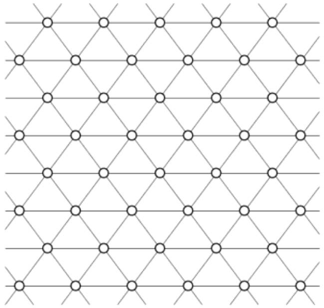
Systems Biology Across Scales: A Personal View VII. Networks: Models I

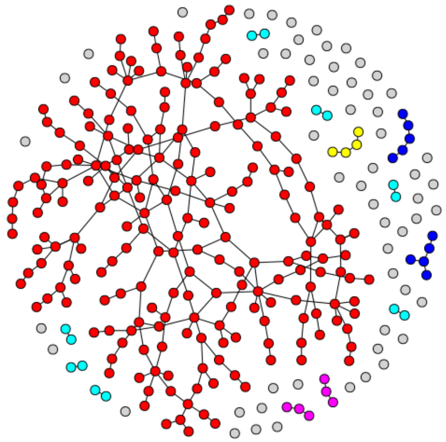
Sitabhra Sinha

IMSc Chennai

Theoretical understanding of networks



- Regular lattice or grid (*Physics*)
  - average path length ~ *N* (no. of nodes)
  - clustering *high*
  - delta function distribution of degree (links/node)



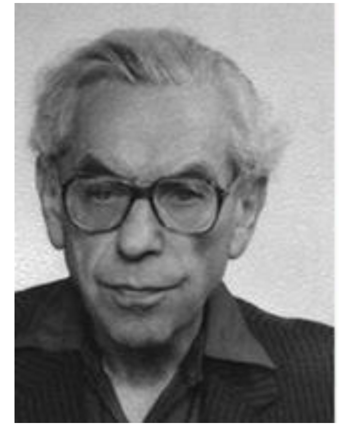
- Random networks (*Graph theory*)

  Also known as Erdos-Renyi networks
  - average path length ~ log *N*
  - clustering low
  - Poisson distribution of degree

Random networks

Erdos-Renyi model (1959): Two closely related probability-based models for generating random networks

Behavior of random networks is typically studied in the limit where the number of nodes N→∞



Paul Erdos
1913-1996

Alfred Renyi
1921-1970

- The G(N,L) model: when any member of a family of all graphs with N nodes and L links is chosen uniformly at random.

Example: G(3,2) comprises three possible networks of three nodes A,B and C such that each of Graph 1: {AB,BC}, Graph2: {AC,AB} and Graph 3: {AC,BC} can be picked with probability 1/3

- The G(N,p) model: when a network is constructed by randomly placing a link between each possible pair of nodes with a probability p ($0<p<1$)

Example: G(3,0.5) comprises the ensemble of all possible networks of three nodes A,B and C such that each of {AB,BC,AC} are inserted or not, based on tossing a fair coin

As $N\rightarrow\infty$, if $p \geq 2\ln(N)/N$ then a network will almost surely be connected.

Percolation & Random networks

A largest connected component (LCC, also referred to as giant component) is a connected component (for directed networks, strongly connected) whose size N_l is a finite fraction of that of the size N of the entire network, even as the network becomes larger and larger, i.e., $\lim_{N\to\infty} N_l / N = c > 0$.

In the G(N,p) random network model, the LCC size is
$N_l = 1$ when p = 0 (no nodes have any links) and $N_l = N$ when p = 1 (clique)
As p is gradually increased from 0 to 1, a phase transition occurs: the fraction N_l / N suddenly increases from 0 to a finite value (>0) at the critical value of p, $p_c = 1/N$

This concept is related to the theory of bond percolation
The Question: Consider a 2-dimensional lattice of N × N sites in which the links between any two neighboring sites is open with probability p [and hence, closed with prob (1 – p)]. What is the probability that a connected path exists from one side of the lattice to the other ?

