Reti Sociali e Diffusione di Epidemie

Carlo PICCARDI e Renato CASAGRANDI

Dipartimento di Elettronica e Informazione,
Politecnico di Milano

email carlo.piccardi@polimi.it
http://home.dei.polimi.it/piccardi

STRUCTURE OF NETWORKS:

- Networks and their representation: examples
- Distance, diameter, degree distribution
- Network models: random, scale-free, small-world networks

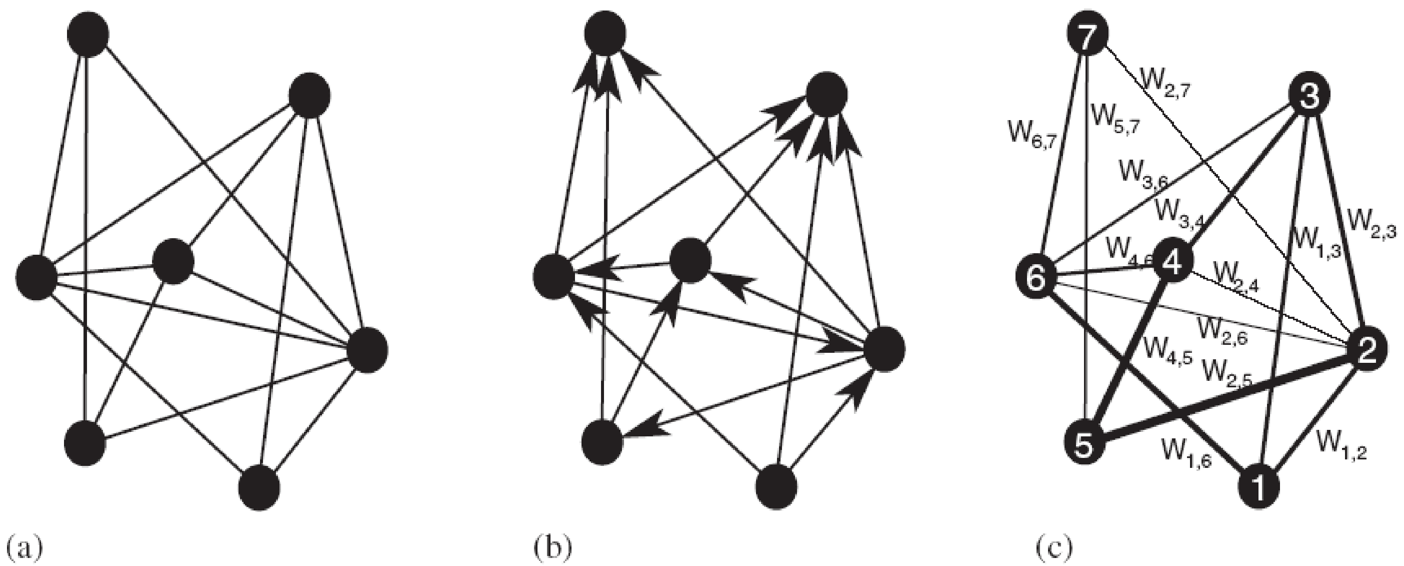
EPIDEMICS ON NETWORKS:

- Infectious diseases: classical and network approaches
- Modeling epidemics on networks

NETWORKS

A network is represented by a graph with N nodes (or vertices) and L links (or edges).

Nodes represent individuals, objects, subsystems, etc.. Links represent interactions, dependencies, communication channels, etc.



A network can be undirected (a,c) or directed (b), weighted (c) or unweighed (a,b).

Networks provide a truly interdisciplinary modeling tool...

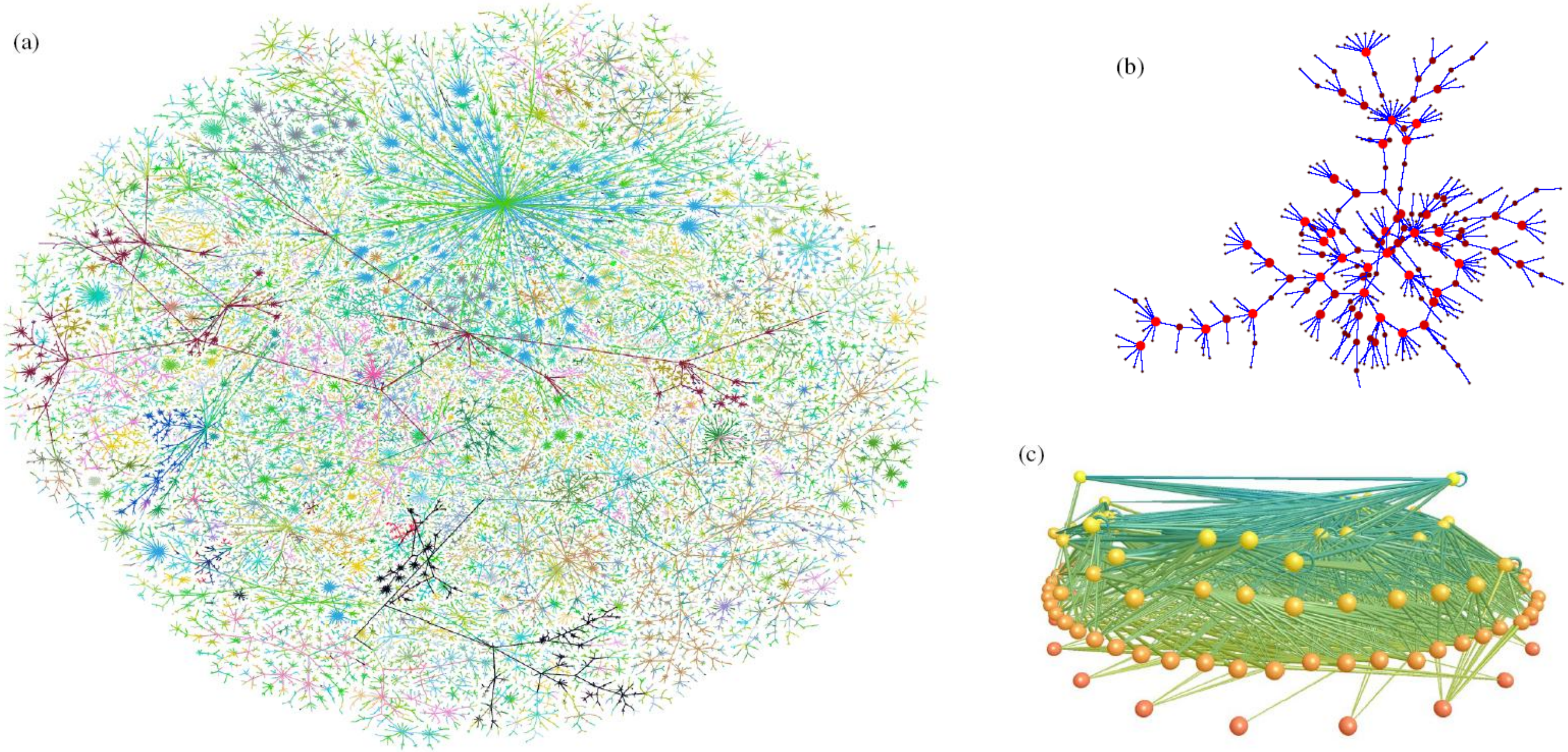


Fig. 1.2 *Three examples of the kinds of networks that are the topic of this review.* (a) *A visualization of the network structure of the Internet at the level of "autonomous systems"—local groups of computers each representing hundreds or thousands of machines. Picture by Hal Burch and Bill Cheswick, courtesy of Lumeta Corporation.* (b) *A social network, in this case of sexual contacts, redrawn from the HIV data of Potterat et al.* [341]. (c) *A food web of predator-prey interactions between species in a freshwater lake* [271]. *Picture courtesy of Richard Williams.*

Social networks...

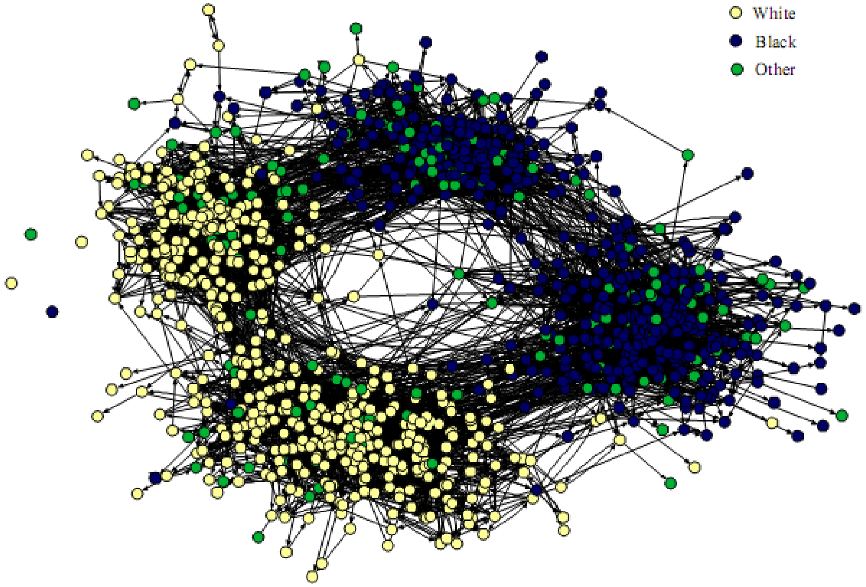
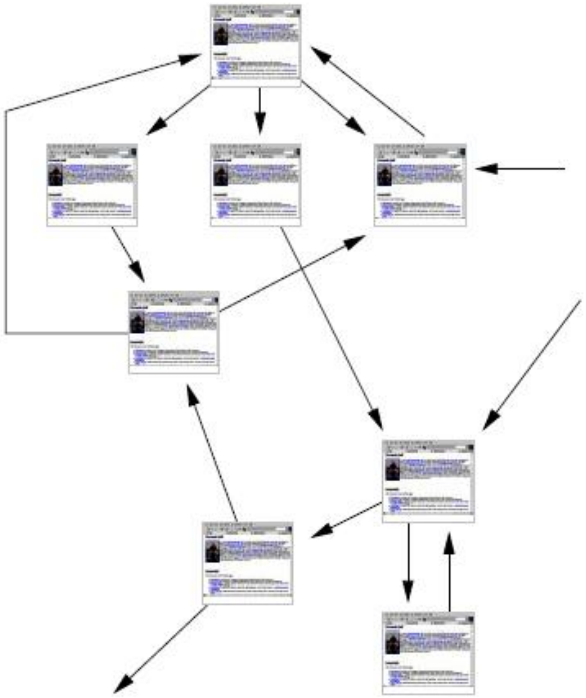


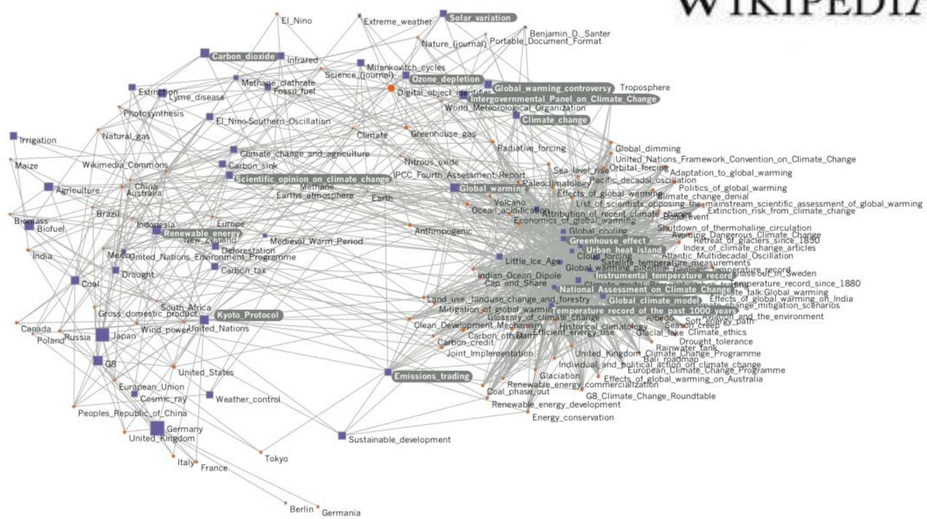
Fig. 3.4 *Friendship network of children in a U.S. school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not vice versa. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom is between middle school and high school, i.e., between younger and older children. Picture courtesy of James Moody.*

Information networks...

