Modeling of Complex Networks
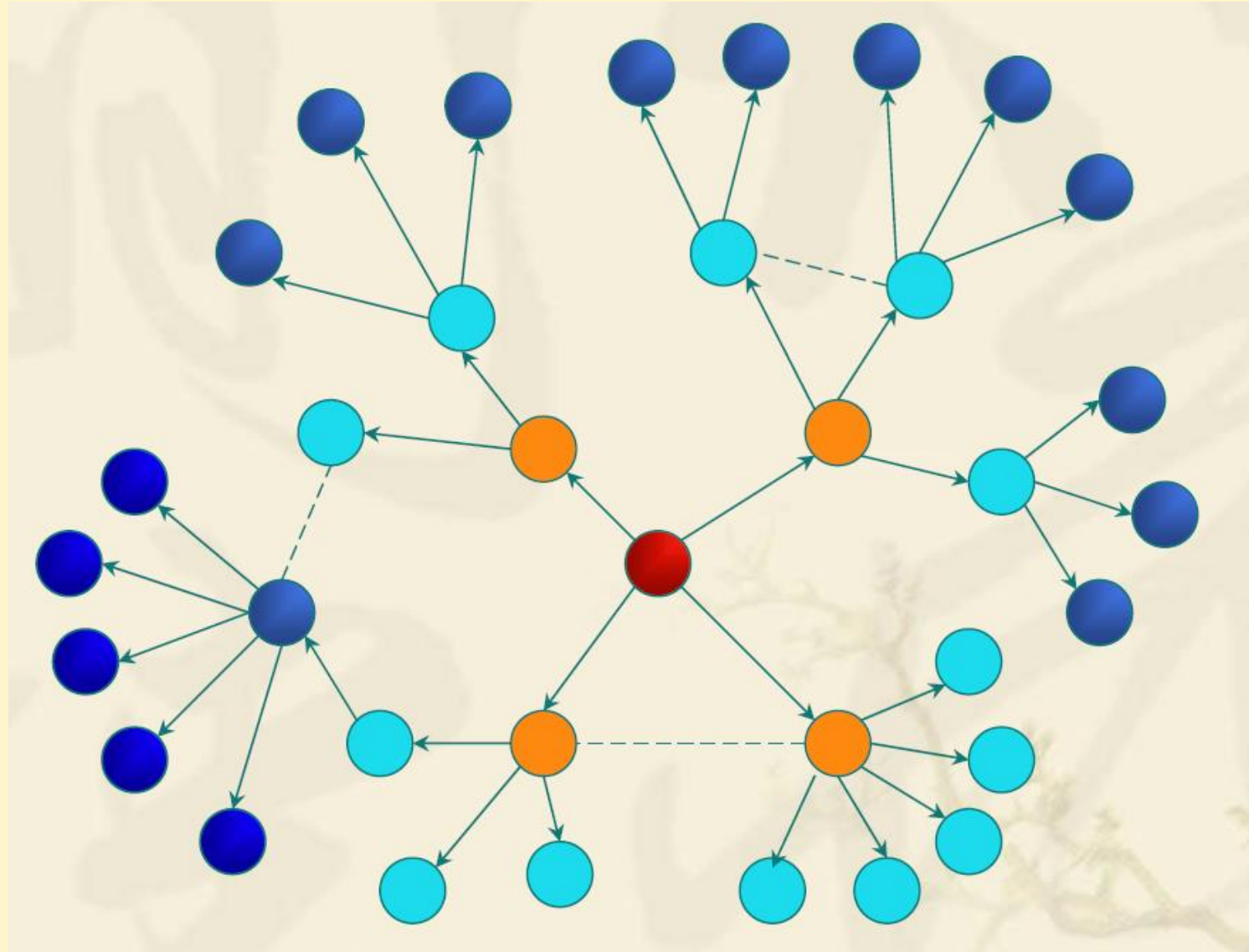
Lecture 5: Spreading Dynamics

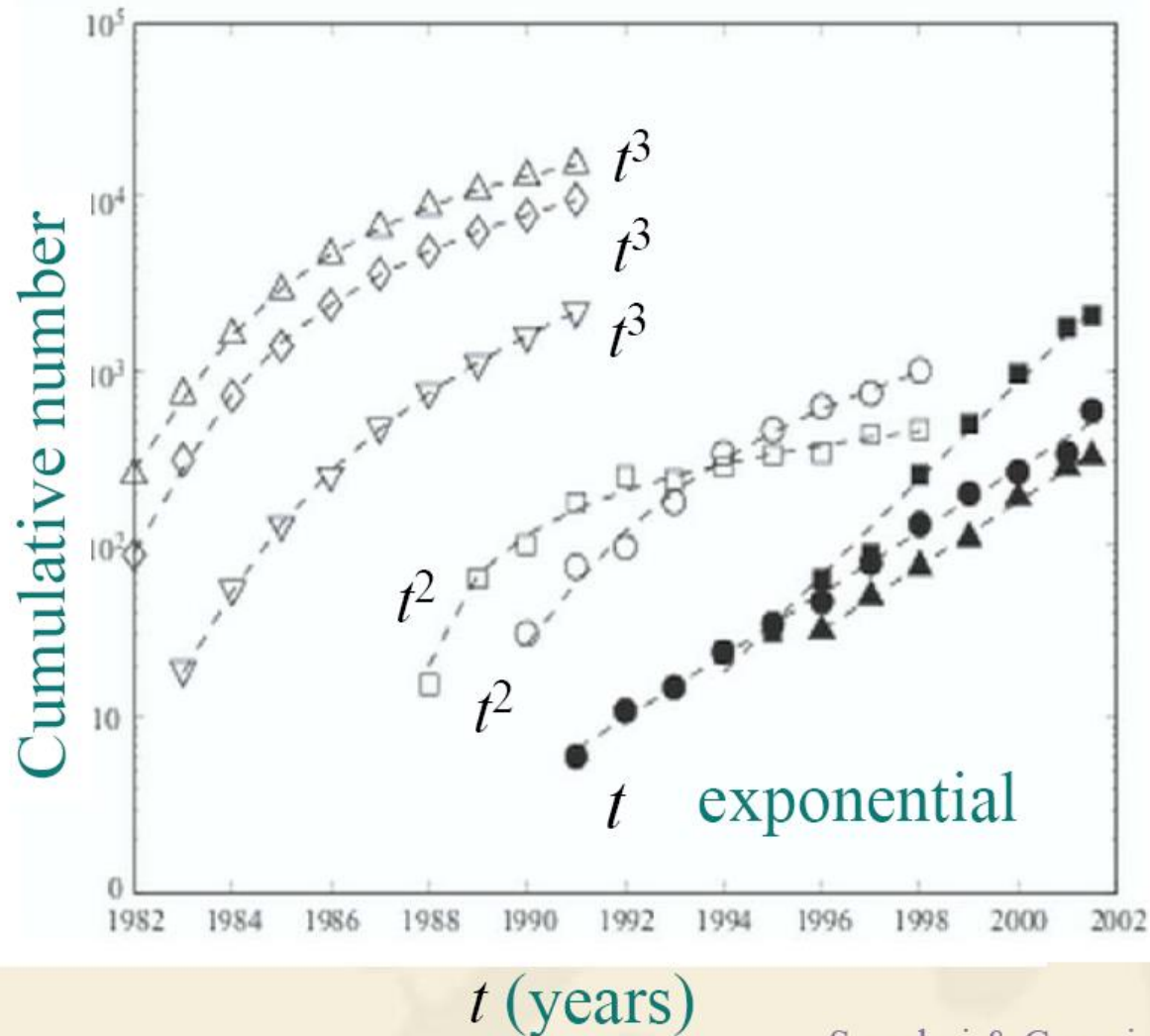### S8101003Q-01 (Sem A, Fall 2023)

Epidemic Spreading

Biological Network Example: AIDS Epidemics
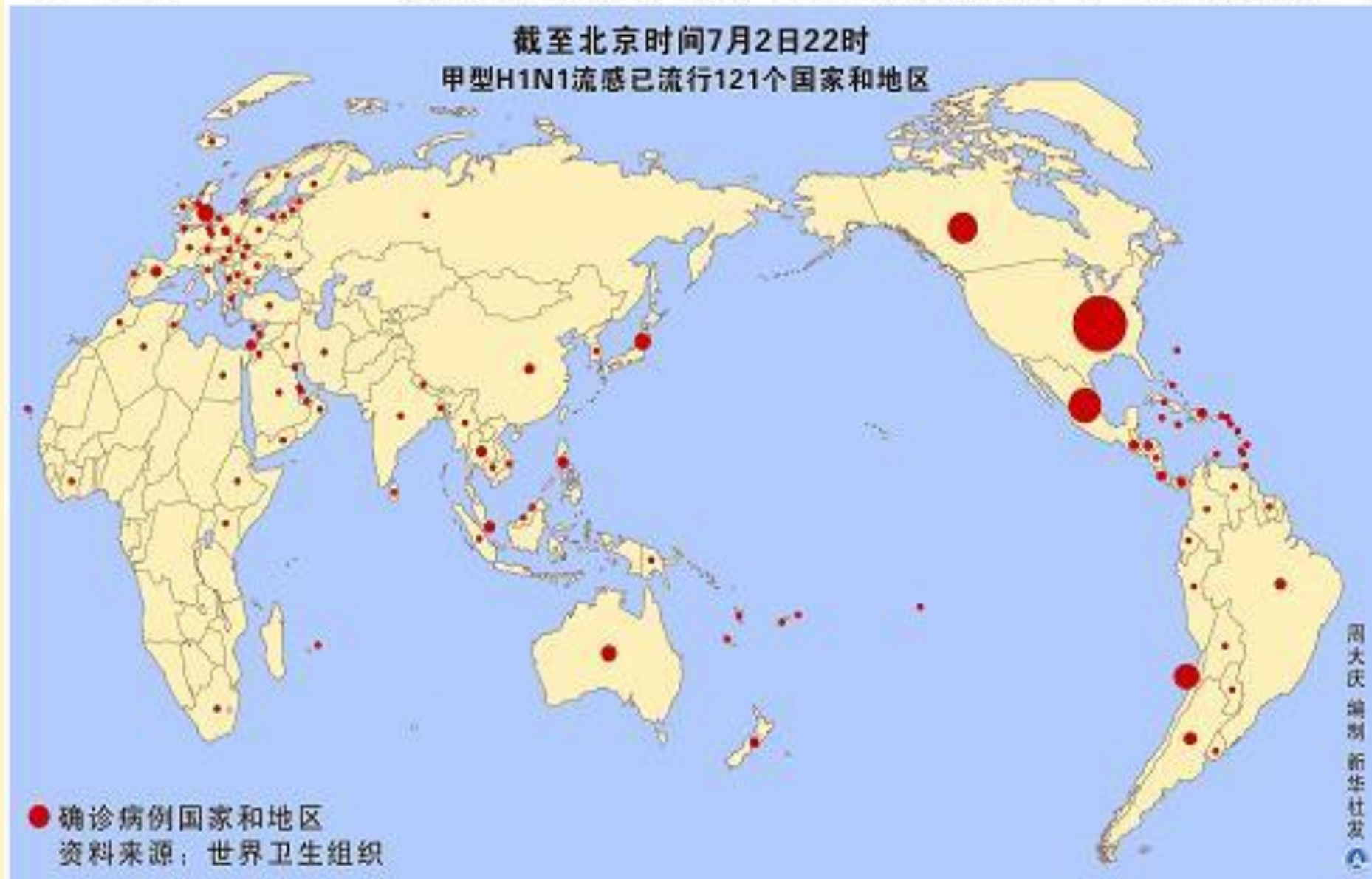

