

THE ROLE OF CONNECTIVITY PATTERNS IN COMPUTER VIRUS SPREADING

Alessandro Vespignani (ICTP-Trieste)

Romualdo Pastor-Satorras (Tech. Univ. Barcelona)

Recently :

Yamir Moreno Vega (ICTP)

Alexei Vazquez (SISSA-ICTP)

DIMACS

Two levels

•**Microscopic level**

Researchers who disassemble and try to kill off new viruses.

Corresponds to the quest for new vaccines and medicines

•**Macroscopic level**

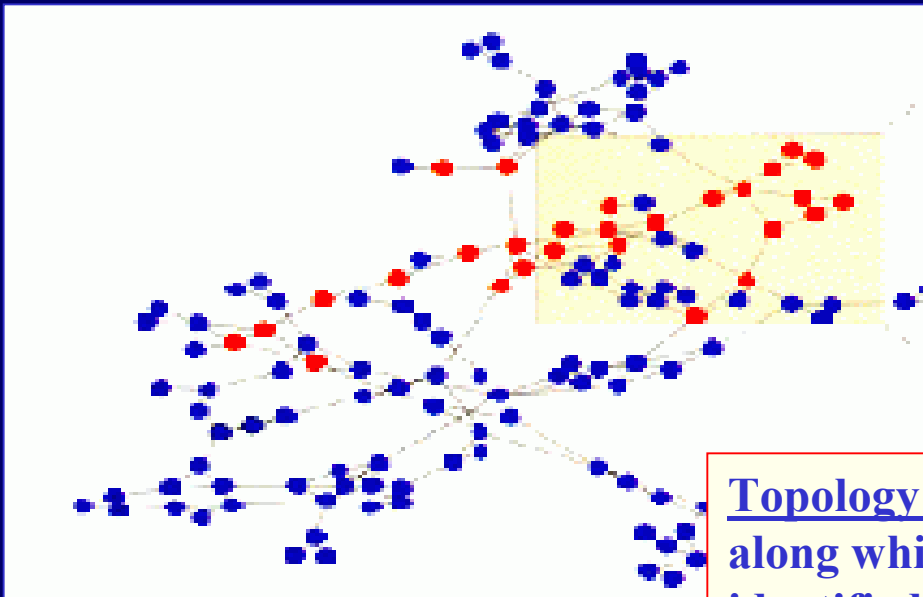
Statistical analysis and modeling of epidemiological data in order to find informations and policies aimed at lowering epidemic outbreaks

Macroscopic prophylaxis , Vaccination campaigns

Mathematical models of epidemics

Coarse grained description of individuals and their state

- Individuals exist only in few states:
- **Healthy or Susceptible * Infected * Immune * Dead**
- Particulars on the infection mechanism on each individual are neglected.



Topology of the system: the pattern of contacts along which infections spread in population is identified by a network

- Each node represents an individual
- Each link is a connection along which the virus can spread

The Susceptible-Infected-Susceptible (SIS) model

- Each node is infected with rate ν if connected to one or more infected nodes
- Infected nodes are recovered (cured) with rate δ without loss of generality $\delta=1$ (sets the time scale)
- Definition of an effective spreading rate $\lambda=\nu/\delta$

re-infection is possible.

Dynamical Mean-Field equation for the order parameter ρ = density of infected nodes

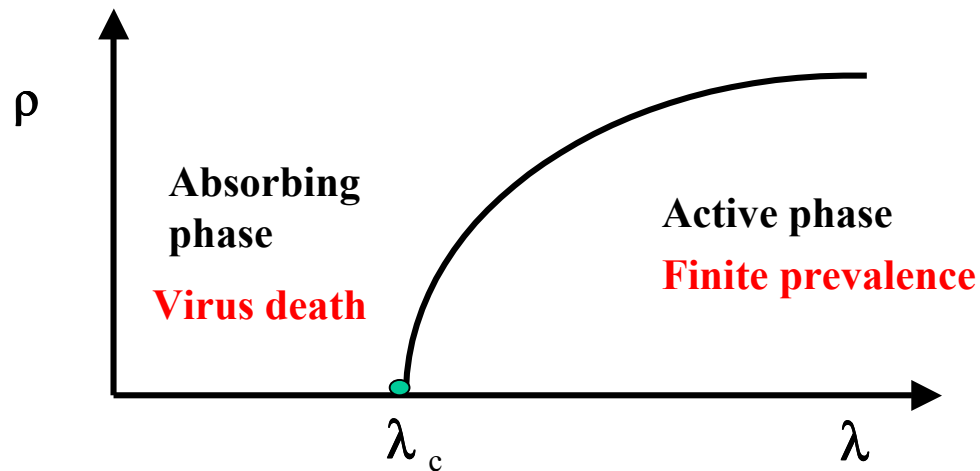
$$\partial_t \rho(t) = -\rho(t) + \lambda \langle k \rangle \rho(t) [1 - \rho(t)] + h.o. .$$