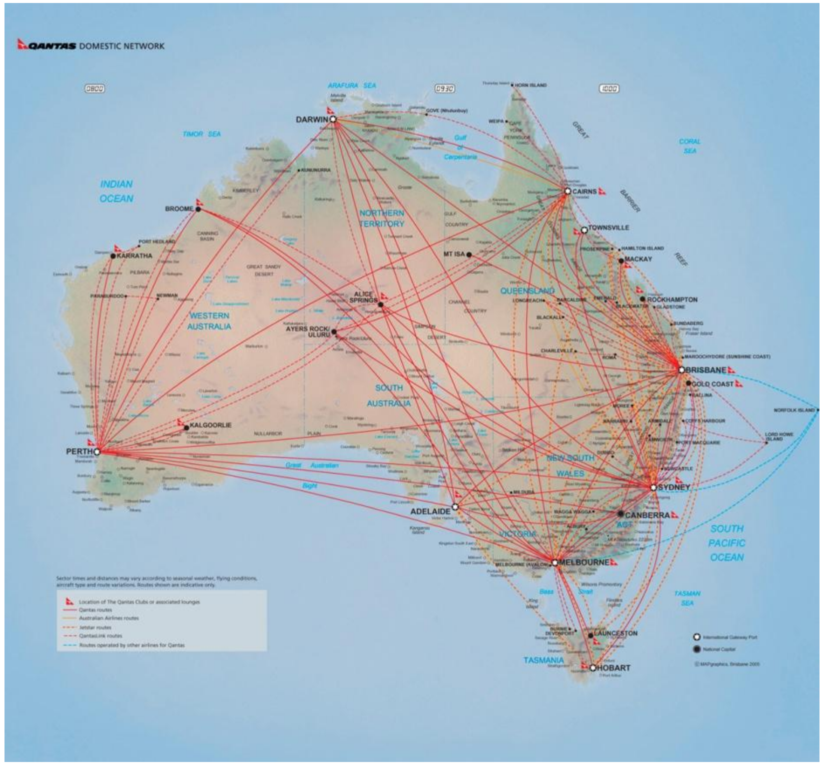


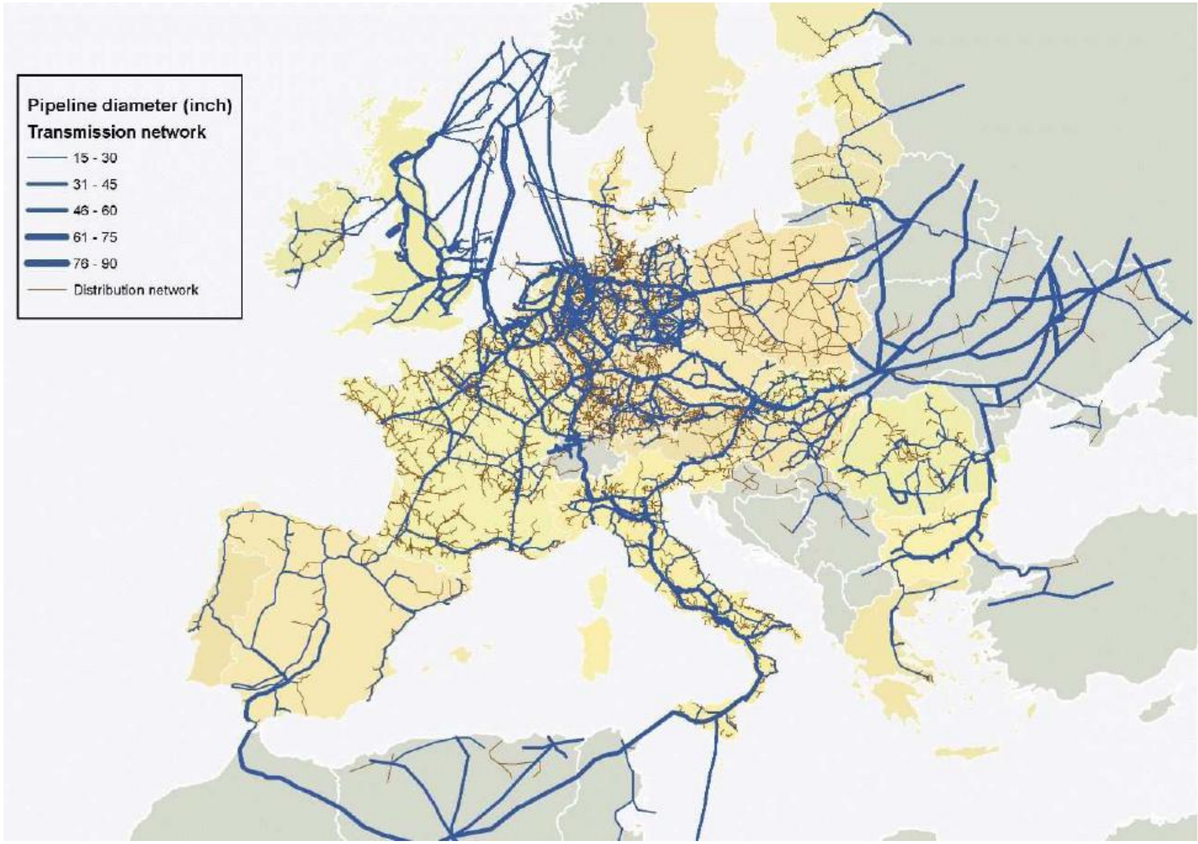
World–Wide Web

WIKIPEDIA



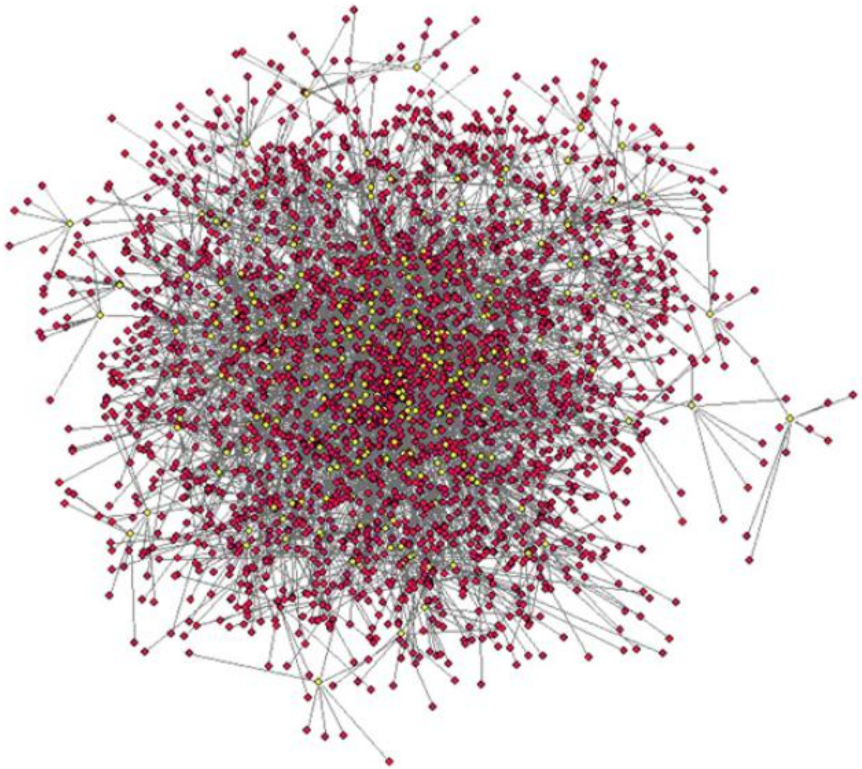
Transportation networks...



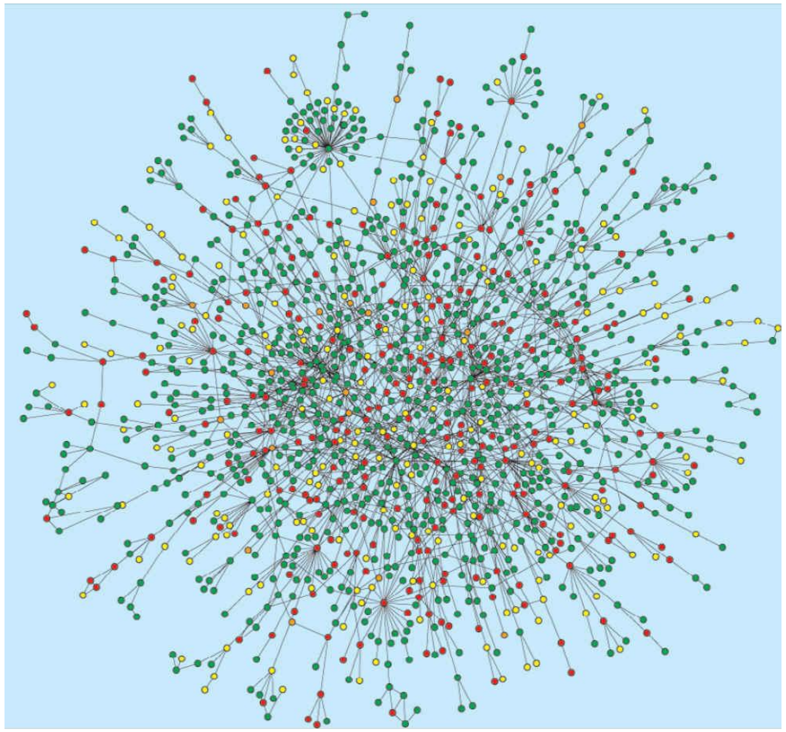


Similar topological structures are found in very different contexts:

⇨ common theories, methodologies, algorithms



The "directors network" of the Italian companies

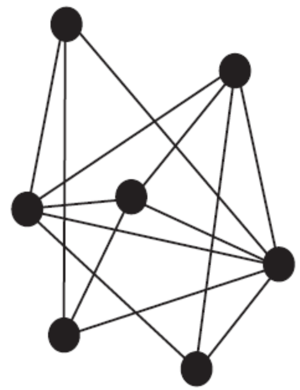


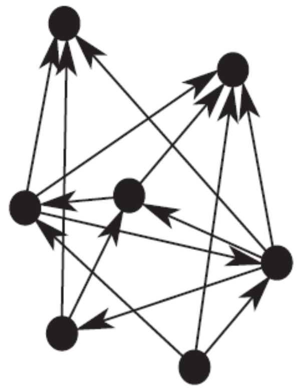
The protein interaction network of yeast

ADJACENCY MATRIX

An unweighed network is completely described by the $N \times N$ adjacency matrix $A = [a_{ij}]$:

$$a_{ij} = 1 \text{ if the link } i \to j \text{ exists,}$$
$$a_{ij} = 0 \text{ otherwise}$$





$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

A is symmetrical if the network is undirected, asymmetrical if the network is directed.

Typically A is a sparse matrix (many nodes, few edges), often very sparse.