Szendroi & Czanyi, Proc. Royal Society of London, B 2004
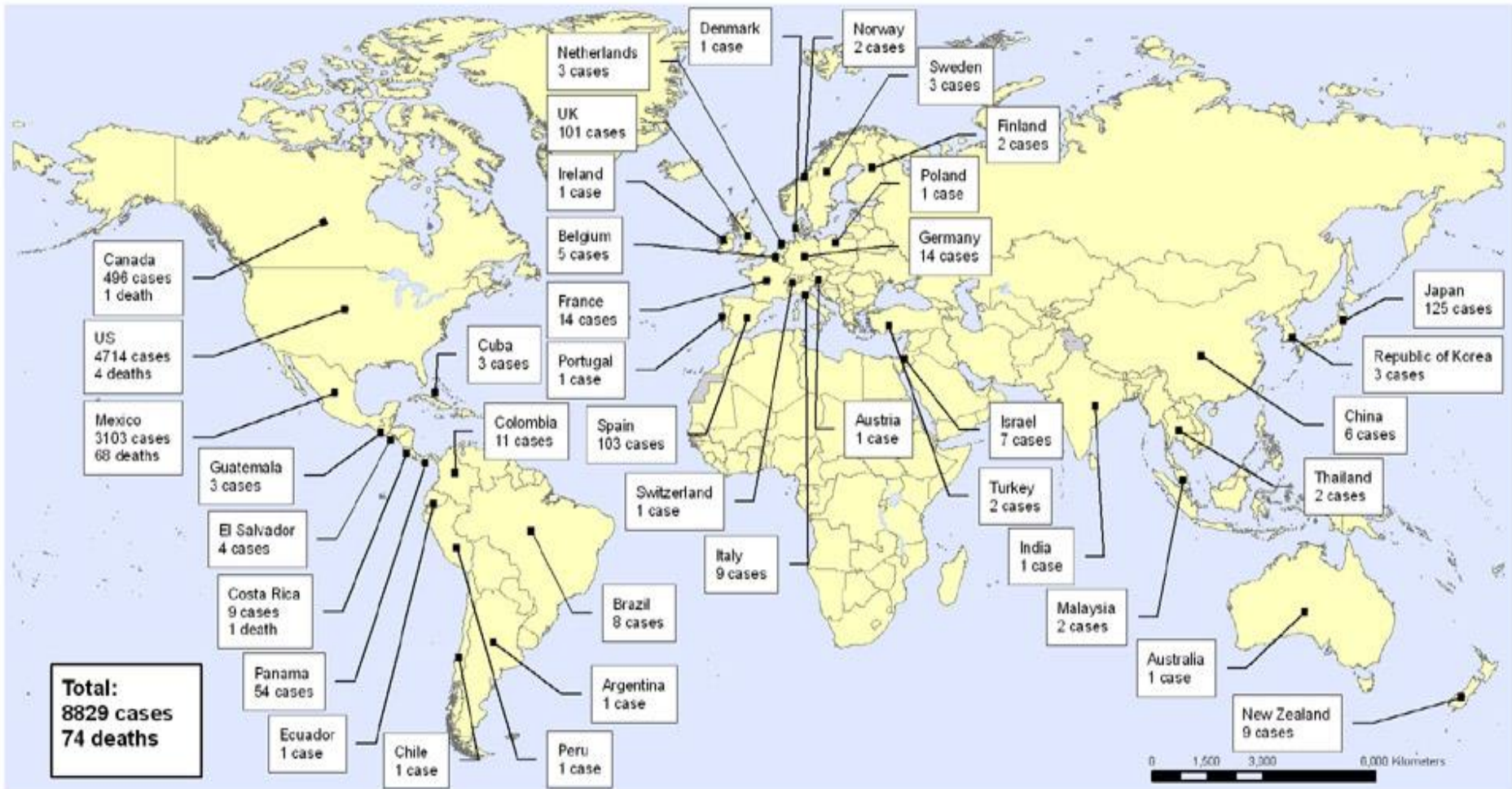
甲型 H1N1 流感疫情全球流行示意图

截至北京时间7月2日22时
甲型H1N1流感已流行121个国家和地区

● 确诊病例国家和地区
资料来源：世界卫生组织

New Influenza A (H1N1),
Number of laboratory confirmed cases and deaths as reported to WHO

Status as of 18 May 2009
6:00 GMT

Denmark 1 case
Netherlands 3 cases
Norway 2 cases
Sweden 3 cases
UK 101 cases
Finland 2 cases
Ireland 1 case
Poland 1 case
Belgium 5 cases
Germany 14 cases
France 14 cases
Japan 125 cases
Canada 496 cases 1 death
Republic of Korea 3 cases
US 4714 cases 4 deaths
Cuba 3 cases
Portugal 1 case
Austria 1 case
Israel 7 cases
China 6 cases
Mexico 3103 cases 68 deaths
Colombia 11 cases
Spain 103 cases
Turkey 2 cases
Thailand 2 cases
Guatemala 3 cases
Switzerland 1 case
India 1 case
El Salvador 4 cases
Italy 9 cases
Costa Rica 9 cases 1 death
Malaysia 2 cases
Brazil 8 cases
Australia 1 case
Panama 54 cases
Argentina 1 case
New Zealand 9 cases
Ecuador 1 case
Chile 1 case
Peru 1 case

Total:
8829 cases
74 deaths

The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

Map produced: 18 May 2009 6:10 GMT

Data Source: World Health Organization
Map Production: Public Health Information and Geographic Information Systems (GIS)
World Health Organization

World Health Organization

© WHO 2009. All rights reserved

9/19/2023

5

Computer Virus Spreading

- The first virus capable of infecting PCs was perhaps the Brain virus developed in Pakistan in 1986.

- I-Love-You bug began in Philippines on May 4, 2000, and spread across the world in one day (traveling through Hong-Kong to Europe then to USA), infecting more than 78 million computers worldwide in just 4 days, i.e., about 10% of all computers connected to the Internet, and causing about US$5.5 billions in damage.

- It was estimated that there were more than 48,000 identified viruses on the Internet worldwide in year 2000 alone.

- In 2000-2001, well-known viruses: Lover Letter, Nimda, and Sircam …

- It was estimated that more than 80% computers were attacked by viruses in China in 2004.

- Also in 2004, the infamous Worm Sasser attacked several hundred thousands of computers over the world within just a couple of weeks.

- In 2009, Conficker (worm) remotely installs software on infected machines.

- ……Many other examples … still today! [VirusList]

Computer Viruses and Worms

- Computer viruses are usually referred to as some small computer programs that can reproduce themselves by infecting other programs and computers, which continuously grow and spread out.

- When a virus is active inside a computer, it is able to copy itself in many different ways into the codes of some programs of the computer.

- When the infected computer program is run into another computer, typically the code of the virus is executed first thus continues to infect other programs in the new computer.

- This process repeats endlessly, leading to the collapse of a local or even global network of computers eventually, causing tremendous technological and economical disasters.

Computer Viruses and Worms

Major computer viruses may be roughly classified into:

- ➤ **Boot-sector viruses,** which infect the boot sectors of floppies and hardware devices

- ➤ **File viruses,** which infect application programs

- ➤ **Macro viruses,** which infect data files directly

- ➤ **Hybrid types of viruses,** in a certain combination of the above basic ones, which infect some special technologies or applications such as Java, ActiveX, and HTML, etc.

Computer Viruses and Worms

- The first computer worm was created in 1988 by the now-infamous Robert Tappan Morris [book]

- Computer worms are most aggressive cyber-organisms (larger and more sophisticated programs) with much more powerful abilities to attack computers.

- They shut down the Internet e-mail systems that got clogged with infected e-mails propagating from the worm.

- A worm is capable of sending itself to all e-mail addresses in the e-mail address book of the computer, which received an infected e-mail. This makes worms very effective in spreading over the Internet.

An Example

<< See all Yahoo! Tech News

Koobface, Other Worms Target Facebook Friends (NewsFactor)

Posted on Thu Mar 5, 2009 11:31AM EST
Add articles about technology to your My Yahoo! [+] MY Y!

- As Facebook works to make itself more relevant and timely for its growing member base with a profile page makeover, attackers seem to be working overtime to steal the identities of the friends, fans and brands that connect though the social-networking site.

Indeed, Facebook has seen five different security threats in the past week. According to Trend Micro, four new hoax applications are attempting to trick members into divulging their usernames and passwords. And a new variant of the Koobface worm is running wild on the site, installing malware on the computers of victims who click on a link to a fake YouTube video.

The Koobface worm is dangerous. It can be dropped by other malware and downloaded unknowingly by a user when visiting malicious Web sites, Trend Micro reports. When attackers execute the malware, it searches for cookies created by online social networks. The latest variant is targeting Facebook, but earlier variants have also plagued MySpace.

RECENT TECH NEWS STORIES
Check out some fancy new kitchen gadgets
Delaware County Times Thursday 05th March, 02:32:55 AM

Along the border, the bulletproofing business is boomin ...
Austin American-Statesman Friday 06th March, 08:38:56 AM

Border violence leads to armor-plated vehicles
Park Hills Daily Journal Friday 06th March, 07:42:36 AM

Mathematical Modeling
of
Epidemic Spreading

Epidemic Models

Consider a population of N connected individuals, described by a network of N nodes.

Let the number of healthy nodes be $S(t)$, infected nodes be $I(t)$, and recovered nodes be $R(t)$ at time $t \geq 0$, initially $S(0) = S_0$, $I(0) = I_0$, and $R(0) = R_0 = 0$.

For convenience, one may normalize these variables by dividing them by N, so they also have the meaning of "density" (percentage), with $S(t) + I(t) + R(t) = 1$ for all $t \geq 0$.

Assume that the lifetime of a virus or disease is much shorter than the lifetime of the nodes in the population; therefore, birth and death of nodes are not taken into consideration in the following study.

Suppose that the probability of a node becoming "infected" from "susceptible" is v, and the probability of a node being cured and becomes "susceptible" again is δ. Then, define the virus *effective spreading rate* as

$$\lambda = \frac{v}{\delta} \tag{5-1}$$

One extreme case is that all cured nodes immediately become susceptible again, which corresponds to $\delta = 1$, thus $\lambda = v = $ probability of a node becomes infected.

SI Model

Suppose $R(t) = 0$ for all $t \geq 0$. Consider the rate of increase in the density of infected nodes, from time t to time $t + \Delta t$, which is given by

$$\frac{I(t + \Delta t) - I(t)}{\Delta t} \approx [\lambda S(t)] I(t)$$

In the limit $\Delta t \to 0$, one has $\dfrac{dI(t)}{dt} = \lambda S(t) I(t)$. Since $S(t) = 1 - I(t)$, this becomes the so-called Logistic model:

$$\frac{dI(t)}{dt} = \lambda [1 - I(t)] I(t), \qquad I(0) = I_0 \tag{5-2}$$

This equation has a solution

$$I(t) = \frac{1}{1 + \left(I_0^{-1} - 1\right) e^{-\lambda t}} \tag{5-3}$$

Clearly, $I(t) \to 1$ as $t \to \infty$, meaning that all nodes are infected eventually. This is true if no infected nodes are cured and then recovered in the process.

SIS Model

Again, consider the case of $R(t) = 0$. But, now, suppose that the rate of recovery from "infected" back to "susceptive" is δ, namely, $S(t) = \delta I(t)$. It then follows from (5-2) that

$$\frac{dI(t)}{dt} = \lambda[1 - I(t)]I(t) - \delta I(t), \qquad I(0) = I_0 \tag{5-4}$$

This equation has a solution

$$I(t) = \begin{cases} \left\{ \left[\dfrac{1}{1 - \dfrac{\delta}{\lambda}} + \left(I_0^{-1} - \dfrac{1}{1 - \dfrac{\delta}{\lambda}} \right) e^{-\left(1 - \frac{\delta}{\lambda}\right)\lambda t} \right]^{-1} \right\}, & \delta \neq \lambda \\[2em] \dfrac{1}{\lambda t + I_0^{-1}}, & \delta = \lambda \end{cases} \tag{5-5}$$

There are three cases:

(i) $\delta > \lambda$: This means that recovering is faster than infecting. It follows from (5-5) that $I(t) \rightarrow 0$ as $t \rightarrow \infty$, implying that the virus will die out eventually.

(ii) $\delta < \lambda$: This means that recovering is slower than infecting. It follows from (5-5) that $I(t) \rightarrow \left(1 - \dfrac{\delta}{\lambda} \right)$ as $t \rightarrow \infty$, implying that a certain fraction of nodes will remain being infected. In particular, if $\delta \ll \lambda$ then $I(t) \sim 1$ as $t \rightarrow \infty$, implying that almost all nodes will be infected eventually.

