approx} -(n-1)\frac{c}{n-1}(1-u) = -c(1-u) \;\Rightarrow\; u = e^{-c(1-u)} \;\Rightarrow\; S = 1 - e^{-cS}$$

vertices in giant component

$$S = 1 - e^{-cS}$$

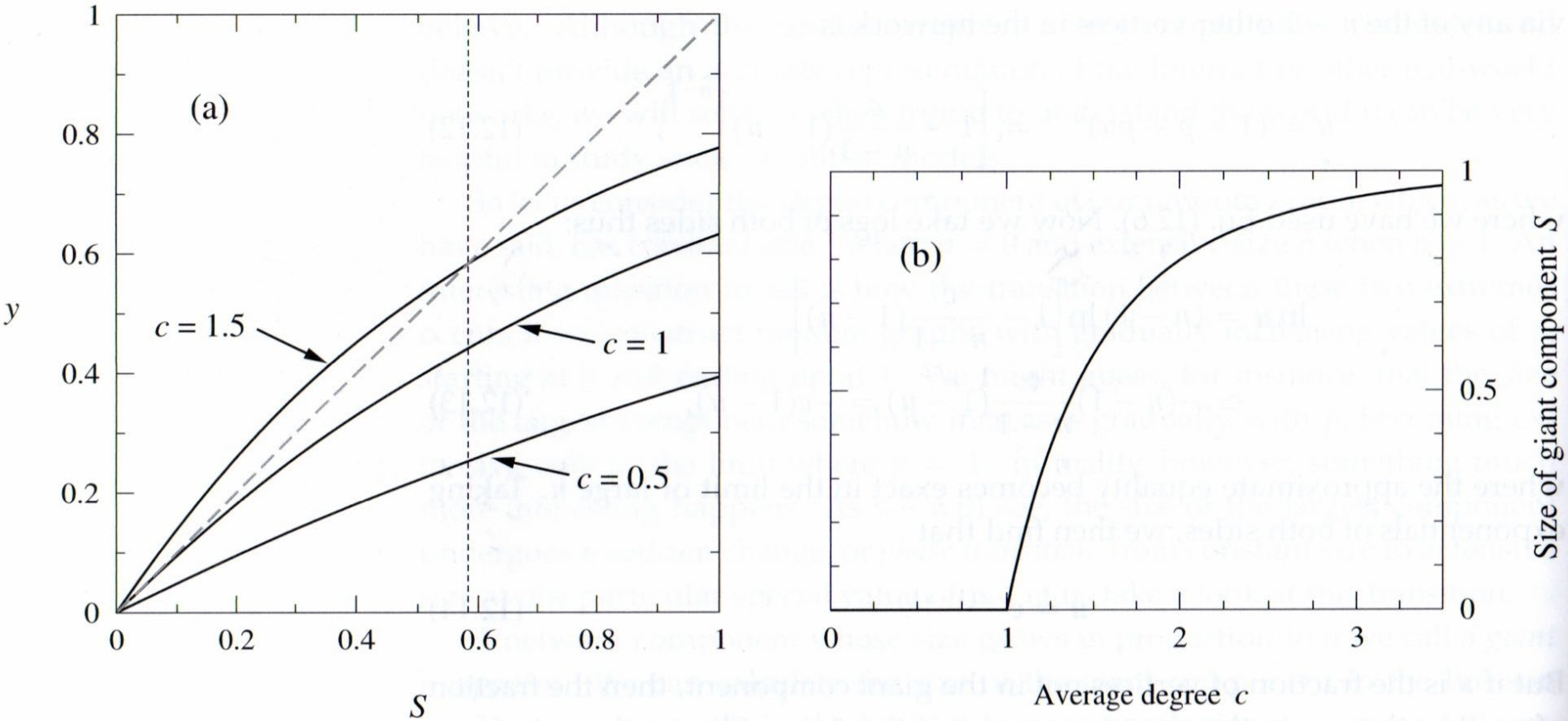


Figure 12.1: Graphical solution for the size of the giant component. (a) The three curves in the left panel show $y = 1 - \mathrm{e}^{-cS}$ for values of c as marked, the diagonal dashed line shows $y = S$, and the intersection gives the solution to Eq. (12.15), $S = 1 - \mathrm{e}^{-cS}$. For the bottom curve there is only one intersection, at $S = 0$, so there is no giant component, while for the top curve there is a solution at $S = 0.583\ldots$ (vertical dashed line). The middle curve is precisely at the threshold between the regime where a non-trivial solution for S exists and the regime where there is only the trivial solution $S = 0$. (b) The resulting solution for the size of the giant component as a function of c.

=> Demo in Matlab

Newman "Networks, An Introduction"

	Medline	Physics E-print Archive complete	astro-ph	cond-mat	hep-th	SPIRES	NCSTRL
total papers	2163923	98502	22029	22016	19085	66652	13169
total authors	1520251	52909	16706	16726	8361	56627	11994
first initial only	1090584	45685	14303	15451	7676	47445	10998
mean papers per author	6.4(6)	5.1(2)	4.8(2)	3.65(7)	4.8(1)	11.6(5)	2.55(5)
mean authors per paper	3.754(2)	2.530(7)	3.35(2)	2.66(1)	1.99(1)	8.96(18)	2.22(1)
collaborators per author	18.1(1.3)	9.7(2)	15.1(3)	5.86(9)	3.87(5)	173(6)	3.59(5)
size of giant component	1395693	44337	14845	13861	5835	49002	6396
first initial only	1019418	39709	12874	13324	5593	43089	6706
as a percentage	92.6(4)%	85.4(8)%	89.4(3)	84.6(8)%	71.4(8)%	88.7(1.1)%	57.2(1.9)%
2nd largest component	49	18	19	16	24	69	42
clustering coefficient C	0.066(7)	0.43(1)	0.414(6)	0.348(6)	0.327(2)	0.726(8)	0.496(6)
mean distance	4.6(2)	5.9(2)	4.66(7)	6.4(1)	6.91(6)	4.0(1)	9.7(4)
maximum distance	24	20	14	18	19	19	31

Table 1: Summary of results of the analysis of seven scientific collaboration networks. Numbers in parentheses give an estimate of the error on the least significant figures.