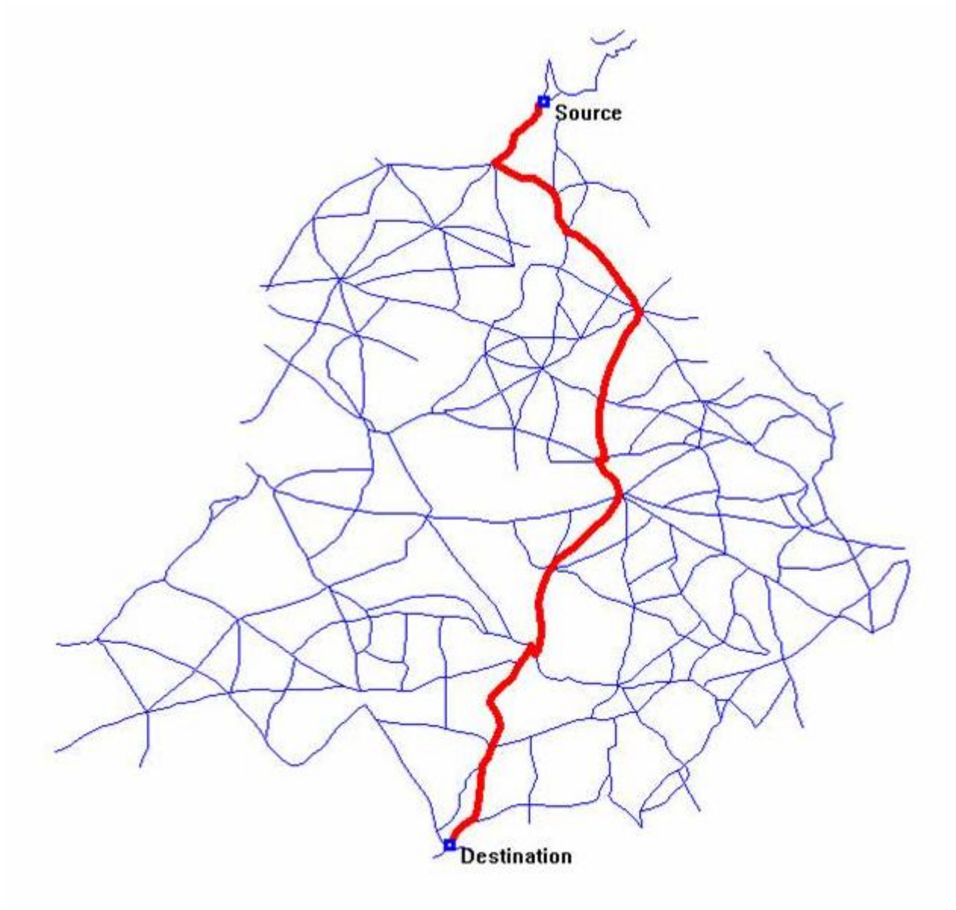
DISTANCE AND DIAMETER

The distance d_{ij} is the length (measured in number of links) of the shortest path connecting $i \to j$.

For a connected network, the diameter D and the average distance d are:

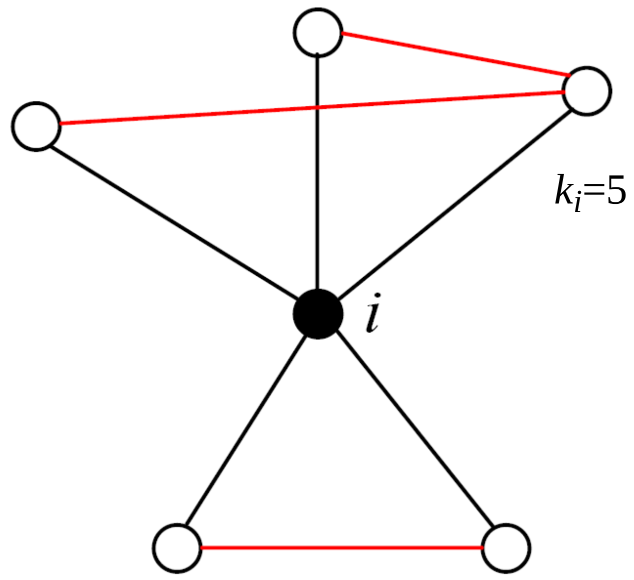
$$D = \max_{i,j} d_{ij}$$

$$d = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i,j\,(i \neq j)} d_{ij}$$



DEGREE AND DEGREE DISTRIBUTION

In an undirected network, the degree k_i of node i is the number of links connected to i (=the number of neighbors of i):



$$k_i = \sum_j a_{ij}$$

If the network is directed, we must distinguish between in-, out-, and total degree of node i.

The degree distribution $P(k)$ of a network specifies the fraction of nodes having exactly degree k (=the probability that a randomly selected node has degree k):

$$P(k) = \frac{\#\,\text{nodes with degree } k}{N} \quad , \qquad \sum_k P(k) = 1$$

It is often more practical to consider the cumulative degree distribution:

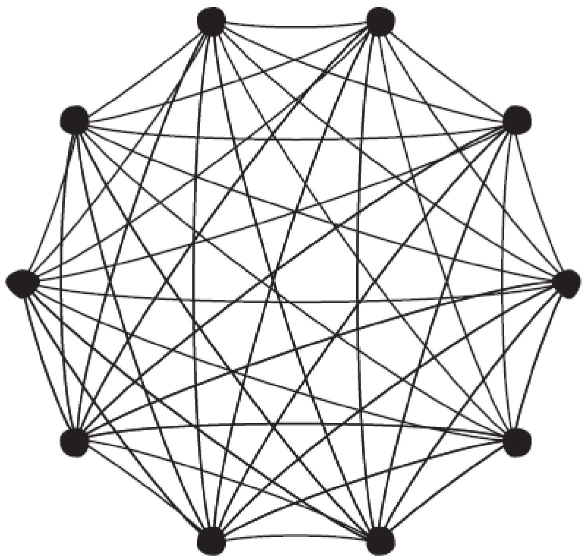
$$\overline{P}(k) = \frac{\#\,\text{nodes with degree} \geq k}{N} = \sum_{h=k}^{k_{\max}} P(h) \quad , \qquad \overline{P}(k_{\min}) = 1$$

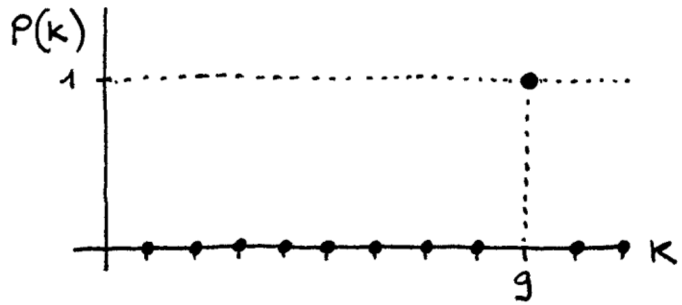
The first moment of the degree distribution $P(k)$ is the average degree:

$$< k >= \sum\nolimits_k kP(k) = \frac{1}{N} \sum\nolimits_i k_i = \frac{2L}{N}.$$

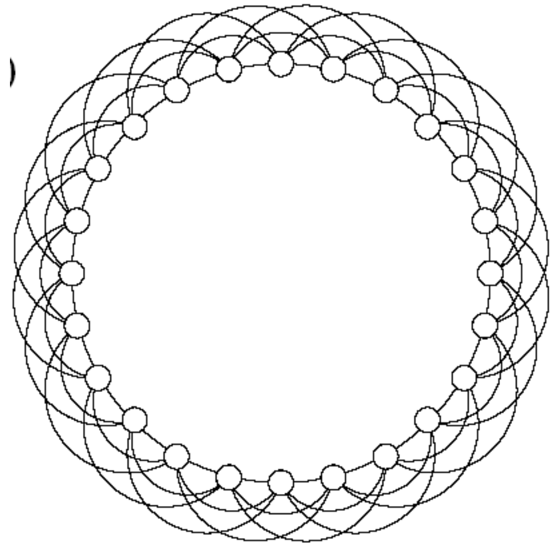
In a (strictly) homogeneous network all nodes have the same degree.

Example: a complete (=all-to-all) network with $N = 10$ and $k_i = \langle k \rangle = 9 \ \forall i$.

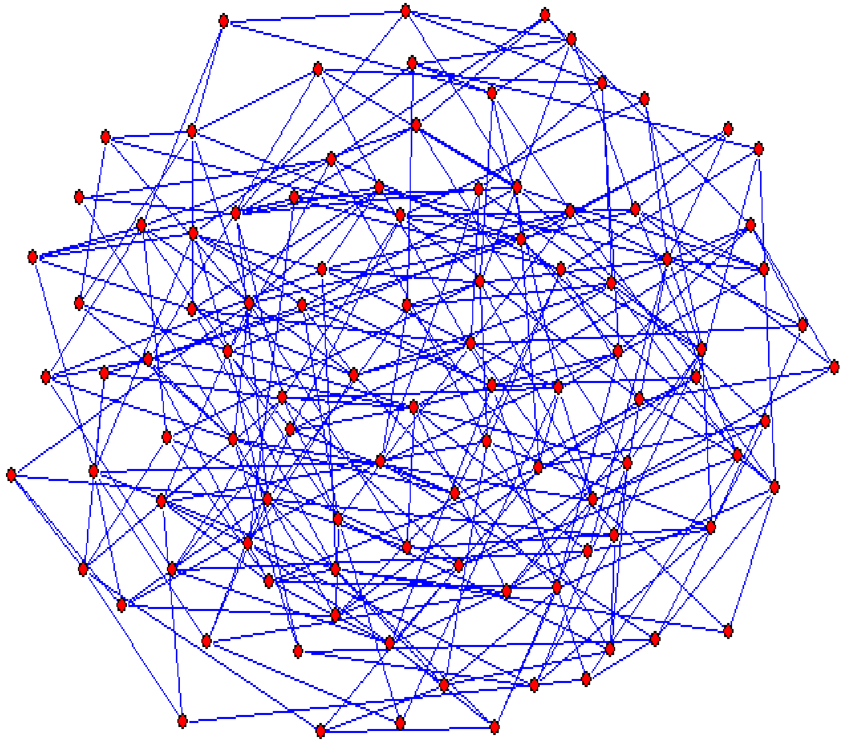




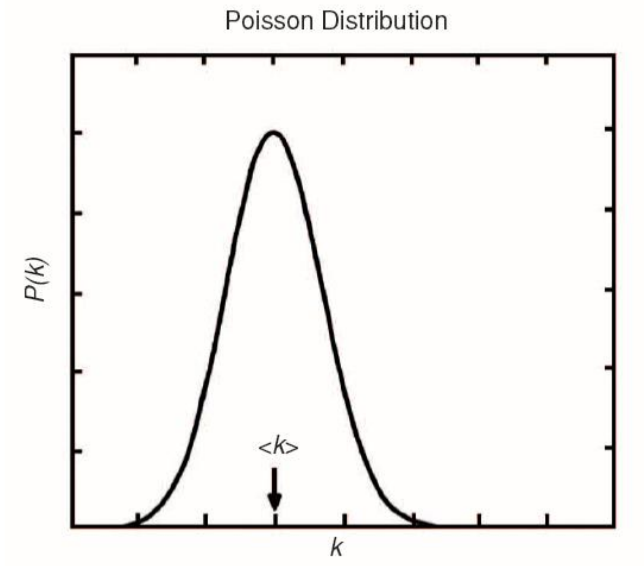
Example: a homogeneous network with $P(6) = 1$.



RANDOM NETWORKS



This is a random (or Erdös-Rényi) network, obtained by letting $N = 100$ and connecting $L = 300$ randomly extracted pairs (hence $< k >= 2L / N = 2 \times 300 / 100 = 6$).

