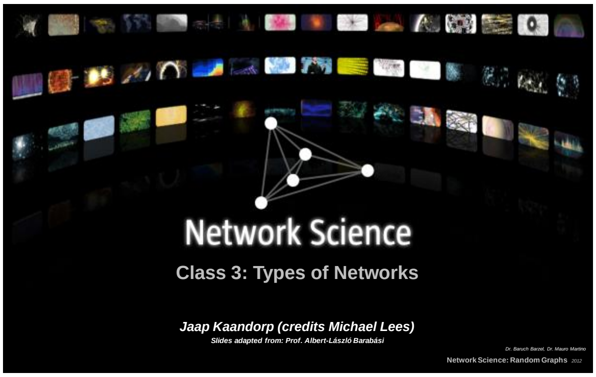


1

Networks can be characterized by their form

We will try to cover (briefly) a selection of these Networks (explore more in Labs):

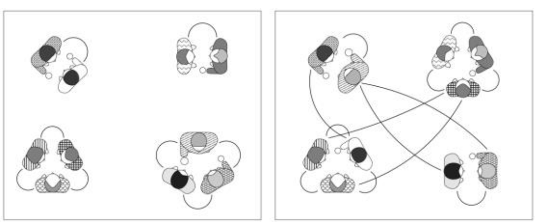
- **Random Networks** – edges are randomly created between nodes
- **Watts Strogatz model** - Small world Networks
- **Scale Free Networks** – a few nodes with many connections

We cover all this in one lecture – so it's introductory, there is a great deal more for you to discover!

Hidden slides are there for your reference, you can also read either of the books.

2

RANDOM NETWORK MODEL

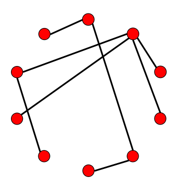


3

RANDOM NETWORK MODEL

Pál Erdös
(1913-1996)

Alfréd Rényi
(1921-1970)



Erdös-Rényi model (1960)

Connect with probability p

p=1/6 N=10

⟨k⟩ ~ 1.5

4

RANDOM NETWORK MODEL

Definition:

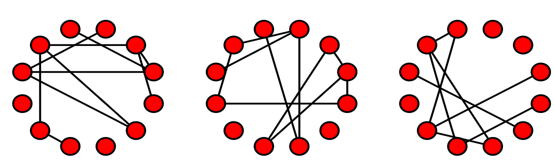
A **random graph** is a graph of N labeled nodes where each pair of nodes is connected by a preset probability **p**.

We will call is G(N, p).

5

RANDOM NETWORK MODEL

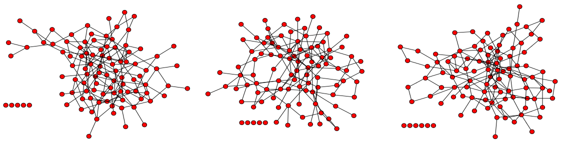
p=1/6
N=12



6

RANDOM NETWORK MODEL

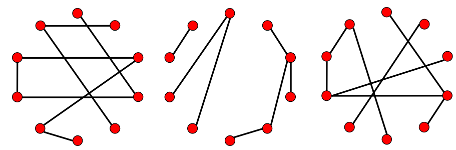
p=0.03
N=100



7

RANDOM NETWORK MODEL

N and ***p*** do not uniquely define the network– we can have many different realizations of it. **How many?**



N=10
p=1/6

The probability to form a *particular* graph **G(N,p)** is

$$P(G(N,p)) = p^L (1-p)^{\frac{N(N-1)}{2} - L}$$

That is, each graph **G(N,p)** appears with probability **P(G(N,p))**.

8

RANDOM NETWORK MODEL

P(L): the probability to have exactly *L* links in a network of *N* nodes and probability *p*:

The maximum number of links in a network of N nodes.

$$P(L) = \underbrace{\binom{\frac{N(N-1)}{2}}{L}}_{\text{Number of different ways we can choose L links among all potential links.}} \underbrace{p^L (1-p)^{\frac{N(N-1)}{2} - L}}_{\text{Choose } L \text{ links with probability } p \text{ and remainder } (L_{max} - L) \text{ with probability } 1-p}$$

Binomial distribution...

9

Binomial distribution (intermezzo)

the probability of x success in N trials

$$P_x(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$$

Binomial coefficient $\binom{N}{x} = \frac{N!}{x!(N-x)!}$

- The mean <*x*> in a distribution (1° moment)

$$< x >= \sum_{x=0}^{N} x \binom{N}{x} p^x (1-p)^{N-x} = Np$$

- The variance (2° moment):

$$\sigma^2 =< x^2 >= \sum_{x=0}^{N} x^2 p_x = p(1-p)N + p^2 N^2$$

- The standard deviation:

$$\sigma_x = [p(1-p)N]^{\frac{1}{2}}$$

10

RANDOM NETWORK MODEL

P(L): the probability to have a network of exactly *L* links

$$P(L) = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2} - L}$$

- The average number of links <*L*> in a random graph

$$< L >= \sum_{L=0}^{\frac{N(N-1)}{2}} L P(L) = p \frac{N(N-1)}{2} \qquad < k >= 2L/N = p(N-1)$$

- The variance:

$$\sigma^2 = p(1-p) \frac{N(N-1)}{2}$$

11

RANDOM NETWORK MODEL

- Remember (complete graph)

$$L_{\max} = \frac{N(N-1)}{2}$$

- Remember (undirected) graph

$$L = \frac{1}{2} \sum_{i=1}^{N} k_i$$

$$< k >= \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}$$

- if we put <k> into <L>

$$< k >= 2 < L > / N = p(N-1)$$

- The expected number of links in a
- random graph

$$< L >= \sum_{L=0}^{\frac{N(N-1)}{2}} L P(L) = p \frac{N(N-1)}{2}$$

12

RANDOM NETWORK MODEL

•Variance

$$s^2 = p(1-p)\frac{N(N-1)}{2}$$

•We can increase p=0 to p=1.0, the random network becones dense

•Average links

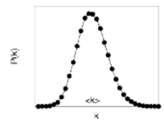
$$<L>=0..<L>=L_{\max}$$

•Average degree

$$<k>=0..<k>=N-1$$

13

DEGREE DISTRIBUTION OF A RANDOM GRAPH



$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Select k nodes from N-1

probability of having *k* edges

probability of missing N-1-k edges

$$<k>= p(N-1) \qquad s_k^2 = p(1-p)(N-1)$$

$$\frac{\sigma_k}{<k>} = \left[\frac{1-p}{p}\frac{1}{(N-1)}\right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

As the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of <k>.

