

Virus Propagation on Time-Varying Networks: Theory and Immunization Algorithms

B. Aditya Prakash^{*}, Hanghang Tong^{*}, Nicholas Valler⁺, Michalis Faloutsos⁺,
and Christos Faloutsos^{*}

^{*}Computer Science Department, Carnegie Mellon University
{badityap,htong,christos}@cs.cmu.edu
⁺Department of Computer Science, University of California - Riverside
{nvaller,michalis}@cs.ucr.edu

Abstract. Given a contact network that changes over time (say, day vs night connectivity), and the SIS (susceptible/infected/susceptible, flu like) virus propagation model, what can we say about its epidemic threshold? That is, can we determine when a small infection will “take-off” and create an epidemic? Consequently then, which nodes should we immunize to prevent an epidemic? This is a very real problem, since, e.g. people have different connections during the day at work, and during the night at home. Static graphs have been studied for a long time, with numerous analytical results. Time-evolving networks are so hard to analyze, that most existing works are simulation studies [5].

Specifically, our contributions in this paper are: (a) we formulate the problem by approximating it by a Non-linear Dynamical system (NLDS), (b) we derive the **first** closed formula for the epidemic threshold of time-varying graphs under the SIS model, and finally (c) we show the usefulness of our threshold by presenting efficient heuristics and evaluate the effectiveness of our methods on synthetic and real data like the MIT reality mining graphs.

1 Introduction

The goal of this work is to analytically study the epidemic spread on time-varying graphs. We focus on time-varying graphs that follow an alternating connectivity behavior, which is motivated by the day-night pattern of human behavior. Note that our analysis is not restricted to two graphs: we can have an arbitrary number of alternating graphs. Furthermore, we focus on the SIS model [19], which

This material is based upon work supported by the Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053, the National Science Foundation under Grants No. CNS-0721736 and CNS-0721889 and a Sprint gift. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, the U.S. Government, the National Science Foundation, or other funding parties. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

resembles a flu-like virus, where healthy nodes get the virus stochastically from their infected neighbors, and infected nodes get cured with some probability and become susceptible again. The SIS model can be also used in modeling many different types of dynamical processes as well, for example, modeling product penetration in marketing [36].

More specifically, the inputs to our problem are: (a) a set of T alternating graphs, (b) the infectivity of the virus and the recovery rate (β, δ for the SIS model), (d) k number of “vaccinations”. We want to answer two questions (rigorously defined in Section 3):

- Q1. Can we say whether a small infection can “take-off” and create an epidemic under the SIS model (i.e. determine the so-called epidemic threshold)?
- Q2. What is an effective and fast way to vaccinate people to minimize the spread of the virus?

While epidemic spreading on static graphs has been studied extensively (e.g. see [19, 2, 33, 8]), virus propagation on time-varying graphs has received little attention. Moreover, most previous studies on time-varying graphs use only simulations [5]. We review in more detail the previous efforts in Section 2.

We are arguably the first to study virus propagation *analytically* on arbitrary, and *time-varying* graphs. In more detail, the contributions of our work can be summarized in the following points:

1. We formulate the problem, and show that it can be approximated with a *Non-Linear Dynamical System* (NLDS).
2. We give the **first** closed-formula for the epidemic threshold, involving the first eigenvalue of the so-called *system-matrix* (see Theorem 2). The system-matrix combines the connectivity information (the alternating adjacency matrices) and the characteristics of the virus (infectivity and recovery rate).
3. We show the importance of our threshold by using it to develop and evaluate several immunization policies on real data like the MIT Reality Mining graph.

The rest of the paper is organized as follows: We review related work in Section 2, explain the formal problem definitions in Section 3, and describe the proofs for the threshold and illustrate the theorem in Section 4. We then discuss various immunization policies in Section 5 and present experimental evaluations in Section 6. We discuss and provide additional explanations in Section 7 and finally conclude in Section 8.

2 Related Work

In this section, we review the related work, which can be categorized into three parts: epidemic threshold, immunization algorithms and general graph mining.

Epidemic Thresholds. The class of epidemiological models that are most widely used are the so-called *homogeneous models* [3, 29, 2]. A homogeneous

model assumes that every individual has equal contact to others in the population, and that the rate of infection is largely determined by the density of the infected population. Kephart and White [22, 23] were among the first to propose epidemiology-based models (hereafter referred to as the KW model) to analyze the propagation of computer viruses. The KW model provides a good approximation of virus propagation in networks where the contact among individuals is sufficiently homogeneous. However, there is overwhelming evidence that real networks (including social networks [10], router and AS networks [12], and Gnutella overlay graphs [35]) deviate from such homogeneity - they follow a power law structure instead.

Pastor-Satorras and Vespignani studied viral propagation for such power-law networks [32, 33]. They developed an analytic model for the Barabási-Albert (BA) power-law topology [4]. However, their derivation depends on some assumptions which does not hold for many real networks [25, 12]. Pastor-Satorras et al. [33] also proposed an epidemic threshold condition, but this uses the “mean-field” approach, where all graphs with a given degree distribution are considered equal. There is no particular reason why all such graphs should behave similarly in terms of viral propagation. Newman [31] studied the epidemic thresholds for multiple competing viruses on special, *random* graphs.

Immunization. Briesemeister et al. [7] focus on immunization of power law graphs. They focus on the random-wiring version (that is, standard preferential attachment), versus the “highly clustered” power law graphs. Their simulation experiments on such synthetic graphs show that such graphs can be more easily defended against viruses, while random-wiring ones are typically overwhelmed, despite identical immunization policies.

Cohen et al. [9] studied the *acquaintance* immunization policy (see Section 5 for a description of this policy), and showed that it is much better than random, for both the SIS as well as the SIR model. For power law graphs (with no rewiring), they also derived formulae for the critical immunization fraction, above which the epidemic is arrested. Madar et al. [27] continued along these lines, mainly focusing on the SIR model for scale-free graphs. They linked the problem to bond percolation, and derived formulae for the effect of several immunization policies, showing that the “acquaintance” immunization policy is the best. Both works were analytical, without studying any real graphs.