In the stationary state $\partial_t \rho = 0$, we have

$$\rho[-1 + \lambda \langle k \rangle (1 - \rho)] = 0$$

Definition of the epidemic threshold $\lambda_c = \langle k \rangle^{-1}$

$$\begin{aligned} \rho &= 0 && \text{if } \lambda < \lambda_c, \\ \rho &\sim \lambda - \lambda_c && \text{if } \lambda > \lambda_c, \end{aligned}$$



- **Non-equilibrium phase transition**
- **SIS** model is a variation of the contact process
- **epidemic threshold = critical point**
- **prevalence ρ = order parameter**

Similar models with immunity and death (removal) can be defined (SIR etc)

The epidemic threshold is a general result

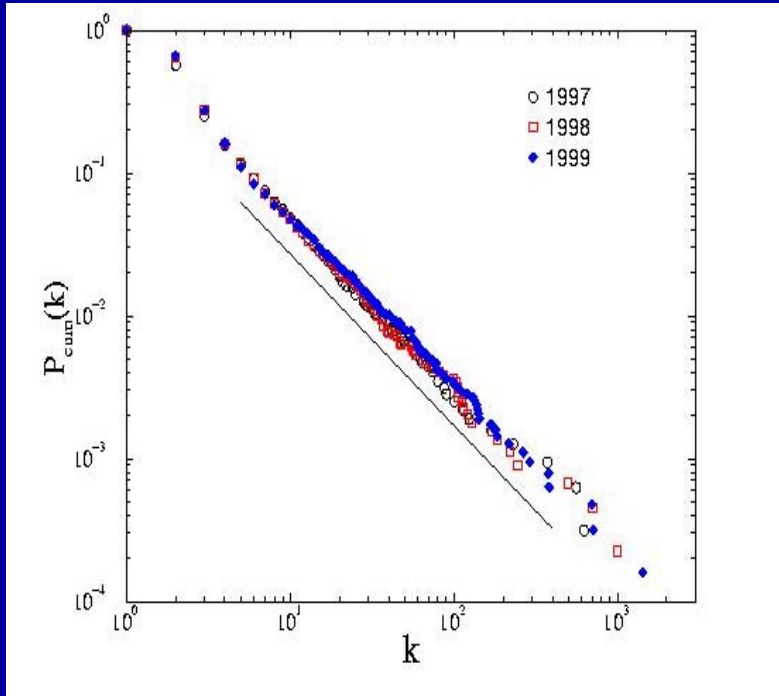


The question of thresholds in epidemics is central

Scale-free networks

Main properties

- complex network
- preferential attachment
- local clustering



Connectivity distribution

- $P(k)$ = probability that a node has k links

$$P(k) \sim k^{-\gamma}$$

