Virus Propagation: Progress Report

Leonid Kontorovich
CALD, SCS
Carnegie Mellon University
lkontor@cs.cmu.edu

Pradeep Ravikumar
CALD, SCS
Carnegie Mellon University
pradeepr@cs.cmu.edu

March 24, 2005

Abstract

We plan to derive closed-form solutions and steady-state characteristics of a recent virus-propagation model. We also address more tractable tasks such as deriving the above for approximations of the system, such as linearizations, and for specific graph topologies.

1 Introduction

The dynamics of social networks has attracted considerable interest in varied communities such as biology, sociology and more recently computer science. The varied applications include analyzing computer-virus propagation in computer networks; studying the propagation of a meme (idea) among a network of people for viral marketing; and the propagation of an infection for epidemic control. These problems could all be generalized to the basic task of analyzing the propagation of a node-state in large interconnected network of nodes.

The basic dynamical system we consider in this project is the discrete, non-linear system of Chakrabarti et al [8] (*described later on*). The key questions we explore about this system are:

- Is there an analytical (closed-form) expression for the dynamics of the system? In particular, we seek analytical expressions for the probability of a node being infected as a function of time and other model parameters. In the event of not being able to derive the same for general graphs, we would then focus on specific topologies such as star graphs.
- Given a particular initial state, does the system then reach a steady state or not? Note that it is also possible for the system to oscillate between multiple states or even exhibit irregular behavior.
- If the system does reach a steady state, what are its characteristics? In particular, what is the number of infected nodes?

- Under what conditions does the system reach a particular steady state? The steady state we investigate is the dying out of the epidemic. In particular, what are the epidemic thresholds of certain model parameters above which the epidemic dies off exponentially?
- We also look at approximations to the dynamical system (such as linearization), and explore the same questions as above.

The outline of the report is as follows. Section 2 gives the background and survey of previous modelling techniques. Section 3 describes our work in answering the questions detailed above, and Section 5 gives the plan of further activities.

2 Survey

There is a rich tradition of modeling biological, and in particular epidemiological phenomena via network graphs; see the comprehensive survey in [3]. In fact, Kephart and White appear to be the first to apply the SIS model to the transmission of computer viruses in a network. They work in the Erdös-Rényi random graph model, where the graph has N nodes and edge probability p. Each edge is assumed to have an infection rate β and each node a cure rate δ. For large N, the authors examine $i(t) \equiv I(t)/N$, where $I(t)$ is the number of infected nodes at time t. Under a deterministic approximation, and with appropriate assumptions, the following differential equation is derived:

$$\frac{di}{dt} = \beta p(N-1)i(1-i) - \delta i, \tag{1}$$

with solution

$$i(t) = \frac{i_0(1-\rho')}{i_0 + (1-\rho'-i_0)e^{-(\beta'-\delta)t}}, \tag{2}$$

where $\beta' = \beta p(N-1)$, $\rho' \equiv \delta/\beta'$ and $i_0 \equiv i(t=0)$.

The authors also provide a more detailed (but still approximate) probabilistic analysis. Let $p(I,t)$ be the probability that there are I infected individuals at time t. This probability evolves in time according to

$$\frac{\partial p(I,t)}{\partial t} = -p(I,t)[R_{I\to I_+} + R_{I\to I_-}] + p(I_+,t)R_{I_+\to I} + p(I_-,t)R_{I_-\to I}, \tag{3}$$

where $I_\pm \equiv I \pm 1$ and $R_{a\to b}$ is the rate of transition from state a to state b. Under various independence assumptions, these rates are approximated by

$$R_{a\to a+1} = a(1-a/N)\beta' \tag{4}$$

and

$$R_{a\to a-1} = a\delta. \tag{5}$$

Substituting these into (3), the authors obtain

$$\frac{\partial p(I,t)}{\partial t} = -p(I,t)[I(1-i)\beta' + \delta I] + p(I_+,t)I_+\delta + p(I_-,t)[I_-(1-i_-)\beta'] \tag{6}$$

where $i \equiv I/N$ and $i_\pm = I_\pm/N$. This system of $N+1$ coupled linear differential equations is straightforward to solve, given the authors' observation that the matrix is tridiagonal.