	Network	Type	n	m	c	S
Social	Film actors	Undirected	449 913	25 516 482	113.43	0.980
	Company directors	Undirected	7 673	55 392	14.44	0.876
	Math coauthorship	Undirected	253 339	496 489	3.92	0.822
	Physics coauthorship	Undirected	52 909	245 300	9.27	0.838
	Biology coauthorship	Undirected	1 520 251	11 803 064	15.53	0.918
	Telephone call graph	Undirected	47 000 000	80 000 000	3.16	
	Email messages	Directed	59 812	86 300	1.44	0.952
	Email address books	Directed	16 881	57 029	3.38	0.590
	Student dating	Undirected	573	477	1.66	0.503
	Sexual contacts	Undirected	2 810			
Information	WWW `nd.edu`	Directed	269 504	1 497 135	5.55	1.000
	WWW AltaVista	Directed	203 549 046	1 466 000 000	7.20	0.914
	Citation network	Directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	Directed	1 022	5 103	4.99	0.977
	Word co-occurrence	Undirected	460 902	16 100 000	66.96	1.000
Technological	Internet	Undirected	10 697	31 992	5.98	1.000
	Power grid	Undirected	4 941	6 594	2.67	1.000
	Train routes	Undirected	587	19 603	66.79	1.000
	Software packages	Directed	1 439	1 723	1.20	0.998
	Software classes	Directed	1 376	2 213	1.61	1.000
	Electronic circuits	Undirected	24 097	53 248	4.34	1.000
	Peer-to-peer network	Undirected	880	1 296	1.47	0.805
Biological	Metabolic network	Undirected	765	3 686	9.64	0.996
	Protein interactions	Undirected	2 115	2 240	2.12	0.689
	Marine food web	Directed	134	598	4.46	1.000
	Freshwater food web	Directed	92	997	10.84	1.000
	Neural network	Directed	307	2 359	7.68	0.967

Two giant components in $G(n,p)$?

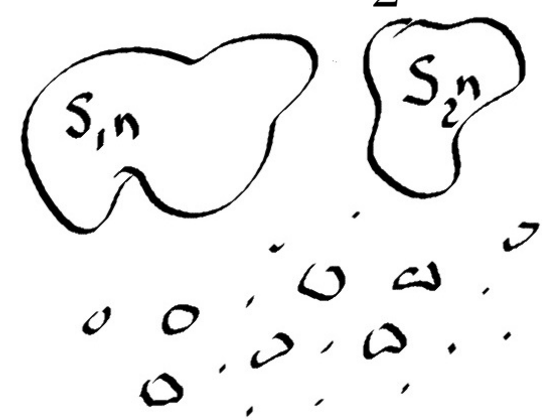
- Generate G with $p = c/(n-1)$
- Suppose that after adding edges with prob p, we have 2 giant components that cover fractions of nodes S_1, and S_2
- S_1 and S_2 remain separate with probability

$$q = (1-p)^{S_1 S_2 n^2} = (1 - \frac{c}{n-1})^{S_1 S_2 n^2}$$

$$\Rightarrow \quad \ln q = S_1 S_2 \lim_{n\to\infty} (n^2 \ln(1 - \frac{c}{n-1})) = S_1 S_2(-c(n+1) + \frac{1}{2}c^2)$$

$$q = q_0 e^{-cS_1 S_2 n},$$

where q_0 is a constant, i.e., $q \overset{n\to\infty}{\rightarrow} 0$



Conclusion: In the limit of large *n*, the probability of existence of two separate giant components goes to zero.

• Alright ... we have only one giant component. What about the sizes of small components?

π_s is the probability that randomly chosen node belongs to a small component of size s.

- We cannot normalize π_s to unity because some nodes may belong to the giant component, i.e.,

$$\sum_{s=0}^{\infty} \pi_s = 1 - S.$$

fraction of nodes in giant component

- **Observation**: small components are likely to be trees.
  Consider a small tree component of s nodes. The total number of places we can add an extra edge to is $\binom{s}{2} - (s-1)$

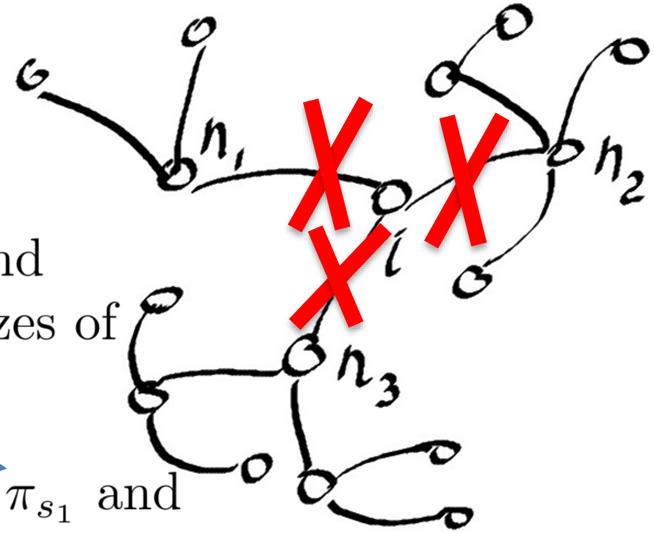
edges in tree

edge prob

$$\text{Average total number of added edges } \frac{1}{2}(s-1)(s-2) \cdot \frac{c}{n-1} \xrightarrow{n\to\infty} 0$$

the component is still tree

Calculation of π_s (the probability that randomly chosen node belongs to a small component of size s).