(iii) $\delta = \lambda$: This is a threshold, determining whether or not the virus will spread.

SIR Model

Now, suppose $R(t) > 0$, namely, some nodes will be cured therefore recovered and become susceptive again. In this case, one can similarly establish the following model:

$$\begin{cases} \dfrac{dI(t)}{dt} = \lambda S(t)I(t) - \delta I(t) \\[2mm] \dfrac{dS(t)}{dt} = -\lambda S(t)I(t) \\[2mm] \dfrac{dR(t)}{dt} = \delta I(t) \\[2mm] I(0) = I_0, \quad S(0) = S_0, \quad R(0) = 0 \end{cases} \tag{5-6}$$

By qualitative analysis or numerical solutions, one can find the following:

(i) $S_0 > \dfrac{\delta}{\lambda}$: Virus will spread out for some time, but eventually die out.

(ii) $S_0 \leq \dfrac{\delta}{\lambda}$: Virus will die our quickly.

Consider a population of N individuals, called nodes hereafter, each is either in the Susceptive (S), or Infected (I), or Recovered (R) state.

Let $\rho_{S,k}(t)$, $\rho_{I,k}(t)$ and $\rho_{R,k}(t)$ be the densities of the S, I and R states of nodes of degree k at time t, respectively.

Then, these densities are related by means of the normalization condition

$$\rho_{S,k}(t) + \rho_{I,k}(t) + \rho_{R,k}(t) = 1 \text{ for all } t \geq 0$$

Define the total number of recovered nodes at time t by

$$R(t) = \sum_k P(k)\rho_{R,k}(t)$$

where $P(k)$ is the node-degree distribution. Denote its prevalence (steady state) by

$$R_\infty = \lim_{t \to \infty} R(t)$$

Threshold Theory in Epidemic Spreading
The classical SIR epidemic spreading model

At the mean-field level, for undirected random-graph type of uncorrelated spare networks, the above densities satisfy the following system of differential equations:

$$\frac{d\rho_{I,k}(t)}{dt} = -\rho_{I,k}(t) + \lambda k \rho_{S,k}(t)\Theta(t)$$

$$\frac{d\rho_{S,k}(t)}{dt} = -\lambda k \rho_{S,k}(t)\Theta(t) \tag{1}$$

$$\frac{d\rho_{R,k}(t)}{dt} = \rho_{I,k}(t)$$

where λ is the effective spreading rate and $\Theta(t)$ represents the average density of infected nodes pointed at by any given edge.

Threshold Theory in Epidemic Spreading
The classical SIR epidemic spreading model

To find $\Theta(t)$ ---

In general, the probability that an edge points to an infected node with degree k is proportional to $kP(k)\rho_{I,k}(t)$

Notice that the infected node pointed by the edge has previously received virus through an edge that cannot be used for virus transmission anymore (since its originating node had been recovered), the above probability should be modified to be $(k-1)P(k)\rho_{I,k}$ Consequently, one has

$$\Theta(t) = \frac{1}{\langle k \rangle} \sum_k (k-1)P(k)\rho_{I,k}(t) \qquad (2)$$

A combination of system (1) and equation (2), along with the natural initial conditions

$$\rho_{R,k}(0) = 0, \quad \rho_{I,k}(0) = \rho_k^0 \text{ and } \rho_{S,k}(0) = 1 - \rho_k^0$$

completely define the SIR model of virus spreading on any random-graph type of uncorrected network with the node-degree distribution $P(k)$.

Assume that initially infected nodes are uniformly distributed, namely, $\rho_k^0 = \rho^0$ for all k at time $t = 0$. In this case, in the limit $\rho^0 \to 0$, namely, starting from very few infected nodes, one has $\rho_{I,k}^0(0) \approx 0$ and $\rho_{S,k}(0) \approx 1$. Under these assumptions, the second and third equations in system (1) can be directly integrated, yielding

$$\rho_{S,k}(t) = e^{-\lambda k \phi(t)} \quad \text{and} \quad \rho_{R,k}(t) = \int_0^t \rho_{I,k}(\tau)d\tau \tag{3}$$

where

$$\phi(t) = \int_0^t \Theta(\tau)d\tau = \frac{1}{\langle k \rangle} \sum_k (k-1)P(k)\rho_{R,k}(t)$$

which, based on (3), gives

$$\frac{d\phi(t)}{dt} = \frac{1}{\langle k \rangle} \sum_k (k-1)P(k)\rho_{I,k}(t)$$

$$= \frac{1}{\langle k \rangle} \sum_k (k-1)P(k)[1 - \rho_{R,k}(t) - \rho_{S,k}(t)]$$

$$= 1 - \frac{1}{\langle k \rangle} - \phi(t) - \frac{1}{\langle k \rangle} \sum_k (k-1)P(k)e^{-\lambda k \phi(t)} \tag{4}$$

Threshold Theory in Epidemic Spreading
The classical SIR epidemic spreading model

If equation (4) can be solved, then one can obtain the total epidemic prevalence (steady-state) $\rho_{S,k}(\infty)$ as a function of $\phi_\infty = \lim_{t\to\infty}\phi(t)$, so that by $\rho_{R,k}(\infty) = 1 - \rho_{S,k}(\infty)$ one obtains

$$R_\infty = \sum_k P(k)\left[1 - e^{-\lambda k \phi_\infty}\right]$$

However, for a general node-degree distribution $P(k)$, equation (4) actually cannot be solved. Nevertheless, one can still obtain some useful information about the asymptotic behavior of the virus spreading. Since $\rho_{I,k}(\infty) = 0$ and consequently $\lim_{t\to\infty} d\phi(t)/dt = 0$, from equation (4) one obtains the following steady-state equation:

$$\phi_\infty = 1 - \frac{1}{\langle k \rangle} - \frac{1}{\langle k \rangle}\sum_k (k-1)P(k)e^{-\lambda k \phi_\infty}$$

which has a fixed-point solution $\phi_\infty = 0$.

In order to have a nonzero solution for ϕ_∞, which would mean a prevalence $R_\infty > 0$, it is necessary that

$$\frac{d}{d\phi_\infty}\left[1 - \frac{1}{\langle k \rangle} - \frac{1}{\langle k \rangle}\sum_k (k-1)P(k)e^{-\lambda k \phi_\infty}\right]_{\phi_\infty = 0} \geq 1$$

This condition implies that

$$\frac{\lambda}{\langle k \rangle}\sum_k k(k-1)P(k) \geq 1$$

yielding the epidemic threshold

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \qquad (5)$$

below which $R_\infty = 0$ and above which $R_\infty > 0$.

Virus Spreading Threshold
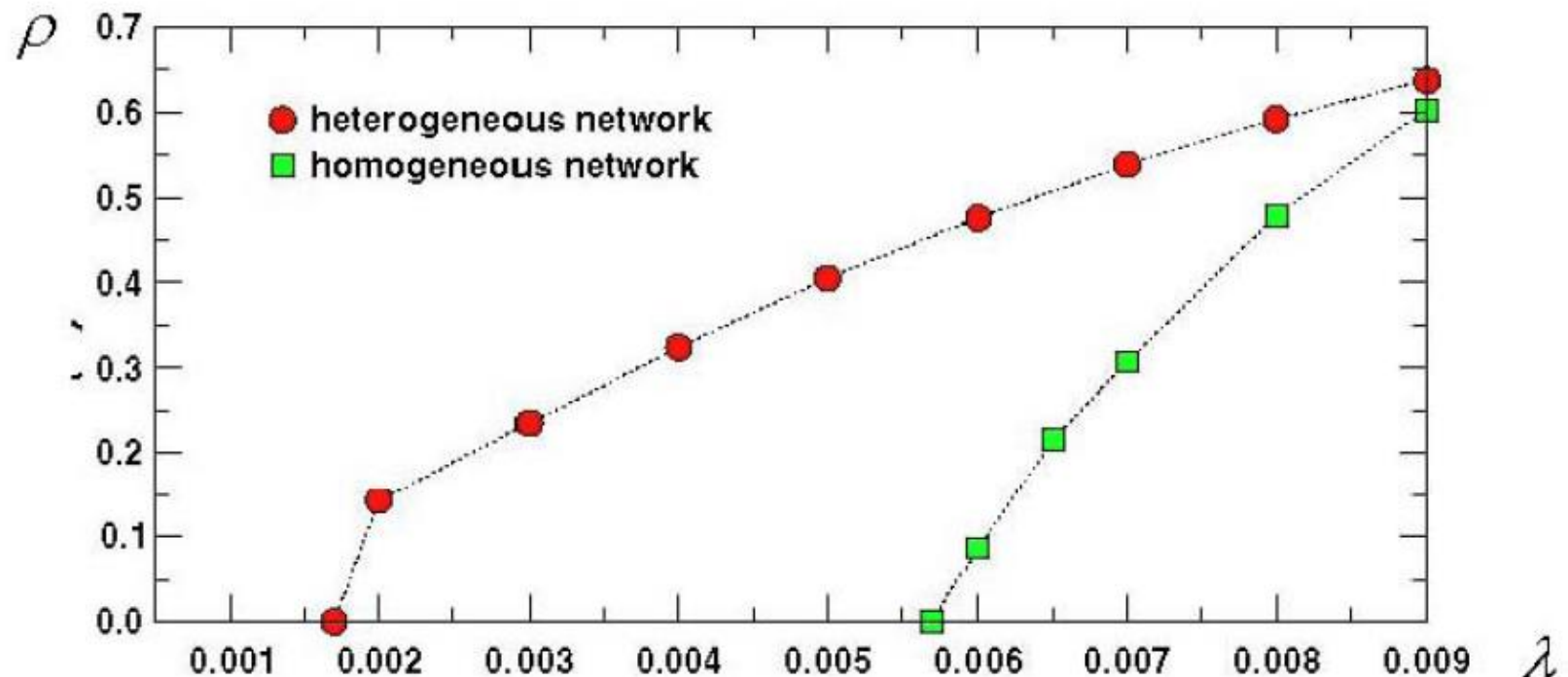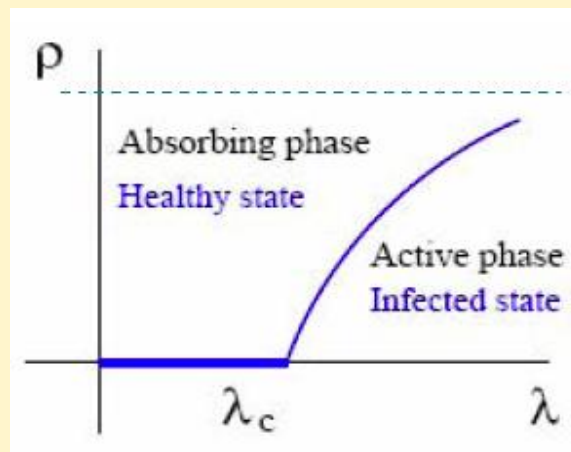
Observations

- This implies that there is a finite epidemic threshold $\lambda_c > 0$:
  - over this threshold, the infected node can propagate the disease to the whole network, eventually leading to a certain equilibrium state, and in this situation the network is said to be in the active phase;
  - under this threshold, the number of infected nodes decays exponentially, therefore disease will not be able to spread out, and in this situation the network is said to be in the absorbing phase, as illustrated at the last figure.

- p: the density of infected population

Epidemic Threshold on Homogenous Networks

Consider the SIS model: S → I → S → I → S →...