14

Binomial distribution -> Poisson distribution I

•Most real networks

$$<k> << N$$

•We can approximate the degree distribuition using the Poisson distribution:

$$P(k) = e^{-<k>}\frac{<k>^k}{k!}$$

15

Binomial distribution -> Poisson distribution II

In

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

•We can rewrite

$$\binom{N-1}{k} = \frac{(N-1)!}{k!(N-1-k)!} = \frac{(N-1)(N-1-1)(N-1-2)...(N-1-k+1)(N-1-k)!}{k!(N-1-k)!} \cong \frac{(N-1)^k}{k!}$$

•We can simplify

$$\ln[(1-p)^{(N-1)-k}] = (N-1-k)\ln(1-p)$$

•Use $<k>= p(N-1)$ $\qquad p = \frac{<k>}{N-1}$

$$\ln[(1-p)^{(N-1)-k}] = (N-1-k)\ln(1-p) = (N-1-k)\ln(1-\frac{<k>}{N-1})$$

16

Binomial distribution -> Poisson distribution III

We have $\ln[(1-p)^{(N-1)-k}] = (N-1-k)\ln(1-p) = (N-1-k)\ln(1-\frac{<k>}{N-1})$

•Use Taylor series

$$\ln(1+x) = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - ... \qquad \text{for} \quad |x| \le 1$$

•We get

$$\ln[(1-p)^{(N-1)-k}] \cong -(N-1-k)\frac{<k>}{N-1} \cong -<k> \qquad \text{if} \quad N >> k$$

•Approximately

$$(1-p)^{(N-1)-k}] \cong e^{-<k>}$$

17

Binomial distribution -> Poisson distribution IV

Combining

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} = \frac{(N-1)^k}{k!} p^k e^{-<k>} =$$

$$\frac{(N-1)^k}{k!}\left(\frac{<k>}{N-1}\right)^k e^{-<k>} = e^{-<k>}\frac{<k>^k}{k!}$$

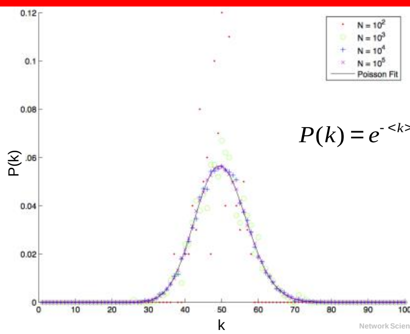
$$\binom{N-1}{k} = \frac{(N-1)^k}{k!}$$

$$p = \frac{<k>}{N-1}$$

$$(1-p)^{(N-1)-k}] \cong e^{-<k>}$$

18

DEGREE DISTRIBUTION OF A RANDOM GRAPH



$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Network Science: Random Graphs 2012

20

DEGREE DISTRIBUTION OF A RANDOM NETWORK

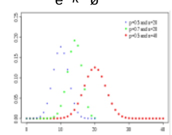
Exact Result
-binomial distribution-

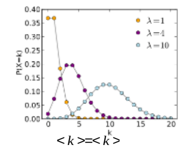
$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Large N limit
-Poisson distribution-

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Probability Distribution Function (PDF)





$$\langle k \rangle = (N-1)p$$

$$\langle k^2 \rangle = p(1-p)(N-1) + p^2(N-1)^2$$

$$\sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = [p(1-p)(N-1)]^{1/2}$$

$$\langle k \rangle = \langle k \rangle$$

$$\langle k^2 \rangle = \langle k \rangle (1 + \langle k \rangle)$$

$$\sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = \langle k \rangle^{1/2}$$

Network Science: Random Graphs 2012

21

NODES HAVE COMPARABLE DEGREES IN RANDOM NETWORKS

What does it mean? Continuum formalism:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

If we consider a network with average degree <k> then the probability to have a node whose degree exceeds a degree k_0 is:

$$P(k > k_0) = \int_{k_0}^{\infty} e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} dk$$

For example, with <k>=10,
•the probability to find a node whose degree is at least twice the average degree is 0.00158826.
•the probability to find a node whose degree is at least ten times the average degree is 1.79967152 × 10^{-13}
•the probability to find a node whose degree is less than a tenth of the average degree is 0.00049
See http://www.stud.feec.vutbr.cz/~xvapen02/vypocty/po.php

What does it mean? Discrete formalism:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

$$\frac{\sigma_k}{\langle k \rangle} = \left[\frac{1-p}{p} \frac{1}{(N-1)} \right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

- The probability of seeing a node with very high of very low degree is exponentially small.
- Most nodes have comparable degrees.
- The larger the size of a random network, the more similar are the node degrees

22

NO OUTLIERS IN A RANDOM SOCIETY

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

According to sociological research, for a typical individual k ~1,000

The probability to find an individual with degree k>2,000 is 10^{-27}.

Given that N ~10^9, the chance of finding an individual with 2,000 acquaintances is so tiny that such nodes are virtually inexistent in a random society.

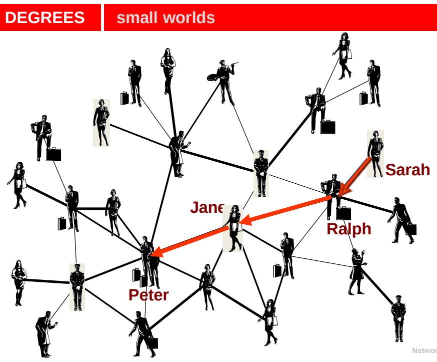
→a random society would consist of mainly average individuals, with everyone with roughly the same number of friends.

→It would lack outliers, individuals that are either highly popular or recluse.