$$(2 < \gamma \leq 3)$$

- $\langle k \rangle = \text{const}$
- $\langle k^2 \rangle \rightarrow \infty$

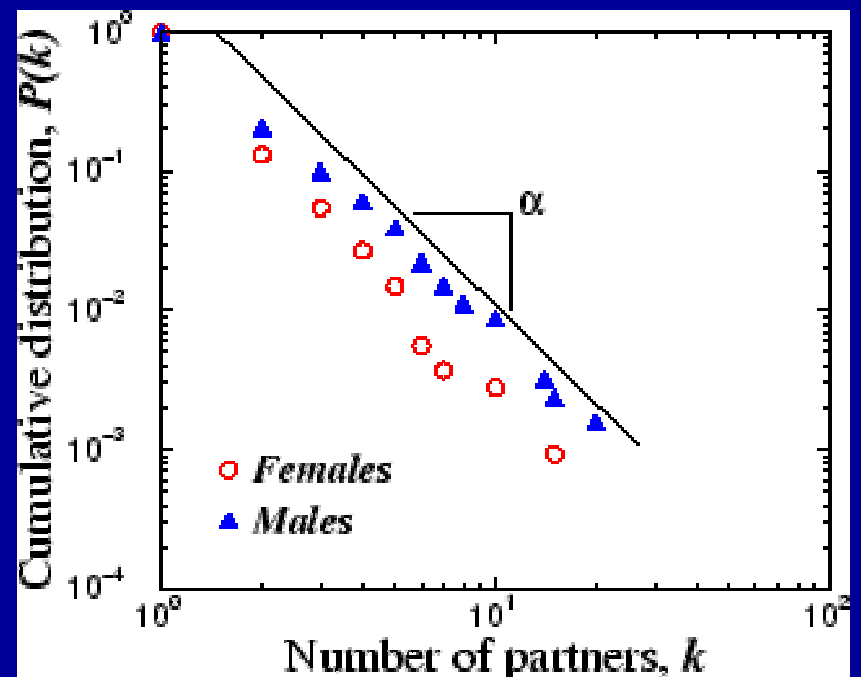
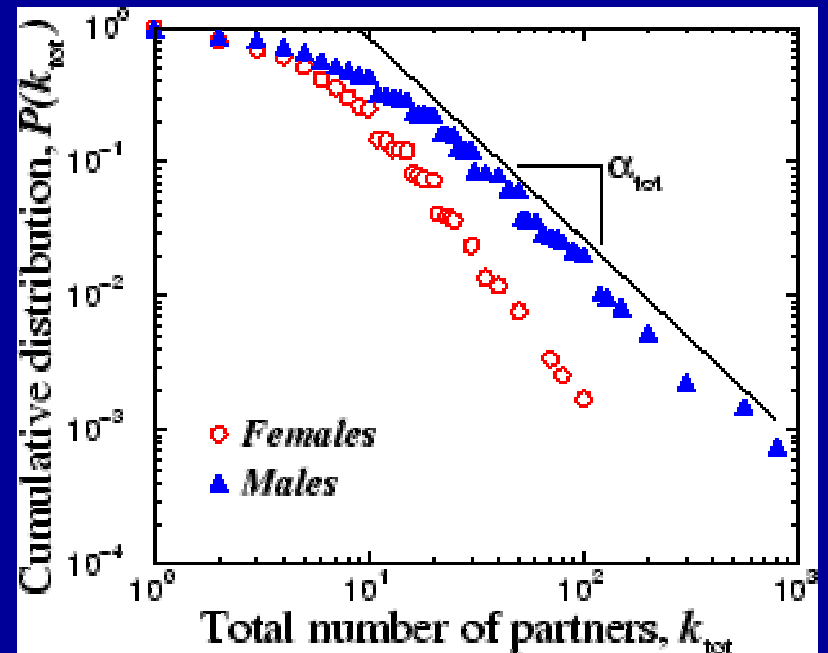
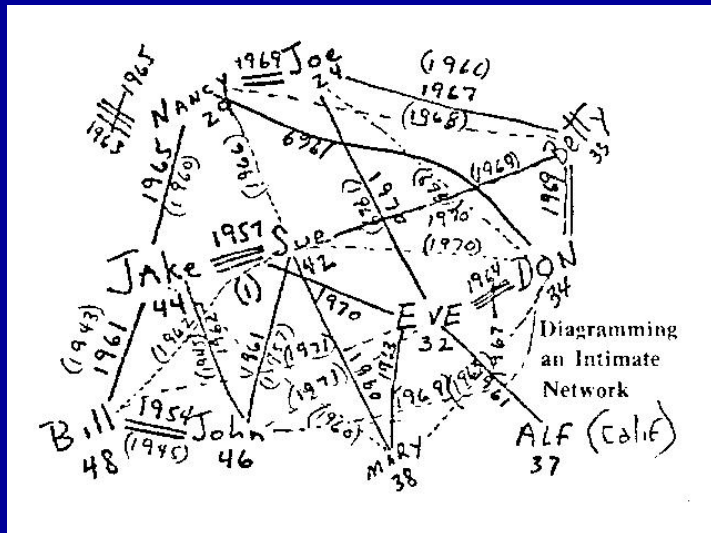
Scale-free properties



Diverging fluctuations

The web of Human sexual contacts

[Liljeros et al., Nature (2001)]



Natural computer virus

- DNS-cache computer viruses
- Routing tables corruption

Data carried viruses

- ftp, file exchange, etc.

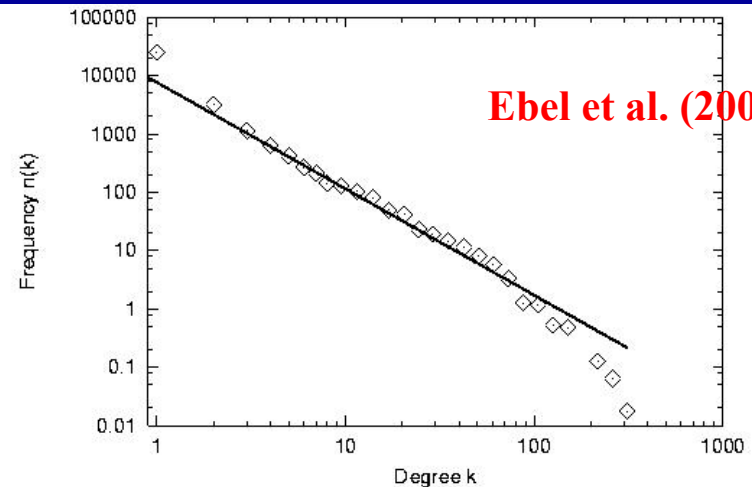
→ Internet topology
(scale-free)

Computer worms

- e-mail diffusing
- self-replicating



E-mail network



How to generate scale-free graph

Growth : at each time step a new node is added with m links to be connected with previous nodes

Preferential attachment: The probability that a new link is connected to a given node is proportional to the number of node's links.

by Barabasi & Albert
(1999)