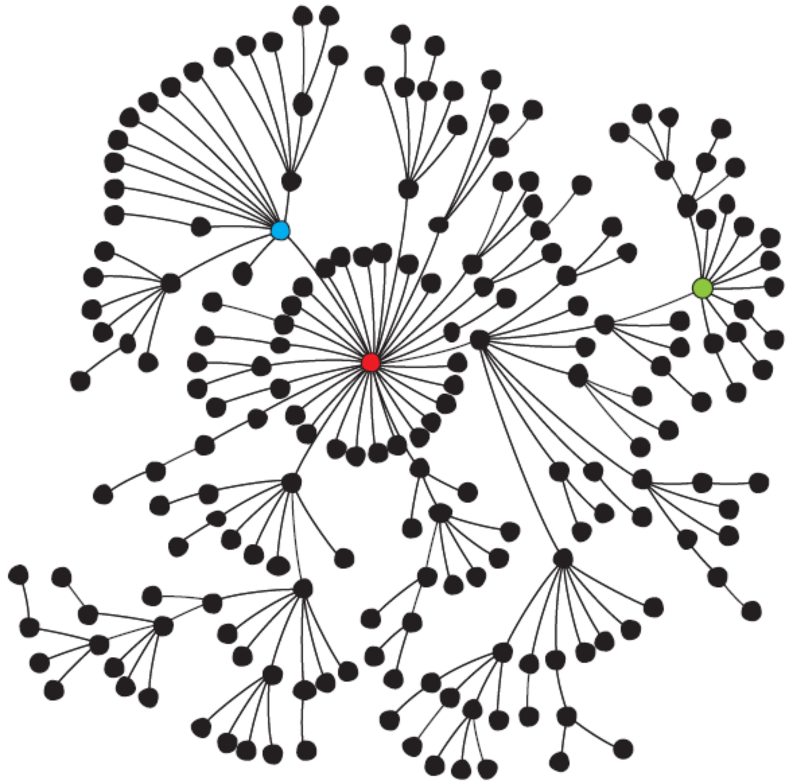


For large N, the degree is Poisson-distributed $P(k) = e^{-<k>} < k >^k / k!$, hence:

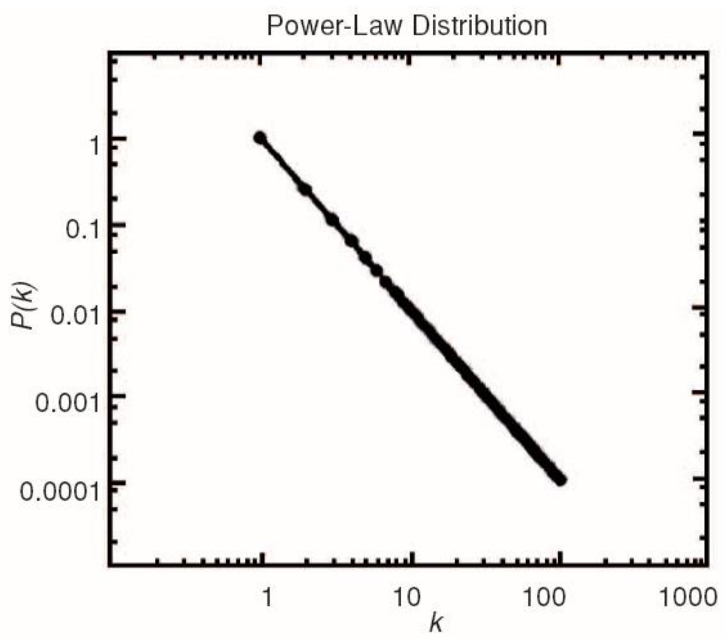
- the "typical" scale of node degree is $k_i =< k >$
- node degrees have small fluctuations around $< k >$
- the network is "almost homogeneous"

“SCALE-FREE” NETWORKS

This is a scale-free network, obtained by adding one node at a time, and connecting it preferentially (=with higher probability) to nodes with higher degree (Barabási-Albert algorithm).



The network contains few very connected nodes ("hubs") and many scarcely connected nodes.



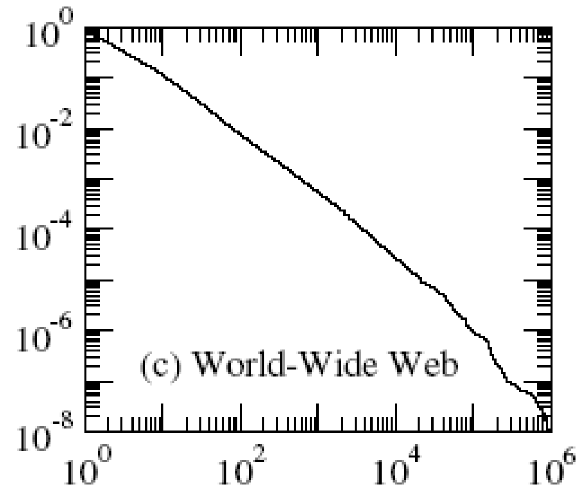
For large N, the degree distribution is a power-law function $P(k) \approx k^{-\alpha}$:

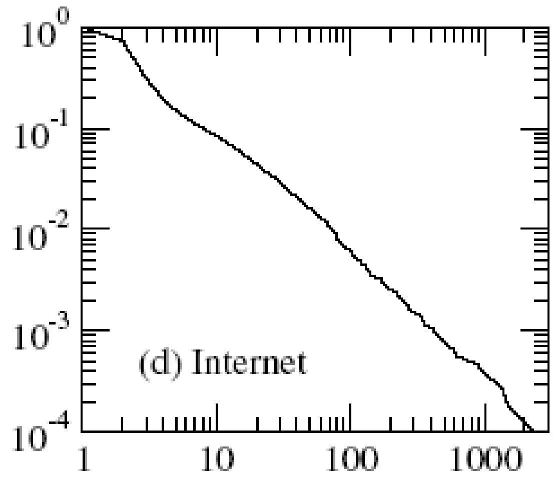
⇨ node degrees have large fluctuations around $<k>$: there is no "typical" scale of node degree

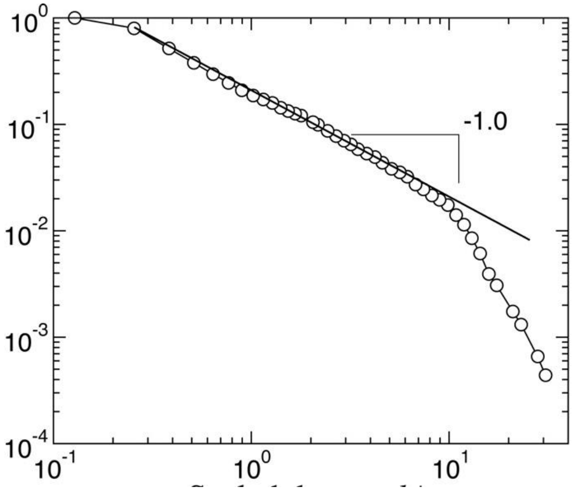
⇨ the network is strongly heterogeneous:

if $\alpha \leq 3$ the second moment $<k^2> = \sigma^2 + <k>^2$ diverges with N (“heavy tail”)

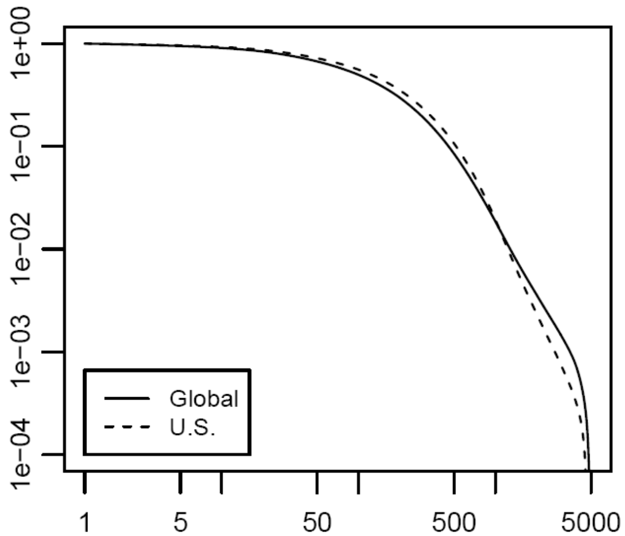
Some examples of (cumulative) degree distribution:







the air transportation network



Facebook (721 million nodes, May 2011)

Robustness & Fragility: A scale-free network is

- robust with respect to failures: if a node is removed at random (with all its links), the connected fraction of the network remains large and the average distance remains small.

- fragile with respect to attacks: if nodes are removed starting from those with highest degree, the connectivity rapidly decays.

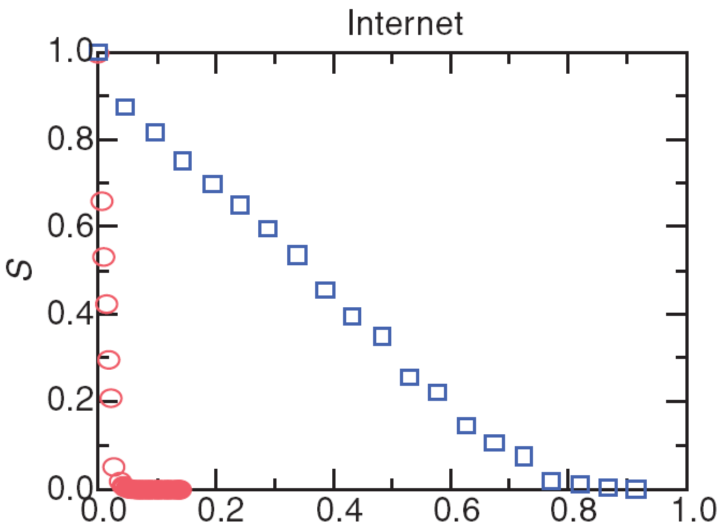
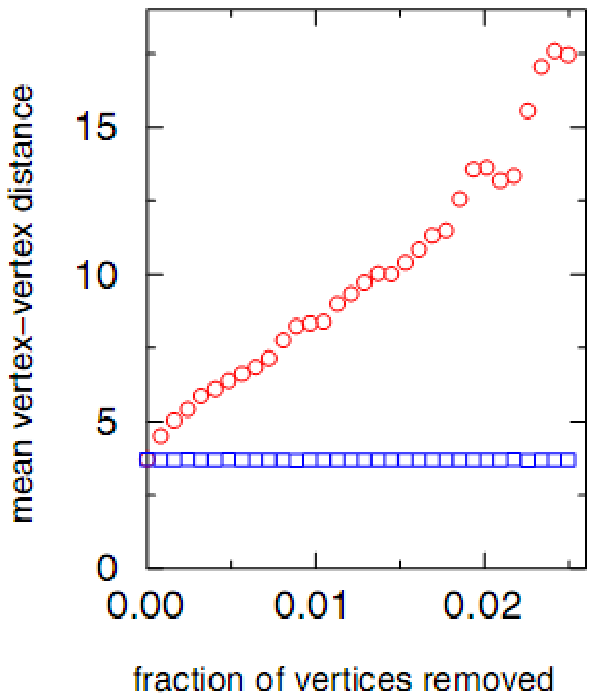


Figure 11. The relative size S of the largest cluster in the Internet, when a fraction f of the domains are removed [25]. □ , random node removal; ○, preferential removal of the most connected nodes.



"SMALL-WORLD" EFFECT

In typical real-world networks, the average distance $d = < d_{ij} >$ turns out to be surprisingly small.

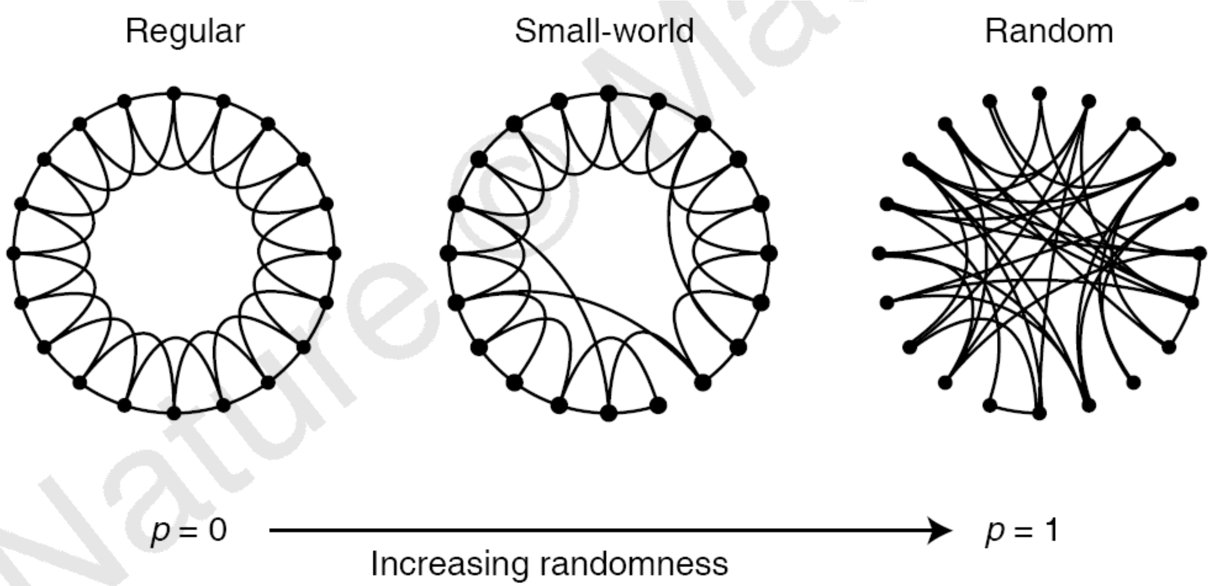
Empirically, it is observed that

$$d \approx \log N$$

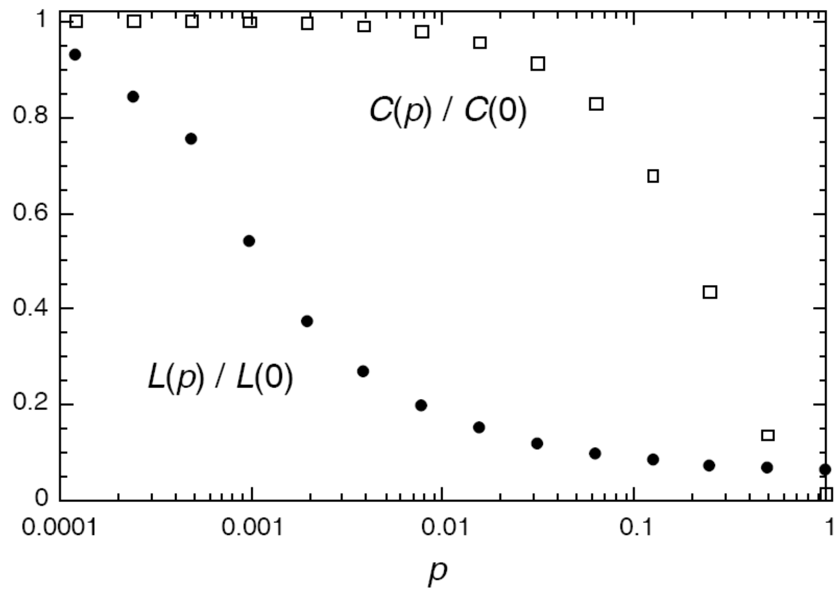
i.e. d increases "slowly" with N ("small-world" effect).