- Consider node i in a small (tree) component
- ... and modified network with deleted i.
  In the modified network, prob p is the same and in the limit of n the changes are negligible. Sizes of gc and sc will be indistinguishable for same p.
- Suppose $d(i) = k$ and $\Pr[n_1 \in \text{ sc of size } s_1] = \pi_{s_1}$ and

$$\Pr[\forall j \in N(i)\ n_j \in \text{ sc of size } s_j] = \Pi_{j=1}^{k} \pi_{s_j}$$

Since $\sum_{j \in N(i)} s_j = s - 1$ we have

Kronecker delta

$p_k = e^{-c}\frac{c^k}{k!}$

$$\Pr[s|k] = \sum_{s_1=1}^{\infty} \cdots \sum_{s_k=1}^{\infty} \left(\Pi_{j=1}^{k} \pi_{s_j}\right) \delta(s-1, \sum_j s_j)$$

$$\pi_s = \sum_{k=0}^{\infty} p_k \Pr[s|k] = e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \sum_{s_1=1}^{\infty} \cdots \sum_{s_k=1}^{\infty} \left(\Pi_{j=1}^{k} \pi_{s_j}\right) \delta(s-1, \sum_j s_j)$$

$$\pi_s = \sum_{k=0}^{\infty} p_k \Pr[s|k] = e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} \sum_{s_1=1}^{\infty} \cdots \sum_{s_k=1}^{\infty} \left(\Pi_{j=1}^{k} \pi_{s_j}\right) \delta(s-1, \sum_j s_j)$$

One way to evaluate π_s is by using generating function

$$h(z) = \sum_{s=1}^{\infty} \pi_s z^s \Rightarrow \langle s \rangle = \frac{\sum_s s\pi_s}{\sum_s \pi_s} = h'(1)/(1-S) = 1/(1-c+cS).$$

see Newman's book, pp 412-413

Average size of the small components in a random model does not grow with the number of vertices.

Average component size

$$R = \frac{2}{2-c+cS}$$

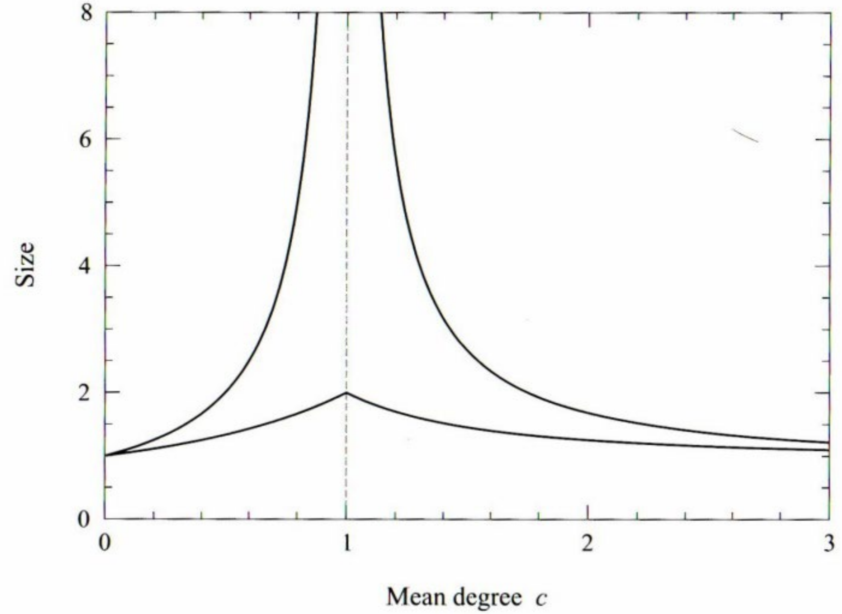


Figure 12.4: Average size of the small components in a random graph. The upper curve shows the average size $\langle s \rangle$ of the component to which a randomly chosen vertex belongs, calculated from Eq. (12.34). The lower curve shows the overall average size R of a component, calculated from Eq. (12.40). The dotted vertical line marks the point $c = 1$ at which the giant component appears. Note that, as discussed in the text, the upper curve diverges at this point but the lower one does not.

Distribution of component sizes

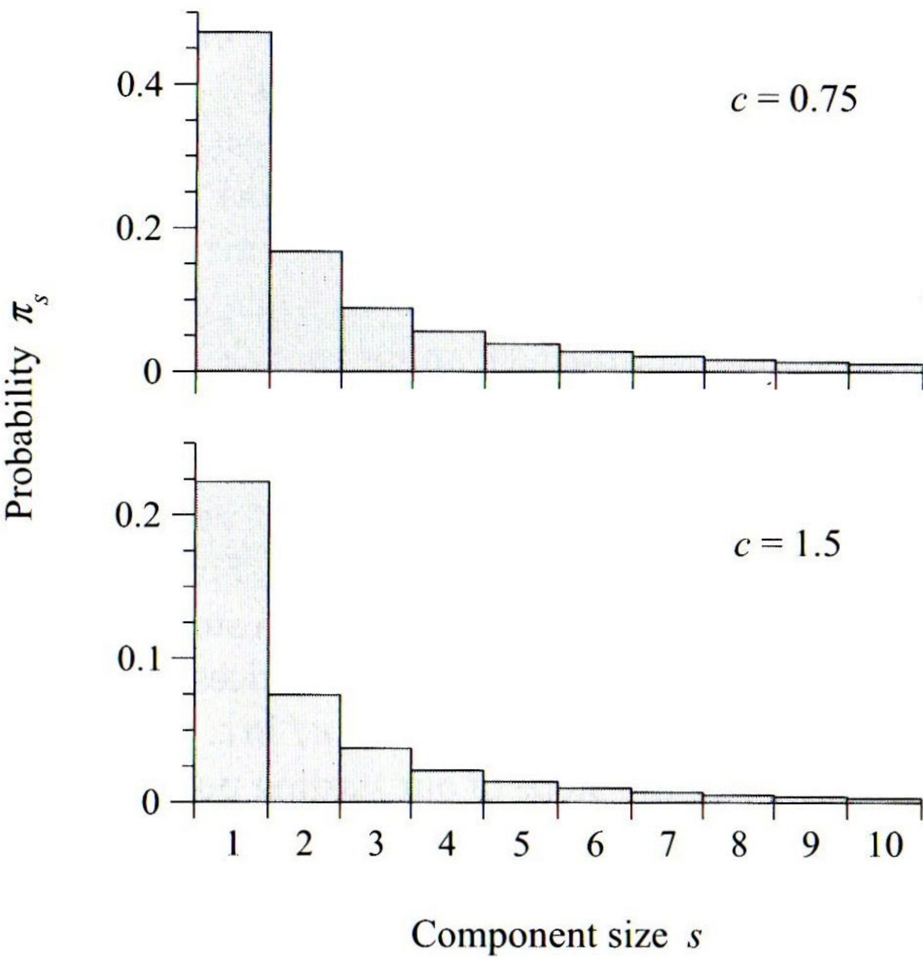
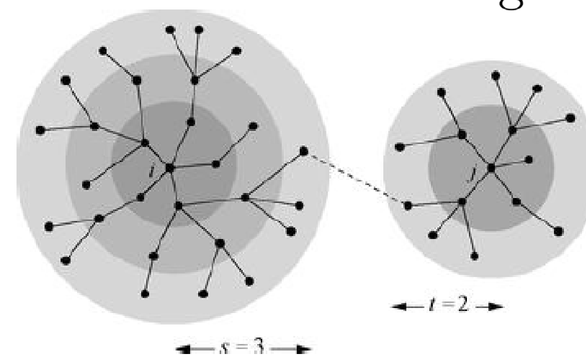