One problem with Kephart et al's work is that they base their network-model on the Erdös-Rényi random graph, which does not represent most modern networks such as the internet, which are of a more scale-free nature. In particular, the graph-generation model they consider falls under the general category of *homogeneous* models, which could be loosely defined as those that assume the same "average" behavior for each node. They also assume that every node has equal contact with other nodes in the graph, and that the rate of infection is dependent on the proportion of infected nodes in the graph.
In a more recent work [4, 5], Pastor-Satorras and Vespignani consider an infinite graph, where the probability that a node has k neighbors behaves as a power law:

$$P(k) \sim k^{-\gamma} \tag{7}$$

with $2 \leq \gamma \leq 3$. They derive a mean-field equation for the evolution of ρ_k, the relative density of infected nodes with connectivity k:

$$\frac{\partial \rho_k(t)}{\partial t} = -\rho_k(t) + \lambda k(1-\rho_k(t))\Theta(\lambda) \tag{8}$$

where λ is the infection rate, and $\Theta(\lambda)$ is the probability that any given link points to an infected node. Under the stationarity assumption $\frac{\partial \rho_k(t)}{\partial t} = 0$, the authors find

$$\rho_k = \frac{k\lambda\Theta(\lambda)}{1+k\lambda\Theta(\lambda)}. \tag{9}$$

Since the probability that a link points to a node with s links is proportional to $sP(s)$, the authors obtain

$$\Theta(\lambda) = \sum_k \frac{kP(k)\rho_k}{\sum_s sP(s)}, \tag{10}$$

which enables them to find $\Theta(\lambda)$ by solving a consistency equation. The overall infection density is simply $\rho = \sum_k P(k)\rho_k$. Approximating k as a continuous variable and integrating (10), the authors write $\Theta(\lambda)$ as

$$\Theta(\lambda) = \frac{e^{-1/(m\lambda)}}{m\lambda}(1-e^{-1/(m\lambda)})^{-1} \tag{11}$$

(here, m is the constant in $P(k) = 2m^2k^{-3}$) which yields the first-order approximation

$$\rho \approx 2e^{-1/(m\lambda)}. \tag{12}$$

The authors provide empirical evidence for their calculations.

However, Pastor-Satorras et al's work [4, 5] suffers the twin drawbacks of firstly, being limited to power law networks and secondly, using an "averaging" approach wherein all graphs with a given degree distribution are considered equal. Chakrabarti et al [8, 1] address both these issues with their model. Their model holds for arbitrary graphs, and is not limited to particular topologies or graph-generation models. Secondly, their model takes into account the entire adjacency matrix of the graph, and does not just consider particular characteristics of the graph such as the degree-distribution. The latter suffers from the problem that all graphs with the same characeteristic would be considered the same as far as the model is concerned. But, for example, as in Pastor-Satorras et al's model, graphs with the same degree distribution might have different topologies, and it is not clear that they should all have the same dynamics such as epidemic thresholds, etc.

Chakrabarti et al have a discrete time model, in which an infected node i tries to infect its neighbours with probability β. Any node i has a probability δ of being cured at any time step. The analytic parameter of the model is $p_{i,t}$, which denotes the probability that the node i is infected at time t. Their analytical model is then given by

$$1 - p_{i,t} = (1 - p_{i,t-1})\zeta_{i,t} + \delta p_{i,t}\zeta_{i,t}$$

where $\zeta_{i,t}$ is the probability that a node i is not infected by its neighbours at time t. This is in turn expressed in terms of the analytic parameters of time $t - 1$,

$$\zeta_{i,t} = \prod_{j \in NBR(i)} (p_{j,t-1}(1 - \beta) + (1 - p_{j,t-1})) = \prod_{j \in NBR(i)} (1 - \beta * p_{j,t-1})$$

While they do not derive a closed-form expression for $p_{i,t}$, they give a necessary and sufficient threshold condition for the epidemic to die out

$$\frac{\beta}{\delta} < \frac{1}{\lambda_{1,A}}$$

where $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix A of the network. Thus, if $\frac{\beta}{\delta} \geq \frac{1}{\lambda_{1,A}}$, the epidemic does not die out. They prove this by using dynamical systems theory [2], in particular the necessary and sufficient conditions for stationarity in such systems.