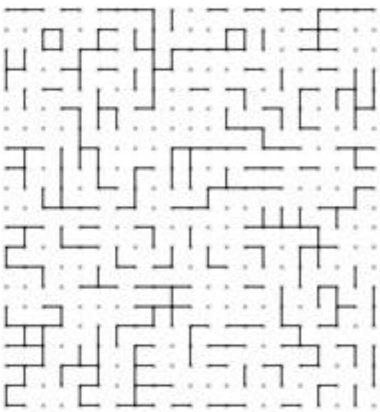


Image: http://www.suffos.bam.de

[Percolation theory has origins in understanding the process of transport through porous medium, e.g., of toxic chemical molecules through the filtering agent of a gas mask]

Path length & Clustering in Random networks

The average path length in the random network is $\langle L \rangle \sim \log \langle N \rangle / \log \langle k \rangle$

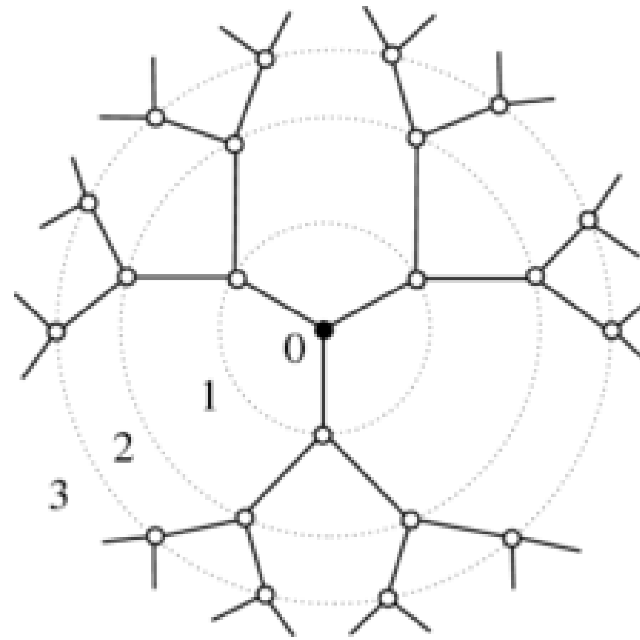


Image: http://i.stack.imgur.com

Intuition:
Locally, a random network G(N,p) with very small p – as cycles or closed loops involving only a few nodes are unlikely – will be approximately like a tree

The average number of neighbors located at distance d away from a node is :

$N_d = \langle k \rangle^d$

$\Rightarrow N = \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 + \ldots + \langle k \rangle^d \sim \langle k \rangle^d$

The average clustering coefficient in a G(N,p) random network is approximately $\langle C \rangle \sim p \approx \langle k \rangle / N$

This is because if you randomly select a node i and look at two neighboring nodes j and k connected to it, the probability that j & k will be connected is just p

Degree distribution of Random networks

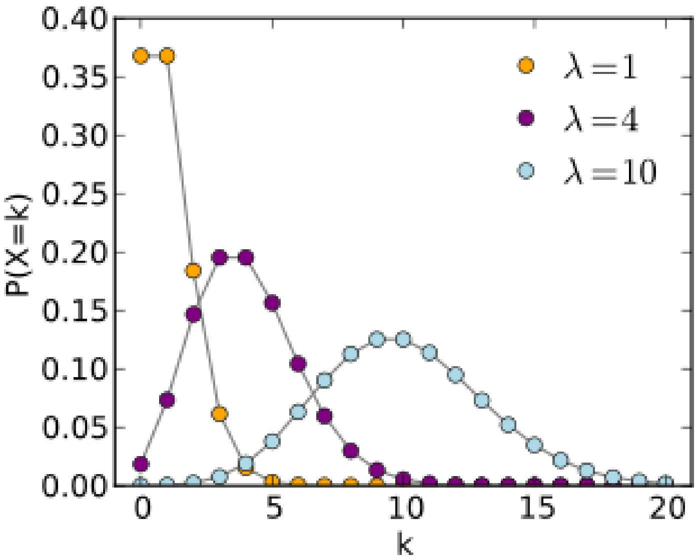
The G(N,p) model:

A given node in the network is connected with independent probability p to each of the N – 1 other nodes.
Thus the probability of being connected to k (and only k) other nodes is $p^k(1-p)^{N-1-k}$
There are ${}^{N-1}C_k$ ways to choose those k other vertices, and hence the total probability of being connected to exactly k others is
$p_k = {}^{N-1}C_k\, p^k(1-p)^{N-1-k}$
which is the Binomial distribution having mean Np and variance Np(1 – p)