- Suppose that the probability of a node becoming "infected" from being "susceptible" is v, and the probability of a node being cured and becomes "susceptible" again is δ. Define the virus effective spreading rate as

$$\lambda = \frac{v}{\delta}$$

- Let the density of infected population at time t be $\rho(t)$, with
$$\rho(t) \rightarrow \rho \ (\ t \rightarrow \infty\ )$$

Epidemic Threshold on Homogenous Networks

Assume:

- Homogeneity: The network is homogeneous (such as an random-graph network or a small-world network), For simplicity, assume that each node has degree $k_i \approx \langle k \rangle$

- Homogenous mixing: The infection strength is proportional to the population density; or equivalently, both v and δ are constant

- The lifetime of the disease is much shorter than the lifetime of the nodes in the population, so birth and death of nodes are not taken into consideration

Epidemic Threshold on Homogenous Networks

Then, the epidemic spreading equation is:

$$\frac{\partial \rho(t)}{\partial t} = -\rho(t) + \lambda \langle k \rangle \rho(t)[1 - \rho(t)]$$

It yields

$$\rho = \begin{cases} 0 & \lambda < \lambda_c \\ \dfrac{\lambda - \lambda_c}{\lambda} & \lambda \geq \lambda_c \end{cases}$$

with the spreading threshold

$$\lambda_c = \frac{1}{\langle k \rangle}$$

ρ

Absorbing phase
Healthy state

Active phase
Infected state

λ_c λ
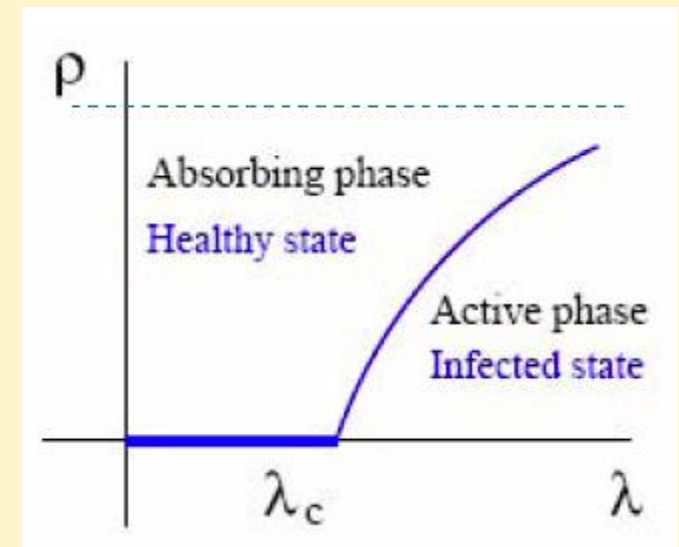
Epidemic Threshold on Homogenous Networks

- Define the virus surviving probability $P_s(t)$ to be the percentage of viruses that are alive since their birth till the current time t

- Figure shows the computer virus surviving probabilities in the 50 months from February 1996 to March 2000, attacked by boot-sector, file and macro viruses, respectively.
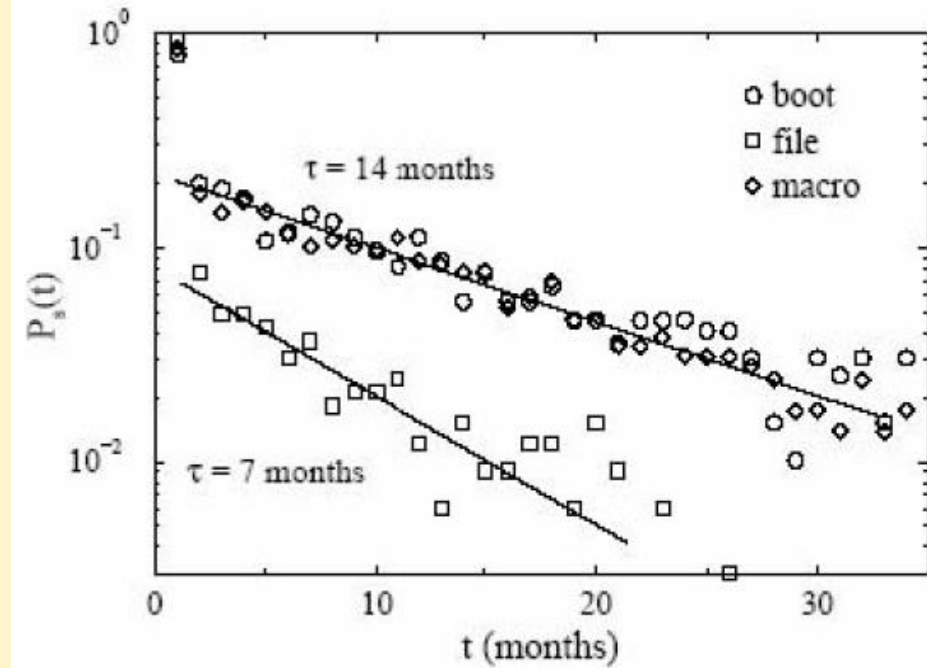
(Pastor-Satorras and Vespignani, 2001)

Observations

- The above statistical data show that the computer virus on the real Internet actually did not propagate as fast and wide as one may imagine. If these strains of viruses were propagating over a homogeneous network then the virus prevalence has to be lower than the epidemic threshold since the surviving probabilities are so low (see Figure),

- Yet the virus effective spreading rate must be higher than this epidemic threshold since the lifetimes of these viruses are so long (see Figure).

- These contradictory phenomena imply that the computer network must be inhomogeneous.
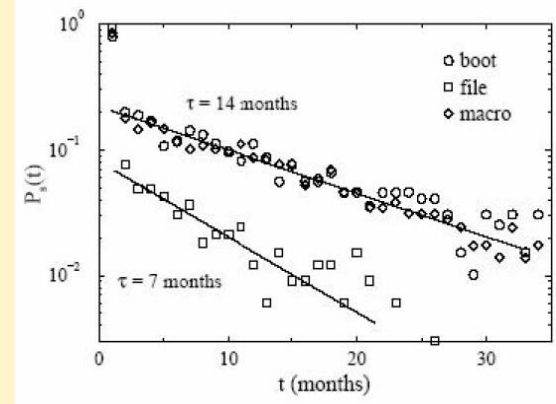
Epidemic Threshold on Scale-Free Networks

The virus spreading equation is modified to:

$$\frac{\partial \rho_k(t)}{\partial t} = -\rho_k(t) + \lambda k[1 - \rho_k(t)]\Theta\left(\rho_k(t)\right)$$

Here, $\Theta(\rho_k(t))$ is the probability of an edge connecting to an infected degree-k node.

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

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

→ Higher-degree nodes have → as $N \rightarrow \infty \rightarrow$ $k \rightarrow \infty$
higher probabilities to be $\rightarrow \lambda_c \rightarrow 0$
infected.

Epidemic Threshold on Scale-Free Networks
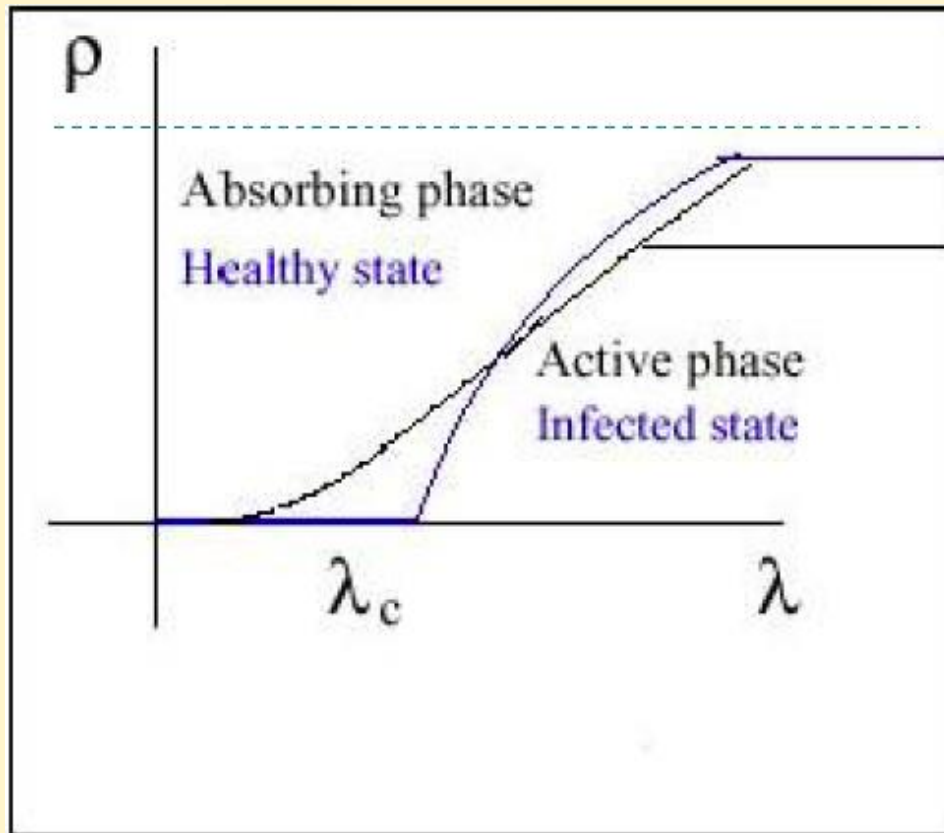
Random graphs

Scale-free networks

λ is the spreading rate

λ_c is the threshold

- Classical theory: $\lambda_c > 0$
- Scale-free network: $\lambda_c = 0$

(Pastor-Satorras, 2001)

Epidemic Threshold on Scale-Free Networks

- It shows that the effective spreading rate λ on the BA network tends to zero continuously and smoothly, implying that it virtually has no positive threshold λc, so for any $\lambda > 0$ virus can easily propagate throughout the network and finally reach a steady state. This demonstrates the fragility of scale-free networks against virus spreading.

- In the last chapter, it was shown that the Internet is rapidly growing following a scale-free manner with the degree distribution in a power-law form. Therefore, it is not surprising to see that computer viruses can spread all over the Internet easily and quickly.

- Yet, fortunately, on the real Internet, it is also found that the λ value is very small ($\lambda << 1$), leading to a low propagation speed. This is also consistent with the observation that the epidemic threshold of a scale-free network vanishes fairly slowly, as the size of the network expands even quickly towards infinity.

Epidemic Threshold on Scale-Free Networks

In a BA network, ρ and $1/\lambda$ have a power-law relation
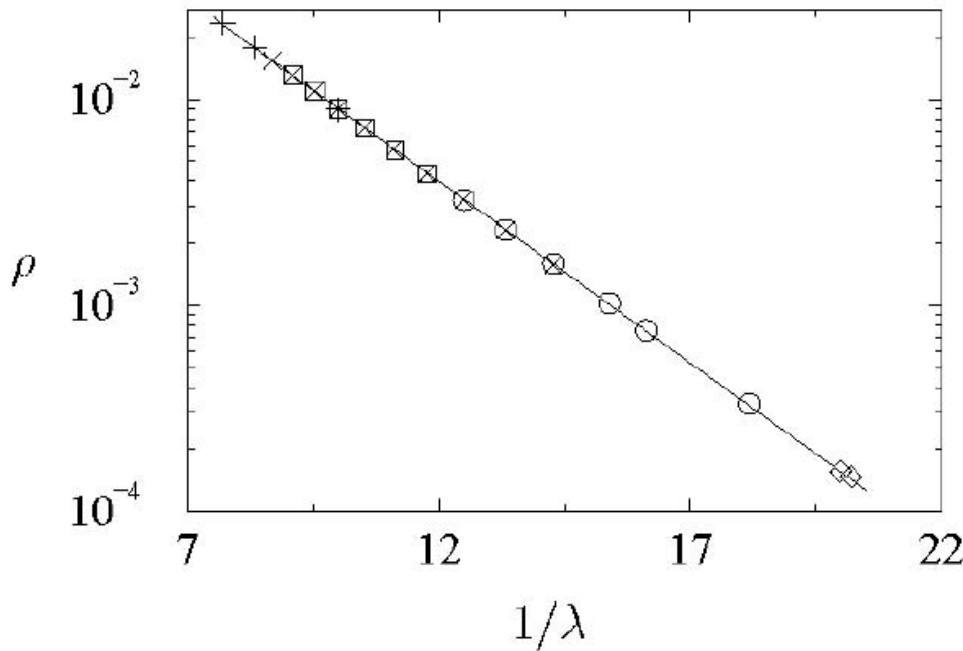
$+$ $N=10^5$

Square $N=5\text{x}10^5$

x $N=10^6$

Circle $N=5\text{x}10^6$

A BA network with different sizes

It implies that
$\rho\,(\lambda) \sim exp(-c\,/\,\lambda)$ for a constant $c > 0$, independent of the size of the network

(Pastor-Satorras and Vespignani, 2003)