Hayashi et al. [18] study the case of a growing network, and they derive analytical formulas for such power law networks (no rewiring). They introduce the SHIR model (Susceptible, Hidden, Infectious, Recovered), to model computers under e-mail virus attack and derive the conditions for extinction under random and under targeted immunization, always for power law graphs with no rewiring.

Thus, none of the earlier related work focus on epidemic thresholds for *arbitrary, real* graphs, with only exceptions of [37, 8], and its follow-up paper by Ganesh et al. [13]. However, even these works [37, 8, 13] assume that the underlying graph is fixed, which is unrealistic in many applications.

General Graph Mining. Graph mining is a very active research area in recent years. Representative works include patterns and “laws” discovery e.g.,

power law distributions [12, 26], small world phenomena [30, 1], and numerous other regularities. Among them, there is a lot of research interest in studying dynamic processes on large graphs, (a) blogs and propagations [17, 24, 21, 34], (b) information cascades [6, 14, 16] and (c) marketing and product penetration [36]. These dynamic processes are all closely related to virus propagation.

In sum, to the best of our knowledge, including comprehensive epidemiological texts [2, 3] and well-cited surveys [19], we are the **first** to analytically study virus propagation on *arbitrary, real* and *time-varying* graphs.

3 Problem Definitions

Table 1 lists the main symbols used in the paper. Following standard notation, we use capital bold letters for matrices (e.g. \mathbf{A}), lower-case bold letters for vectors (e.g. \mathbf{a}), and calligraphic fonts for sets (e.g. \mathcal{S}) and we denote the transpose with a prime (i.e., \mathbf{A}' is the transpose of \mathbf{A}). In this paper, we focus on un-directed un-weighted graphs which we represent by the adjacency matrix.

Also we deal only with the SIS virus propagation model in the paper. The SIS model is the susceptible/infected/susceptible virus model where β is the probability that an infected node will transmit the infection over a link connected to a neighbor and δ is the probability with which an infected node cures itself and becomes susceptible again. Please see [19] for a detailed discussion on SIS and other virus models.

Consider a setting with clearly different behaviors say, day/night, each characterized by a corresponding adjacency matrix (\mathbf{A}_1 for day, \mathbf{A}_2 for night), then what is the epidemic threshold under a SIS virus model? What are the best nodes to immunize to prevent an epidemic as much as possible? More generally, the problems we are tackling can be formally stated as follows:

Problem 1. Epidemic Threshold

Given: (1) T alternating behaviors, characterized by a set of T graphs $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2 \dots \mathbf{A}_T\}$; and (2) the SIS model [8] with virus parameters β and δ ;
Find: A condition, under which the infection will die out exponentially quickly (regardless of initial condition).

Problem 2. Immunization

Given: (1) T alternating behaviors, characterized by a set of T graphs $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2 \dots, \mathbf{A}_T\}$; and (2) the SIS model with virus parameters β and δ and (3) k vaccines;
Find: The best- k nodes for immunization.

We will next solve Problem 1 while we discuss Problem 2 later in Section 5.

Table 1. Symbols

Symbol	Definition and Description
A, B, ...	matrices (bold upper case)
$\mathbf{A}(i, j)$	element at the i^{th} row and j^{th} column of A
$\mathbf{A}(i, :)$	i^{th} row of matrix A
$\mathbf{A}(:, j)$	j^{th} column of matrix A
I	standard $n \times n$ identity matrix
a, b, ...	column vectors
$\mathcal{I}, \mathcal{J}, \dots$	sets (calligraphic)
n	number of nodes in the graphs
T	number of different alternating behaviors
$\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_T$	T corresponding size $n \times n$ symmetric alternating adjacency matrices
β	virus transmission probability in the SIS model
δ	virus death probability in the SIS model
$\lambda_{\mathbf{M}}$	first eigen-value (in absolute value) of a matrix M
$\mathbf{u}_{\mathbf{M}}$	corresponding first eigen-vector (for $\lambda_{\mathbf{M}}$) of a matrix M
$p_{i,t}$	probability that node i is infected at time t
\mathbf{p}_t	$\mathbf{p}_t = (p_{1,t}, p_{2,t}, \dots, p_{n,t})'$
\mathbf{p}_{2t+1}	probability of infection vector for odd days
\mathbf{p}_{2t}	probability of infection vector for even days
η_t	the expected number of infected nodes at time t

4 Epidemic Threshold on Time-varying Graphs

To simplify discussion, we consider $T = 2$ in Problem 1 with \mathcal{A} to consist of only two graphs: G_1 with the adjacency matrix \mathbf{A}_1 for the odd time-stamps (the ‘days’) and G_2 with the adjacency matrix \mathbf{A}_2 for the even time-stamps (the ‘nights’). Our proofs and results can be naturally extended to handle any arbitrary sequence of T graphs.

4.1 The NLDS

We first propose to approximate the infection dynamics by a Non-linear dynamical system (NLDS) representing the evolution of the probability of infection vector (\mathbf{p}_t) over time. We can compute the probability $\zeta_t(i)$ that node i does not receive any infections at time t . A node i won’t receive any infection if either any given neighbor is not infected or it is infected but fails to transmit the infection with probability $1 - \beta$. Assuming that the neighbors are *independent*, we get:

$$\begin{aligned}
 \zeta_{2t+1}(i) &= \prod_{j \in \mathcal{N}_{\mathcal{E}_1}(i)} (p_{j,2t+1}(1 - \beta) + (1 - p_{j,2t+1})) \\
 &= \prod_{j \in \{1..n\}} (1 - \beta \mathbf{A}_1(i, j) p_{j,2t+1})
 \end{aligned} \tag{1}$$

where $\mathcal{NE}_1(i)$ is the set of neighbors of node i in the graph G_1 with adjacency matrix \mathbf{A}_1 . Similarly, we can write $\zeta_{2t+2}(i)$ as:

$$\begin{aligned}\zeta_{2t}(i) &= \prod_{j \in \mathcal{NE}_2(i)} (p_{j,2t}(1-\beta) + (1-p_{j,2t})) \\ &= \prod_{j \in \{1..n\}} (1 - \beta \mathbf{A}_2(i, j) p_{j,2t})\end{aligned}\quad (2)$$