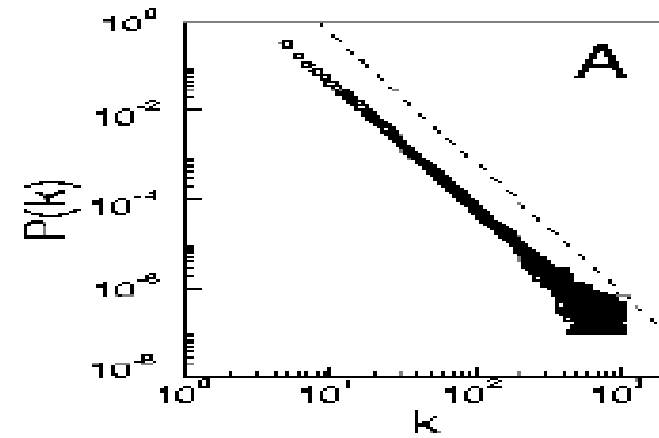
The BA model

The preferential attachment is following the probability distribution :

$$P(k_i) = \frac{k_i}{\sum_j k_j},$$

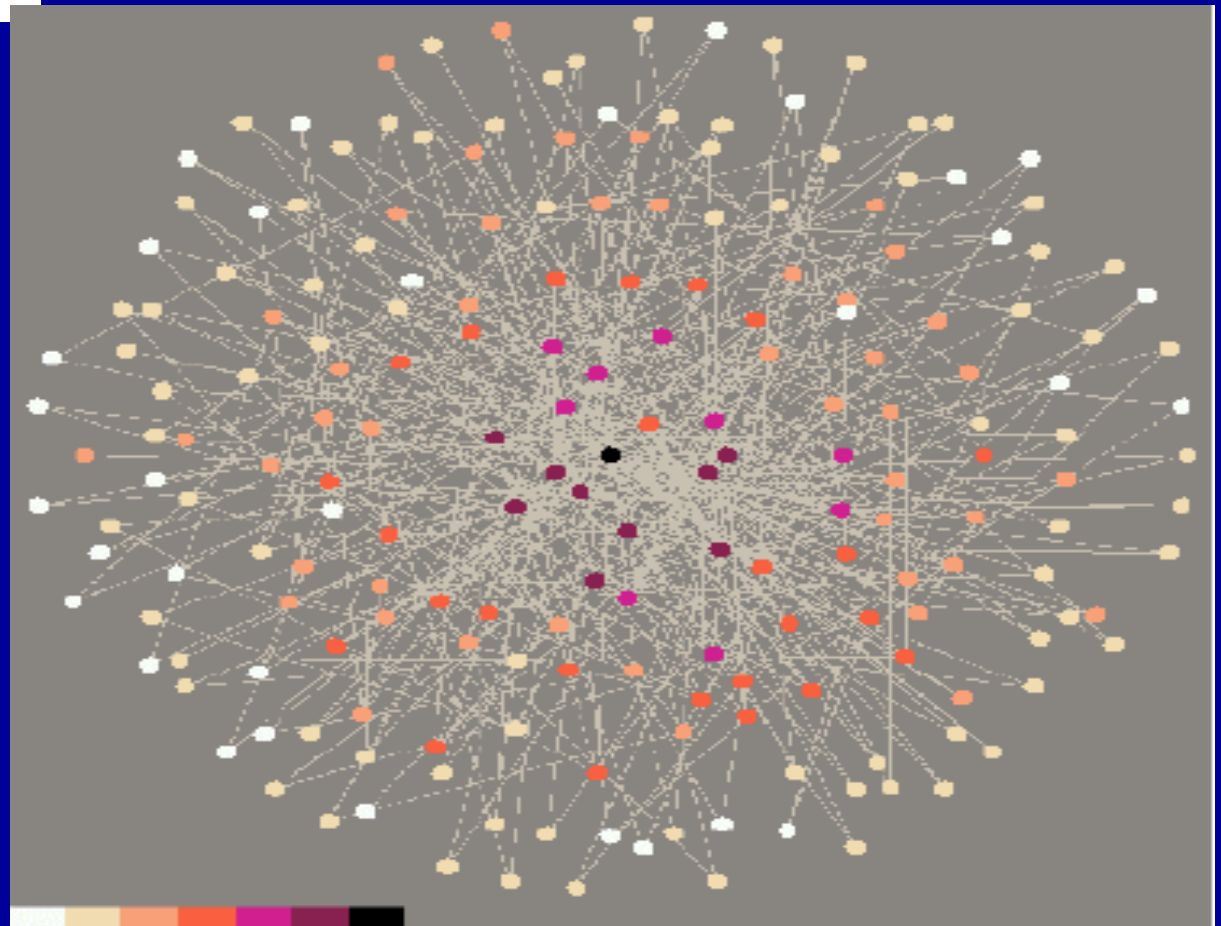
The generated connectivity distribution is