Epidemic Threshold on Scale-Free Networks

In the same BA network: evolution of virus surviving probability
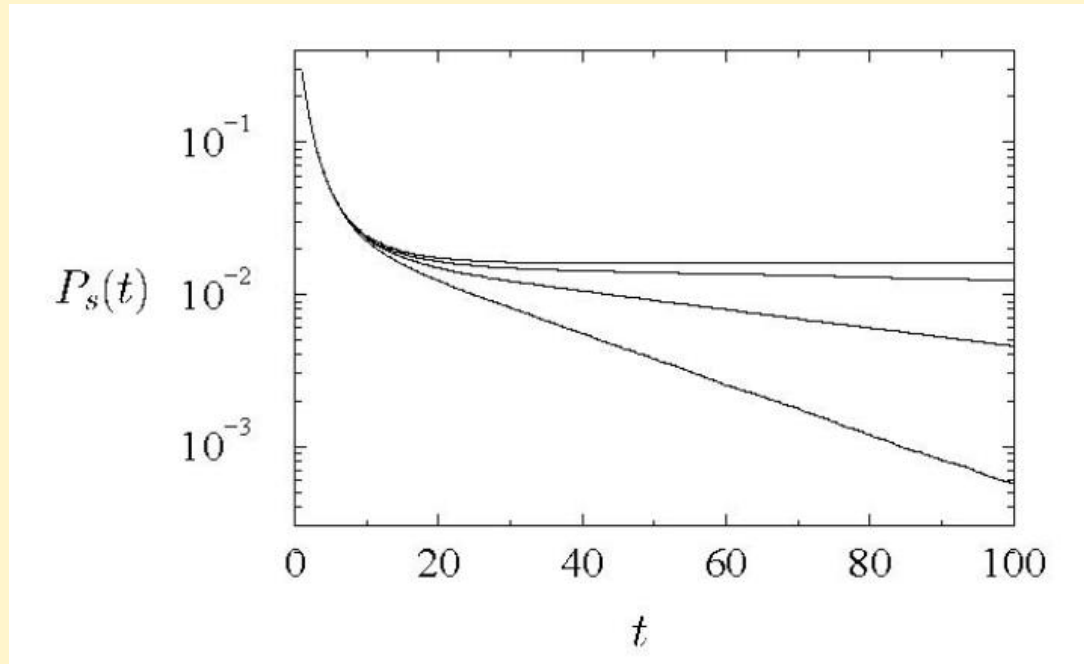
Network sizes increase from top to bottom

A BA network with different sizes

It indicates that on any finite-sized scale-free network, epidemic will eventually die out. This is reasonable because there is a finite probability of all nodes being cured from the infection at the same time as epidemic propagates on the network.
Moreover, the bigger the network is, the slower the disease dies out.

(Pastor-Satorras and Vespignani, 2003)

Epidemic Threshold on Scale-Free Networks

Finite-size effect (for a not-too- large N (so, not-too-large k):

It shows a comparison of the epidemic threshold values on a finite-size scale-free network and on a corresponding homogeneous network, λ_c versus λ_c^H

It can be seen that the threshold of the former is only about 1/10 of the latter, and that as k_c (i.e., N) increases, the threshold can still tend to zero, implying the fragility of finite-sized scale-free networks against virus spreading.
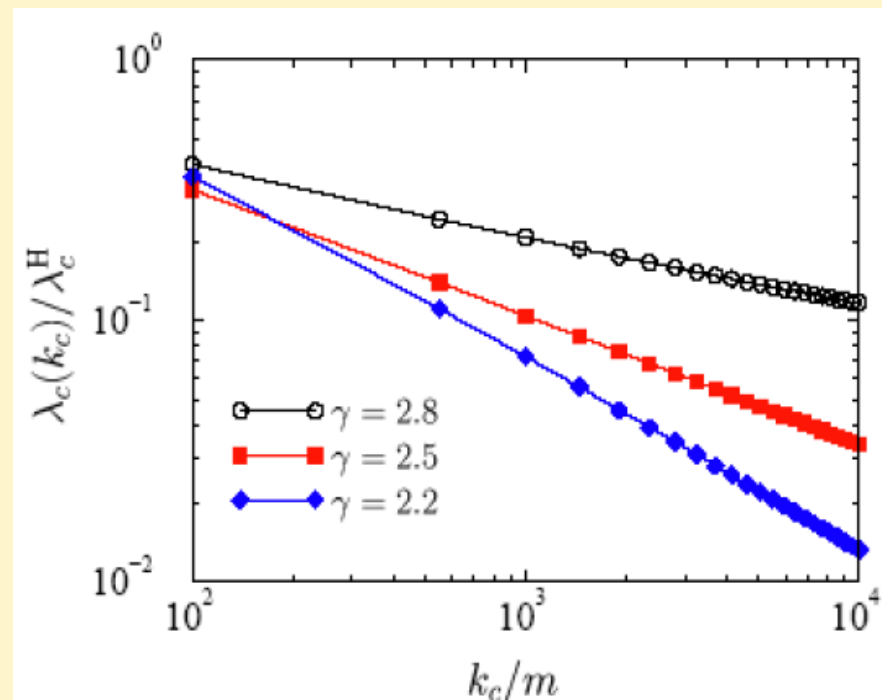
(Pastor-Satorras and Vespignani, 2003)

Immunization on Scale-Free Networks

- Since scale-free networks are fragile to virus attacks, causing wide and serious outbreak, immunization becomes especially important for this type of networks

- There are three effective immunization strategies:

➢ Random Immunization (or, Uniform Immunization)
➢ Targeted Immunization (or, Selective Immunization)
➢ Acquaintance Immunization

Random Immunization

- Let the immunity (density of immunized nodes) be denoted by g.

- Random immunization strategy: the immunization threshold is

$$g_c = 1 - \frac{\lambda_c}{\lambda}$$

where $\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$ is the effective epidemic threshold, giving

$$g_c = 1 - \frac{1}{\lambda}\frac{\langle k \rangle}{\langle k^2 \rangle}$$

Clearly, $N \rightarrow \infty \rightarrow k \rightarrow \infty \rightarrow \lambda_c \rightarrow 0 \rightarrow g_c \rightarrow 1$

→ In order to immune a scale-free network, almost every node has to be immunized.

→ random immunization for scale-free networks is inefficient and expensive

Targeted Immunization

Targeted immunization strategy: to immune the most highly connected nodes, thereby reducing or completely blocking virus spreading

For BA scale-free networks, the immunization threshold is

$$g_c \sim e^{-\frac{2}{m\lambda}}$$

→ a very small immunization threshold can be obtained for a very large range of effective spreading rate λ
→ for scale-free networks, targeted immunization has a much smaller immunization threshold than random immunization, hence is more effective

Satorras and Vespignani (2003)
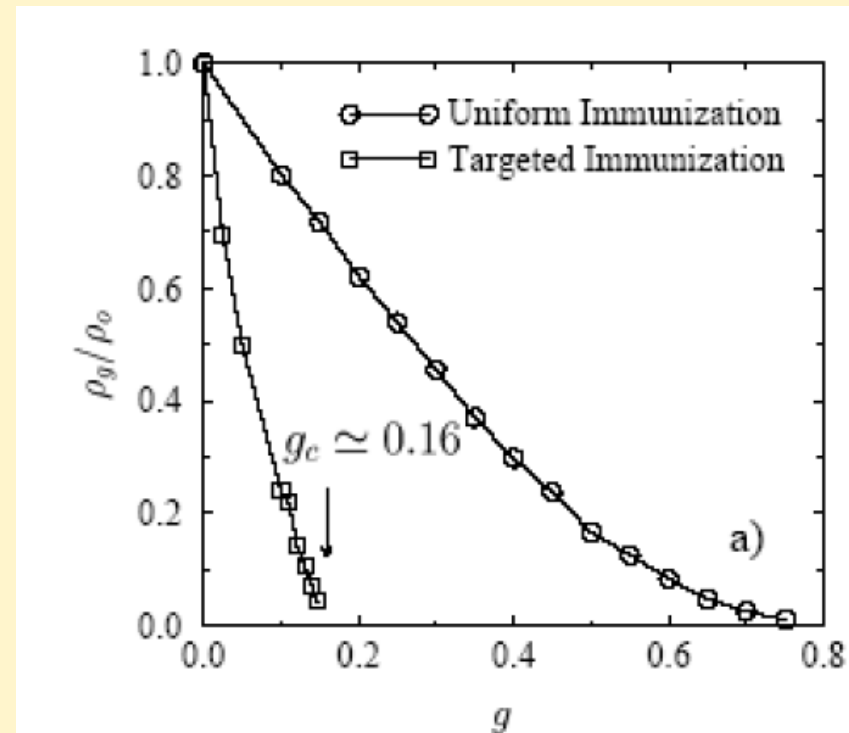
Acquaintance Immunization

Idea:

1. → 2. → 3. → ......
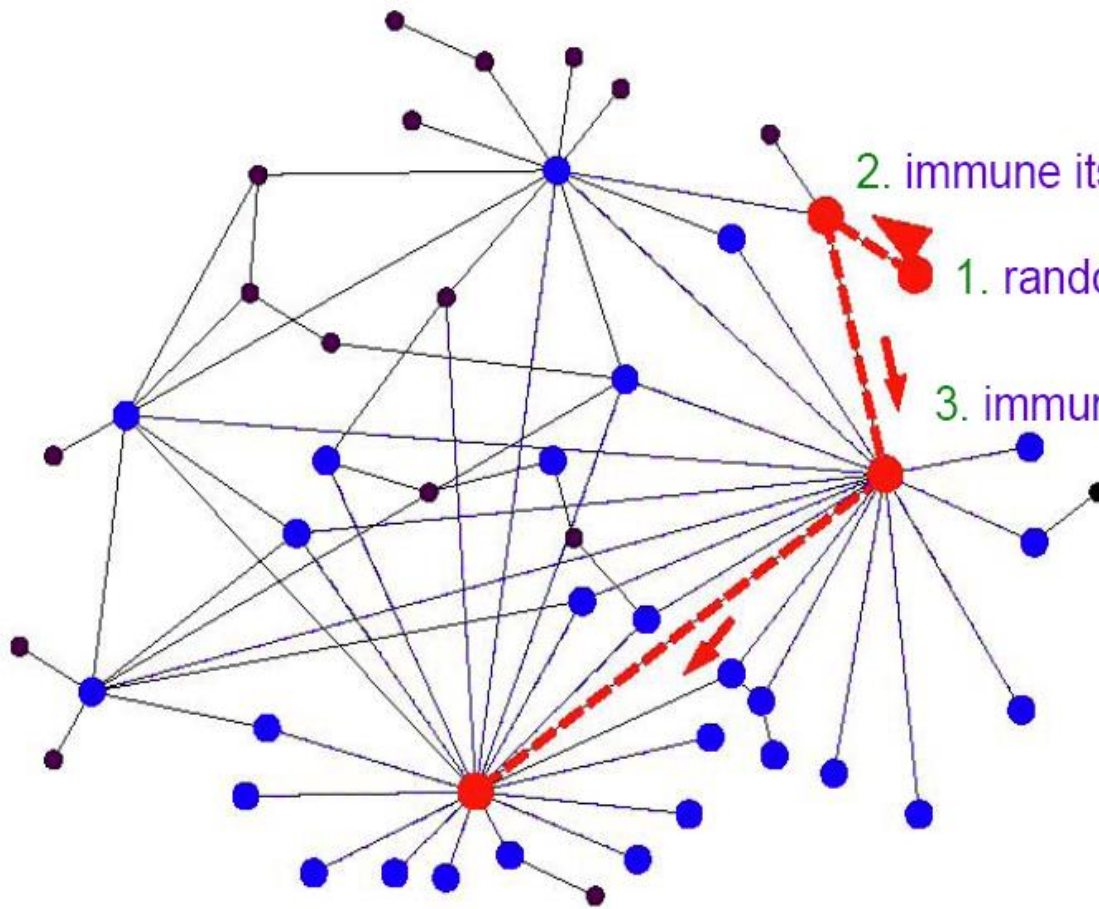
2. immune its biggest neighbor