Figure 12.5: Sizes of small components in the random graph. This plot shows the probability π_s that a randomly chosen vertex belongs to a small component of size s in a Poisson random graph with $c = 0.75$ (top), which is in the regime where there is no giant component, and $c = 1.5$ (bottom), where there is a giant component.

- path lengths
  - Intuition: avg number of nodes s steps away from random i is c^s. We reach all vertices when $c^s \approx n$, i.e., $s \approx \ln n / \ln c$.
  - Problem: this argument doesn't work when s is large.
  - Consider two starting vertices i and j with their s- and $t-$distance neighborhoods, respectively, when s, t are small



1. if edge exists between surfaces then one can show that there are edges between larger surfaces

$$\implies \Pr[d_{ij} > s+t+1] \approx \text{prob } \nexists \text{ edge between two surfaces } c^s, \text{ and } c^t \text{ when } t \text{ is small}$$

2. There are on avg $c^s \times c^t$ pairs of nodes, s.t. one lies on each surface and each pair is connected with prob $p = c/(n-1)$
   i.e., $\Pr[d_{ij} > s+t+1] = (1-p)^{c^{s+t}} = (1-c/n)^{c^{l-1}}$ or $\ln \Pr[d_{ij} > l] = c^{l-1} \ln(1-c/n) \approx -c^l/n$

$l = s+t+1$

Networks	# of nodes	Diameter					Modularity				
		Observed	Expected	% difference	Z-score[14]	P-value	Observed	Expected	% difference	Z-score[14]	P-value
Characters in "Les Miserables"[1]	77	2.64	2.50	5.6	3.58	0.0003	0.56	0.29	93.4	30.12	$<10^{-4}$
Words in "David Copperfield"[2]	112	2.54	2.48	2.3	1.81	0.0703	0.31	0.29	4.8	1.67	0.0949
Dolphins[3]	62	3.36	2.70	24.3	14.40	$<10^{-4}$	0.53	0.37	40.8	11.59	$<10^{-4}$
Political blogs[4]	1224	2.74	2.59	5.7	23.5	$<10^{-4}$	0.43	0.14	206.9	189.27	$<10^{-4}$
Co-authorship[5]	7610	7.03	5.42	29.6	64.70	$<10^{-4}$	0.81	0.49	64.9	12.50	$<10^{-4}$
Football[6]	115	2.51	2.23	12.5	54.30	$<10^{-4}$	0.60	0.28	119.2	44.68	$<10^{-4}$
Power[7]	4941	18.99	8.32	128.3	14.30	$<10^{-4}$	0.93	0.73	28.5	105.10	$<10^{-4}$
Airline[8]	810	3.06	2.61	17.4	3.53	0.0004	0.31	0.13	130.0	114.70	$<10^{-4}$
Electronic circuits[9]	512	6.86	5.64	21.6	12.40	$<10^{-4}$	0.81	0.63	28.6	35.96	$<10^{-4}$
Protein-protein interaction[10]	1870	6.81	5.78	17.8	9.19	$<10^{-4}$	0.81	0.72	13.2	18.23	$<10^{-4}$
Neural[11]	297	2.46	2.35	4.5	3.38	0.0007	0.40	0.22	80.0	51.26	$<10^{-4}$
Transcriptional regulatory[12]	3459	3.72	3.39	9.7	3.60	0.0003	0.60	0.47	29.5	58.29	$<10^{-4}$
Metabolic[13]	563	8.78	6.54	34.3	18.67	$<10^{-4}$	0.84	0.73	14.5	14.72	$<10^{-4}$

[1]The network of coappearances of characters in Victor Hugo's novel "Les Miserables". Nodes represent characters and edges connect any pair of characters that appear in the same chapter.
[2]The network of common adjective and noun adjacencies for the novel "David Copperfield" by Charles Dickens. Nodes represent the most commonly occurring adjectives and nouns in the book.
[3]The network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand.
[4]The network of political blogs. Nodes represent blogs and edges are the links between blogs.
[5]The network of scientists posting preprints on the high-energy theory archive at www.arxiv.org, 1995–1999. Nodes are authors and edges connect coauthors.
[6]The network of American football games between Division IA colleges during regular season Fall 2000. Nodes are teams and edges connect teams that contest in a game.
[7]The network of the Western States Power Grid of the United States. Nodes are power plants, stations and households, and edges are powerlines.
[8]The network of scheduled air line connections in United States, 2005. Nodes are airports and edges are scheduled direct flights.
[9]Electronic circuits. Nodes are electronic elements and edges are electronic connections.
[10]The protein-protein interaction network of the budding yeast *S. cerevisiae*. Nodes are proteins and edges connect proteins that interact with each other.
[11]The neural network for the worm *C. elegans*. Nodes are neurons and edges link neurons that connect.
[12]The transcriptional regulatory network of the budding yeast *S. cerevisiae*. Nodes are genes and edges connect genes that regulate one another.
[13]The metabolic network of the bacterium *E. coli*. Nodes are metabolites and edges connect metabolites that can be converted by a biochemical reaction.
[14]*Z*-score, number of standard deviations by which the observation deviates from the expectation.
doi:10.1371/journal.pone.0005686.t001