Network Science: Random Graphs 2012

23

SIX DEGREES small worlds



Frigyes Karinthy, 1929
Stanley Milgram, 1967

Network Science: Random Graphs 2012

24

SIX DEGREES 1967: Stanley Milgram

HOW TO TAKE PART IN THIS STUDY

1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

2. DETACH ONE POSTCARD. FILL IT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

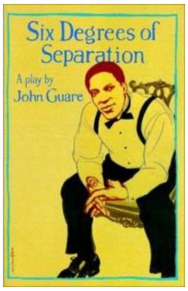
3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder to a friend, relative or acquaintance, but it must be someone you know on a first name basis.

Network Science: Random Graphs 2012

26

SIX DEGREES | 1991: John Guare



"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice…. It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds."

Network Science: Random Graphs 2012

27

WWW: 19 DEGREES OF SEPARATION

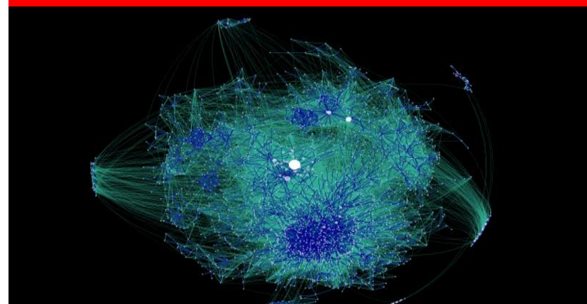


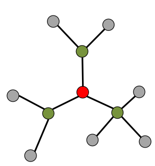
Image by **Matthew Hurst**
Blogosphere

Network Science: Random Graphs 2012

28

DISTANCES IN RANDOM GRAPHS

Random graphs tend to have a tree-like topology with almost constant node degrees.



- nr. of first neighbours: $N_d \cong \langle k \rangle^1$
- nr. of second neighbours: $N_d \cong \langle k \rangle^2$
- nr. of neighbours at distance d: $N_d \cong \langle k \rangle^d$
- estimate maximum distance:

$$N = 1 + \langle k \rangle + \langle k \rangle^2 + \ldots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^d \quad \Rightarrow \quad d = \frac{\log N}{\log \langle k \rangle}$$

Network Science: Random Graphs 2012

29

DISTANCES IN RANDOM GRAPHS | compare with real data

$$l_{\max} = \frac{\log N}{\log \langle k \rangle}$$

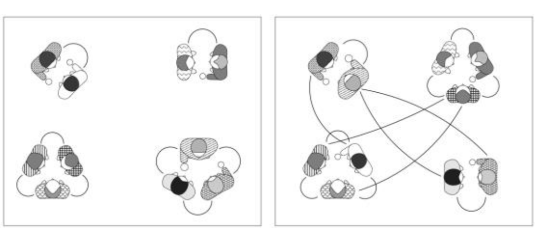
Network	Size	⟨k⟩	l	l_{rand}	C	C_{rand}	Reference	Nr
www, site level, undir	153127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015-6209	3.52-4.11	3.7-3.76	6.36-6.18	0.18-0.3	0.001	Yook e al., 2001a, Pastor-Satorras et al., 2001	2
Movie actors	225226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz,1998	3
LANL co-authorship	52909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE eo-authorship	1520251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabasi et al, 2001	8
Neurosci. co-authorship	209293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabasi et al, 2001	9
E. coli, sustrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Sole, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Sole, 2000	13
Words, co-occurrence	460902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Sole, 2001	14
Words, synonyms	22311	13.48	4.5	3.84	0.7	0.0006	Yook et al. 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C.Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Given the huge differences in scope, size, and average degree, the agreement is excellent.

Network Science: Random Graphs 2012

30

EVOLUTION OF A RANDOM GRAPH



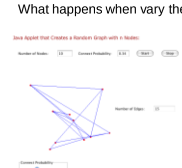
Network Science: Random Graphs 2012

31

EVOLUTION OF A RANDOM NETWORK

Until now we focused on the static properties of a random graph with fixed p value.

What happens when vary the parameter p?

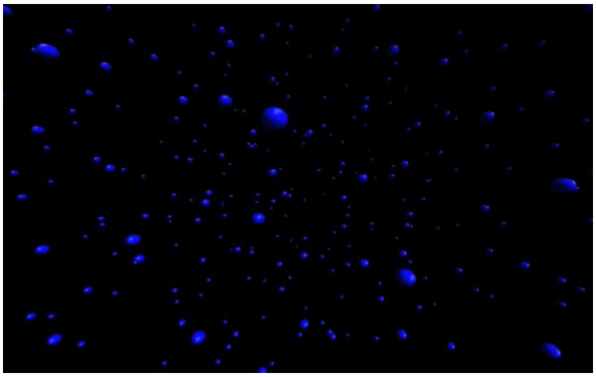


GOTO *http://cs.gmu.edu/~astavrou/random.html*

Choose Nodes=100.
Note that the p goes up in increments of 0.001, which, for N=100, L=pN(N-1)/2~p*50,000, i.e. each increment is really about 50 new lines.

Network Science: Random Graphs 2012

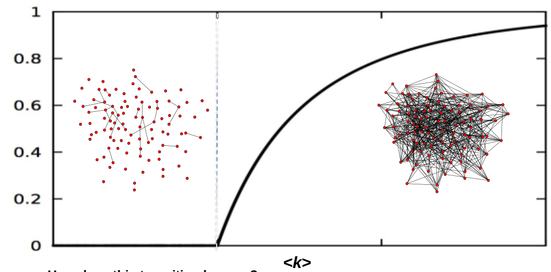
32



33

EVOLUTION OF A RANDOM NETWORK

disconnected nodes ➔ NETWORK.



How does this transition happen?

34

THE PHASE TRANSITION TAKES PLACE AT <k>=1

Let us denote with $u = 1 - \frac{N_g}{N}$, i.e. the fraction of nodes that are NOT part of the giant component (GC) N_g.

