Knowledge and Information Systems manuscript No.
KAIS-12-3483R1

B. Aditya Prakash · Deepayan Chakrabarti · Nicholas Valler · Michalis Faloutsos · Christos Faloutsos

Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks

Received: Feb 29, 2012 / Revised: May 07, 2012 / Accepted: May 13, 2012

Abstract Given a network of who-contacts-whom or who-links-to-whom, will a contagious virus/product/meme spread and 'take-over' (cause an epidemic) or die-out quickly? What will change if nodes have partial, temporary or permanent immunity? The epidemic threshold is the minimum level of virulence to prevent a viral contagion from dying out quickly and determining it is a fundamental question in epidemiology and related areas. Most earlier work focuses either on special types of graphs or on specific epidemiological/cascade models. We are the first to show the G2-threshold (twice generalized) theorem, which nicely de-couples the effect of the topology and the virus model. Our result unifies and includes as special case older results and shows that the threshold depends on the first eigenvalue of the connectivity matrix, (a) for any graph and (b) for all propagation models in standard literature (more than 25, including *H.I.V.*). Our discovery has broad implications for the vulnerability of real, complex networks, and numerous applications, including viral marketing, blog dynamics, influence propagation, easy answers to 'what-if' questions, and simplified design and evaluation of immunization

B. Aditya Prakash
Computer Science Department,
5000 Forbes Avenue,
Carnegie Mellon University
Pittsburgh PA 15213 USA
E-mail: badityap@cs.cmu.edu

Deepayan Chakrabarti
Yahoo! Research
E-mail: deepay@fb.com

Nicholas Valler and Michalis Faloutsos
University of California - Riverside
E-mail: {nvaller, michalis}@cs.ucr.edu

Christos Faloutsos
Carnegie Mellon University
E-mail: christos@cs.cmu.edu

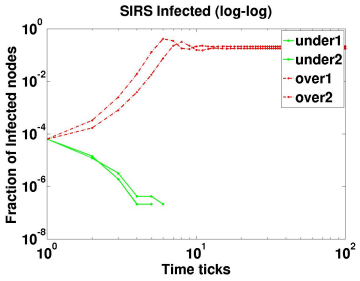


Fig. 1 Qualitatively different infection time-series curves (Fraction of Infected population vs Time) for the SIRS model (temporary immunity, like pertussis) on a large contact-network. What is the condition that separates the two regimes - red (epidemic) vs green (extinction)?

policies. We also demonstrate our result using extensive simulations on real networks, including on one of the biggest available social-contact graphs containing more than *31 million* interactions among more than *1 million* people representing the city of Portland, Oregon, USA.

Keywords Epidemics · Cascades · Virus Propagation Models · Arbitrary graphs · Tipping points

1 Introduction

Given a social or computer network, where the links represent who has the potential to infect whom, what can we say about its epidemic threshold? That is, can we determine whether a small infection can 'take-off' and create an epidemic? What will change if the nodes have permanent, temporary or no immunity? Both the underlying contact-network (or the population structure) and the particular cascade (propagation) model should intuitively play an important role in the spread of contagions (viruses/memes/products). Finding the epidemic threshold for an arbitrary network is an important and fundamental question in epidemiology and related areas. For instance, Figure 1 shows the simulation output after running the SIRS model (Susceptible-Infectious-Recovered-Susceptible which models diseases with temporary immunity like pertussis) on a large contact-network for different values of the virulence of the virus (achieved by tuning the parameters of the model). We can clearly see two different regimes - the fast die-out green regime and the steady-state epidemic red regime. Our paper deals with finding the condition which separates these two regimes in SIRS, as well as in *all* other virus propagation models in standard literature [25; 16], on *arbitrary* contact-networks.

Much of previous work focuses on either special types of graphs (typically cliques [30], block-structure and hierarchical graphs [26] and random power-law graphs [44]) or on specific epidemiological models [10]. We unify and include as special-case older results in two orthogonal directions and show:

– *De-coupling*: the threshold condition *separates* the effect of topology and the virus model,

- *Arbitrary Topology*: the threshold depends on the first eigenvalue of the connectivity matrix,
- *Arbitrary VPM*: the threshold depends on one constant that completely characterizes the virus propagation model (VPM)

Our result has numerous applications and immediate implications (see § 8) including easy answers to 'what-if' questions and simplified design and evaluation of immunization policies. Moreover, a variety of dynamic processes on graphs are modeled like epidemic spreading and hence our result applies to many of them. For example, the linear-cascade model [29] is essentially the SIR model (Susceptible-Infected-Recovered, models chicken pox, see Figure 2 (left inset) for state diagram); also, so-called threshold models (like Granovetter's model [22]) in sociology are similar in reality to cascade models [15]. In contrast to harmful viruses, the propagation of some contagions may in fact be *desirable* e.g. dissemination of a product or an idea in a network of individuals. For example, the Bass model [5] fits product adoption data using parameters for pricing and marketing effects. However it ignores topology; it simply assumes that all adopters have equal probability of influencing non-adopters. Instead, using our result, a more refined picture can be constructed of when a product gains massive adoption on a social network (equivalent to an "epidemic").