So, $p_{i,2t+1}$ and $p_{i,2t+2}$ are:

$$\begin{aligned}1 - p_{i,2t+1} &= \delta p_{i,2t} + (1 - p_{i,2t}) \zeta_{2t}(i) \\ \Rightarrow p_{i,2t+1} &= 1 - \delta p_{i,2t} - (1 - p_{i,2t}) \zeta_{2t}(i)\end{aligned}\quad (3)$$

and

$$\begin{aligned}1 - p_{i,2t+2} &= \delta p_{i,2t+1} + (1 - p_{i,2t+1}) \zeta_{2t+1}(i) \\ \Rightarrow p_{i,2t+2} &= 1 - \delta p_{i,2t+1} - (1 - p_{i,2t+1}) \zeta_{2t+1}(i)\end{aligned}\quad (4)$$

Note that we can write our NLDS as:

$$\mathbf{p}_{2t+1} = \mathbf{g}_2(\mathbf{p}_{2t}) \quad (5)$$

$$\mathbf{p}_{2t+2} = \mathbf{g}_1(\mathbf{p}_{2t+1}) \quad (6)$$

where \mathbf{g}_1 and \mathbf{g}_2 are corresponding non-linear functions as defined by Equations 3 and 4 (\mathbf{g}_1 depends only on \mathbf{A}_1 and \mathbf{g}_2 on \mathbf{A}_2).

We have the following theorem about the asymptotic stability of a NLDS at a fixed point:

Theorem 1. (Asymptotic Stability, e.g. see [20]) *The system given by $\mathbf{p}_{t+1} = \mathbf{g}(\mathbf{p}_t)$ is asymptotically stable at an equilibrium point \mathbf{p}^* , if the eigenvalues of the Jacobian $J = \nabla \mathbf{g}(\mathbf{p}^*)$ are less than 1 in absolute value, where,*

$$J_{k,l} = [\nabla \mathbf{g}(\mathbf{p}^*)]_{k,l} = \left. \frac{\partial p_{k,t+1}}{\partial p_{l,t}} \right|_{\mathbf{p}_t = \mathbf{p}^*}$$

The fixed point of our interest is the $\mathbf{0}$ vector which is the state when all nodes are susceptible and not infected. We want to then analyze the stability of our NLDS at $\mathbf{p}_{2t} = \mathbf{p}_{2t+1} = \mathbf{0}$. From Equations 5 and 6, we get:

$$\left. \frac{\partial \mathbf{p}_{2t+2}}{\partial \mathbf{p}_{2t+1}} \right|_{\mathbf{p}_{2t+1} = \mathbf{0}} = (1 - \delta) \mathbf{I} + \beta \mathbf{A}_1 = \mathbf{S}_1 \quad (7)$$

$$\left. \frac{\partial \mathbf{p}_{2t+1}}{\partial \mathbf{p}_{2t}} \right|_{\mathbf{p}_{2t} = \mathbf{0}} = (1 - \delta) \mathbf{I} + \beta \mathbf{A}_2 = \mathbf{S}_2 \quad (8)$$

Any eigenvalue $\lambda_{\mathbf{S}_1}^i$ of \mathbf{S}_1 and $\lambda_{\mathbf{S}_2}^i$ of \mathbf{S}_2 ($i = 1, 2, \dots$) is related to the corresponding eigenvalue $\lambda_{\mathbf{A}_1}^i$ of \mathbf{A}_1 and $\lambda_{\mathbf{A}_2}^i$ of \mathbf{A}_2 as:

$$\lambda_{\mathbf{S}_1}^i = (1 - \delta) + \beta \lambda_{\mathbf{A}_1}^i \quad (9)$$

$$\lambda_{\mathbf{S}_2}^i = (1 - \delta) + \beta \lambda_{\mathbf{A}_2}^i \quad (10)$$

Recall that as \mathbf{A}_1 and \mathbf{A}_2 are symmetric real matrices (the graphs are undirected), from the Perron-Frobenius theorem [28], $\lambda_{\mathbf{A}_1}$ and $\lambda_{\mathbf{A}_2}$ are real and positive. So, from Equations 9 and 10 $\lambda_{\mathbf{S}_1}$ and $\lambda_{\mathbf{S}_2}$ are also real and positive.

4.2 The Threshold

We are now in a position to derive the epidemic threshold. First, we have the following lemma:

Lemma 1. *If $\lambda_{\mathbf{S}} < 1$, then \mathbf{p}_{2t} dies out exponentially quickly; and $\mathbf{0}$ is asymptotically stable for \mathbf{p}_{2t} , where $\mathbf{S}_1 = (1 - \delta)\mathbf{I} + \beta\mathbf{A}_1$, $\mathbf{S}_2 = (1 - \delta)\mathbf{I} + \beta\mathbf{A}_2$ and $\mathbf{S} = \mathbf{S}_1 \times \mathbf{S}_2$.*

Proof. Since $\mathbf{p}_{2t+2} = g_1(g_2(\mathbf{p}_{2t}))$ (from Equations 5 and 6), we have

$$\begin{aligned} \frac{\partial \mathbf{p}_{2t+2}}{\partial \mathbf{p}_{2t}} \Big|_{\mathbf{p}_{2t}=\mathbf{0}} &= \left(\frac{\partial \mathbf{p}_{2t+2}}{\partial \mathbf{p}_{2t+1}} \times \frac{\partial \mathbf{p}_{2t+1}}{\partial \mathbf{p}_{2t}} \right) \Big|_{\mathbf{p}_{2t}=\mathbf{0}} \\ &= \left(\frac{\partial \mathbf{p}_{2t+2}}{\partial \mathbf{p}_{2t+1}} \Big|_{\mathbf{p}_{2t+1}=\mathbf{0}} \right) \times \left(\frac{\partial \mathbf{p}_{2t+1}}{\partial \mathbf{p}_{2t}} \Big|_{\mathbf{p}_{2t}=\mathbf{0}} \right) \\ &= \mathbf{S}_1 \mathbf{S}_2 = \mathbf{S} \end{aligned} \quad (11)$$

The first equation is due to chain-rule, second equation is because $\mathbf{p}_{2t} = \mathbf{0}$ implies $\mathbf{p}_{2t+1} = \mathbf{0}$; and the final step is due to Equations 7 and 8.