For a node *i* to be part of the GC, it needs to connect to it via another node *j*. If *i* is NOT part of the GC, that could happen for two reasons:

Case A: node *i* does not connect to node *j*, Probability: *1-p*

Case B: node *i* connects to *j*, but *j* is not connected to the GC: Probability: *pu*

Total probability that *i* is not part of the GC via node *j* is: *1-p+pu*

The probability that *i* is not linked to the GC via *any other node* is $(1-p+pu)^{N-1}$

Hence:

$$u = (1 - p + pu)^{N-1}$$

For any *p* and *N* this equation provides the size of the giant component as $N_g=N(1-u)$

35

EVOLUTION OF A RANDOM GRAPH

$$u = (1 - p + pu)^{N-1}$$

Using p=<k>/(N-1) and taking the log of both sides and using *<k><<N* we obtain:

$$\ln u = (N-1)\ln\left[1 - \frac{<k>}{N-1}(1-u)\right] \approx -(N-1)\frac{<k>}{N-1}(1-u) = -<k>(1-u)$$

$\boxed{\ln[1-x] \approx -x}$

Taking an exponential of both sides we obtain

$$u = e^{-<k>(1-u)}$$

Or, if we denote with S the fraction of nodes in the giant component, $S=N_{GC}/N$, i.e. S=1-u

$$S = 1 - e^{-<k>S}$$

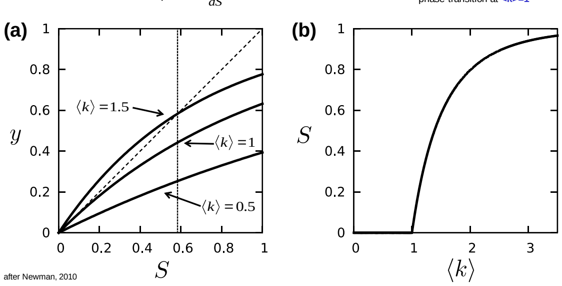
Erdos and Renyi, 1959

36

THE PHASE TRANSTION IN A RN TAKES PLACE AT <k>=1

$S = 1 - e^{-<k>S}$ S: the fraction of nodes in the giant component, $S=N_g/N$

Phase transition point: $\frac{d}{dS}(1 - e^{-<k>S}) = 1$ $<k> e^{-<k>S} = 1$ Set S=0, we obtain a phase transition at **<k>=1**

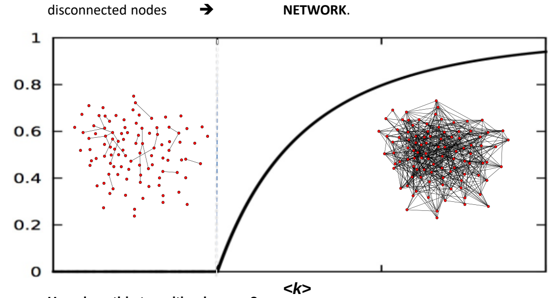


after Newman, 2010

37

EVOLUTION OF A RANDOM NETWORK

disconnected nodes ➔ NETWORK.



How does this transition happen?

38

EVOLUTION OF A RANDOM NETWORK

Probability that a randomly selected node belongs to a cluster of size *s*:

$$p(s) = \frac{e^{-<k>s}(<k>s)^{s-1}}{s!}$$

$$\langle k\rangle^{s-1} = \exp[(s-1)\ln\langle k\rangle]$$

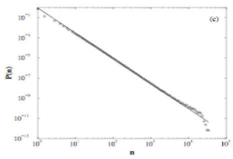
Use Stirling approximation:

$$s! = \sqrt{2\pi s}\left(\frac{s}{e}\right)^{s} \qquad p(s) = \frac{s^{s-1}}{s!}e^{-\langle k\rangle s + (s-1)\ln\langle k\rangle}$$

$$p(s) \sim s^{-3/2}e^{-(\langle k\rangle - 1)s + (s-1)\ln\langle k\rangle}$$

At the critical point <k>=1

$$p(s) \sim s^{-3/2}$$

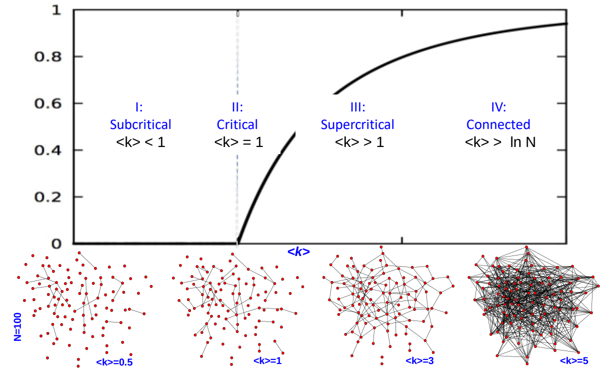


The distribution of cluster sizes at the critical point, displayed in a log-log plot. The data represent an average over 1000 systems of sizes
The dashed line has a slope of

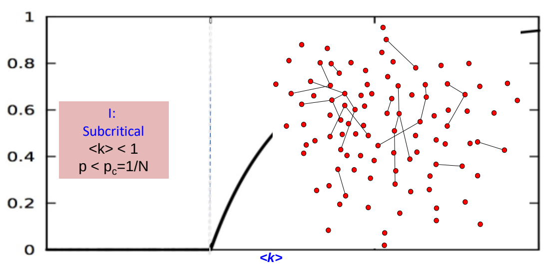
$$-\tau_n = -2.5$$

Network Science: Random Graphs 2012

39



42

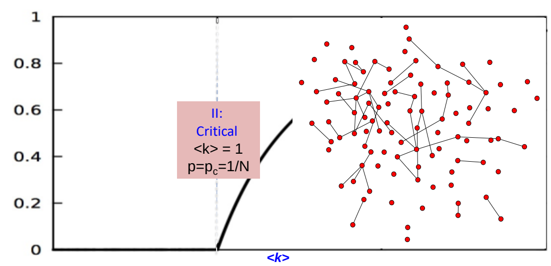


No single giant component.