	Network	Type	n	m	z	ℓ
Social	film actors	undirected	449 913	25 516 482	113.43	3.48
	company directors	undirected	7 673	55 392	14.44	4.60
	math coauthorship	undirected	253 339	496 489	3.92	7.57
	physics coauthorship	undirected	52 909	245 300	9.27	6.19
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92
	telephone call graph	undirected	47 000 000	80 000 000	3.16	
	email messages	directed	59 912	86 300	1.44	4.95
	email address books	directed	16 881	57 029	3.38	5.22
	student relationships	undirected	573	477	1.66	16.01
	sexual contacts	undirected	2 810			
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18
	citation network	directed	783 339	6 716 198	8.57	
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87
	word co-occurrence	undirected	460 902	17 000 000	70.13	
Technological	Internet	undirected	10 697	31 992	5.98	3.31
	power grid	undirected	4 941	6 594	2.67	18.99
	train routes	undirected	587	19 603	66.79	2.16
	software packages	directed	1 439	1 723	1.20	2.42
	software classes	directed	1 377	2 213	1.61	1.51
	electronic circuits	undirected	24 097	53 248	4.34	11.05
	peer-to-peer network	undirected	880	1 296	1.47	4.28
Biological	metabolic network	undirected	765	3 686	9.64	2.56
	protein interactions	undirected	2 115	2 240	2.12	6.80
	marine food web	directed	135	598	4.43	2.05
	freshwater food web	directed	92	997	10.84	1.90
	neural network	directed	307	2 359	7.68	3.97

Watts and Strogatz (1998) demonstrated that adding a few long-distance connections to a regular network yields a dramatic decrease of d.



p = fraction of links rewired



If we increase p from zero the average distance rapidly decreases.

STRUCTURE OF NETWORKS:

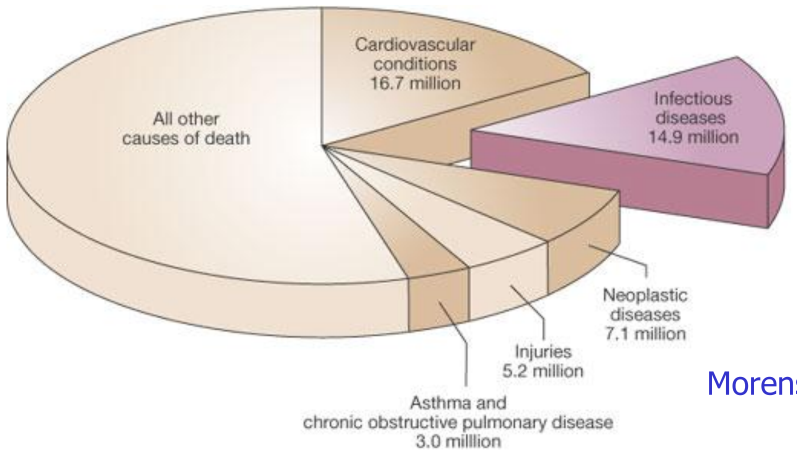
- Networks and their representation: examples
- Distance, diameter, degree distribution
- Network models: random, scale-free, small-world networks

EPIDEMICS ON NETWORKS:

- Infectious diseases: classical and network approaches
- Modeling epidemics over networks

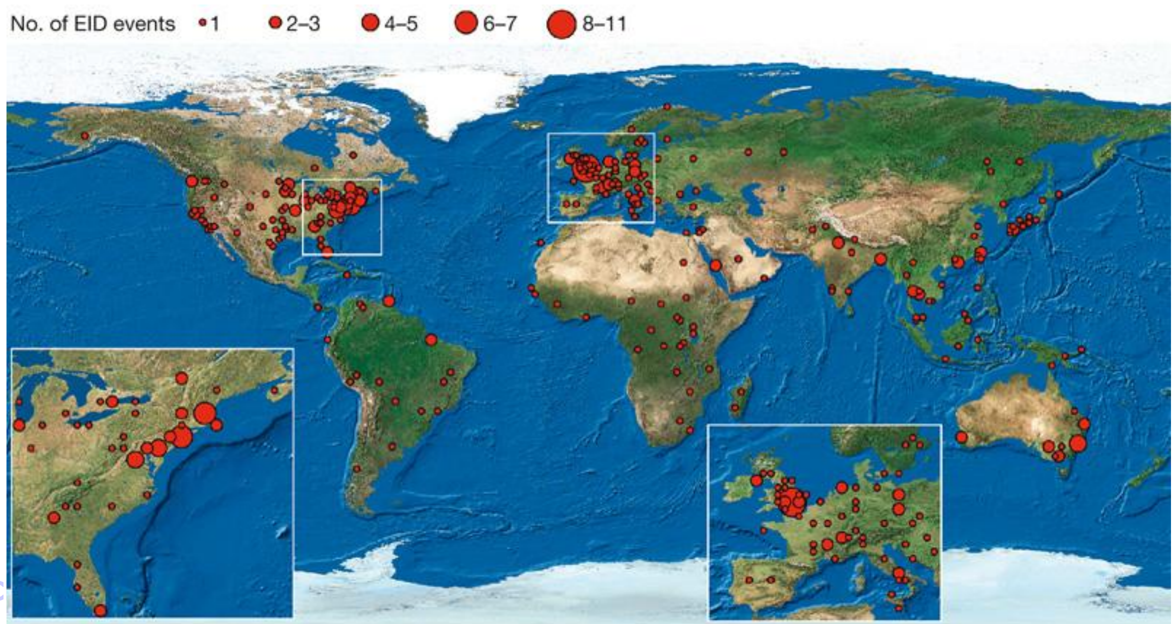
The importance of infectious diseases

More than 25% of annual deaths worldwide are caused by infectious diseases



Infectious diseases	Annual deaths (million)
Respiratory infections	3.96
HIV/AIDS	2.77
Diarrhoeal diseases	1.80
Tuberculosis	1.56
Vaccine-preventable childhood diseases	1.12
Malaria	1.27
STDs (other than HIV)	0.18
Meningitis	0.17
Hepatitis B and C	0.16
Tropical parasitic diseases	0.13
Dengue	0.02
Other infectious diseases	1.76

Morens et al (2004) *Nature* **430**:242

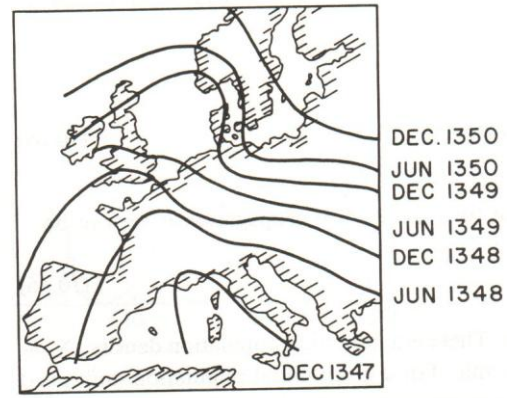


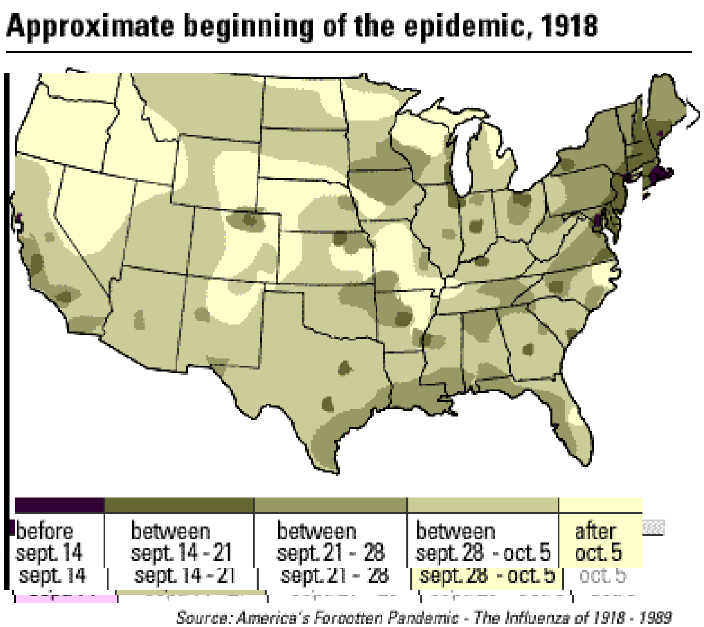
335 new diseases emerged from 1940 to 2004

Jones et al (2008) *Nature* **451**:990

The importance of infectious diseases

Victimes de la peste de 1349
(Gilles de Muisit, 1272-1353)





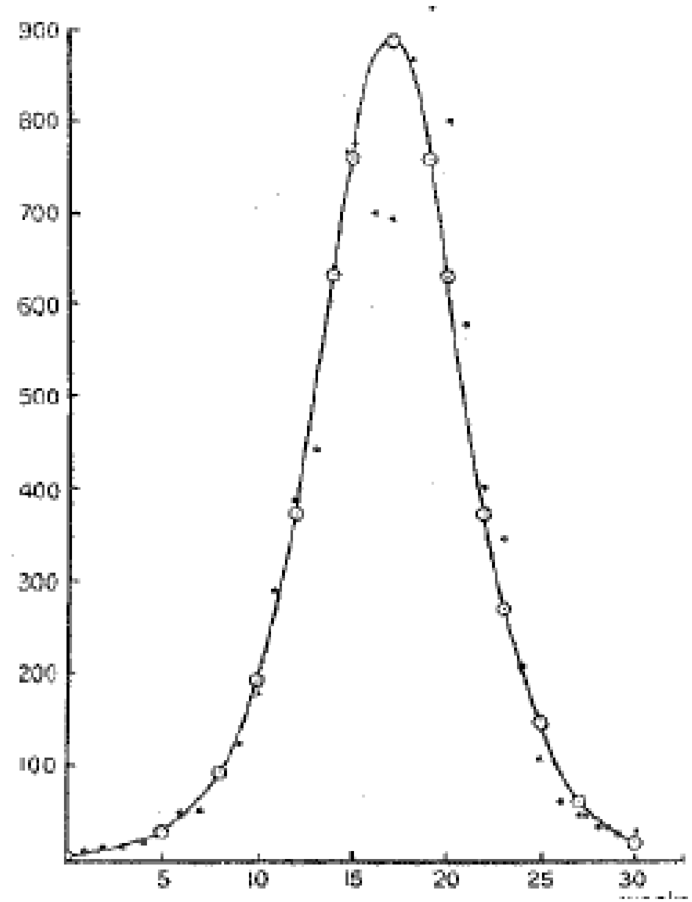
Classical modelling approaches

Homogeneous mixing among N individuals

y_t is "the total number who are ill" at time t (infectious, thus infectives)

x_t "denotes the number of individuals still unaffected" (susceptibles)

z_t is "the number who have been removed by recovery and death" (recovered)



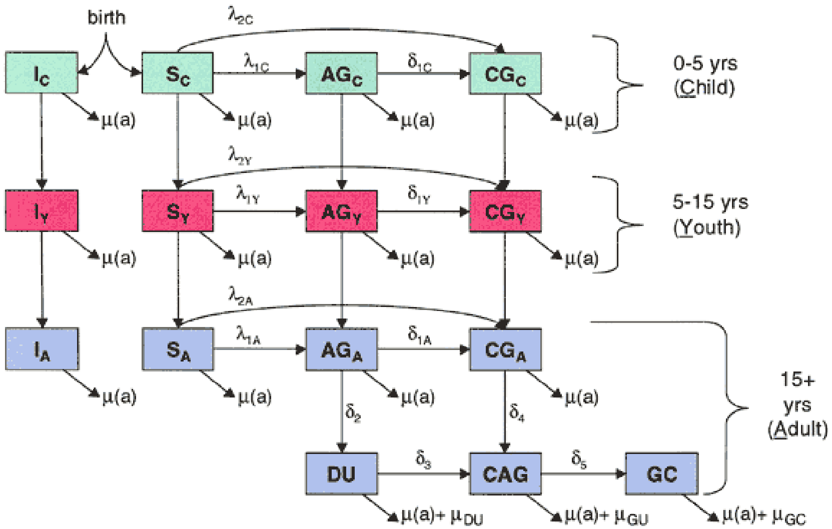
$$\left.\begin{aligned} \frac{dx}{dt} &= -\kappa xy \\ \frac{dy}{dt} &= \kappa xy - ly \\ \frac{dz}{dt} &= ly \end{aligned}\right\}$$

of infection.) This follows since the chance of an infection is proportional to the number of infected on the one hand, and to the number not yet infected on the other.