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



Connectivity distribution

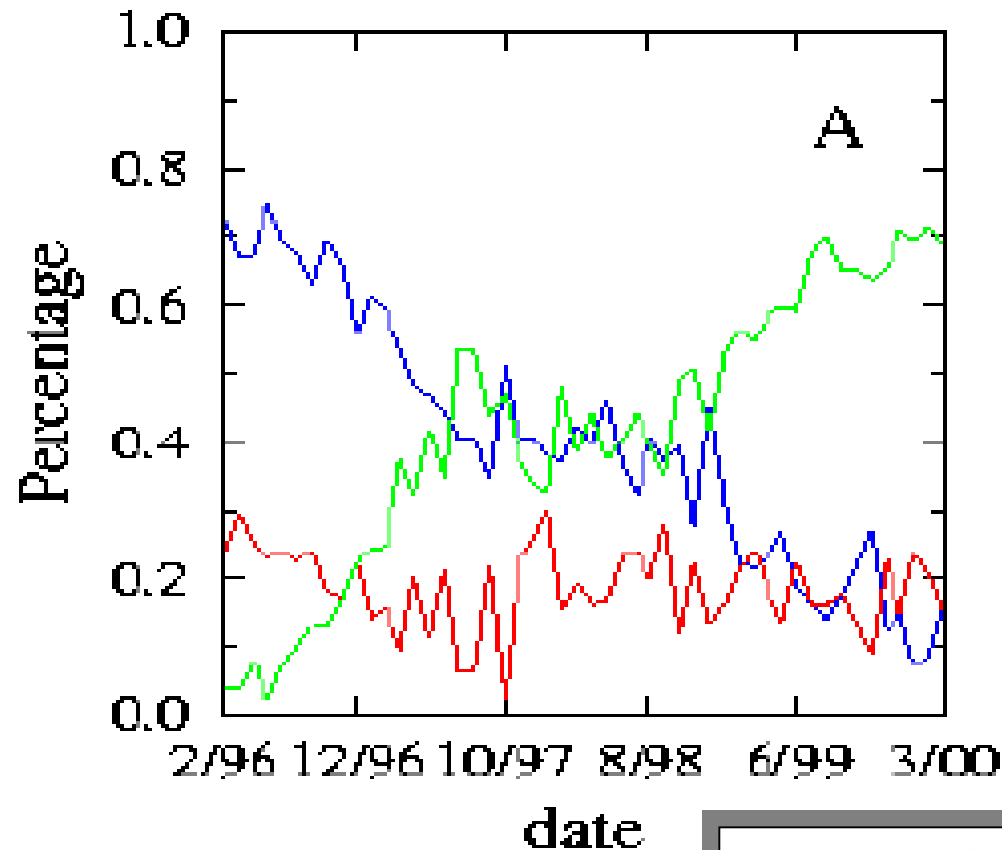
BA network



Strain data analysis

We analyzed homogeneous groups of viruses

- effective parameters
- similar properties within the strain

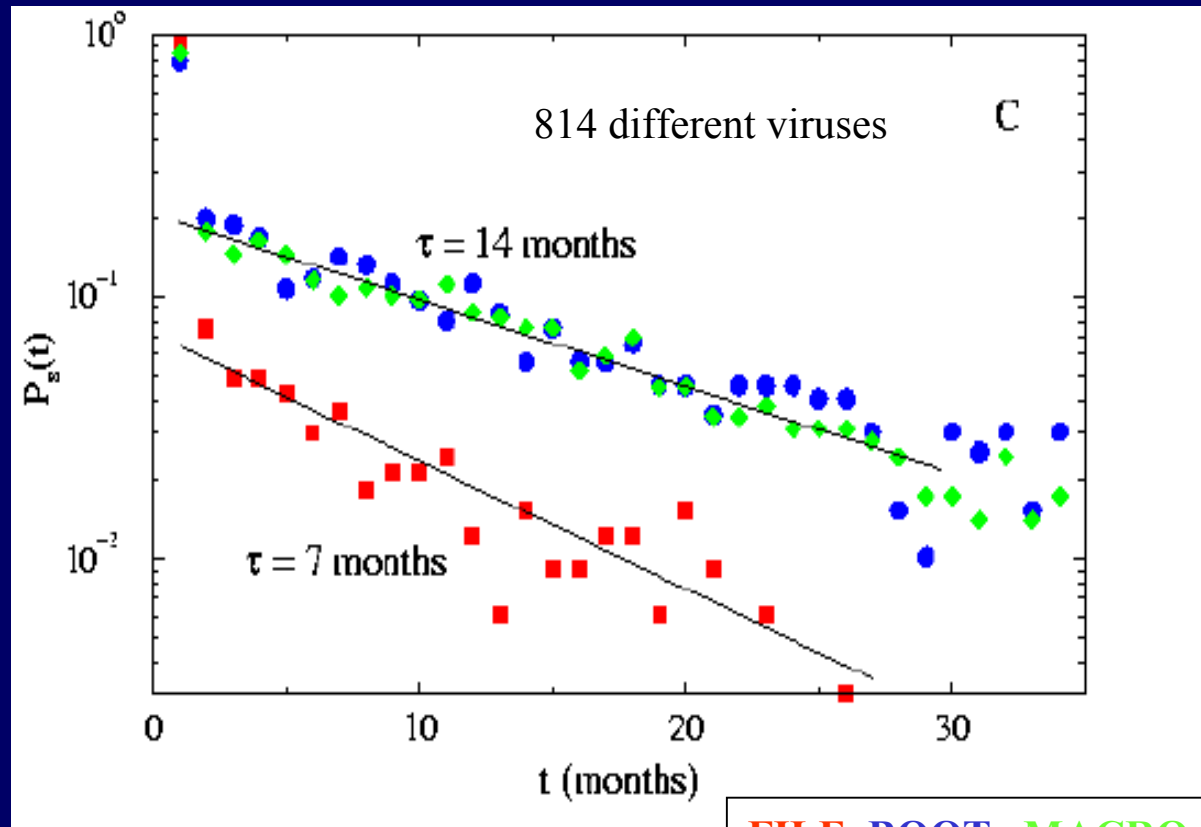


- In green MACRO viruses
- In red FILE viruses
- In blue BOOT viruses

• **Survival probability**
 $P_s(t)$ = fraction of
viruses still in the wild
at time t after their
birth

$$P_s(t) \sim \exp(-t/\tau)$$

τ = average lifetime
(characteristic time)
of the virus strain



FILE BOOT MACRO

**Average lifetime extremely long compared to the virus
rates time scale**

- Anti-virus software is delivered in a few hours after the first detection
- I Love You virus is still present in the wild list after two years