1. randomly select one node

3. immune its biggest neighbor

Immunization on Scale-Free Networks

Acquaintance immunization strategy: to randomly select a fraction of nodes from the population, and for each node in the fraction select the biggest one of its neighbors for immunization

This scheme is a localized strategy -- only requires information about the randomly selected node and its neighbors, but not the global network
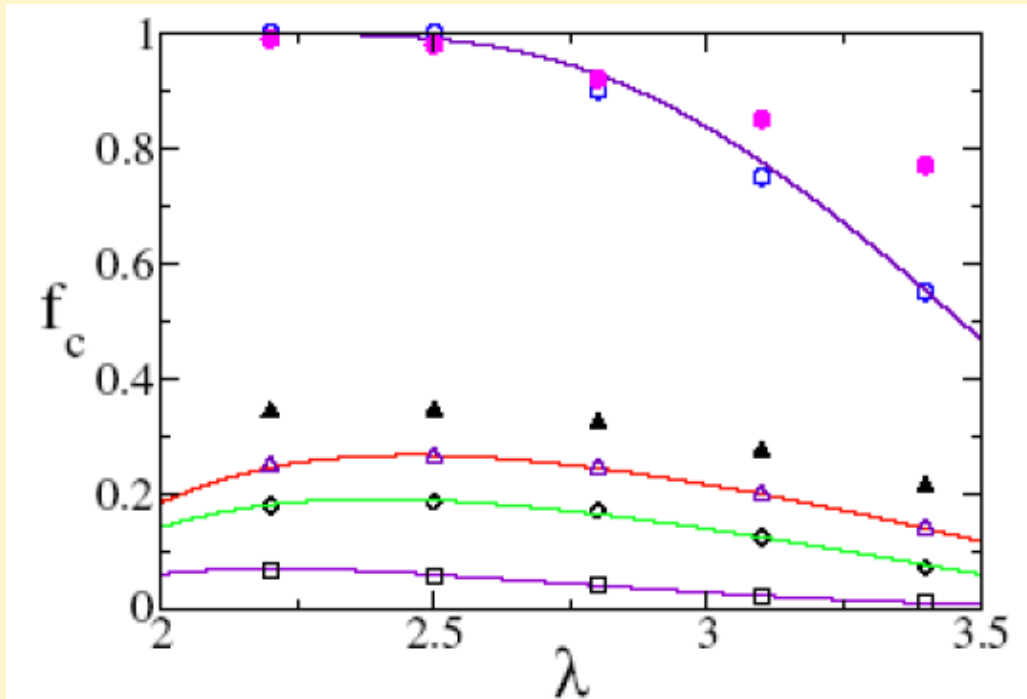
empty circles – random
empty triangles – acquaintance
empty diamonds – two-acquaintance
empty squares – targeted

solid circle and solid triangles
– assortative networks
(random)   (acquaintance)
f_c- immunization threshold

Madar et al. (2004)

Computer Virus Spreading over the Internet

Now, let N be the number of susceptible nodes in the network, $a = a(t)$ be the fraction of infected nodes at time t , and K be the rate of infection

Since the total number of infected nodes is Na, and an infected node infects a total of $K(1 - a)$ nodes each time, the rate of total nodes to be infected in the next short time duration, dt, will be $(Na)K(1 - a)$, giving

$$\frac{d(Na)}{dt} = (Na)K(1-a) \qquad \rightarrow \qquad \frac{da}{dt} = Ka(1-a)$$

which is the logistic equation, with solution

$$a = \frac{e^{K(t-t_0)}}{1+e^{K(t-t_0)}}$$

Clearly, $t \rightarrow \infty$ $\rightarrow$ $a \rightarrow 1$ $\rightarrow$ all nodes will be infected eventually

This formula of a can be used to predict a computer virus spreading

Computer Virus Spreading over the Internet
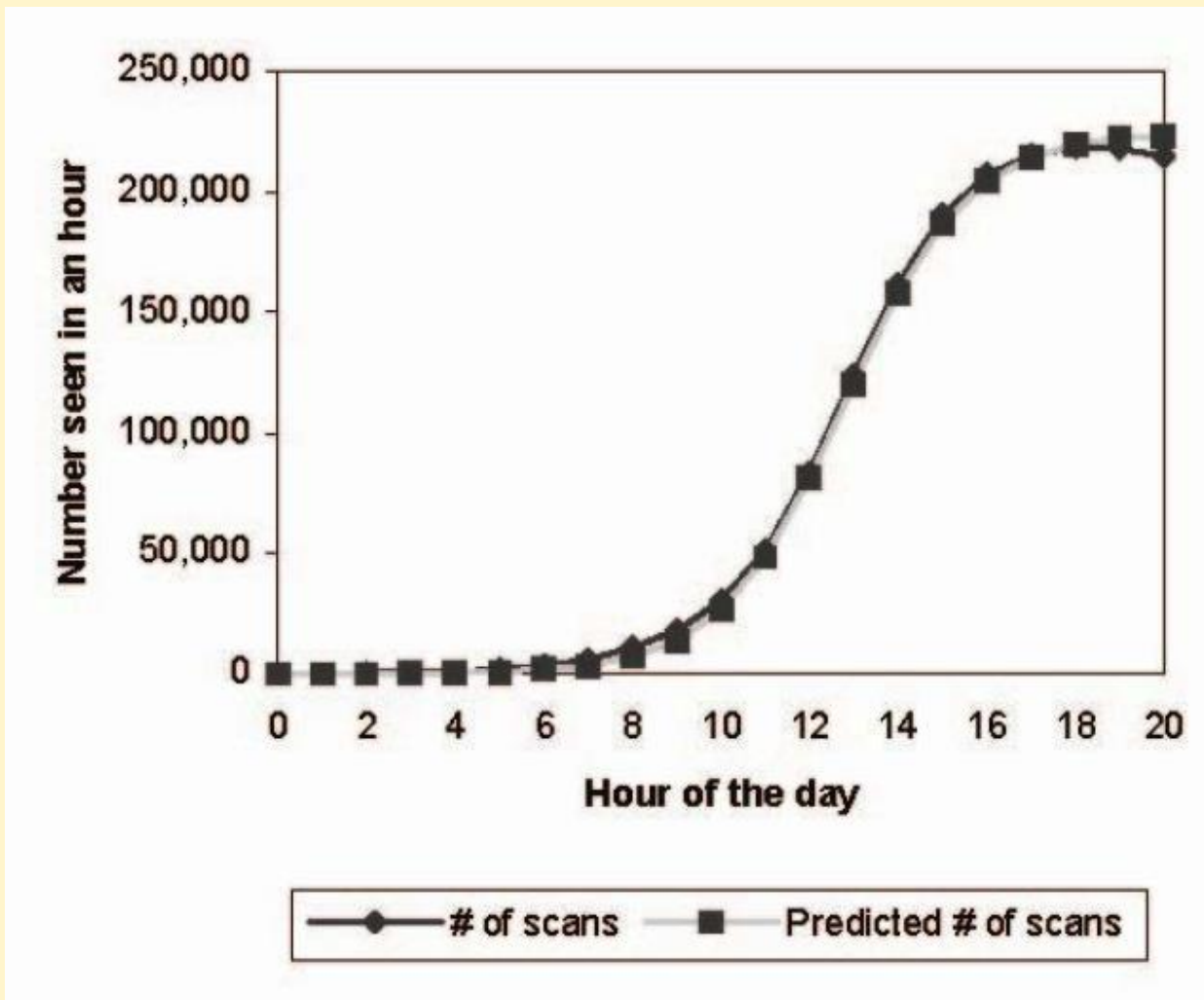
Code-Red Worm: Hourly probe rate data during the reemergence on 1 August 2002 (staniford et al., 2002)

Modeling the Computer Worms
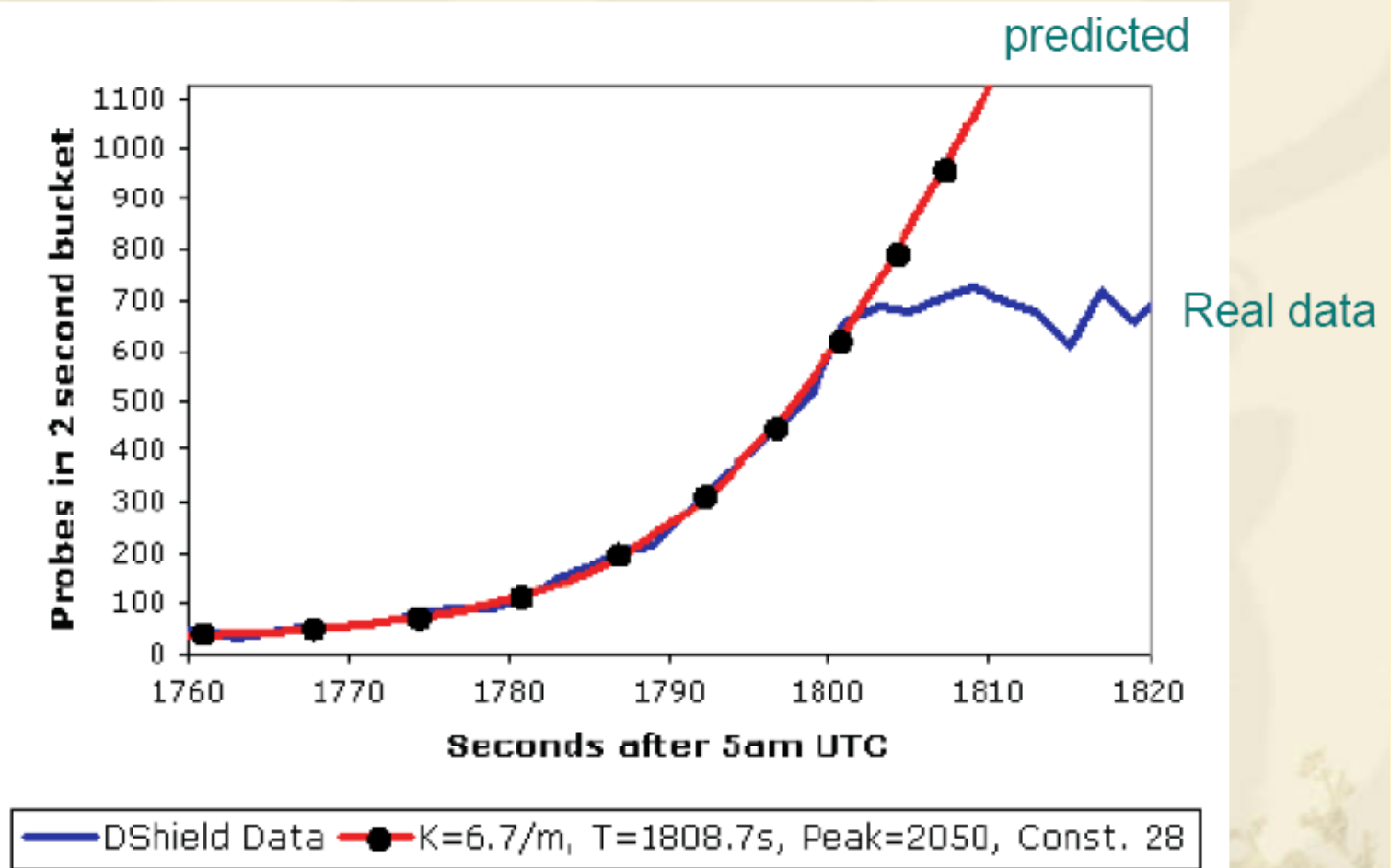
Slammer Worm: Real data versus model prediction
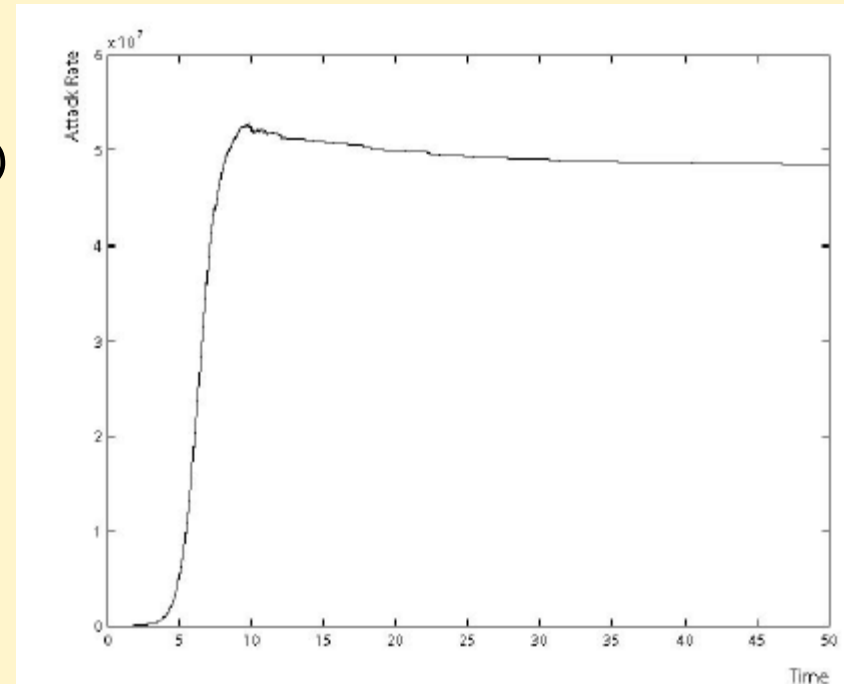1 July 2001 (staniford et al., 2002)