As N becomes large with p being extremely small ($\rightarrow$0), such that $Np = \langle k \rangle = \lambda$ is finite, this tends to the Poisson distribution

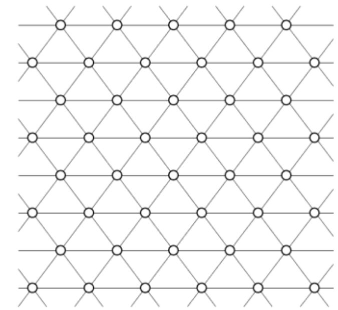
$$P(k) = e^{-\lambda} (\lambda^k/k!)$$

Both the mean and variance is given by λ.
For large values of λ this converges to the bell-shaped Gaussian or Normal distribution

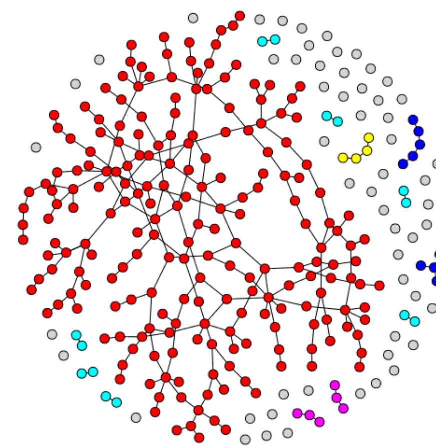


Source: wikipedia

Theoretical understanding of networks



- Regular lattice or grid (*Physics*)
  - average path length ~ N (no. of nodes)
  - clustering *high*
  - delta function distribution of degree (links/node)