Therefore, using Theorem 1, we get that if $\lambda_{\mathbf{S}} < 1$, we have that $\mathbf{0}$ is asymptotically stable for \mathbf{p}_{2t} .

We now prove that \mathbf{p}_{2t} in fact goes down exponentially to $\mathbf{0}$ if $\lambda_{\mathbf{S}} < 1$. To see this, after linearizing both g_1 and g_2 at $\mathbf{p}_{2t} = \mathbf{p}_{2t+1} = \mathbf{0}$, we have

$$\begin{aligned} \mathbf{p}_{2t+2} &\leq \mathbf{S}_1 \mathbf{p}_{2t+1} \\ \mathbf{p}_{2t+1} &\leq \mathbf{S}_2 \mathbf{p}_{2t} \end{aligned} \quad (12)$$

Doing the above recursively, we have

$$\mathbf{p}_{2t} \leq (\mathbf{S}_1 \mathbf{S}_2)^t \mathbf{p}_0 = (\mathbf{S})^t \mathbf{p}_0 \quad (13)$$

Let η_t be the expected number of infected nodes at time t . Then,

$$\begin{aligned} \eta_{2t} = |\mathbf{p}_{2t}|_1 &\leq |(\mathbf{S})^t \mathbf{p}_0|_1 \\ &\leq |(\mathbf{S})^t|_1 |\mathbf{p}_0|_1 = |(\mathbf{S})^t|_1 \eta_0 \\ &\leq \sqrt{n} |(\mathbf{S})^t|_2 \eta_0 = \sqrt{n} \lambda_{\mathbf{S}}^t \eta_0 \end{aligned} \quad (14)$$

Therefore, if $\lambda_{\mathbf{S}} < 1$, we have that η_{2t} goes to zero exponentially fast. \square

The above lemma provides the condition for the even time-stamp probability vector to go down exponentially. But, the next lemma shows that this condition is enough to ensure that even the odd time-stamp probability vector to go down exponentially.

Lemma 2. *If $\lambda_{\mathbf{S}} < 1$, then \mathbf{p}_{2t+1} dies out exponentially quickly; and $\mathbf{0}$ is asymptotically stable for \mathbf{p}_{2t+1} , where $\mathbf{S}_1 = (1 - \delta)\mathbf{I} + \beta\mathbf{A}_1$, $\mathbf{S}_2 = (1 - \delta)\mathbf{I} + \beta\mathbf{A}_2$ and $\mathbf{S} = \mathbf{S}_1 \times \mathbf{S}_2$.*

Proof. Doing the same analysis as in Lemma 1, we can see that the condition for \mathbf{p}_{2t+1} to be asymptotically stable and die exponentially quickly is:

$$\lambda_{\mathbf{S}_2 \times \mathbf{S}_1} < 1 \quad (15)$$

Now note that as \mathbf{S}_1 and \mathbf{S}_2 are invertible: $\mathbf{S}_1 \times \mathbf{S}_2 = \mathbf{S}_1 \times \mathbf{S}_2 \times \mathbf{S}_1 \times \mathbf{S}_1^{-1}$. But this implies that $\mathbf{S}_2 \times \mathbf{S}_1$ is similar to $\mathbf{S}_1 \times \mathbf{S}_2$ (matrix \mathbf{P} is similar to \mathbf{Q} if $\mathbf{P} = \mathbf{B}\mathbf{Q}\mathbf{B}^{-1}$, for some invertible \mathbf{B}). We know that similar matrices have the same spectrum [15], thus $\mathbf{S}_2 \times \mathbf{S}_1$ and $\mathbf{S}_1 \times \mathbf{S}_2$ have the same eigenvalues. Hence, the condition for exponential die out of \mathbf{p}_{2t+1} and asymptotic stability is the same as that for \mathbf{p}_{2t} which is $\lambda_{\mathbf{S}} < 1$. \square

Lemma 1 and Lemma 2 imply that this threshold is well-defined in the sense that the probability vector for both the odd and even time-stamps go down exponentially. Thus we can finally conclude the following theorem:

Theorem 2. (Epidemic Threshold) *If $\lambda_{\mathbf{S}} < 1$, then \mathbf{p}_{2t} and \mathbf{p}_{2t+1} die out exponentially quickly; and $\mathbf{0}$ is asymptotically stable for both \mathbf{p}_{2t} and \mathbf{p}_{2t+1} , where $\mathbf{S}_1 = (1 - \delta)\mathbf{I} + \beta\mathbf{A}_1$, $\mathbf{S}_2 = (1 - \delta)\mathbf{I} + \beta\mathbf{A}_2$ and $\mathbf{S} = \mathbf{S}_1 \times \mathbf{S}_2$. Similarly for any general T , the condition is:*

$$\boxed{\lambda_{\prod_i \mathbf{S}_i} < 1} \quad (16)$$

where $\forall i \in \{1, 2, \dots, T\}$ $\mathbf{S}_i = (1 - \delta)\mathbf{I} + \beta\mathbf{A}_i$.