Modeling the Computer Worms

A Compartment-Based Model of Computer Worms

- Let N_i be the number of susceptible hosts in the ith AS and a_i be the fraction of the infected hosts among them
- Let K be the (constant) average rate of spreading spread
- Then, the model is

$$da_i / dt = \left[a_i K \frac{N_i}{N} + \sum_{\substack{j=1 \\ j \neq i}} \frac{N_j}{N} a_j K \right] (1 - a_i)$$

$$i = 1, 2, ..., N$$

Serazzi and Zanero (2004)

Spreading Models of Email Viruses

- **Typical examples:**
  - Milessa (1999)
  - I-Love-You (2000)
  - Nimda (2001)
  - Win32/Sircam (2001)
- Chinese versions:
  - Worm.Klez.cn.b (2002)
  - Worm.SoBig.c (2003)
  - Worm.Mimail.C (2003)
  - SCO bomb (2004)
  - ……
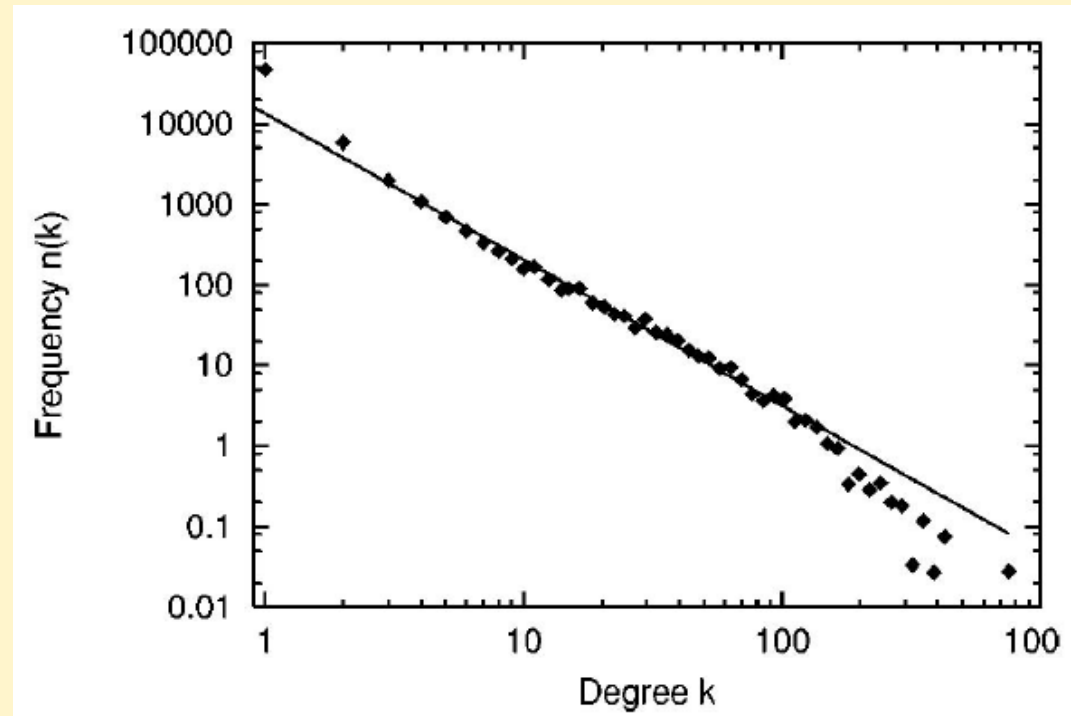
Node-degree distribution of the email network in Kiel University, Germany, 2002

$$N = 59812, \langle k \rangle = 2.88, \gamma = -1.81$$

Ebel et al (2002)

Reinfection versus Non-Reinfection

- **Reinfection** means that a user will send out virus-contaminated emails whenever he opens a virus-contaminated email attachment

Reinfection vs. Non-reinfection
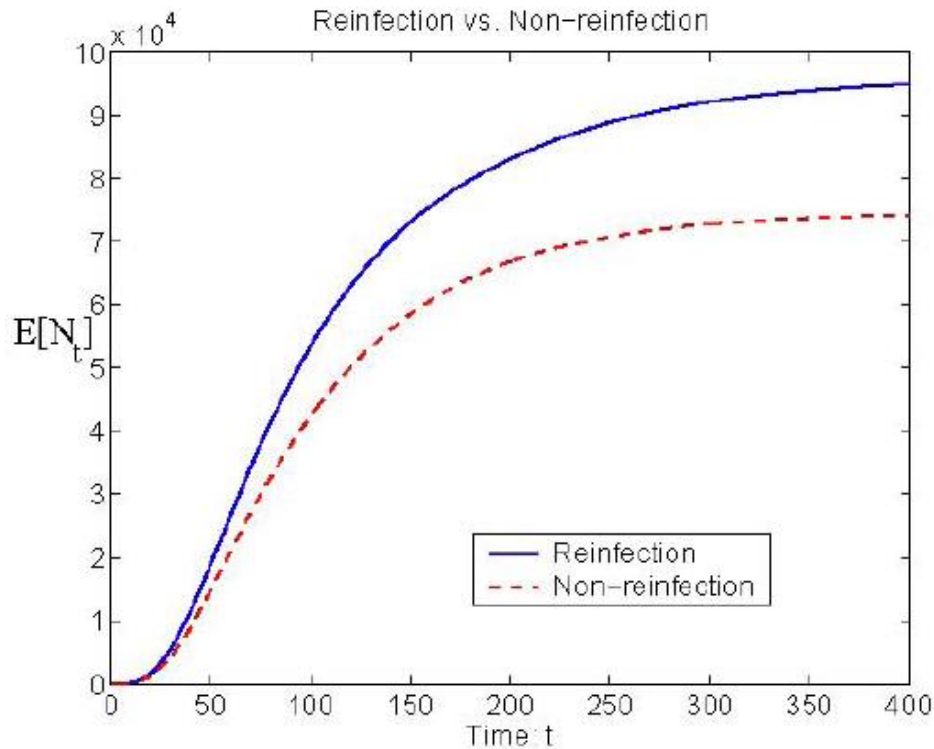
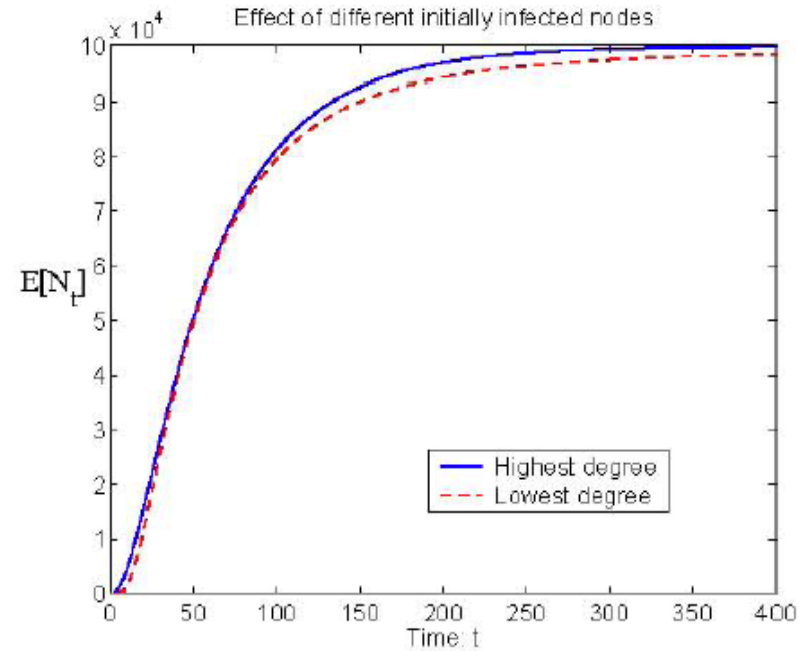$E[N_t]$ - expectation of total number of infected nodes at time t

Zou et al (2003)

→In the case of reinfection, more users are being infected

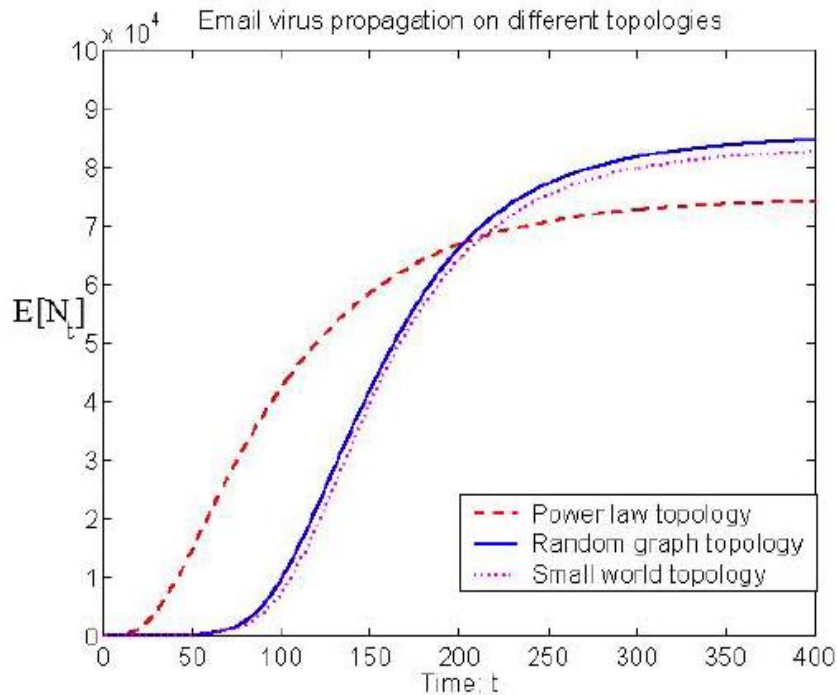# Reinfection versus Non-Reinfection
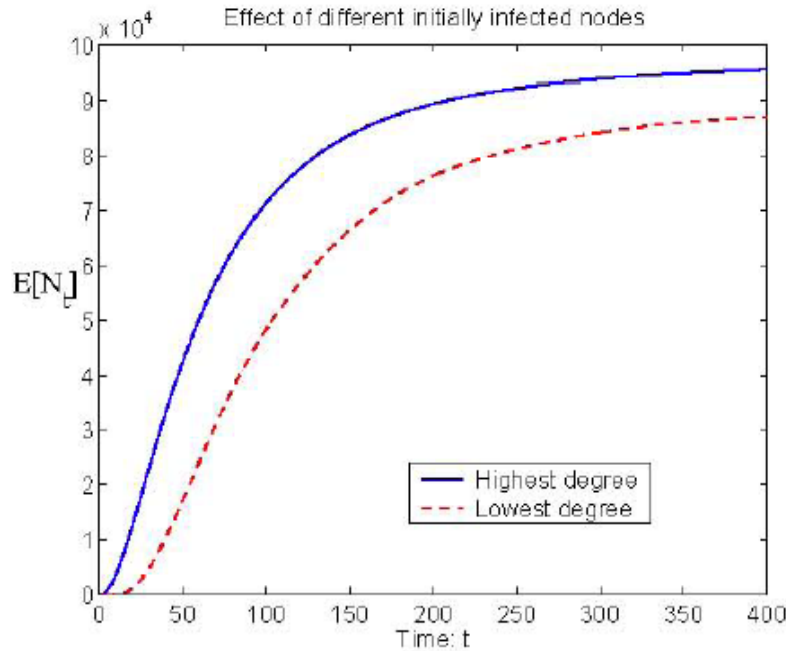
a. Network with average degree 8   b. Network with average degree 20