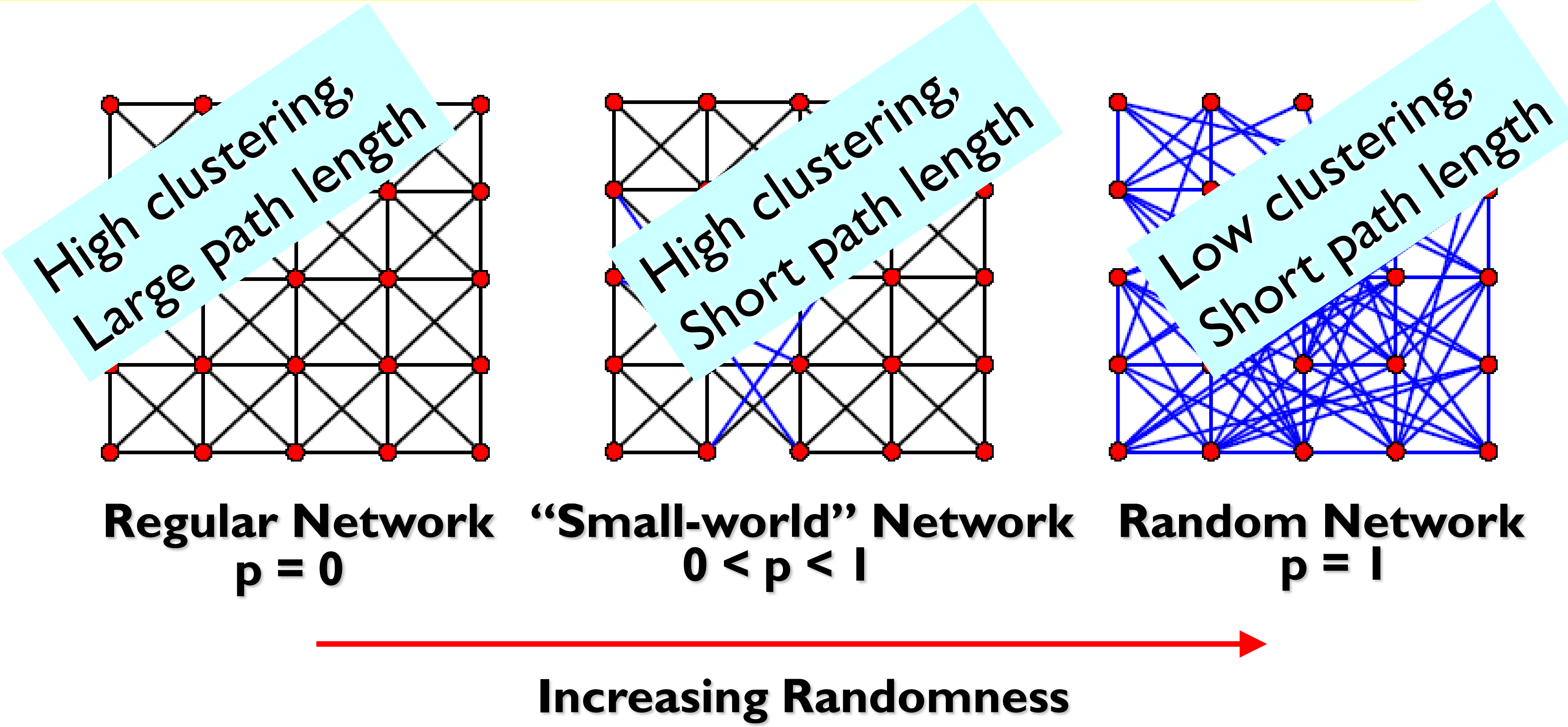


- Random networks (*Graph theory*)

  Also known as Erdos-Renyi networks
  - average path length ~ log N
  - clustering low
  - Poisson distribution of degree

Empirical networks are not random – many have certain structural patterns

Example: *small-world* networks



p: fraction of random, long-range connections

Watts and Strogatz (1998): Many biological, technological and social networks have connection topologies that lie between the two extremes of completely regular and completely random.

"It's a *small world*": The Milgram Experiment

eweb.furman.edu

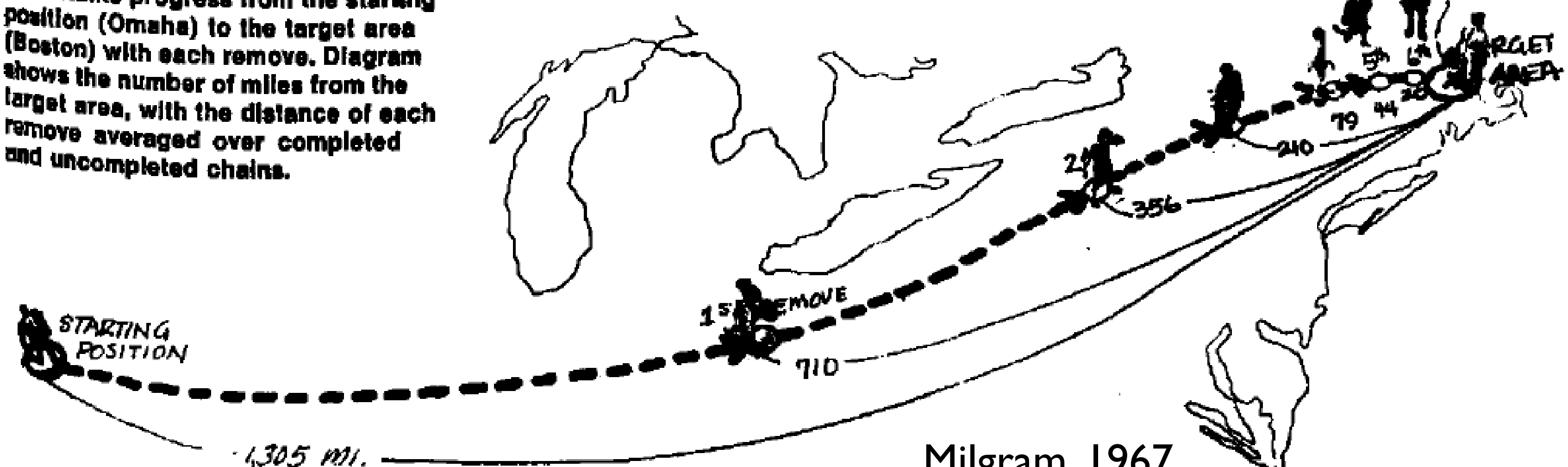
The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.