N-L isolated clusters, cluster size distribution is exponential

The largest cluster is a tree, its size ~ *ln N*

43

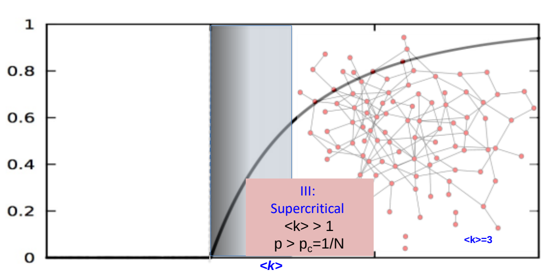


Unique giant component: $N_G \sim N^{2/3}$
→contains a vanishing fraction of all nodes, $N_G/N \sim N^{-1/3}$
→Small components are trees, GC has loops.

Cluster size distribution: $p(s) \sim s^{-3/2}$

A jump in the cluster size:
N=1,000 → ln N~ 6.9; $N^{2/3}$~95
N=7 10^9 → ln N~ 22; $N^{2/3}$~3,659,250

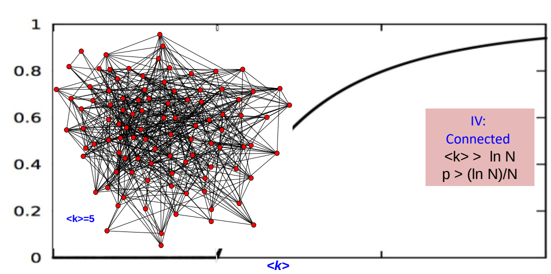
44



Unique giant component: $N_G \sim (p-p_c)N$
→GC has loops.

Cluster size distribution: exponential

45



Only one cluster: N_G=N
→GC is dense.
Cluster size distribution: None

Network Science: Random Graphs 2012

46

IV:
Connected
<k> > ln N
p > (ln N)/N

The probability that a node does *not* connect to the giant component is
$(1-p)^{N_G} \sim (1-p)^N$

The expected number of such nodes is:

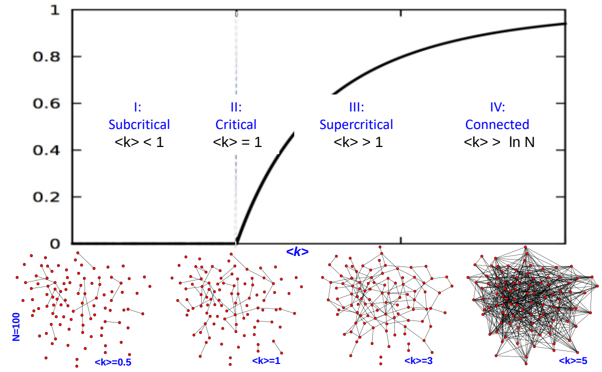
$$C = N(1-p)^N = N\left(1-\frac{Np}{N}\right)^N \approx Ne^{-Np} \qquad \left(1-\frac{x}{n}\right)^n \approx e^{-x}$$

For a sufficiently large *p* we are left with only one disconnected node, i.e. C=1.

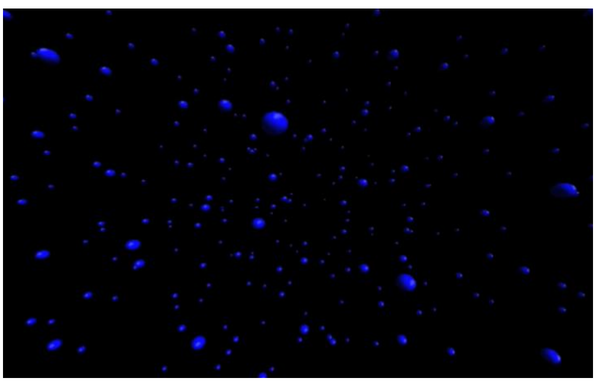
$$Ne^{-Np} = 1 \qquad \ln N - Np = 0$$

$$p \approx \frac{\ln N}{N} \qquad \langle k \rangle \approx \ln N$$

47



48



49

Are real networks like random graphs?

52

ARE REAL NETWORKS LIKE RANDOM GRAPHS?

As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory.

Note that once we have N and <k> for a random network, from it we can derive every measurable property. Indeed, we have:

Average path length: $\langle l_{rand} \rangle \approx \frac{\log N}{\log \langle k \rangle}$

Clustering Coefficient: $C_{rand} = p = \frac{\langle k \rangle}{N}$

Degree Distribution: $P_{rand}(k) \approx C^k_{N-1} p^k (1-p)^{N-1-k}$

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

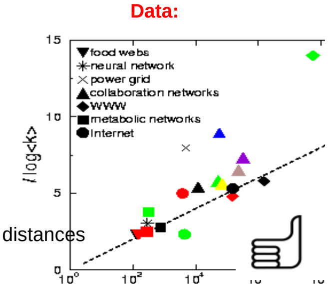
53

PATH LENGTHS IN REAL NETWORKS

Prediction:

$$l_{rand} = \frac{\log N}{\log \langle k \rangle}$$

Data:



Real networks have short distances like random graphs.

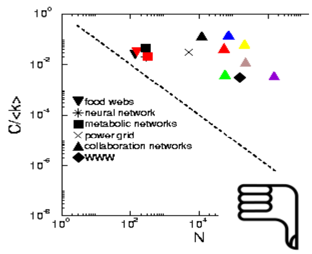
54

CLUSTERING COEFFICIENT

Prediction:

$$C_{rand} = \frac{\langle k \rangle}{N}$$

Data:



C_{rand} underestimates with orders of magnitudes the clustering coefficient of real networks.

Network Science: Random Graphs

55

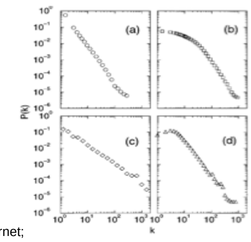
THE DEGREE DISTRIBUTION

Prediction:

$$P_{rand}(k) @ C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

Data:

$$P(k) \gg k^{-g}$$



(a) Internet;
(b) Movie Actors;
(c) Coauthorship, high energy physics;
(d) Coauthorship, neuroscience

Network Science: Random Graphs

56

ARE REAL NETWORKS LIKE RANDOM GRAPHS?

As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory.