3 Proposed Work

The epidemiological model we consider is that of Chakrabarti et al [8]. Their model is basically an instance of the SIS (Susceptible Infected Susceptible) model, in which each node is either Susceptible (S) or Infective (I). A susceptible node is currently healthy, but can be infected by its infected neighbours. An infective node can be cured and go back to being susceptible.

As reviewed in Section 2, Chakrabarti et al propose the following dynamical system,

$$1 - p_{i,t} = (1 - p_{i,t-1})\zeta_{i,t} + \delta p_{i,t}\zeta_{i,t}$$

The system is thus coupled, and non-linear and a closed form expression for $p_{i,t}$ for the general topology case seems infeasible. For insights about the convergence properties of the system, we first look at specific graph topologies.

3.1 Special case: the star topology

We consider the case of a star graph – that is, a graph with a single "hub" node, and all the other nodes as its "spokes". Suppose the graph has $n+1$ nodes: the hub is numbered 0 and the spokes are numbered 1 through n. Then Eq. 7 from Chakrabarti et al. becomes:

$$\begin{aligned} p_{0,t} &= 1 - (1 - p_{0,t-1})\prod_{j=1}^{n}(1 - \beta p_{j,t-1}) - \delta \prod_{j=1}^{n}(1 - \beta p_{j,t-1}) && (13) \\ p_{i,t} &= 1 - (1 - p_{i,t-1})(1 - \beta p_{0,t-1}) - \delta(1 - \beta p_{0,t-1}), \qquad 0 < i \leq n+1. && (14) \end{aligned}$$

Let us consider the case $n = 1$; we define $x_t \equiv p_{0,t}$ and $y_t \equiv p_{1,t}$. We give a complete solution for this case:

Theorem 1 *Let $\binom{x_t}{y_t}$ evolve according to*

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = F \begin{pmatrix} x_t \\ y_t \end{pmatrix} \tag{15}$$

where

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 - (1-x)(1-\beta y) - \delta x(1 - \beta y) \\ 1 - (1-y)(1-\beta x) - \delta y(1 - \beta x) \end{pmatrix}. \tag{16}$$

Then, for all $x, y \in [0, 1]$,

(i) The iterated map $F^t\binom{x}{y}$ converges to a fixed point; namely, the limit $\binom{x^*}{y^*} = \lim_{t\to\infty} F^t\binom{x_0}{y_0}$ exists.

(ii) $\binom{x^*}{y^*}$ is satisfies

$$x^* = y^* = \begin{cases} \frac{\beta-\delta}{\beta(1-\delta)}, & \delta < \beta \\ 0, & \delta \geq \beta \end{cases}. \tag{17}$$

(ii) The convergence of $\binom{x_t}{y_t}$ to $\binom{x^*}{y^*}$ is exponentially fast.

Here we explicitly exclude the degenerate case $\beta = \delta = 1$, for which $\binom{x_{t+1}}{y_{t+1}} = \binom{y_t}{x_t}$ and the map does not converge.

Proof 1 *First we show that x_t and y_t quickly become very close in value. It is immeidate from (16) that*

$$|x_t - y_t| = |\beta + \delta - 1| \cdot |x_t - y_t|. \tag{18}$$

Since $\kappa = |\beta + \delta - 1| < 1$, we have $|x_{t+1} - y_{t+1}| = \kappa^t |x_0 - y_0| \to 0$ exponentially fast. This greatly simplifies matters – it means that we need only keep track of one quantity, x_t, which evolves according to

$$x_{t+1} = f(x_t) = 1 - (1 - x_t)(1 - \beta x_t) - \delta x_t(1 - \beta x_t). \tag{19}$$

Now

$$\begin{aligned} \left|f(x) - f(x')\right| &= \left|(x - x')(1 - \delta + \beta(1 + (\delta - 1)x + (\delta - 1)x'))\right| & (20)\\ &\leq |x - x'| \max\{|1 - \delta + \beta|, |1 - \delta + \beta(2d - 1)|\} & (21)\\ &\leq \lambda|x - x'| & (22) \end{aligned}$$

where $\lambda < 1$ is a constant. This means that f is a contraction on the complete metric space $[0, 1]$, so by the Banach Fixed Point theorem[1]*, the iterated map $f^t(x)$ converges to a unique fixed point x^*, regardless of starting value x_0. Convergence is achieved at the exponential rate of λ^t.*