http://complexnt.blogspot.com

Generating Functions and Degree Distributions

The *generating function* (gf) for the probability distribution p_k is the polynomial

$$g(z) = \sum_{k=0}^{\infty} p_k z^k.$$

If we know gf for p_k then we can recover the values of p_k by differentiating

$$p_k = \frac{1}{k!} \frac{d^k g}{dz^k}\bigg|_{z=0}$$

Example: $k = 0, 1, 2$ with the respective $p_k = \frac{1}{2}, \frac{7}{16}, \frac{1}{16}$ for all k then

$$g(z) = \frac{1}{2} + \frac{7}{16} z + \frac{1}{16} z^2$$

Example: k follows Poisson distribution, i.e., $p_k = e^{-c} \frac{c^k}{k!}$

$$g(z) = e^{-c} \sum_{k=0}^{\infty} \frac{c^k}{k!} z^k = e^{c(z-1)}$$

Power-law distributions $p_k = Ck^{-\alpha},\ \alpha > 0,\ k > 0$

Reminder: C is calculated from normalization condition, i.e., $C = 1/\zeta(\alpha)$

$$p_k = \begin{cases} 0 & k = 0 \\ k^{-\alpha}/\zeta(\alpha) & k > 0 \end{cases} \implies g(z) = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha} z^k = \frac{Li_\alpha(z)}{\zeta(\alpha)}$$

Polylogarithm

Since we are interested in differentiating $g(z)$ note that

$$\frac{\partial Li_\alpha(z)}{\partial z} = \frac{Li_{\alpha-1}(z)}{z}$$

Some properties of $g(z)$

- $g(1) = 1$
- $\langle k \rangle = g'(1)$, $\langle k^2 \rangle = \left[\left(z \frac{d}{dz} \right)^2 g(z) \right]_{z=1}$, ... , $\langle k^m \rangle = \left[\left(z \frac{d}{dz} \right)^m g(z) \right]_{z=1}$
- Choose m integers k_i from $p_k \Rightarrow \Pr[\text{chosing particular set of values } \{k_i\}] = \prod_i p_{k_i}$

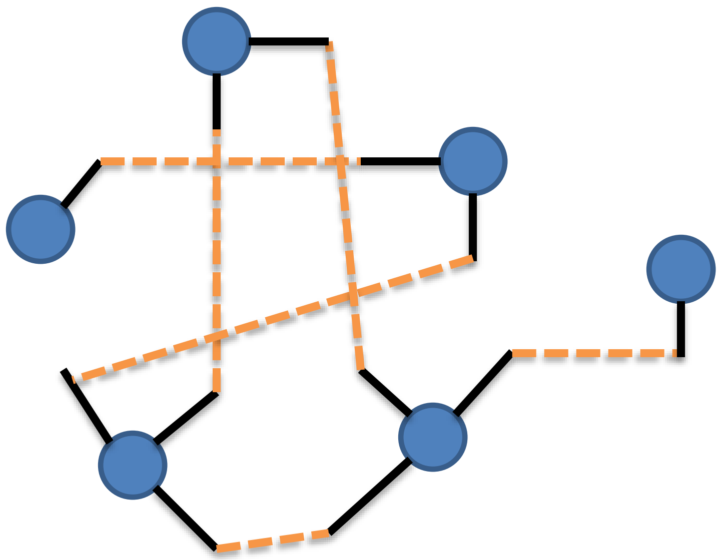
$$\pi_s = \Pr[\sum_{i=1}^{m} k_i = s] = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} \delta(s, \sum_i k_i) \prod_{i=0}^{m} p_{k_i} \Rightarrow$$

$$h(z) = \sum_{s=0}^{\infty} \pi_s z^s = \cdots = \left(\sum_{k=0}^{\infty} p_k z_k \right)^m = (g(z))^m$$

drawn values add to a specific sum *s*

Random Graphs and Configuration Model

Degrees: 1, 1, 2, 2, 3, 3



1. Add *n* nodes
2. Add initial *d(i)* stubs to each *i*
3. Connect stubs iteratively

Problems? Total degree is even; Can create self-loops, multi-edges

Configuration Model

Multi-edges: Probability of adding an edge between i and j with degrees k_i, and k_j is

$$p_{ij} = \frac{k_i k_j}{2m-1}$$

in the limit we can omit -1

Probability of second edge is $(k_i - 1)(k_j - 1)/2m$

Expected number of multiedges in conf model

$$\frac{1}{2(2m)^2}\sum_{ij} k_i k_j (k_i - 1)(k_j - 1) = \frac{1}{2\langle k\rangle^2 n^2}\sum_i k_i(k_i - 1)\sum_j k_j(k_j - 1) = \frac{1}{2}\left[\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\right]^2$$

Similar result for self-edges

$$\sum_i p_{ii} = \sum_i \frac{k_i(k_i - 1)}{4m} = \frac{\langle k^2\rangle - \langle k\rangle}{2\langle k\rangle}$$

Conclusion? Expected number of multi-edges remains constant as network grows.

Expected number of common neighbors

$$n_{ij} = \sum_l \frac{k_i k_l}{2m}\frac{k_j(k_l - 1)}{2m} = \frac{k_i k_j}{2m}\frac{\sum_l k_l(k_l - 1)}{n\langle k\rangle} = p_{ij}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}$$

i is connected to *l*

j is connected to *l* given *il*

Random graphs with given expected degree

$\forall i \in V$ define parameter c_i. Then edge probability

$$p_{ij} = \begin{cases} c_i c_j / 2m & i \neq j \\ c_i^2 / 4m & i = j \end{cases}, \text{ where } \sum_i c_i = 2m$$

average number of edges in network

$$\sum_{i \leq j} p_{ij} = \sum_{i<j} \frac{c_i c_j}{2m} + \sum_i \frac{c_i^2}{4m} = m$$

average degree

$$\langle k_i \rangle = 2p_{ii} + \sum_{j \neq i} p_{ij} = \frac{c_i^2}{2m} + \sum_{j \neq i} \frac{c_i c_j}{2m} = \sum_j \frac{c_i c_j}{2m} = c_i$$

More properties of random model

Excess degree distribution is the probability distribution, for a vertex reached by following an edge, of the number of other edges attached to that vertex.