Stanley Milgram (1933-1984), US social psychologist

Arbitrarily selected individuals in Nebraska were asked to generate acquaintance chains (knowing on first name basis) connecting them to a target individual in Boston

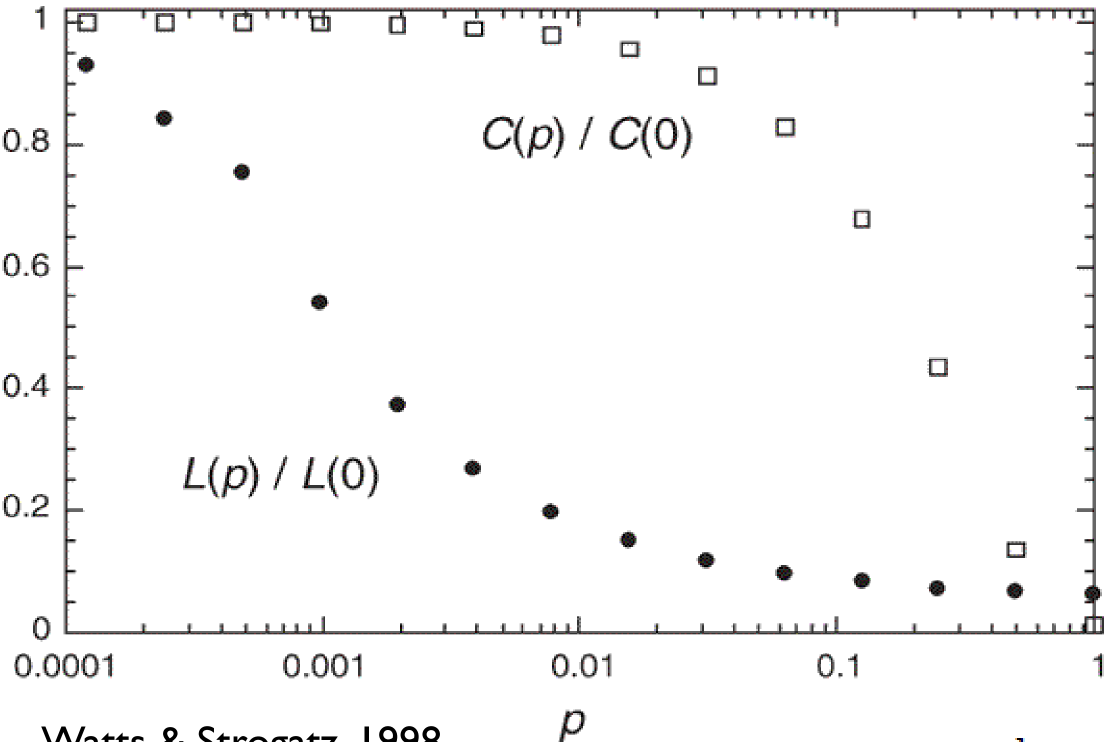
In one experiment, 64 of the 296 chains initiated eventually reached the target – the mean number of intermediaries between source and target being slightly larger than 5

$\Rightarrow$ Six degrees of separation



Milgram, 1967

"Small world": Local properties of regular networks but global properties of random networks



Watts & Strogatz, 1998

Characteristic path length

$$L = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}$$

d_{ij} : shortest distance between nodes i and j

Alternatively $\ell^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}$

Clustering coefficient

$$C = \frac{1}{n} \sum_i C_i$$

where

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}$$

Alternatively $$C = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}}$$

Small-world networks can be highly clustered (like regular networks), yet have small characteristic path lengths (as in random networks).

Epidemics on "Small world"