Note that once we have N and <k> for a random network, from it we can derive every measurable property. Indeed, we have:

Average path length: $< l_{rand} > \gg \frac{\log N}{\log\langle k \rangle}$ 

Clustering Coefficient: $C_{rand} = p = \frac{\langle k \rangle}{N}$ 

Degree Distribution: $P_{rand}(k) @ C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$ 

Network Science: Random Graphs

57

ARE REAL NETWORKS LIKE RANDOM GRAPHS?

Network	N	L	<k>	ln N
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	186,936	8.08	10.04
Actor Network	212,250	3,054,278	28.78	12.27
Yeast Protein Interactions	2,018	2,930	2.90	7.61

Tab 3.1 Barabassi book

Network Science: Random Graphs

59

ARE REAL NETWORKS LIKE RANDOM GRAPHS?

As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory.

Note that once we have N and <k> for a random network, from it we can derive every measurable property. Indeed, we have:

Is there a GC if: $\langle k \rangle > \ln N$? 

Network Science: Random Graphs

60

IS THE RANDOM GRAPH MODEL RELEVANT TO REAL SYSTEMS?

(A) Problems with the random network model:
-the degree distribution differs from that of real networks
-the giant component in most real network does NOT emerge through a phase transition
-the clustering coefficient in most systems will now vanish, as predicted by the model.
-evolution of the network is different, in reality usually <k> < ln N

(B) Most important: we need to ask ourselves, are real networks random?

The answer is simply: NO

Hence it is **IRRELEVANT.**

A→ **WRONG** B → **IRRELEVANT**

There is no network in nature that we know of that would be described by the random network model.

Network Science: Random Graphs

61

IF IT IS WRONG AND IRRELEVANT, WHY DID WE DEVOT TO IT A FULL CLASS?

It is the reference model for the rest of the class.

It will help us calculate many quantities, that can then be compared to the real data, understanding to what degree is a particular property the result of some random process.

Organizing principles: patterns in real networks that are shared by a large number of real networks, yet which deviate from the predictions of the random network model.

In order to identify these, we need to understand how would a particular property look like if it is driven entirely by random processes.

While WRONG and IRRELEVANT, it will turn out to be extremly USEFUL!

Network Science: Random Graphs

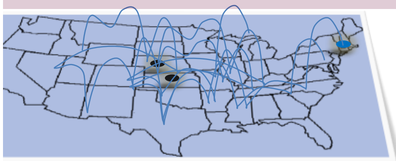
62

65

Milgram's Six Degrees

The first chain letters
The destination: Boston, Massachusetts
Starting Points: Omaha, Nebraska & Wichita, Kansas

SIX DEGREES



Travers and Milgram, Sociometry **32**,425 (1969)

66

Clustering vs. Randomness

A network can be a small world as long as clustering can be ignored

Where should we place the social network?

Clustered

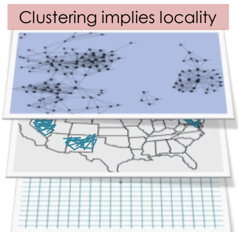
Random

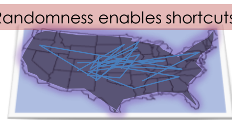
73

What we Really Mean by Clustering

Clustering implies locality

Randomness enables shortcuts





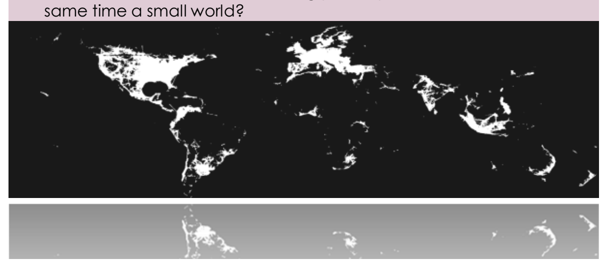
Locally Structured

Random

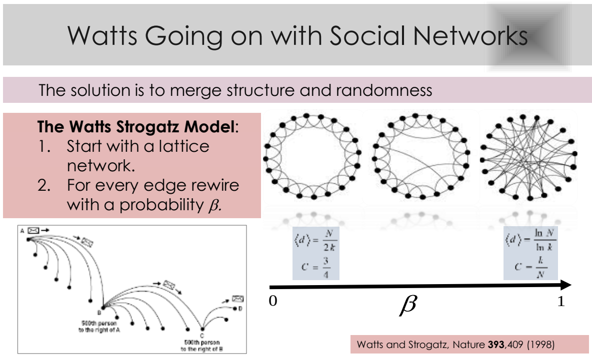
74

Watts Going on with Social Networks

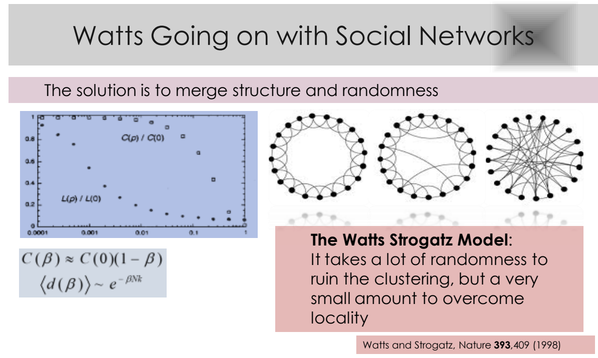
Could a network which is so strongly locally structured be at the same time a small world?