Epidemic spreading on Scale-Free networks

- Highly connected nodes are statistically significant $\langle k^2 \rangle = \infty$
- Connectivity fluctuations must be included

Relative density $\rho_k(t)$ of infected nodes with given connectivity k

$$\partial_t \rho_k(t) = -\rho_k(t) + \lambda k [1 - \rho_k(t)] \Theta(\rho(t)),$$

$\Theta(\rho(t))$ = Prob. that any given link points to an infected node.

Annihilation term

Creation term

- Θ is function of the average density of infected nodes
- Links point with higher probability to highly connected nodes

[Pastor Satorras & Vespignani,
PRL 86, 3200 (2001)]

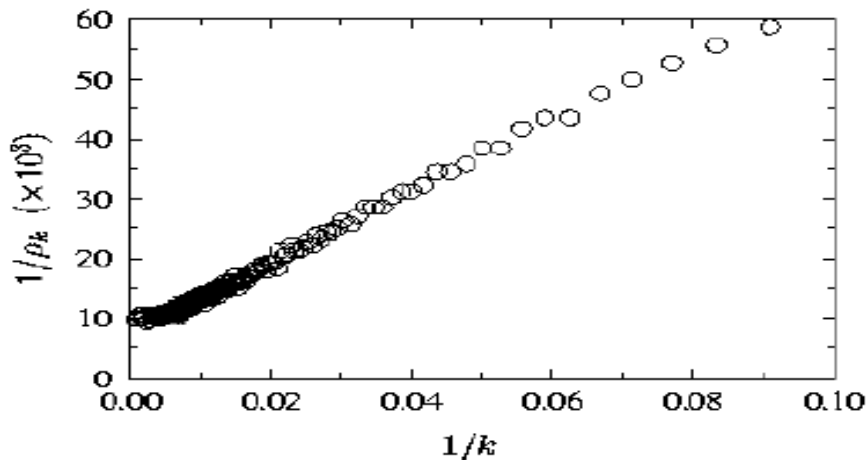
Stationary state

In the stationary state $\partial_t \rho_k(t) = 0$ we have that $\Theta(\rho) \Rightarrow \Theta(\lambda)$.

MF equations yield

$$\rho_k = \frac{k\lambda\Theta(\lambda)}{1 + k\lambda\Theta(\lambda)}.$$

Simulations in a BA network



- Higher is the node connectivity and higher is the probability to be in an infected state
- Strong inhomogeneity

- A link is more likely connected to a node with high connectivity

- The probability that a link points to a node with s links is proportional to $sP(s)$.

The probability of pointing to an infected node is



$$\Theta(\lambda) = \sum_k \frac{kP(k)\rho_k}{\sum_s sP(s)}$$

ρ_k themselves are a function of $\Theta(\lambda)$

Self-consistent equation

Finally the equation for the order parameter is

$$\rho = \sum_k P(k)\rho_k$$

In the case of the BA-model ($P(k) = 2mk^{-3}$), we consider k as a continuous variable and $\langle k \rangle = 2m$.

The first self-consistent equation is

$$\Theta(\lambda) = m\lambda\Theta(\lambda) \int_m^\infty \frac{1}{k^3} \frac{k^2}{1 + k\lambda\Theta(\lambda)}$$

which yields the solution

$$\Theta(\lambda) = \frac{e^{-1/m\lambda}}{\lambda m} (1 - e^{-1/m\lambda})^{-1}$$

The order parameter equation is

$$\rho = 2m^2\lambda\Theta(\lambda) \int_m^\infty \frac{1}{k^3} \frac{k}{1 + k\lambda\Theta(\lambda)},$$

Obtaining

$$\rho = 2e^{-1/m\lambda} + h.o.$$

- **Absence of any epidemic threshold (critical point)**

- **Active state for any value of λ**

- **The infection pervades the system whatever spreading rate**

- **In infinite systems the infection is infinitely persistent (indefinite stationary state)**