The initially infected nodes are more important in a sparsely connected network (with $<k>=8$) than a densely connected network (with $<k>=20$) of the same size and the same type
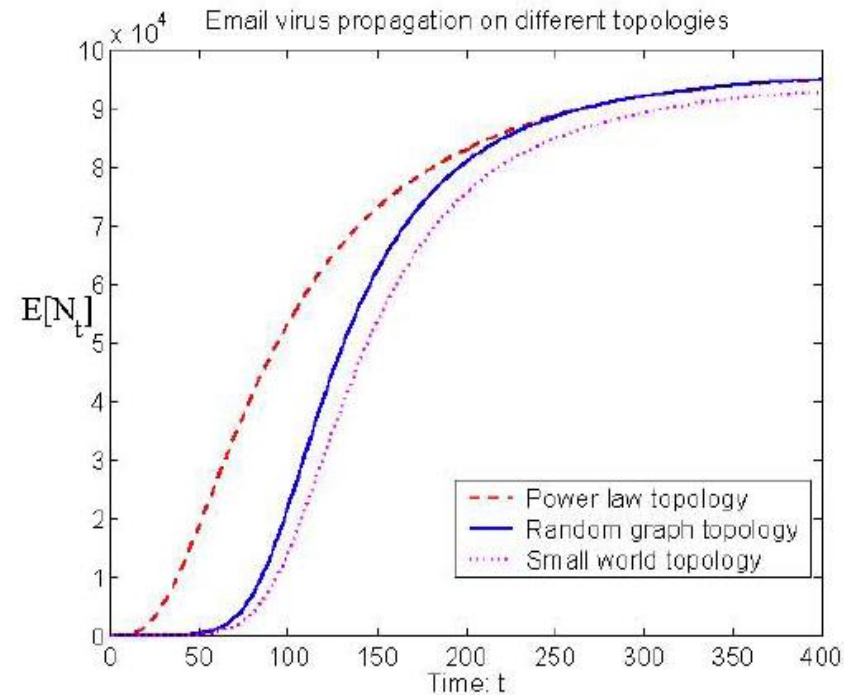
zou et al. (2003)

Reinfection versus Non-Reinfection

Network topology affects email viruses spreading:
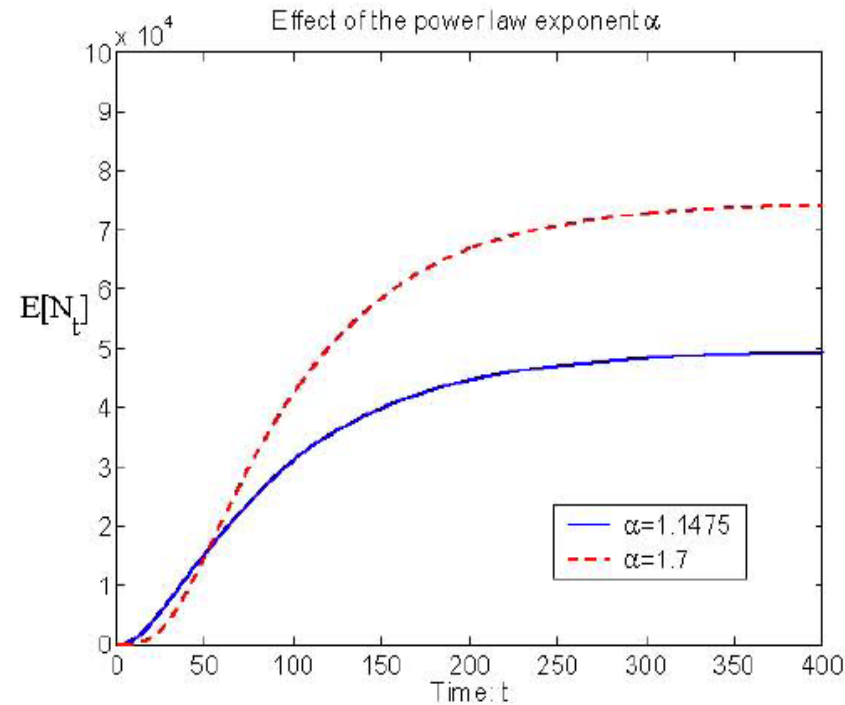

a. Non-reinfection case

b. Reinfection case

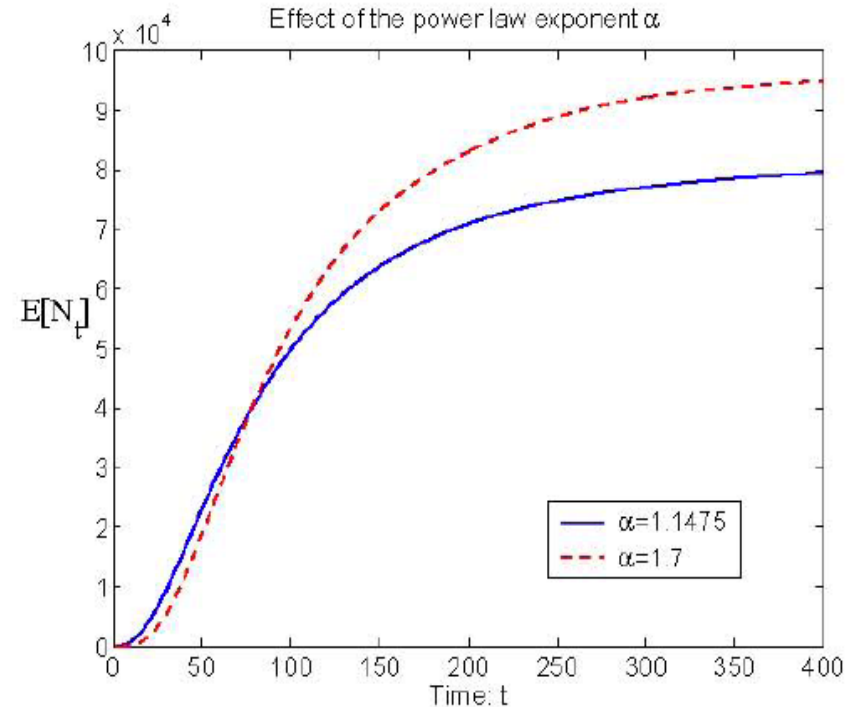Email viruses propagate much faster in scale-free networks than in the other two

zou et al. (2003)

Reinfection versus Non-Reinfection

Power-law exponent affects email viruses spreading:

a. Non-reinfection case
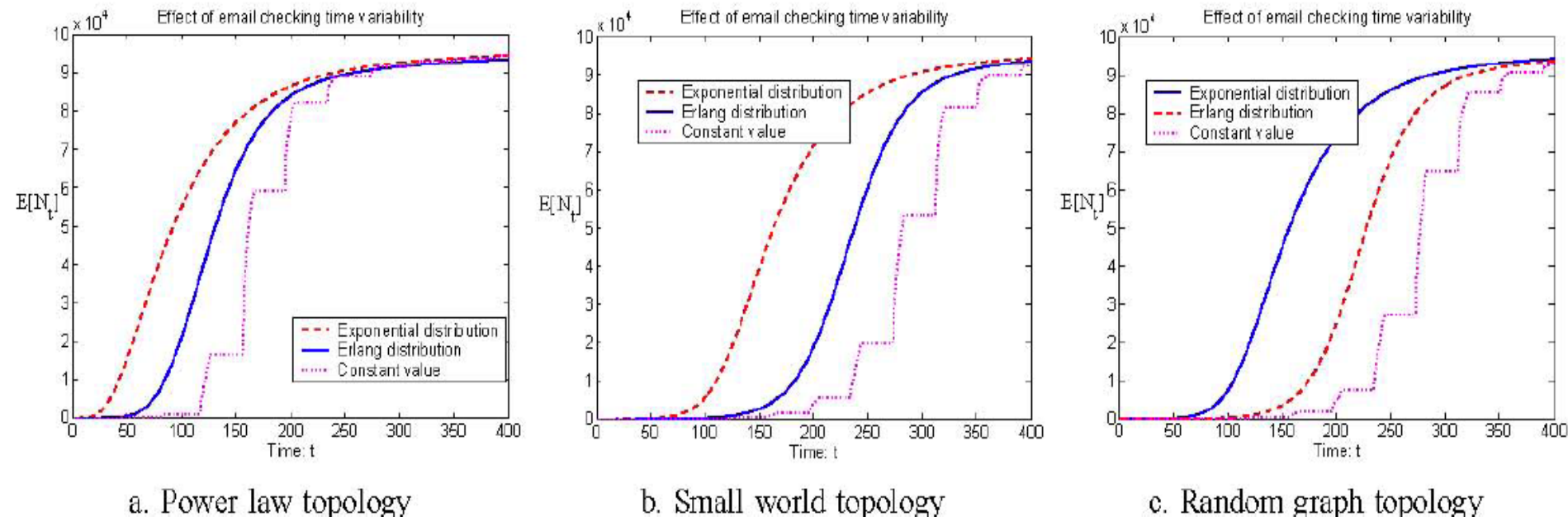
b. Reinfection case

Networks with larger exponents have more nodes to be infected

zou et al. (2003)

Reinfection versus Non-Reinfection

By changing the distribution of the email checking times, different effects on the average number of infected nodes may vary.

Effects of email checking time:

a. Power law topology   b. Small world topology   c. Random graph topology

The more variable the distribution has, the faster the virus propagates

zou et al. (2003)

On the contrary-
Virus Spreading Can also Change the Network Topology

It has been seen that network topologies significantly affect the virus spreading patterns over the networks. On the contrary, when a computer virus propagates over a network, it traverses part of the network therefore could affect the topology of this part of the network as the medium of virus spreading. Of course, another part of the network over which the virus never travels will be left unchanged. In most cases, the topology of the virus-traveling network is determined by the spread and replication of virus. Most viruses and worms spread through different formats at different rates to different extents, which often change the spreading-path topologies on the underlying networks quite significantly.
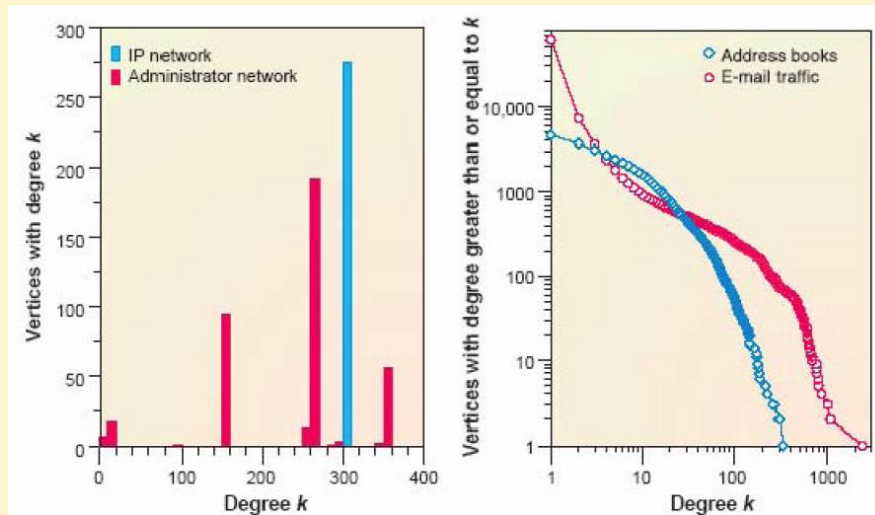
Balthrop et al. (2004)

It is desirable to have a control strategy that is immune to the topological changes