We call \mathbf{S} as the *system-matrix* of the system; thus, the first eigenvalue of the system-matrix determines whether a given system is below threshold or not.

4.3 Salient Points

Sanity check: Clearly, when $T = 1$, the system is equivalent to a static graph system with \mathbf{A}_1 and virus parameters β, δ . In this case the threshold is (from Theorem 2) $\lambda_{(1-\delta)\mathbf{I}+\beta\mathbf{A}_1} < 1 \Rightarrow \beta\lambda_{\mathbf{A}_1}/\delta < 1$ i.e. we recover the known threshold in the static case [8].

A conservative condition: Notice that from Equations 7 and 8 and Theorem 1, for our NLDS to be fully asymptotically stable at $\mathbf{0}$ (i.e. \mathbf{p}_t decays monotonically), we need the eigenvalues of both \mathbf{S}_1 and \mathbf{S}_2 be less than 1 in absolute value. Hence, $\beta\lambda/\delta < 1$ where $\lambda = \max(\lambda_{\mathbf{A}_1}, \lambda_{\mathbf{A}_2})$ is sufficient for full stability. Intuitively, this argument says that the alternating sequence of graphs can not be worse than static case of having the best-connected graph of the two repeated indefinitely. Let $\lambda_{\mathbf{A}_1} > \lambda_{\mathbf{A}_2}$. Consider a sequence of graphs $\mathcal{S} = \{\mathbf{A}_1, \mathbf{A}_1 \dots\}$ repeating indefinitely instead of our alternating $\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_1, \mathbf{A}_2, \dots\}$ sequence. Clearly, if an infection dies exponentially in \mathcal{S} , then it will die exponentially in our original alternating sequence as well because $\lambda_{\mathbf{A}_1} > \lambda_{\mathbf{A}_2}$. The set \mathcal{S} is essentially just the static graph case: hence, if $\beta\lambda_{\mathbf{A}_1}/\delta = \beta\lambda/\delta < 1$, then $\mathbf{0}$

is asymptotically stable for \mathbf{p}_t . The case when $\lambda_{\mathbf{A}_1} < \lambda_{\mathbf{A}_2}$ is similar. But this notion of a threshold is too stringent and *conservative*: it can happen that a stronger virus can still lead to a general exponential decrease instead of a strict monotonous decrease. This is because we forced the eigenvalues of *both* \mathbf{S}_1 and \mathbf{S}_2 to be less than 1 in absolute value here, when we can probably get away with less. Theorem 2 precisely formalizes this idea and gives us a more practical condition for a general decreasing trend of every corresponding alternating time-stamp values decaying. We illustrate this further in the experiments.

4.4 Experiments

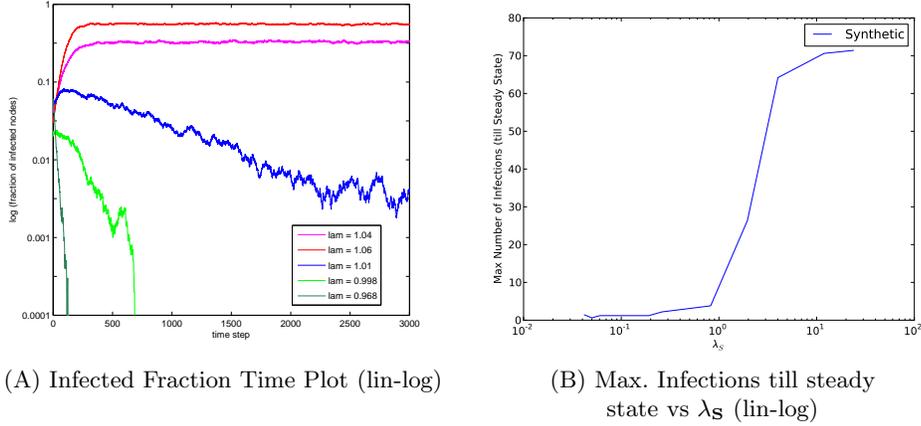


Fig. 1. SIS simulations on our synthetic example (all values averages over 20 runs) (A) Fraction of nodes infected vs Time-stamp (lin-log scale). Note the qualitative difference in behavior *under* (green) and *above* (red) the threshold. Also, note that the green line is *below* the threshold but is actually *above* the conservative threshold ($\beta\lambda/\delta = 1.100$ here). Hence while both \mathbf{p}_{2t} and \mathbf{p}_{2t+1} decrease exponentially separately, but \mathbf{p}_t itself does not monotonously go down. (B) Plot of Max. number of infected nodes till steady state vs λ_S (by varying β) (lin-log). As predicted by our results, notice that there is a sudden ‘take-off’ and a change of behavior of the curve exactly when $\lambda_S = 1$.

Figures 1 and 2 demonstrate our result on a synthetic example and graphs from MIT reality data (more details on the reality mining graphs are in Section 6). In the synthetic example, we have 100 nodes, such that G_1 is a full clique (without self loops) whereas G_2 is a chain. All values are average over several runs of the simulations and the infection is started by infecting 5 nodes. In short, as expected from the theorem, the difference in behavior above, below and at threshold can be distinctly seen in the figures.