Having proved the existence of a fixed point, it remains to compute it. We do this by solving the equation $f(x) = x$; this quadratic polynomial has two roots: $x = 0$ and $x = \frac{\beta-\delta}{\beta(1-\delta)}$. Since f is a concave function, it always lies below its derivative. We have $f'(0) = 1 + \beta - d$, so $f(x) = x$ has a nonzero solution in $[0, 1]$ iff $\delta < \beta$. In the latter case, it is immediate from the geometry of the superimposed graphs of $y = x$ and $y = f(x)$ that the trajectories $x_{t+1} = f(x_t)$ converge monotonically to the nonzero fixed point.

The case of n spokes has proved more complicated. We can immediately observe that all the spokes assume identical values quite rapidly. Indeed, from (13), we have, for $i, j \neq 0$:

$$\begin{aligned} p_{i,t} - p_{j,t} &= (1 - p_{i,t-1})(1 - \beta p_{0,t-1}) - \delta(1 - \beta p_{0,t-1}) - (1 - p_{j,t-1})(1 - \beta p_{0,t-1}) - \delta(1 - \beta p_{0,t-1})\\ &= (p_{i,t-1} - p_{j,t-1})(1 - \beta p_{0,t-1})(1 - \delta) & (23) \end{aligned}$$

so $|p_{i,t} - p_{j,t}| = (1 - \delta)^t |p_{i,0} - p_{j,0}|$. This simplifies matters significantly, for we now need only keep track of x_t (the probability that the hub is infected) and y_t (the probability that a spoke is infected).

These evolve according to

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = F \begin{pmatrix} x_t \\ y_t \end{pmatrix} \tag{24}$$

[1] See Rudin's *Principles of Mathematical Analysis* [7], p. 220 (where he calls this "the contraction principle").

where

$$F\binom{x}{y} = \begin{pmatrix} 1-(1-x)(1-\beta y)^n - \delta x(1-\beta y)^n \\ 1-(1-y)(1-\beta x) - \delta y(1-\beta x) \end{pmatrix}. \tag{25}$$

We are at this point unable to prove that the iterated map F^t converges to a fixed point, though this is what emerges from numerical experiments. Attempts to approximate F by a linear transformation have met abysmal failure. It is a simple matter to solve the equation $F\binom{x}{y} = \binom{x}{y}$ numerically to obtain the fixed points.

Figure 3.1 shows typical evolution trajectories for $p_{i,t}$, for $n = 4$.

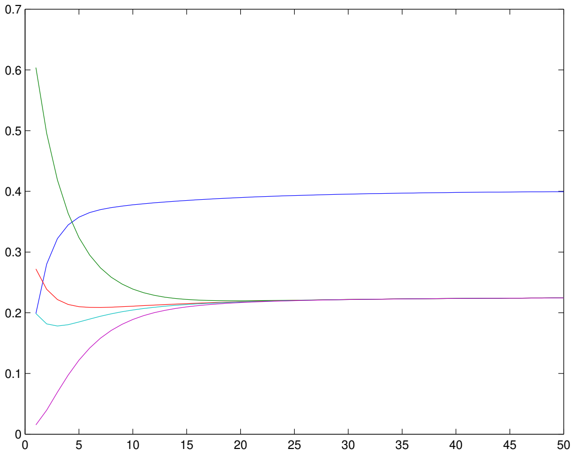

Figure 1: The four spokes converge to $p \approx 0.2$; the hub converges to $p \approx 0.4$.

We hope to make progress on this problem in the near future. We also hope to investigate the grid graph topology.

Approximations of the system: Linearization

Since the convergence analysis for the above non-linear dynamical system seems untenable, one approach would be to look at simpler approximations to the system. One such approximation is a linearization of the system.