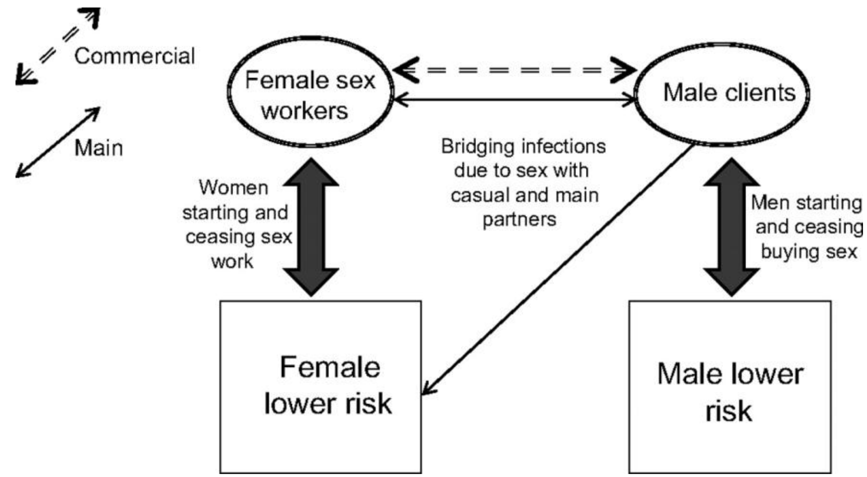
Kermack and McKendrick (1927), Proc. Royal Soc. A **115**: 700

Classical modelling approaches

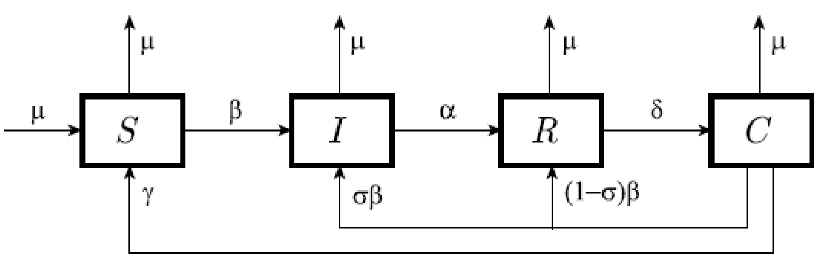
The flexibility of compartmental models



Rupnow et al (2000), *Emerg Infec Dis* **6**: 228



Pickles et al (2010), *Sex Transm Infect* **86**: i33

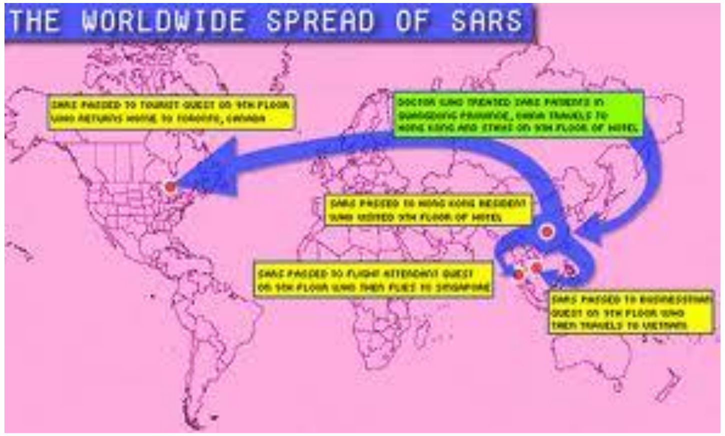


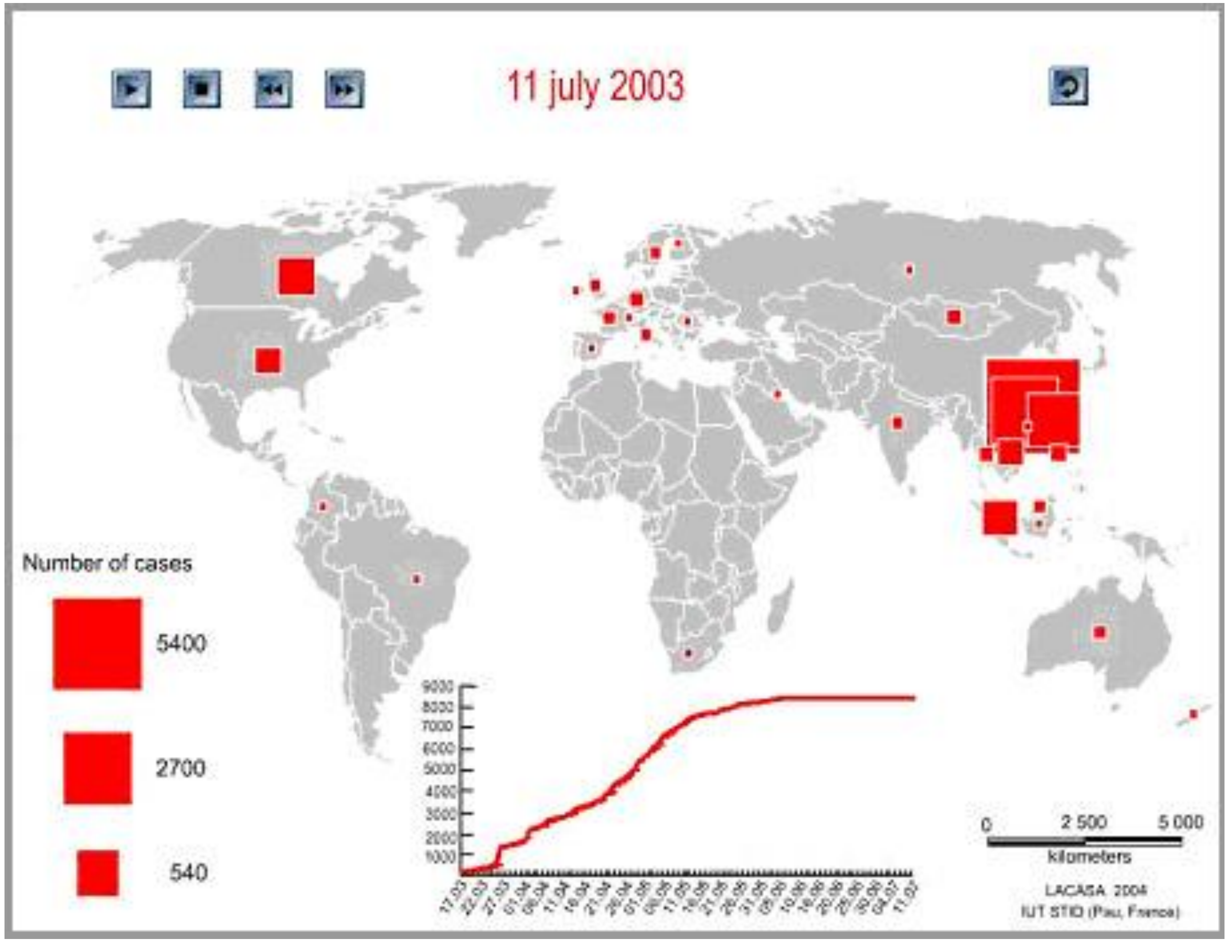
Casagrandi et al (2006), *Math Biosc* **200**: 152

$$\frac{dS_1(t)}{dt} = \lambda_1 - \mu S_1(t) - B_1(t)S_1(t) + C_{21}S_2(t) - C_{12}S_1(t),$$
$$\frac{dI_1(t)}{dt} = B_1(t)S_1(t) - (\mu + \sigma + D_{12})I_1(t) + D_{21}I_2(t),$$
$$\frac{dS_2(t)}{dt} = \lambda_2 - \mu S_2(t) - B_2(t)S_2(t) + C_{12}S_1(t) - C_{21}S_2(t),$$
$$\frac{dI_2(t)}{dt} = B_2(t)S_2(t) - (\mu + \sigma + D_{21})I_2(t) + D_{12}I_1(t),$$
$$\frac{dA(t)}{dt} = \sigma(I_1(t) + I_2(t)) - (\mu + \gamma)A(t),$$