Figures 1(A) and 2(A) show the time-plot of number of infections for λ_S values above and below the threshold. While above threshold the infection reaches a steady state way above the starting point, below threshold it decays fast and

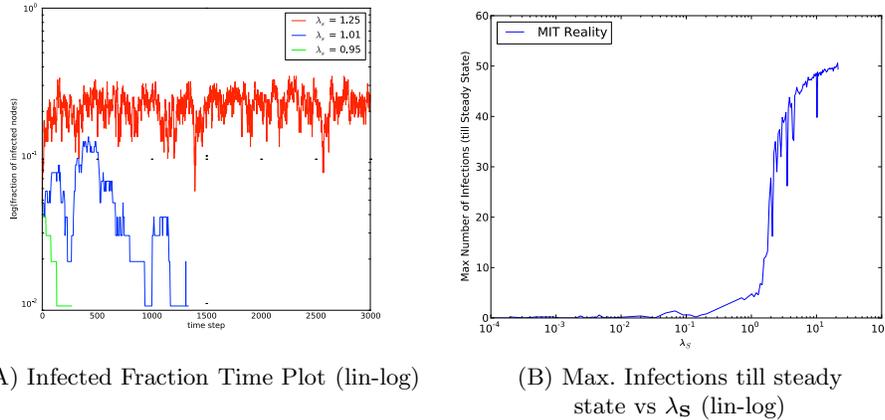


Fig. 2. SIS simulations on the MIT reality mining graphs (all values averages over 20 runs) (A) Fraction of nodes infected vs Time-stamp (lin-log scale). Note the qualitative difference in behavior *under* (green) and *above* (red) the threshold. (B) Plot of Max. number of infected nodes till steady state vs λ_S (by varying β) (lin-log). As predicted by our results, notice that there is a sudden ‘take-off’ and a change of behavior of the curve when $\lambda_S = 1$.

dies out. In case of Figure 1(A), also note the the difference between the conservative threshold and our threshold. The green curve is *below* our threshold but *above* the conservative threshold. But again, as predicted from our theorems, clearly while there are dampening oscillations and the infection decays but \mathbf{p}_t itself does not monotonously go down (and hence the “bumpy” nature of the curve). This exemplifies the practical nature of our threshold and its usefulness as we are more concerned with the general trend of the number of infections curve and not every small ‘bump’ because of the presence of alternating graphs.

Figures 1(B) and 2(B) show a ‘take-off’ plot showing max. number of infections till steady state (intuitively the ‘footprint’) for different values of λ_S (by varying β). As predicted by our theorem, note the sudden steep change and spike in the size of the footprint when $\lambda_S = 1$ in both the plots.

5 Immunization Algorithms

Given the theoretical results in the previous section, can we exploit them to our advantage to ensure effective immunization (Problem 2)?

5.1 Quality Metric

Using our results, we can evaluate the quality of *any* immunization policy. Note that smaller the value of λ_S , lesser are chances of the virus causing any epidemic. Put differently, we want to decrease the λ_S value of the system as much as

possible. Thus, the efficacy of any immunization policy should be measured using the amount of “drop” in $\lambda_{\mathbf{S}}$ it causes and the resulting $\lambda_{\mathbf{S}}$ after immunization.

5.2 Proposed immunization policies

We now discuss some new immunization policies for time-varying graphs, partially motivated by known policies used for static graphs. Again, for ease of exposition we focus our attention only on the $\{\mathbf{A}_1, \mathbf{A}_2\}$ system of Section 3. From the above, it is clear that optimally we should choose that set of k nodes which result in the smallest $\lambda_{\mathbf{S}}$ value possible after immunization. This implies that for each set of k node we test, we need to delete k rows/columns from both \mathbf{A}_1 and \mathbf{A}_2 to get new matrices \mathbf{A}_1^* and \mathbf{A}_2^* and then compute the new $\lambda_{\mathbf{S}}$ value. The number of k sets is $\binom{n}{k}$ and therefore this method is combinatorial in nature and will be intractable even for small graphs. Nevertheless, we call this strategy **Optimal** and show experimental results for this policy too, because this policy will give us the lower-bound on the $\lambda_{\mathbf{S}}$ that can be achieved after k immunizations.

We want policies which are practical for large graphs and at the same time be efficient in lowering the $\lambda_{\mathbf{S}}$ value of the system i.e. they should be close to **Optimal**. Specifically to this effect, we now present several greedy policies as well. As the heuristics are greedy in nature, we only describe how to pick the best *one* node for immunization from a given set of G_1 and G_2 graphs. Our proposed policies are:

Greedy-DmaxA (Highest degree on \mathbf{A}_1 or \mathbf{A}_2 matrices) Under this policy, at each step we select the node with the highest degree considering both the \mathbf{A}_1 or \mathbf{A}_2 adjacency matrices. This is motivated by the degree immunization strategy used for static graphs.

Greedy-DavgA (Highest degree on the “average” matrix) We select the node with the highest degree in the \mathbf{A}_{avg} matrix where $\mathbf{A}_{\text{avg}} = (\mathbf{A}_1 + \mathbf{A}_2)/2$.

Greedy-AavgA (Acquaintance immunization [9] on the average matrix) The “acquaintance” immunization policy works by picking a random person, and then immunizing one of its neighbors at random (which will probably be a ‘hub’). We run this policy on the \mathbf{A}_{avg} matrix.

Greedy-S (Greedy on the system-matrix) This is the greedy strategy of immunizing the node at each step which causes the largest drop in $\lambda_{\mathbf{S}}$ value. Note that even this can be expensive for large graphs as we have to evaluate the first eigenvalue of \mathbf{S} after deleting each node to decide which node is the best.

Optimal Finally, this is the optimal through combinatorial strategy mentioned above of finding the best- k set of nodes which decrease $\lambda_{\mathbf{S}}$ the most.

6 Experimental Evaluations

In this section we present experimental results of applying the various immunization policies discussed previously. We have already demonstrated our theoretical threshold results in Section 4.

6.1 Experimental Setup

We conducted a series of experiments using the MIT Reality Mining data set [11]. The Reality Mining data consists of 104 mobile devices (cellular phones) monitored over a period of nine months (September 2004 - June 2005). If another participating Bluetooth device was within a range of approximately 5-10 meters, the date and time of the contact and the device’s MAC address were recorded. Bluetooth scans were conducted at 5-minute intervals and aggregated into two 12-hour period adjacency matrices (*day* and *night*). The epidemic simulations were accomplished by alternating the *day* and *night* matrices over the period of simulation. All experiments were run on a 64-bit, quad-core (2.5Ghz each) server running a CentOS linux distribution with shared 72 GB of RAM. Simulations were conducted using a combination of Matlab 7.9 and Python 2.6.