75

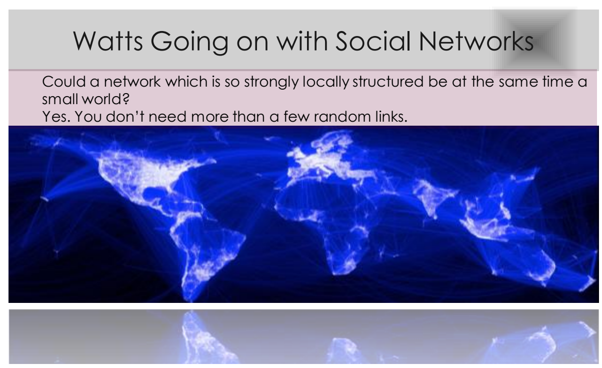


76

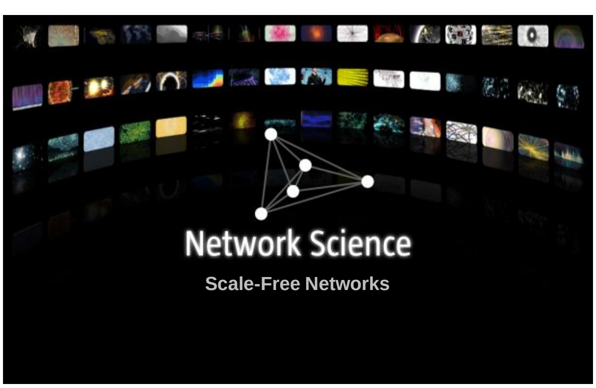


79

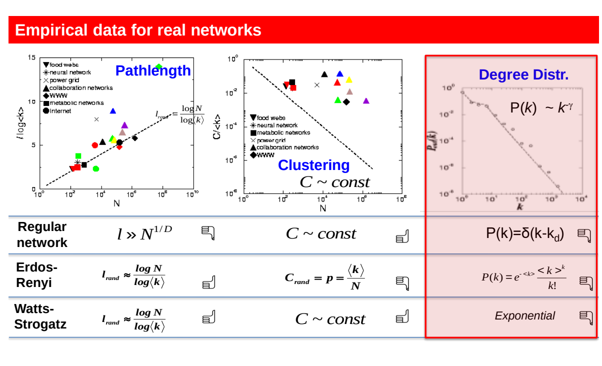
80



81



82



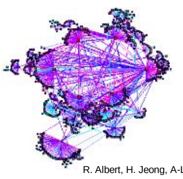
83

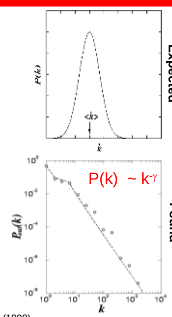
WORLD WIDE WEB

Nodes: **WWW documents**
Links: **URL links**

Over 3 billion documents

ROBOT: collects all URL's found in a document and follows them recursively

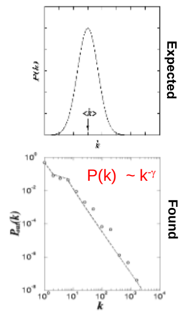


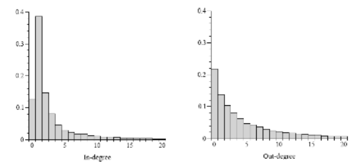


R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

84

Degree distribution of the WWW

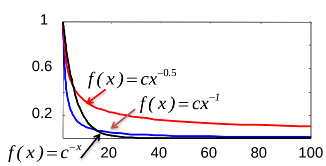




R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

85

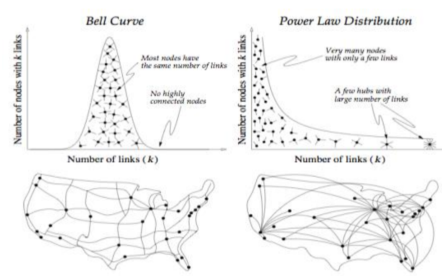
The difference between a power law and an exponential distribution



Above a certain x value, the power law is always higher than the exponential.

86

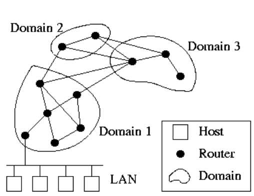
WORLD WIDE WEB

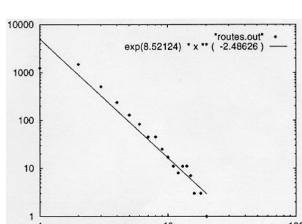


88

INTERNET BACKBONE

Nodes: computers, routers
Links: physical lines



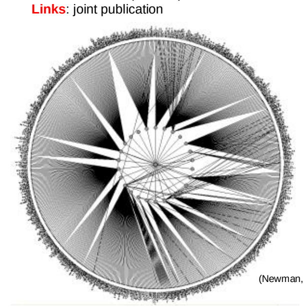


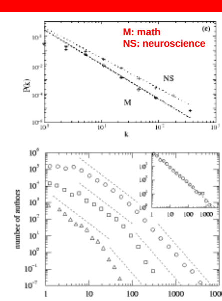
(Faloutsos, Faloutsos and Faloutsos, 1999)

90

SCIENCE COAUTHORSHIP

Nodes: scientist (authors)
Links: joint publication

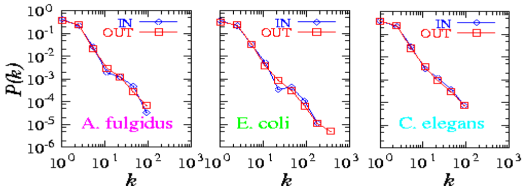




(Newman, 2000, Barabasi et al 2001)

91

METABOLIC NETWORK



Archaea Bacteria Eukaryotes

Organisms from all three domains of life are scale-free!

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

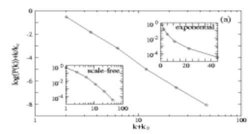
$P_{out}(k) \approx k^{-2.2}$

H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, 407 651 (2000)

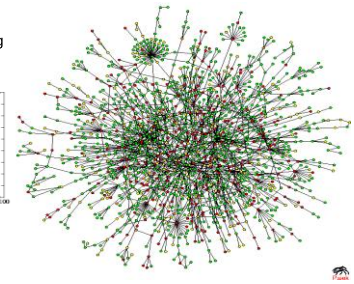
92

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions-binding



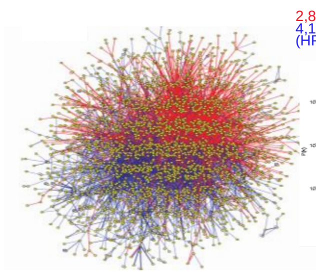
$P(k) \sim (k+k_0)^{-\gamma} \exp(-\frac{k+k_0}{k_\tau})$



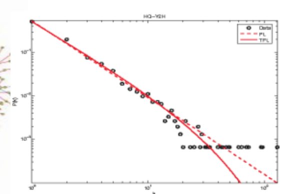
H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

93

HUMAN PROTEIN-PROTEIN INTERACTION NETWORK



2,800 Y2H interactions
4,100 binary LC interactions
(HPRD, MINT, BIND, DIP, MIPS)