Consider the given non-linear dynamical system

$$\vec{P}_t = g(\vec{P}_{t-1})$$

where, $g(\vec{p}) = 1 - (1-(1-\delta)p_i)\prod_j(1-\beta A_{ji}p_j)$.

One method of linearizing g is to look at its first order Taylor expansion around a point $\vec{q}$.

$$\bar{g}(\vec{p}) = g(\vec{q}) + [\nabla g(\vec{q})](\vec{p} - \vec{q})$$

where,

$$[\nabla g(\vec{q})]_{i,i} = (1-\delta)\prod_{j}(1-\beta A_{ji}q_j)$$

$$[\nabla g(\vec{q})]_{i,j} = \beta A_{ji}(1-(1-\delta)q_i)\prod_{k\neq j}(1-\beta A_{ki}q_{ki})$$

Through a shifting of origin if necessary, we can assume that the linear operator defined above is homogeneous, i.e. we get a matrix B s.t. $\bar{g}(\vec{p}) = B\vec{p}$.
This operator has the following convergence property: The operator converges to the origin, irrespective of the starting point, if and only if all its eigenvalues are less than one in the absolute value.
The Taylor expansion around $\vec{q} = \vec{0}$, gives

$$\bar{g}(\vec{p}) = [\beta A + (1-\delta)I]\vec{p}$$

Denoting the eigenvalues of A by $\lambda_{i,A}$, and following the proof of theorem 3 of Chakrabarti et al [1], we can show that the above condition reduces to,

$$\frac{\beta}{\delta} < \frac{1}{\max_i |\lambda_{i,A}|}$$

Noting that the adjacency matrix A is non-negative it follows from the Perron-Frobenius theorem that the $\max_i |\lambda_{i,A}| = \lambda_{1,A}$.

Thus, the necessary and sufficient condition for the operator to converge to the origin is that

$$\frac{\beta}{\delta} < \frac{1}{\lambda_{1,A}}$$

This is the same condition as that for the asymptotic stability of $\vec{0}$ as a fixed point for the original non-linear dynamical system [Theorems 3 and 4 of Chakrabarti et al [1]].

Continuous variant of the dynamical system

Note that the dynamical system considered above is discrete in nature. Due to the roadblocks we faced in analyzing global convergence, and the larger literature for continuous dynamical systems, we also looked at the following continuous variant of the Chakrabarti et al's discrete system:

$$\frac{\partial p_{i,t}}{\partial t} = -\delta p_{i,t}\zeta_{i,t} + (1-p_{i,t})(1-\zeta_{i,t}) \tag{26}$$

or, in more compact form, $p' = f(p)$, where $f(p) = -\delta p_{i,t}\zeta_i + (1-p_i)(1-\zeta_i)$

As before, δ is the curing rate, $p_{i,t}$ is the probability of node i being infected at time t and $\zeta_{i,t}$ is the probability that node i is not infected by its neighbours at time t, which assuming an infection rate of β, is given by

$$\zeta_{i,t} = \prod_{j \in N(i)} (1 - \beta p_{j,t}) = \prod_j (1 - \beta A_{ji} p_{j,t})$$

where $N(i)$ is the set of neighbours of node i in the graph, and A_{ji} is the (j,i)th element of the adjacency matrix A. This generalizes the corresponding dynamical equation of Pastor-Satorras et al,

$$\frac{\partial \rho_k(t)}{\partial t} = -\rho_k(t) + (1 - \rho_k(t))\lambda k \Theta(\lambda) \tag{27}$$

There, the dynamics of the system is expressed in terms of $\rho_k(t)$, the fraction of infected nodes among all nodes with degree k, at time t. Here, it is expressed in terms of $p_{i,t}$, the probability of a single node i being infected.

An important task is that of finding the set of equilibrium points of the system. A point $\vec{p} = \{p_1, \ldots, p_n\}$ is an equilibrium point of equation 26, if it satisfies $f(\vec{p}) = [\frac{\partial p_{i,t}}{\partial t}]|_{\vec{p}} = 0$. In other words,

$$\begin{aligned} \delta p_i \zeta_i &= (1-p_i)(1-\zeta_i) \\ p_i &= \frac{1-\zeta_i}{1-(1-\delta)\zeta_i}, \qquad i = 1, \ldots, n \end{aligned} \tag{28}$$

The equations in (28) give the set of fixed-points, though unfortunately they are not in closed form.