Watts & Strogatz, 1998

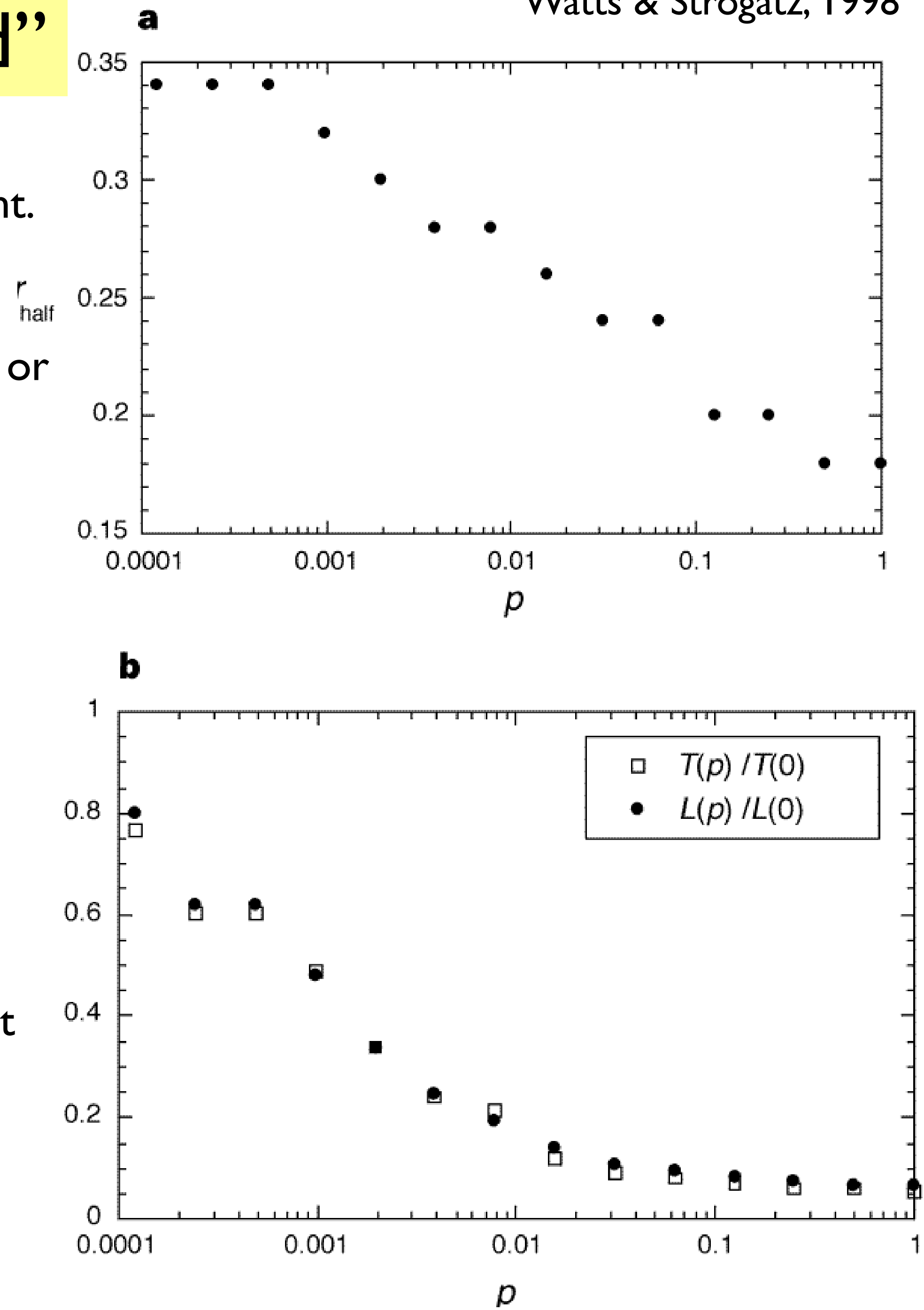
Dynamical process:

- Time t = 0: single infected individual present.
- Each infected agent can infect any of its neighbours with probability r.
- Infected individuals removed (by immunity or death) after unit period of sickness.

Key Results:

- Critical infectiousness r_{half}, at which the disease infects half the population, decreases with p
- Time required for a maximally infectious disease (r = 1) to spread throughout the entire population $T(p)$ has same form as characteristic path length $L(p)$

⇒ rewiring only a few links in the original lattice causes global infection to occur almost as fast as in random network



Implication:

"Control the truck-drivers"

Do *small-world* networks occur in real life ?

	# nodes	Avg degree	Avg path length		Clustering coefficient	
Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006
Power grid	4941	2.67	18.7	12.4	0.08	0.005
C. Elegans	282	14	2.65	2.25	0.28	0.05

Albert & Barabasi, 2003