Epidemic threshold in scale-free networks

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

$$\langle k^2 \rangle \rightarrow \infty$$

$$\lambda_c \rightarrow 0$$

Order parameter
behavior in an
infinite systems

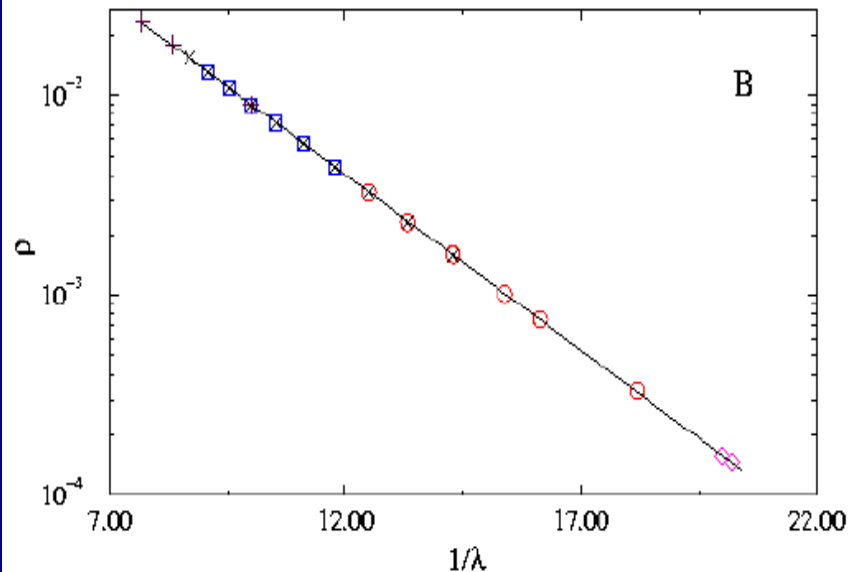
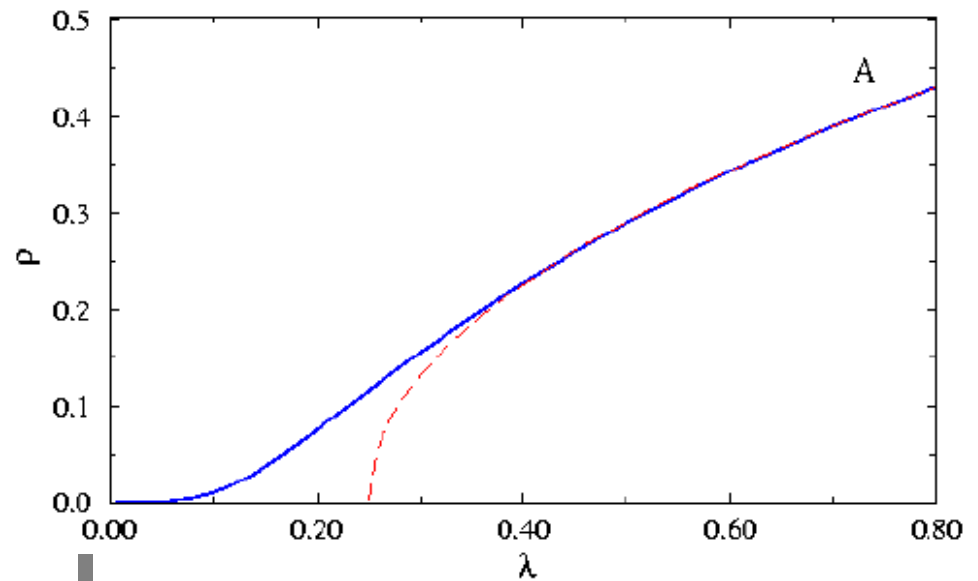
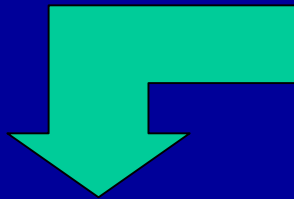