6.2 Results

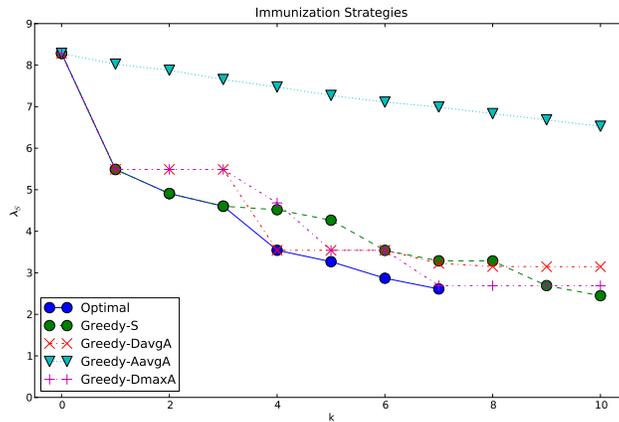


Fig. 3. Experiments on Reality Mining graphs: λ_S vs k for different immunization policies. Lower is better. **Greedy-DavgA** clearly drops the λ_S value aggressively and is close to the **Greedy-Opt**.

Figure 3 shows the λ_S value after immunizing $k = 1, 2, \dots, 10$ nodes using each of the policies outlined in Section 5. As **Optimal** and **Greedy-S** require β and δ as inputs, we set $\beta = 0.5, \delta = 0.1$. Running **Optimal** became prohibitively expensive (> 4 hours on the MIT reality graphs) after $k = 7$ - hence we don’t show data points for $k \geq 8$ for **Optimal**. Moving on to the greedy strategies we find that **Greedy-S** performs the best after $k = 10$ by dropping the λ_S value as aggressively as possible - *equal* to **Optimal** at many places. We find that **Greedy-DavgA** also performs very well. Intuitively this is because the highest degree node in \mathbf{A}_{avg} is very well-connected and hence has a huge effect in reducing the \mathbf{A}_{avg} value (we discuss more about \mathbf{A}_{avg} later in Section 7). At the

same time, **Greedy-DmaxA** is comparable to **Greedy-DavgA** as we find the highest degree among both the graphs: so this highest degree will also mostly have the highest mean degree. Finally, **Greedy-AavgA** drops the λ_S value the least among all the policies. Given the first random choice of node, **Greedy-AavgA** can be “trapped” in the neighborhood of a node far from the best node to immunize, and thus can be forced to make a choice based on the limited local information.

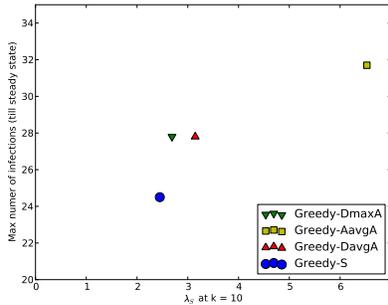


Fig. 4. Scatter plot of Max. infections till steady state and λ_S for different immunization policies after $k = 10$ immunizations. Points closer to the origin are better. All policies perform in accordance to the λ_S values achieved (see Figure 3).

Figure 4 demonstrates the effectiveness of our quality metric i.e. the λ_S value for each immunization policy after $k = 10$ immunizations. It is a scatter plot of Max. infections till steady state and the various λ_S values at the end of the immunizations. So points closer to the origin (minimum footprint and λ_S) represent better policies. Clearly, **Optimal** should theoretically be the closest to the origin (we don’t show it as it didn’t finish). Also as discussed before, **Greedy-AavgA** is the worst and that is demonstrated by its point. From Figure 3 we can see that **Greedy-S** has the least λ_S value after $k = 10$, hence it is closest to the origin and thus has the smallest footprint. Others perform well too, as their final λ_S values were close as well.

To summarize, in our experiments we demonstrated that policies decreasing λ_S the most are the best policies as they result in smaller footprints as well. Also, simple greedy policies were effective and achieved similar λ_S values like expensive combinatorial policies.

7 Discussion

We discuss some pertinent issues and give additional explanations in this section.

Generality of our results: How can we model more complex situations like ‘unequal duration’ behaviors etc.? Note that the alternation period T can be

longer than 2 and we can have *repetitions* in the set \mathcal{A} as well e.g. to represent a weekly-style (work day-weekend) alternation we can have $T = 7$ and $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_1, \dots, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_2\}$. Similarly, we can model situations like unequal duration of ‘day’ and ‘night’ i.e. unequal duration of matrices \mathbf{A}_1 and \mathbf{A}_2 . Say, \mathbf{A}_1 is present for only 8 hours at work, while \mathbf{A}_2 is present for the remaining 16 hours at home. Then, thinking of an hour as our time-step i.e. $T = 24$, the set $\mathcal{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_2, \dots\}$, where \mathbf{A}_1 occurs 8 times in \mathcal{A} while \mathbf{A}_2 occurs 16 times. All the threshold results carry forward seamlessly in all the above cases.

Goodness of the \mathbf{A}_{avg} matrix: We saw that Greedy-DavgA gave very good results and was close to Greedy-S and Optimal. This can be explained with the help of the following lemma.

Lemma 3. $(1 - 2\delta)\mathbf{I} + 2\beta\mathbf{A}_{\text{avg}}$ is a first-order approximation of the \mathbf{S} matrix.