One trivial solution of the above is $\vec{p} = \mathbf{0}$.

An important question is whether a particular equilibrium state is asymptotically stable, that is, if the dynamical system gets close to the equilibrium state, it gets pushed even closer to the equilibrium state. According to dynamical systems theory, a sufficient condition is:

Theorem 1 *Given the dynamical system $p' = f(p)$, and a point $\bar{p}$ in the domain of the system, suppose every eigenvalue of $\nabla f(\bar{p})$ has a real part less than $-c, c > 0$. Then $\bar{p}$ is asymptotically stable, and there is a neighbourhood U of $\bar{p}$ such that for any solution $p(t)$ of the system that starts at any point in U (i.e. $p(0) \in U$), we have that $p(t) \to \bar{p}$ as $t \to \infty$.*
A proof is shown in [2], pages 181-182.

The gradient matrix $[\nabla f(p)]$ is given by

$$[\nabla f(p)]_{ii} = -\delta - 1 + \prod_j (1 - \beta A_{ji} p_{ji})$$

$$[\nabla f(p)]_{ij} = \beta A_{ji}(1-p_i)\prod_{k\neq j}(1-\beta A_{ki}p_{ki}) \tag{29}$$

We are still trying to see if the condition for this matrix of eigenvalues having negative real parts can be simplified further.

For illustration, consider the equilibrium point $\bar{p} = 0$. From 29, $[\nabla f(0)]_{ii} = -\delta$, and $[\nabla f(0)]_{ij} = \beta A_{ij}$. Thus, $B = \nabla f(0) = \beta A - \delta I$.

Denoting the eigenvalues of B by $\lambda_{i,B}$ and those of A by $\lambda_{i,A}$, we observe: $\lambda_{i,B} = \beta\lambda_{i,A} - \delta, \forall i$. The sufficient condition for asymptotic stability from the theorem is

$$\lambda_{i,B} < 0, \forall i$$

This gives,

$$\frac{\beta}{\delta} < \frac{1}{\max_i \lambda_{i,A}}$$

As A is a symmetric real matrix, by the Perron-Frobenius theorem, $\lambda_{1,A} = \max_i \lambda_{i,A}$, and hence we get the following sufficient condition for $p = 0$ being an asymptotically stable equilibrium point:

$$\frac{\beta}{\delta} < \frac{1}{\lambda_{1,A}}$$

This was precisely the condition derived for the *discrete* dynamical system of Chakrabarti et al [1], and thus acts as a sanity check.

An important question is whether the system would converge to *some* fixed state given a particular initial state p_0. For this we need a global convergence theory for continuous non-linear dynamical systems; and we are in the process of obtaining a working understanding of the same from [6].

4 Experiments

This is a theoretical project, but we plan to simulate the system on general random graphs. This would provide the necessary sanity checks in the event we are able to derive analytical expressions for the state of the system, and the epidemic thresholds.

5 Plan of Activities

We plan to study the global convergence theory for continuous dynamical systems in more detail, as covered in chapters 9 and 13 of [2], and chapter 3 of [6]. We also plan to do a more thorough study of fixed-point theory for continuous maps to see if we can derive more tractable conditions under which

Chakrabarti et al's discrete dynamical system converges. Along with working on the general case, we plan to rigorously answer the specified questions for specific simple graphs, in particular the star graph.

References

[1] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. *Journal Version, in preparation*, 2005.

[2] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Chapters 8,9,13. Academic Press, 1974.

[3] Jeffrey O. Kephart and Steve R. White. Directed-graph epidemological models of computer viruses. *IEEE*, pages 343–359, 1991.

[4] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters, 86(14), pp. 3200–3203*, 2001.

[5] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E. 65, 035108(R) (2002)*, 2002.

[6] Lawrence Perko. *Differential Equations and Dynamical Systems*. Chapter 3. Springer, 2000.

[7] W. Rudin. Principles of mathematical analysis. 1976.

[8] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. *22nd Symposium on Reliable Distributed Computing (SRDS2003)*, 2003.