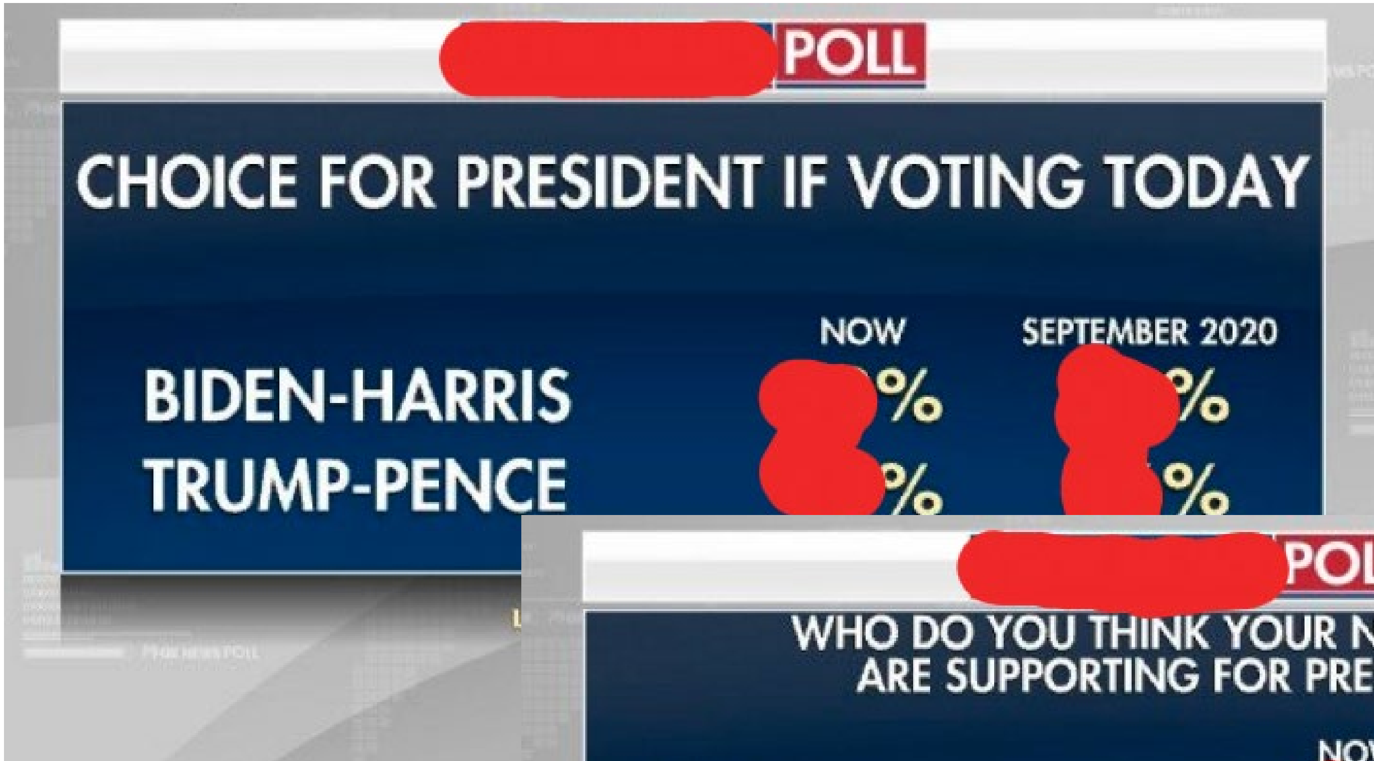
$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$$

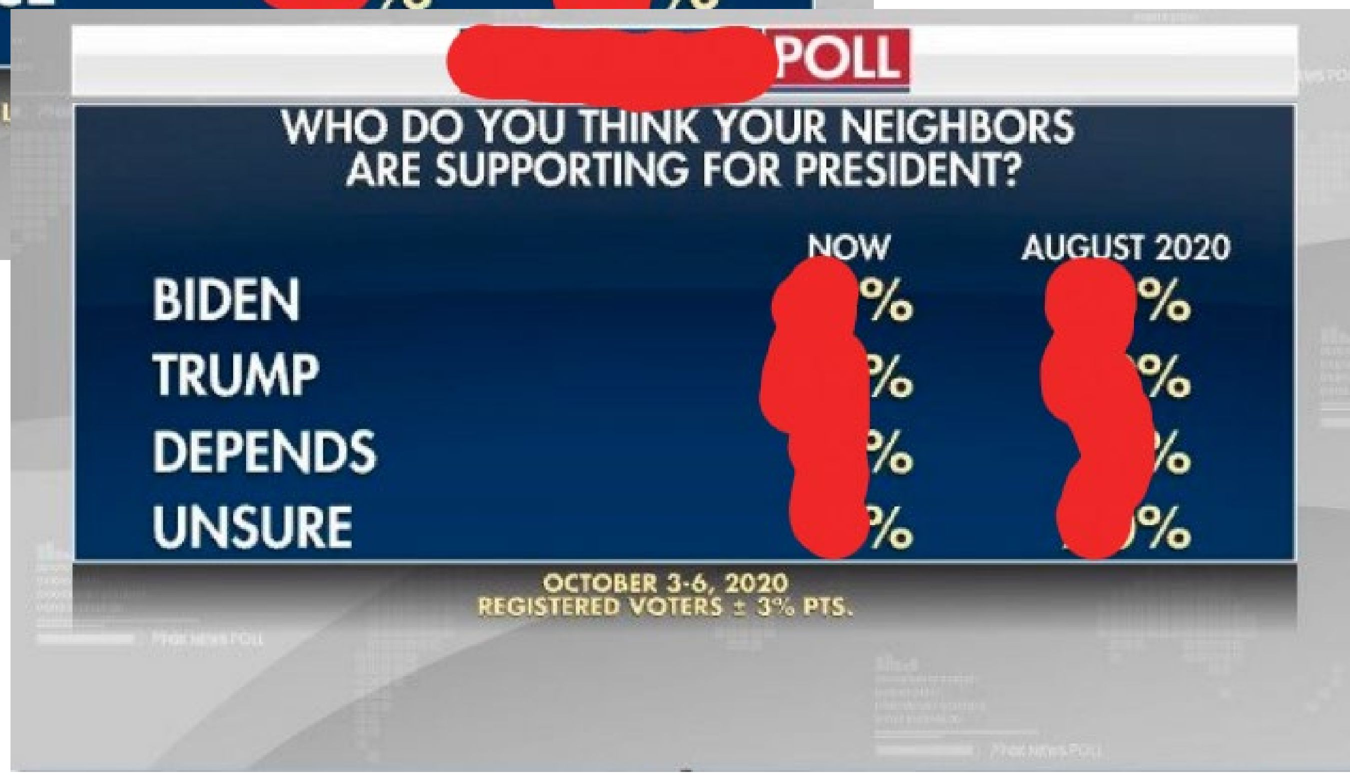
Two academic collaboration networks, in which scientists are connected together by edges if they have coauthored scientific papers, and a snapshot of the structure of the Internet at the autonomous system level.