$$\rho = 2e^{-1/\pi\lambda}$$

$$\rho = 2e^{-1/\pi\lambda}$$

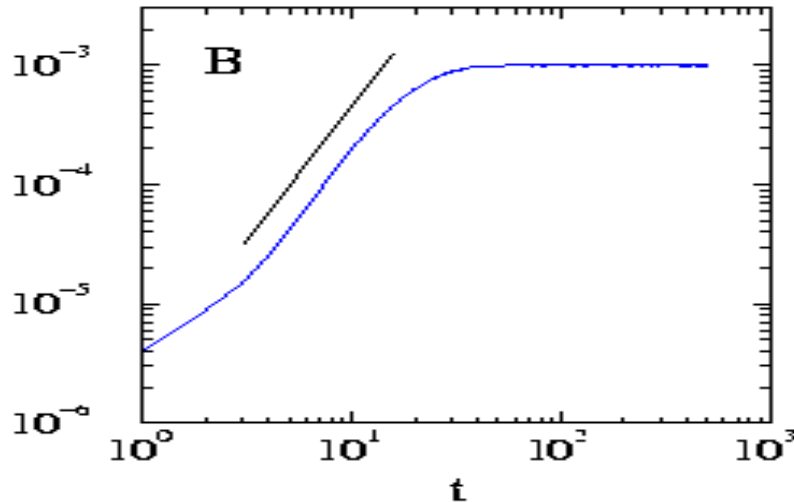
Numerical simulations in a BA network

Zoom in lin-log scale



Network sizes
 $N=10^3$ to $N=10^7$

Spreading of a virus starting from a localized seed

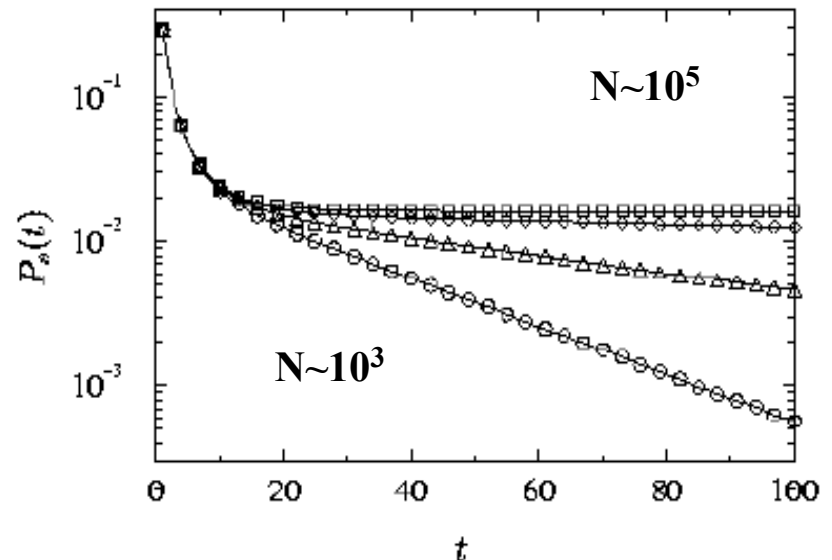


**Time behavior of
the prevalence
(algebraic)**

$$\lambda=0.06$$

**Surviving
probability with the
same definition
used to analyze
data.**

**Exponential behavior with
characteristic time increasing
with the network size**



Results can be generalized to generic connectivity distributions $P(k) \sim k^{-\gamma}$

- If $2 < \gamma \leq 3$ we have absence of an epidemic threshold and no critical behavior.
- If $3 < \gamma \leq 4$ an epidemic threshold appears, but it is approached with vanishing slope (no criticality).
- If $\gamma > 4$ the usual MF behavior is recovered. SF networks are equal to random graph.

Finite size scale-free networks

$$P(k) \sim k^{-\gamma} \exp(-k/k_c)$$

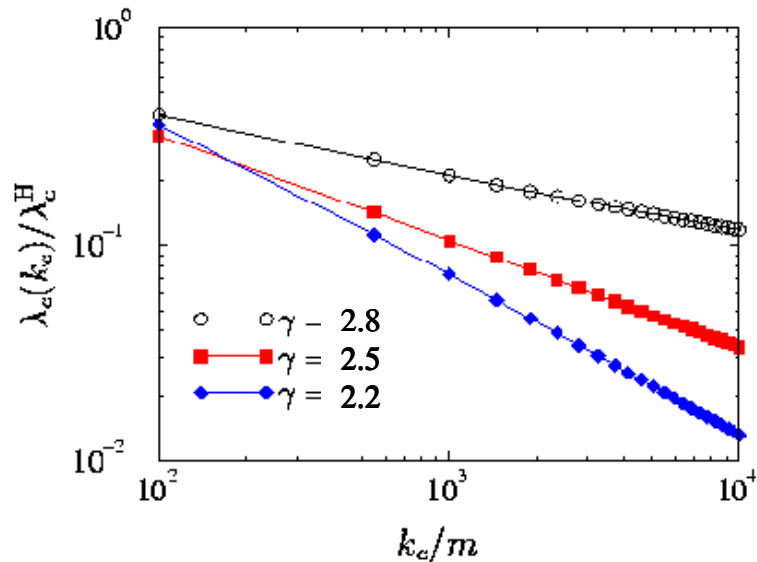
Exponentially bounded

$$P(k) \sim k^{-\gamma} \theta(k - k_c)$$

Hard cut-off



$$\lambda_c \sim k_c^{\gamma-3}$$



Ratio between SF and homogeneous
Epidemic threshold for $k_c, N=10^4$

$$\lambda_c / \lambda_c^H < 10^{-1}$$

MAIN RESULTS FOR S-F NETWORKS

- **Absence of an epidemic/immunization threshold**
- **The network is prone to infections
(endemic state always possible)**
- **Small prevalence for a wide range of spreading rates**
- **Lifetime is related to the network size**

Rationalization of computer virus data

NEXT STEPS



•SIR epidemic outbreaks (non-closed population)

May & Lloyd PRE 2001;

Moreno, Pastor-Satorras & Vespignani EPJB (2002);

Newman cond-mat (2002)

•Latency effects and population heterogeneity

•Modeling and simulations on real Internet maps

•Optimal immunization schemes for real maps



•Finite network effects

•Pollution of food-webs

References :

- **R.Pastor Satorras and A.Vespignani, PRL 86, 3200 (2001)**
- **R.Pastor Satorras and A.Vespignani, PRE 63, 066117 (2001)**
- **R.Pastor Satorras, A. Vazquez and A.Vespignani, PRL 87, 258701 (2001)**
- **R.Pastor Satorras and A.Vespignani, PRE 65, 036104 (2002)**
- **Y. Moreno, R.Pastor Satorras and A.Vespignani, EPJB 26, 521 (2001)**
- **A. Vazquez, R.Pastor Satorras and A.Vespignani, cond-mat/0112400 (2001)**