Greenhalgh et al (2001), *IMA* **18**: 225

The need for network approaches: the small-world effect





The need for network approaches: scale-free networks

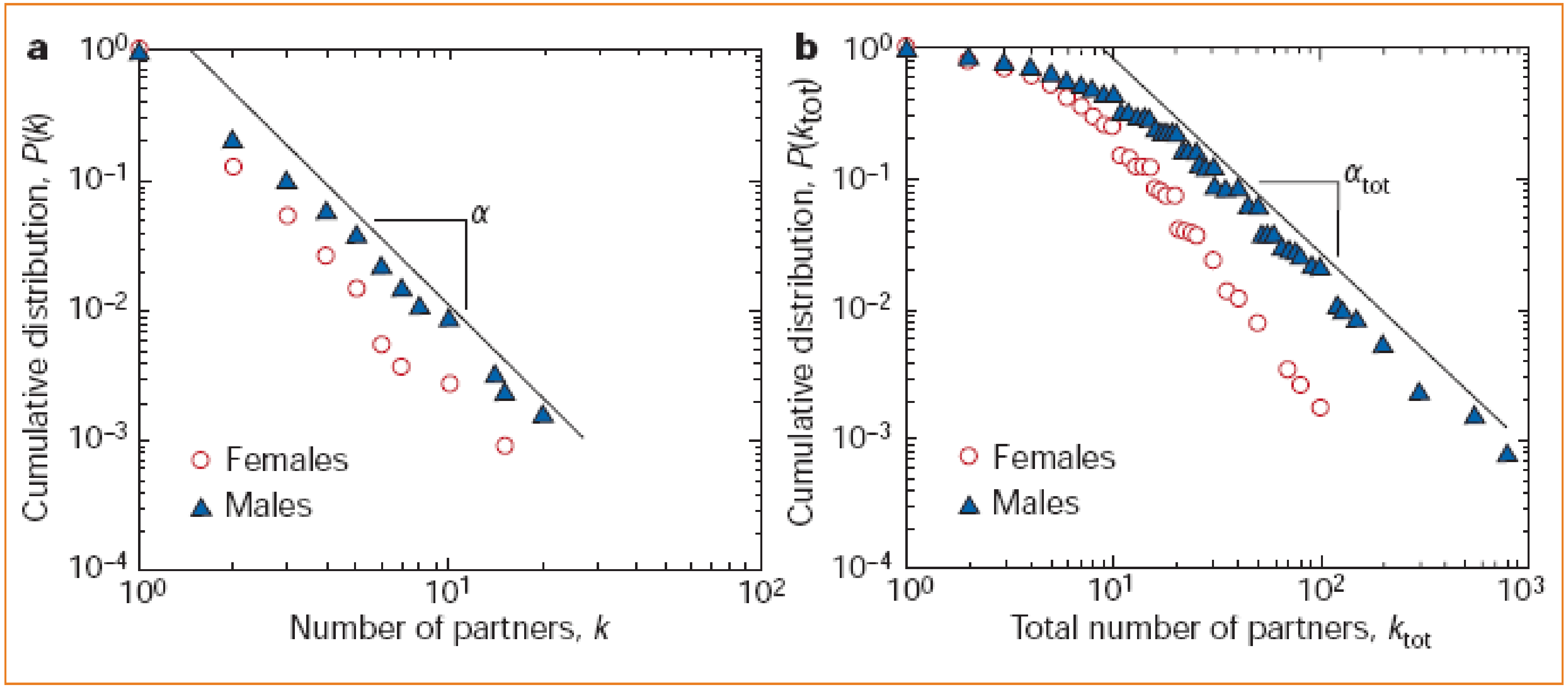


Figure 2 Scale-free distribution of the number of sexual partners for females and males. **a,** Distribution of number of partners, k, in the previous 12 months. Note the larger average number of partners for male respondents: this difference may be due to 'measurement bias' — social expectations may lead males to inflate their reported number of sexual partners. Note that the distributions are both linear, indicating scale-free power-law behaviour. Moreover, the two curves are roughly parallel, indicating similar scaling exponents. For females, $\alpha = 2.54 \pm 0.2$ in the range $k > 4$, and for males, $\alpha = 2.31 \pm 0.2$ in the range $k > 5$. **b,** Distribution of the total number of partners k_{tot} over respondents' entire lifetimes. For females, $\alpha_{tot} = 2.1 \pm 0.3$ in the range $k_{tot} > 20$, and for males, $\alpha_{tot} = 1.6 \pm 0.3$ in the range $20 < k_{tot} < 400$. Estimates for females and males agree within statistical uncertainty.

CONTAGION AND EPIDEMICS ON NETWORKS

Probabilistic cellular automata are used to model the spread of infectious diseases over the network - but also of products' adoption, opinions, etc.

- FINITE STATE SET: node (=individual) i is in state $s^i \in \Sigma = \{1, 2, \ldots, \sigma\}$ at time t

  e.g.:
  Σ={Susceptible, Infected, Recovered}
  in epidemics
  Σ={Non adopter, Adopter}
  in marketing

- LOCAL RULES (=CONTAGION MECHANISM): the next state s^i_{t+1} depends (according to probabilistic rules) on s^i_t and on the state s^j_t of the neighbors

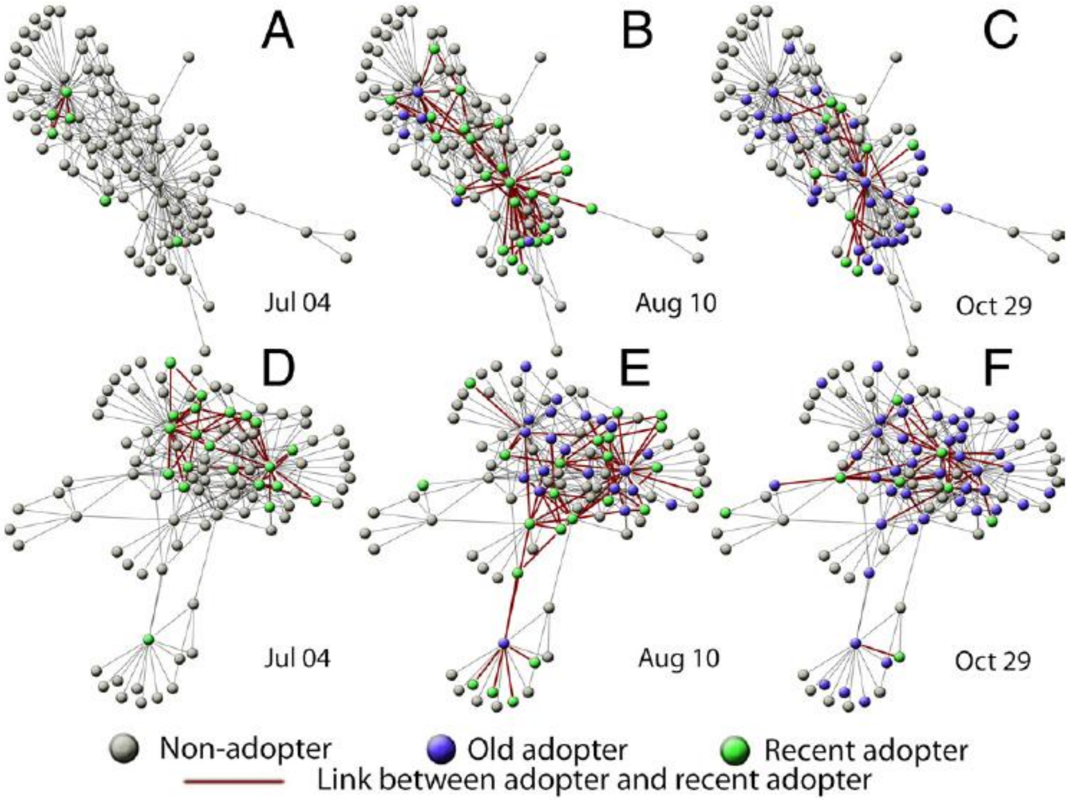


Fig. 1. Diffusion of Yahoo! Go over time. (*A–C* and *D–F*) Two subgraphs of the Yahoo! IM network colored by adoption states on July 4 (the Go launch date), August 10, and October 29, 2007. For animations of the diffusion of Yahoo! Go over time see Movies S1 and S2.

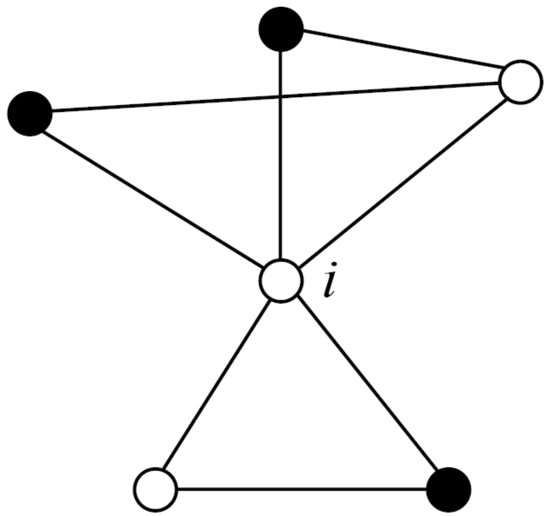
Example: the SIS process

At time t, each node is

- susceptible (S) (= it is wealthy but can potentially be infected), or
- infected (I) (= it is infected and capable of transmitting the infection)

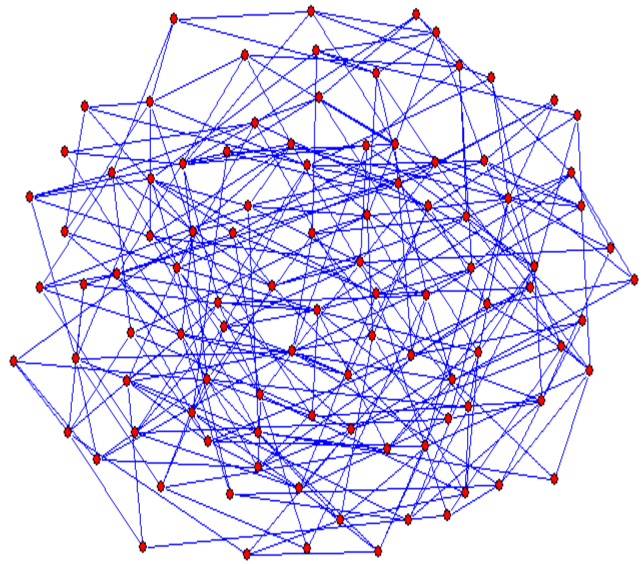
LOCAL RULES:

- infection: a node i in state S becomes I with probability βI_i, i.e. proportional to the number I_i of infected neighbors
- recovering: a node in state I returns S with probability γ



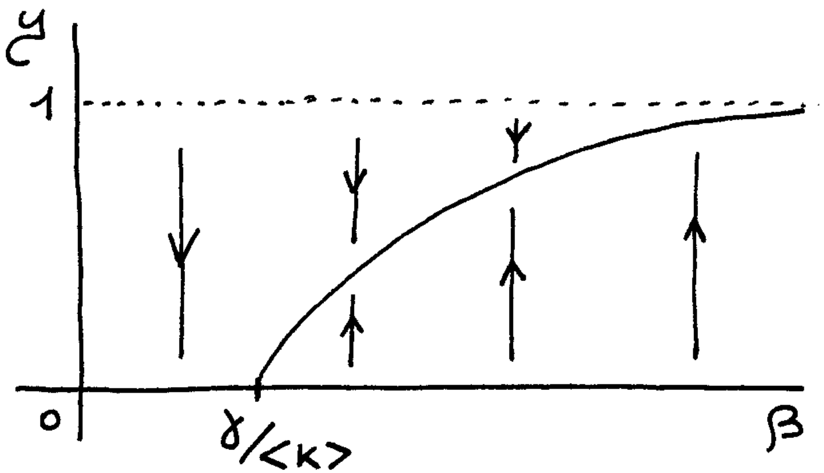
What is the global behavior of the epidemics?

In a homogeneous (or almost homogeneous) network:



- if $\beta \le \dfrac{\gamma}{<k>}$ $\Longrightarrow$ the fraction $y(t)$ of infected tends to 0 (=the epidemics dies out)

- if $\beta > \dfrac{\gamma}{<k>}$ $\Longrightarrow$ the fraction y of infected increases with the transmission rate β



This result is consistent with the classical epidemiology (Kermack and McKendrick, 1927):

No epidemics can survive if the transmission rate is below the epidemic threshold.

Some technical details...

Let $Y(t) \in [0, N]$ *be the number of I and* $y(t) = Y(t)/N \in [0,1]$ *their density (prevalence).*

$$Y(t+1) = Y(t) - \gamma \Delta Y(t) + \beta \Delta \Theta(t)(N - Y(t))$$

where $\Theta(t)$ *is the estimate of the average number of I among the neighbors of any S.*

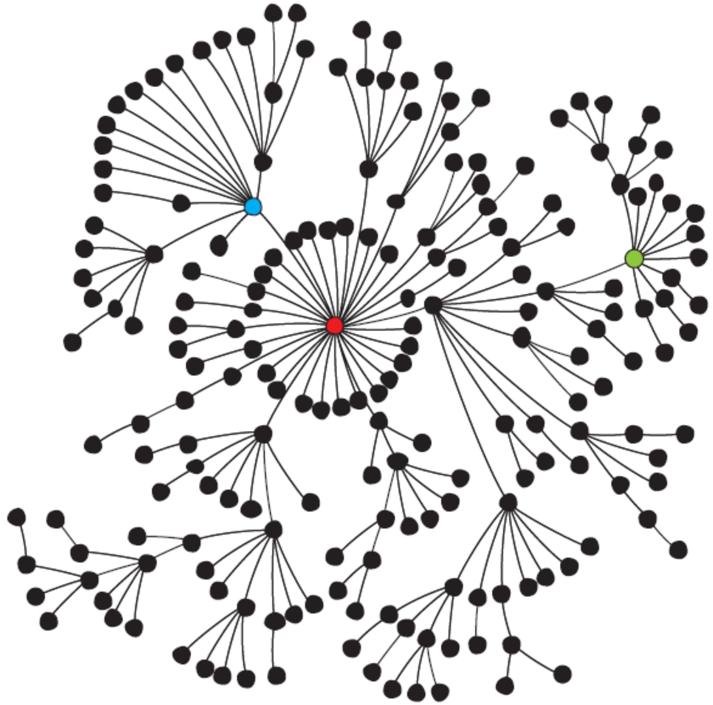
Assuming $\Theta = <k> y(t)$ *(average n. of neighbors* $\times$ *prob. that a neighbor is I) we obtain (for* $\Delta \to 0$*) the classical SIS model:*

$$\dot{y}(t) = -\gamma y(t) + \beta <k> y(t)(1 - y(t))$$

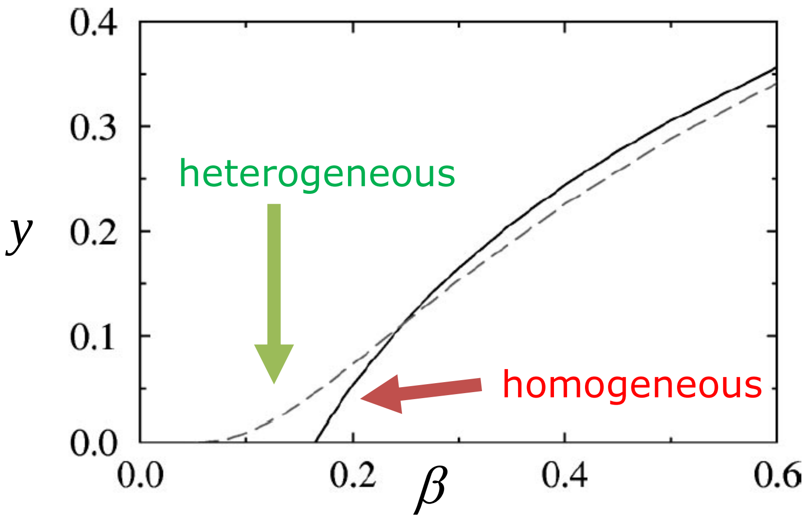
The non-trivial (>0*), asymptotically stable equilibrium* $y = 1 - \gamma/(\beta <k>)$ *exists iff* $\beta > \gamma / <k>$*.*