Network	n	Average degree	Average neighbor degree	$\frac{\langle k^2 \rangle}{\langle k \rangle}$
Biologists	1 520 252	15.5	68.4	130.2
Mathematicians	253 339	3.9	9.5	13.2
Internet	22 963	4.2	224.3	261.5

According to these results a biologist's collaborators have, on average, more than four times as many collaborators as they do themselves. On the Internet, a node's neighbors have more than 50 times the average degree! Note that in each of the cases in the table the configuration model value of $\langle k^2 \rangle / \langle k \rangle$ overestimates the real average neighbor degree.

M. Newman "Networks"


POLL
CHOICE FOR PRESIDENT IF VOTING TODAY
NOW
SEPTEMBER 2020
BIDEN-HARRIS
TRUMP-PENCE


POLL
WHO DO YOU THINK YOUR NEIGHBORS
ARE SUPPORTING FOR PRESIDENT?
NOW
AUGUST 2020
BIDEN
TRUMP
DEPENDS
UNSURE
OCTOBER 3-6, 2020
REGISTERED VOTERS ± 3% PTS.

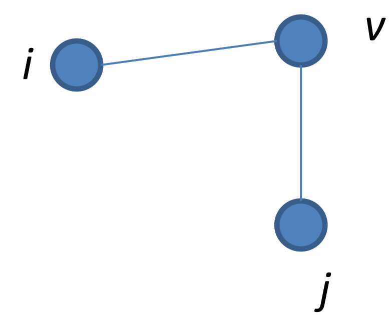
More properties of random model

Excess degree distribution is the probability distribution, for a vertex reached by following an edge, of the number of other edges attached to that vertex.

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$$

Using the excess degree distribution it is easy to compute the clustering coefficient for configuration model

$$C = \sum_{k_i, k_j = 0}^{\infty} q_{k_i} q_{k_j} \frac{k_i k_j}{2m} = \frac{1}{2m} \Big(\sum_{k=0}^{\infty} k q_k \Big)^2 = \cdots = \frac{1}{n} \frac{(\langle k \rangle^2 - \langle k \rangle)^2}{\langle k \rangle^3}$$



Generating Functions and Degree Distributions

For degree and excess degree distributions we define generating functions

$$g_0(z) = \sum_{k=0}^{\infty} p_k z^k \text{ and } g_1(z) = \sum_{k=0}^{\infty} q_k z^k, \text{ respectively}$$

They are not independent

we add zero term because of infinity

$$g_1(z) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k+1) p_{k+1} z^k = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k p_k z^{k-1} = \frac{1}{\langle k \rangle} \frac{\mathrm{d} g_0}{\mathrm{d} z} = \frac{g_0'(z)}{g_0'(1)}$$

Example (Poisson): $p_k = e^{-c} \frac{c^k}{k!} \Rightarrow g_0(z) = e^{c(z-1)}$ and $g_1(z) = e^{c(z-1)}$

Example (power-law): $p_k = Ck^{-\alpha} \Rightarrow g_0(z) = \frac{Li_\alpha(z)}{\zeta(\alpha)}$. Thus,

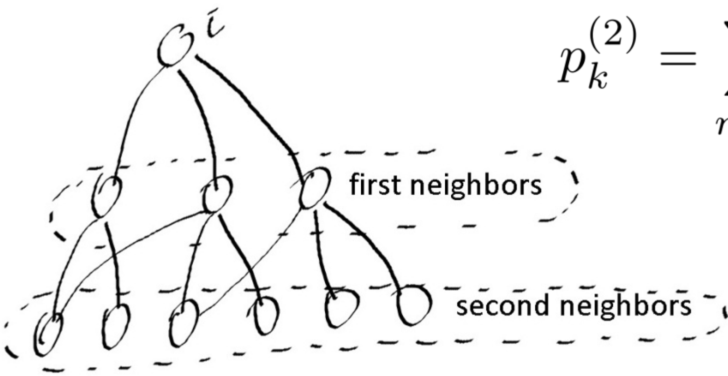
$$g_1(z) = \frac{Li_{\alpha-1}(z)}{zLi_{\alpha-1}(1)} = \frac{Li_{\alpha-1}(z)}{z\zeta(\alpha-1)}$$

Polylogarithm function

$$Li_s(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^s} = z + \frac{z^2}{2^s} + \frac{z^3}{3^s} + \ldots$$