Proof. Note that ($T = 2$),

$$\begin{aligned} \mathbf{S} &= \mathbf{S}_1 \times \mathbf{S}_2 \\ &= ((1 - \delta)\mathbf{I} + \beta\mathbf{A}_1) \times ((1 - \delta)\mathbf{I} + \beta\mathbf{A}_2) \\ &= (1 - \delta)^2\mathbf{I} + (1 - \delta)\beta(\mathbf{A}_1 + \mathbf{A}_2) + \beta^2\mathbf{A}_1\mathbf{A}_2 \\ &\approx (1 - 2\delta)\mathbf{I} + \beta(\mathbf{A}_1 + \mathbf{A}_2) = (1 - 2\delta)\mathbf{I} + 2\beta\left(\frac{\mathbf{A}_1 + \mathbf{A}_2}{2}\right) \end{aligned}$$

where we neglected second order terms involving β and δ . Thus $(1 - 2\delta)\mathbf{I} + 2\beta\mathbf{A}_{\text{avg}}$ is a first-order approximation of the \mathbf{S} matrix. \square

In other words, we can consider the time-varying system to be approximated by a static graph system of the \mathbf{A}_{avg} graph adjacency matrix with a virus of the same strength (β/δ remains the same). The threshold for a static graph with adjacency matrix \mathbf{A} is $\beta\lambda_{\mathbf{A}}/\delta$. So in our static case, we should aim to reduce $\lambda_{\mathbf{A}_{\text{avg}}}$ as much as possible. Any policy which aims to reduce the $\lambda_{\mathbf{A}_{\text{avg}}}$ value will approximate our original goal of dropping the $\lambda_{\mathbf{S}}$ value. Thus, this gives a theoretical justification of why Greedy-DavgA gave good results.

Temporal Immunization: In this paper, we concentrated only on *static* immunization policies - policies where once immunized, a node is ‘removed’ from the contact graphs. While this makes sense for biological vaccinations, in a more complex ‘resource’ oriented scenario where each ‘glove’ protects a person only for the time it is worn, a time-varying immunization policy might be more useful. e.g. we may have finite number of gloves to give and we can change the assignment of gloves during day/night. In this case, it may be better to immunize nurses in hospitals during the day by giving them the gloves but during the night, we can decide to give gloves to restaurant waiters or children, because the nurses are now not well-connected in the contact graph. Our threshold results can trivially estimate the impact of any ‘what-if’ scenarios w.r.t. such temporal immunization algorithms.

8 Conclusion

In this paper, we analytically studied virus-spreading (specifically the SIS model) on arbitrary, time-varying graphs. Given a set of T alternating graphs, modeling e.g. the day/night pattern of human behavior, we ask: (a) what is the epidemic threshold? and (b) what are the best- k nodes to immunize to defend against an epidemic? Our main contributions are:

1. We show how to formulate the problem, namely by approximating it with a *Non-Linear Dynamical System* (NLDS).
2. We give the **first** closed-formula for the threshold, involving the first eigenvalue of the *system-matrix* (see Theorem 2).
3. We use the insight from our threshold to develop and evaluate several immunization policies on real data like MIT reality mining graphs.

Future work can focus on providing bounds for the effectiveness of our immunization heuristics.

References

1. R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the World-Wide Web. *Nature*, 401:130–131, September 1999.
2. R. M. Anderson and R. M. May. *Infectious diseases of humans: Dynamics and control*. Oxford Press, 2002.
3. N. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, London, 1975.
4. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 15 October 1999.
5. C. L. Barrett, K. R. Bisset, S. G. Eubank, X. Feng, and M. V. Marathe. Episim-demics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. *ACM/IEEE Conf. on Supercomputing*, 2008.
6. S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change in informational cascades. *Journal of Political Economy*, 100(5):992–1026, October 1992.
7. L. Briesemeister, P. Lincoln, and P. Porras. Epidemic profiles and defense of scale-free networks. *WORM 2003*, Oct. 27 2003.
8. D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM TISSEC*, 10(4), 2008.
9. R. Cohen, S. Havlin, and D. ben Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 91(24), Dec. 2003.
10. P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, pages 57–66, 2001.
11. N. Eagle, A. Pentland, and D. Lazer. Inferring social network structure using mobile phone data. *Proc. of the National Academy of Sciences*, 106(36), 2009.
12. M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *SIGCOMM*, pages 251–262, Aug-Sept. 1999.
13. A. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM*, 2005.

14. J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 2001.
15. G. H. Golub and C. F. Van-Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 2nd edition, 1989.
16. M. Granovetter. Threshold models of collective behavior. *Am. Journal of Sociology*, 83(6):1420–1443, 1978.
17. D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW '04*, 2004.
18. Y. Hayashi, M. Minoura, and J. Matsukubo. Recoverable prevalence in growing scale-free networks and the effective immunization. *arXiv:cond-mat/0305549 v2*, Aug. 6 2003.
19. H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42, 2000.
20. M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, 1974.
21. D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03*, 2003.
22. J. O. Kephart and S. R. White. Directed-graph epidemiological models of computer viruses. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359, May 1991.
23. J. O. Kephart and S. R. White. Measuring and modeling computer virus prevalence. In *Proceedings of the 1993 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 2–15, May 1993.
24. R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. On the bursty evolution of blogspace. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 568–576, New York, NY, USA, 2003. ACM Press.
25. S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the web for emerging cyber-communities. *Computer Networks*, 31(11-16):1481–1493, 1999.
26. J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proc. of ACM SIGKDD*, pages 177–187, Chicago, Illinois, USA, 2005. ACM Press.
27. N. Madar, T. Kalisky, R. Cohen, D. ben Avraham, and S. Havlin. Immunization and epidemic dynamics in complex networks. *Eur. Phys. J. B*, 38(2):269–276, 2004.
28. C. R. McCuler. The many proofs and applications of perron’s theorem. *SIAM Review*, 42, 2000.
29. A. G. McKendrick. Applications of mathematics to medical problems. In *Proceedings of Edin. Math. Society*, volume 14, pages 98–130, 1926.
30. S. Milgram. The small-world problem. *Psychology Today*, 2:60–67, 1967.
31. M. E. J. Newman. Threshold effects for two pathogens spreading on a network. *Phys. Rev. Lett*, 2005.
32. R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.
33. R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65:035108, 2002.
34. M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. *SIGKDD*, 2002.
35. M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal*, 6(1), 2002.
36. E. M. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, August 2003.
37. Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *SRDS*, 2003.