Rual *et al.* Nature 2005; Stelze *et al.* Cell 2005

94

ACTOR NETWORK

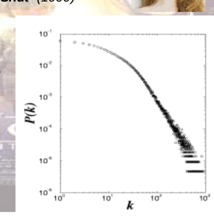
Nodes: actors
Links: cast jointly

IMDb Internet Movie Database

Days of Thunder (1990)
Far and Away (1992)
Eyes Wide Shut (1999)

N = 212,250 actors
⟨k⟩ = 28.78

P(k) ~$k^{-\gamma}$
γ=2.3



95

SCALE-FREE NETWORKS

Many real world networks have a similar architecture:

Scale-free networks

WWW, Internet (routers and domains), electronic circuits, computer software, movie actors, coauthorship networks, sexual web, instant messaging, email web, citations, phone calls, metabolic, protein interaction, protein domains, brain function web, linguistic networks, comic book characters, international trade, bank system, encryption trust net, energy landscapes, earthquakes, astrophysical network…

96

UNIVERSALITY AGAIN

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}
WWW	325 729	4.51	900	2.45	2.1
WWW	4×10^7	7		2.38	2.1
WWW	2×10^8	7.5	4000	2.72	2.1
WWW, site	260 000				1.94
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2
Internet, router*	3888	2.57	30	2.48	2.48
Internet, router*	150 000	2.66	60	2.4	2.4
Movie actors*	212 250	28.78	900	2.3	2.3
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1
Co-authors, math.*	70 975	3.9	120	2.5	2.5
Sexual contacts*	2810			3.4	3.4
Metabolic, *E. coli*	778	7.4	110	2.2	2.2
Protein, *S. cerev.**	1870	2.39		2.4	2.4
Ythan estuary*	134	8.7	35	1.05	1.05
Silwood Park*	154	4.75	27	1.13	1.13
Citation	783 339	8.57			3
Phone call	53×10^6	3.16		2.1	2.1
Words, co-occurrence*	460 902	70.13		2.7	2.7
Words, synonyms*	22 311	13.48		2.8	2.8

Critical phenomena:
Universality means that the exponents are the same for different systems… they are independent of details.

Networks:
The exponents vary from system to system.
Most are between 2 and 3

Universality:
the emergence of common features across different networks. Like the scale-free property.

97

FINITE SCALE-FREE NETWORKS

All real networks are finite → let us explore its consequences.
→ We have an expected maximum degree, K_{max}

Estimating K_{max}

$$\int_{K_{max}}^{\infty} P(k)dk \approx \frac{1}{N}$$

Why: the probability to have a node larger than K_{max} should not exceed the prob. to have one node, i.e. *1/N* fraction of all nodes

$$K_{max} = K_{min} N^{\frac{1}{\gamma-1}}$$

$$\int_{K_{max}}^{\infty} P(k)dk = (\gamma-1)K_{min}^{\gamma-1}\int_{K_{max}}^{\infty} k^{-\gamma}dk = \frac{(\gamma-1)}{(-\gamma+1)}K_{min}^{\gamma-1}\left[k^{-\gamma+1}\right]_{K_{max}}^{\infty} = \frac{K_{min}^{\gamma-1}}{K_{max}^{\gamma-1}} \approx \frac{1}{N}$$

Network Science: Scale-Free Property 2012

98

DISTANCES IN SCALE-FREE NETWORKS

$$K_{max} = K_{min} N^{\frac{1}{\gamma-1}}$$

$$\langle l \rangle \sim \begin{cases} const. & \gamma = 2 \\ \frac{\ln \ln N}{\ln(\gamma-1)} & 2 < \gamma < 3 \\ \frac{\ln N}{\ln \ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases}$$

Ultra Small World ($2 < \gamma < 3$); Small World ($\gamma > 3$)

$\gamma = 2$: Size of the biggest hub is of order O(N). Most nodes can be connected within two layers of it, thus the average path length will be independent of the system size.

$2 < \gamma < 3$: The average path length increases slower than logarithmically. In a random network all nodes have comparable degree, thus most paths will have comparable length. In a scale-free network the vast majority of the path go through the few high degree hubs, reducing the distances between nodes.

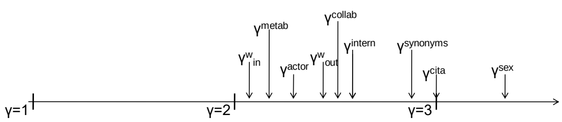
$\gamma = 3$: Some key models produce γ=3, so the result is of particular importance for them. This was first derived by Bollobas and collaborators for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

$\gamma > 3$: The second moment of the distribution is finite, thus in many ways the network behaves as a random network. Hence the average path length follows the result that we derived for the random network model earlier.

Cohen, Havlin Phys. Rev. Lett. 90, 58701(2003); Cohen, Havlin and ben-Avraham, in Handbook of Graphs and Networks, Eds. Bornholdt and Shuster (Willy-VCH, NY, 2002) Chap. 4; Confirmed also by: Dorogovtsev et al (2002), Chung and Lu (2002); (Bollobas, Riordan, 2002; Bollobas, 1985; Newman, 2001

99

SUMMARY OF THE BEHAVIOR OF SCALE-FREE NETWORKS

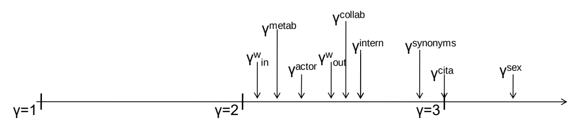


<k²> diverges		<k²> finite
<k> diverges	<k> finite	
Ultra small world behavior		Small world
Regime full of anomalies…	The scale-free behavior is relevant	Behaves like a random network

Network Science: Scale-Free Property 2012

100

SUMMARY OF THE BEHAVIOR OF SCALE-FREE NETWORKS



<k²> diverges		<k²> finite
Regime full of anomalies…	The scale-free behavior is relevant	Behaves like a random network

Why are most exponents in this regime?

Network Science: Scale-Free Property 2012

101

Network Science: Random Graphs 2012

102