$\beta_c = \gamma / <k>$ *is the epidemic threshold.*

In a heterogeneous (e.g., scale-free) network (Pastor-Satorras and Vespignani, 2001):



- the epidemic threshold is $\beta_c = \gamma < k > / < k^2 >$, then it may tend to 0 for large networks ($N \to \infty$)
- $\Longrightarrow$ $y(t)$ never vanishes, whatever the value of the transmission rate β
- the nodes with larger degree are rare but have a large probability of being infected



The epidemics is able to survive with arbitrarily small transmission rate β (but with vanishing prevalence y).

Some technical details...

How can we model the epidemic dynamics when the network is strongly inhomogeneous?

We must model y *separately for each ensemble of nodes having the same degree* k*:*

$$\dot{y}_k(t) = -\gamma y_k(t) + \beta \Theta_k(t)(1 - y_k(t)) \quad , \quad k = k_{\min}, \ldots, k_{\max}$$

$\Theta_k(t) =$*(n. of neighbors* $\times$ *prob. that a neighbor is I)*$= k\tilde{y}(t) = k(\Sigma_h h P(h) y_h(t)) / \langle k \rangle$.

At the equilibrium ($\dot{y}_k = 0$ *) we obtain*

$$y_k = \frac{\beta k \tilde{y}}{\gamma + \beta k \tilde{y}} = \frac{1}{1 + \gamma / (\beta k \tilde{y})}$$

Thus the prevalence y_k *grows with* k *and tends to 1 as* $k \to \infty$ *(=nodes with a very large number of connections are rare but, most likely, they are infected).*

The (global) prevalence is given by

$$y(t) = \Sigma_k P(k) y_k(t)$$

Fighting the disease: uniform immunization

Different ways to fighting diseases

- use of drugs (antivirals) $\gamma \rightarrow \gamma \cdot \rho \qquad \rho > 1$
- immunization (vaccines) $\beta \rightarrow \beta \cdot (1-g)$

  where g is the fraction of immunized nodes

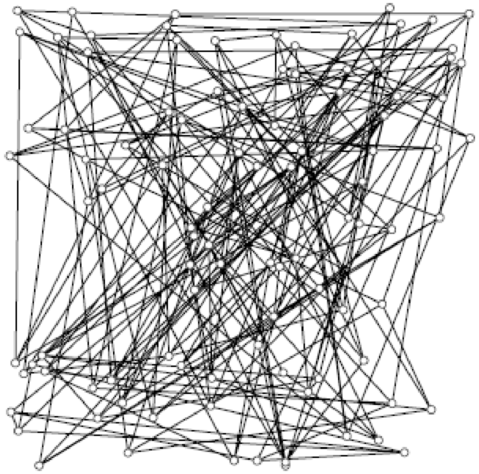
Homogeneous networks

$$\dot{y}(t) = -\gamma y(t) + \beta(1-g)\langle k\rangle y(t)(1-y(t))$$

An immunization threshold does exist

$$g_c \rightarrow 1 - \frac{\gamma}{\beta\langle k\rangle}$$

Homogeneous networks can be completely protected



NOTE: WS too

Fighting the disease: uniform immunization

Different ways to fighting diseases

- use of drugs (antivirals) $\gamma \to \gamma \cdot \rho \qquad \rho > 1$
- immunization (vaccines) $\beta \to \beta \cdot (1 - g)$

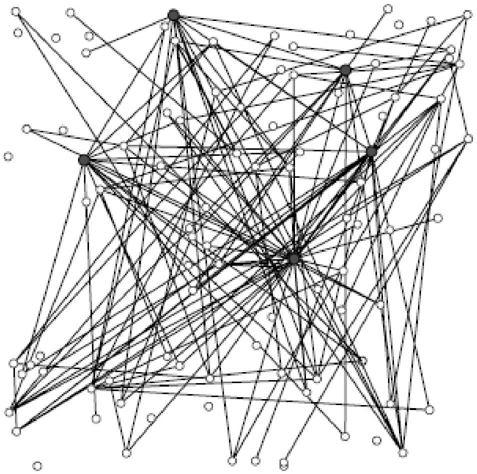
  where g is the fraction of immunized nodes

Heterogeneous networks

$$\dot{y}_k(t) = -\gamma y_k(t) + \beta(1-g)\Theta_k(t)\big(1 - y_k(t)\big) \qquad k = k_{\min}, \ldots, k_{\max}$$

The immunization threshold becomes

$$g_c \to 1 - \frac{\gamma}{\beta} \frac{\langle k \rangle}{\langle k^2 \rangle}$$



Only complete immunization of scale-free networks ($N \to \infty$) ensures disease eradication

Fighting the disease: targeted immunization

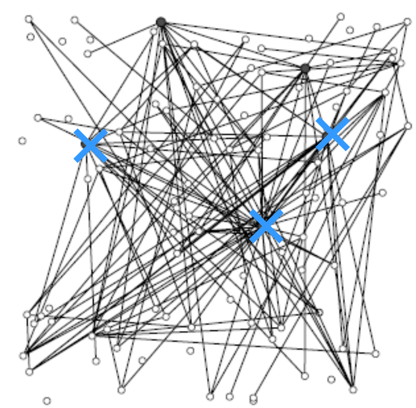
The weakness of highly heterogeneous networks (low resilience to targeted attacks) can become a defensive strategy

Immunize a fraction g of nodes starting from those with highest degree

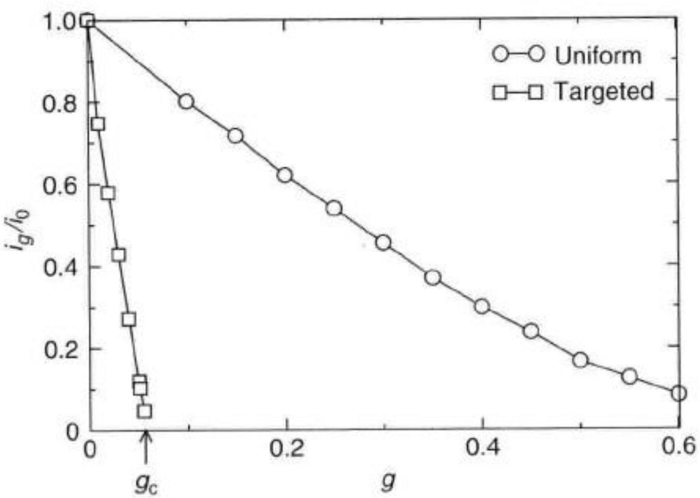
- cut-off $k_c(g)$ for the degree distribution
- removal $r(g)$ of links between immunized and others

$\Rightarrow$ new $P_g(k)$

$$\frac{\langle k \rangle_g}{\langle k^2 \rangle_g} \geq \frac{\beta}{\gamma} \quad \Rightarrow \quad g_c\left(\frac{\beta}{\gamma}\right)$$



Scale-free network $P(k) \sim k^{-3}$



$$g_c \sim \exp\left(-\frac{2\gamma}{k_{\min}\beta}\right)$$

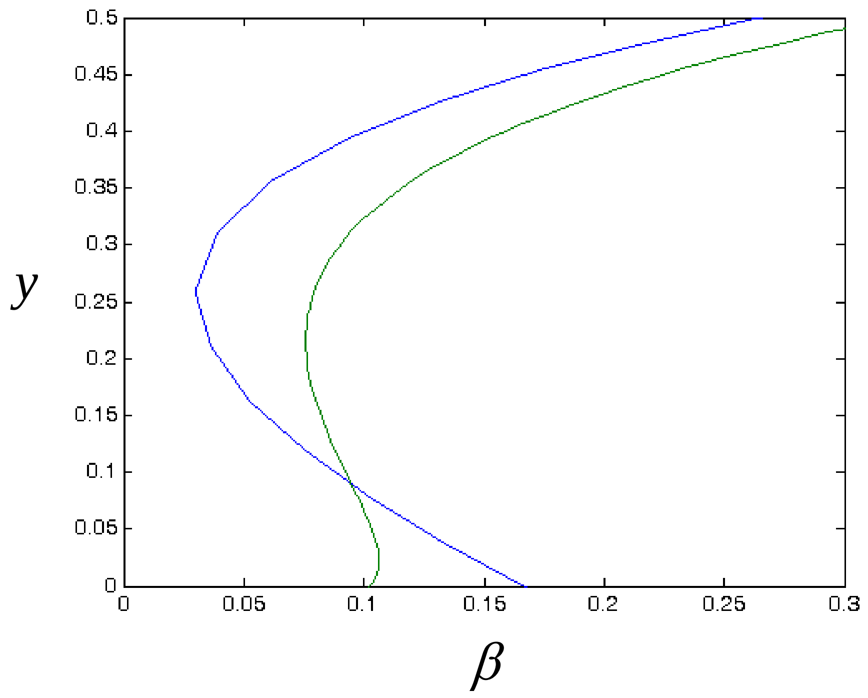
$k_{\min}$ is the minimum degree of the network

Barrat et al (2008), *Cambr Univ Press*

In modo analogo si possono studiare:

- altri tipi di epidemie (SIR, SIRC, virus informatici, …)
- strategie di vaccinazione (p.e. omogenee vs disomogenee)
- propagazione di informazioni, opinioni, prodotti ("word-of-mouth")
- epidemie con densità non infinitesima alla soglia di sopravvivenza:

Per y infinitesimo, la rete scale-free (curva verde) è più efficiente della rete omogenea (curva blu) nel propagare l'epidemia (così come nel SIS), ma per y elevato avviene il contrario.



y

SUGGESTED READINGS (and sources of most figures):

Books:

- S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press (2003)
- Barrat, M. Barthélemy, A. Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press (2008)
- M.E.J. Newman, *Networks: an Introduction*, Oxford University Press (2010)

Survey papers:

- S.H. Strogatz, *Exploring complex networks*, Nature 410 (2001) 268-276
- X.F. Wang, G. Chen, *Complex networks: Small-world, scale-free and beyond,* IEEE Circuits and Systems Magazine (2003) 6-20
- S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, *Complex networks: Structure and dynamics*, Physics Reports 424 (2006) 175–308
- M.D. Konig, S. Battiston, *From graph theory to models of economic networks. A tutorial*, in A.K. Naimzada et al. (eds.), *Networks, Topology and Dynamics*, Springer-Verlag (2009)

Other sources of figures:

Carvalho et al., *Robustness of trans-European gas networks*, Physical Review E 80 (2009) 016106
Liljeros et al., *The web of human sexual contacts*, Nature 411 (2001) 907-908
Ugander et al., *The Anatomy of the Facebook Social Graph*, ArXiv 1111.4503, 2011.