Number of second neighbors of a vertex

Probability that i has exactly k second neighbors



$$p_k^{(2)} = \sum_{m=0}^{\infty} p_m P^{(2)}(k|m)$$

degree distribution

Probability of having *k* second neighbors given m first neighbors

Prob excess degrees of *m* first neighbors take values $j_1, j_2, \ldots, j_m$

$$P^{(2)}(k|m) = \sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} \delta\left(k, \sum_{r=1}^{m} j_r\right) \prod_{r=1}^{\infty} q_{j_r}$$

all sets of values $j_1, j_2, \ldots, j_m$ that sum to k

$$g^{(2)}(z) = \sum_{k=0}^{\infty} p_k^{(2)} z^k = \sum_{k=0}^{\infty} z^k \cdot \sum_{m=0}^{\infty} p_m \sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} \delta\left(k, \sum_{r=1}^{m} j_r\right) \prod_{r=1}^{m} q_{j_r} = \cdots =$$

$$= \sum_{m=0}^{\infty} p_m \cdot \left(\sum_{j=0}^{\infty} q_j z^j\right)^m = g_0(g_1(z))$$

generating function of $p_k^{(2)}$

Conclusion: Once we know generating functions of g_0 and g_1 the generating function of second neighbor distribution is straightforward to calculate. Moreover, this can be extended to

$$g^{(3)}(z) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} p_m^{(2)} P^{(3)}(k|m)\, z^k = \sum_{n=0}^{\infty} p_m^{(2)} (g_1(z))^m = g^{(2)}(g_1(z)) = g_0(g_1(g_1(z)))$$

$$\Longrightarrow g^{(d)}(z) = g^{(d-1)}(g_1(z)) = g_0(g_1(\ldots g_1(z)\ldots))$$

Problem: Sometimes it is difficult to extract explicit probabilities for numbers of second neighbors and it is hard to evaluate *n* derivatives (in order to recover the probabilities).
Solution: calculate the average number of neighbors at distance *d.* At *z=1* of the first derivative we can evaluate the average of a distribution (see Slide 16).

$$\frac{\mathrm{d}g^{(2)}}{\mathrm{d}z} = g_0'(g_1(z))g_1'(z) \overset{z=1,g_1(1)=1}{\Longrightarrow} \underset{\uparrow}{c_2} = g_0'(1)g_1'(1) \overset{g_0'(1)=\langle k\rangle}{\Longrightarrow} g_1'(k) = \sum_{k=0}^{\infty} kq_k =$$

mean number of
second neighbors

$$\frac{1}{\langle k\rangle}\sum_{k=0}^{\infty} k(k+1)p_{k+1} = \frac{1}{\langle k\rangle}(\langle k^2\rangle - \langle k\rangle)$$

Conclusion: $c_2 = \langle k^2\rangle - \langle k\rangle$ and more general

$$c_d = \left(\frac{c_2}{c_1}\right)^{d-1} c_1 \Longrightarrow$$

Condition of giant component's existance
in configuration model is $\langle k^2\rangle - 2\langle k\rangle > 0$

[MR] A critical point for random graphs with given degree sequence

Let's use theory for practical results …

Given a network with **power-law distribution** $p_k = Ck^{-\alpha},\ \alpha > 0,\ k > 0$
Reminder: C is calculated from normalization condition, i.e., $C = 1/\zeta(\alpha)$

$$p_k = \begin{cases} 0 & k = 0 \\ k^{-\alpha}/\zeta(\alpha) & k > 0 \end{cases}$$

This network will have a giant component iff $\langle k^2 \rangle - 2\langle k \rangle > 0$

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+1} = \frac{\zeta(\alpha - 1)}{\zeta(\alpha)}$$

$$\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 p_k = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+2} = \frac{\zeta(\alpha - 2)}{\zeta(\alpha)}$$

$$\Longrightarrow \zeta(\alpha - 2) > 2\zeta(\alpha - 1)$$

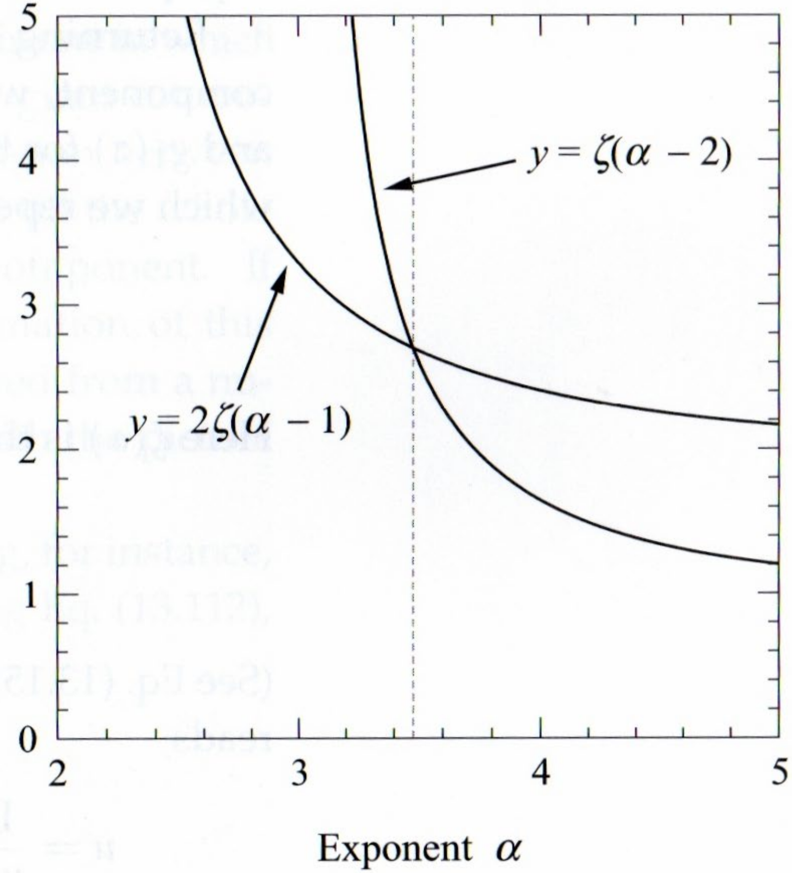


Figure 13.8: Graphical solution of Eq. (13.138). The configuration model with a pure power-law degree distribution (Eq. (13.133)) has a giant component if $\zeta(\alpha - 2) > 2\zeta(\alpha - 1)$. This happens for values of α below the crossing point of the two curves.

Newman “Networks: An Introduction”