Several VPMs have direct applications in modeling computer and email viruses [31; 24]. In these cases, more so than the biological ones, it is easier to get the entire underlying network. Hence our threshold results can be used to make the network more robust by "immunizing" a few carefully chosen computers in the network (like installing a firewall on them). Another application is the efficient spreading of software patches over a computer network. The patches behave like computer worms [51] and can help defend against other malicious worms. Given full knowledge of the router-network involved, we can then estimate how "infectious" the patch-worm has to be (say by increasing the number of probes for possible hosts before dying out) to at least initiate an "epidemic" w.r.t. the patch. Additionally, we can help determine the vulnerability and consequently the cost of not patching parts of the network. Various epidemic models have also been used to model blog cascades which can now be applied to arbitrary graphs e.g. to study the propagation of memes through blogs [36].

The rest of the paper is organized as follows: we first give the related work in § 2, then formulate the problem (§ 3) and state our main result (§ 4), give a proof roadmap and example (§ 5) and then show simulation experiments (§ 6) to demonstrate the result. We discuss the broad implications and many applications of the result in § 7 and § 8. We then conclude (§ 9) and finally give a detailed proof in the Appendix.

2 Related Work

We review related work here, which can be categorized into three parts: epidemic thresholds, information diffusion and cyber-physical infrastructures.

None of these works generalize in two directions: for arbitrary propagation models and arbitrary networks.

2.1 Epidemic Thresholds

Canonical texts for epidemiology include [2; 25]. The most widely-studied epidemiological models include the so-called *homogeneous models* [39], which assume that every individual has equal contact to others in the population and that the rate of infection is determined by the density of the infected population. Kephart and White [30] were among the first to propose epidemiology-based models (the KW model) to analyze the propagation of computer viruses on homogeneous networks. However, there is overwhelming evidence that real networks including social networks, router and AS networks [18] etc. follow a power law structure instead. Pastor-Satorras and Vespignani [44] studied viral propagation for random power-law networks, and showed low or non-existent epidemic thresholds, meaning that even an agent with extremely low infectivity could propagate and persist in the network. They use 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. In a recent work, Castellano and Pastor-Satorras [8] empirically argue that some special family of random power-law graphs have a non-vanishing threshold under the SIR model in the limit of infinite size, but provide no theoretical justification.

Newman [43; 42] mapped the SIR model to a percolation problem on a network and studied thresholds for multiple competing viruses on special random graphs. Finally, Chakrabarti et.al. [10] and Ganesh et.al [19] gave the threshold for the SIS model on arbitrary undirected networks. Hence, *none* of the earlier work focuses on epidemic thresholds for *arbitrary* virus propagation models on *arbitrary*, real graphs.

2.2 Information Diffusion

There is a lot of research interest in studying dynamic processes on large graphs, (a) blogs and propagations [23; 32; 29; 48], (b) information cascades [6; 20; 22; 21; 54] and (c) marketing and product penetration [49; 35]. Competitive cascades have been studied in [45; 47]. Various optimization problems have also been studied on such processes like influence maximization [29; 12; 50] and finding effectors [34]. These dynamic processes are all closely related to virus propagation, with many directly based on epidemiological models [5; 29] e.g. the award-winning linear-cascade model [29] is a *special* case of our model : specifically it is essentially a SIR model with $\delta = 1$ and all our results carry through.

Table 1 Common Terminology

Term	Definition
VPM	virus-propagation model
NLDS	non-linear discrete-time dynamical system
β	attack/transmission probability over a contact-link
δ	healing probability once infected
γ	immunization-loss probability once recovered (in SIRS) or vigilant (in SIV, SEIV)
ϵ	virus-maturation probability once exposed hence, $1-\epsilon$ is the virus-incubation probability
θ	direct-immunization probability when susceptible
$\mathbf{A}$	adjacency matrix of the underlying undirected contact-network
N	number of nodes in the network
λ_1	largest (in magnitude) eigenvalue of $\mathbf{A}$
s	effective strength of a epidemic model on a graph with adjacency matrix $\mathbf{A}$

2.3 Cyber-physical infrastructures

Cascade models have also been applied to real-world networks to understand network robustness in cyber-physical infrastructures, i.e., the ability of a network to continue it's function in light of failures. The exact nature of a network's "function" varies from network-to-network and is typically determined during their design phase. Models related to the spread of disease–similar to those modeled herein–have been used to analyze the robustness of networks against node failures, for instance, see Chakrabarti et al. [9]. Another such work is Buldyrev et al. [7], which explores cascading failures in coupled power and data networks. Their model is based on a percolation model common in statistical physics, and can be shown equivalent to the SIR model we describe later in the paper.

3 Problem Formulation

Table 1 and Table 2 list common terminology and describe some of the epidemic models we will be using in the paper. We use the term 'cascade model' and 'virus propagation model' interchangeably in the paper. We next state formally the problem we address in the paper:

Problem 1 Epidemic Threshold

Given: A undirected unweighted graph G, and a virus propagation model (VPM) and its parameters (e.g. β and δ for SIR).
Find: A condition under which will an infection will die out and not cause an epidemic on the graph.

Table 2 Some Virus Propagation Models (VPMs)

Model	Description
SIS	'susceptible, infected, susceptible' VPM - no immunity, like flu
SIR	'susceptible, infected, recovered' VPM - life-time immunity, like mumps
SIRS	VPM with temporary immunity
SIV	'susceptible, infected, vigilant' VPM - immunization/vigilance with temporary immunity
SEIR	'susceptible, exposed, infected, recovered' VPM - life-time immunity *and* virus incubation
SEIV	VPM with vigilance/immunization with temporary immunity *and* virus incubation

4 Results

The epidemic threshold is usually defined as the minimum level of virulence to prevent a viral contagion from dying out quickly [2; 25; 3; 31]. In order to standardize the discussion of threshold results, we express the threshold in terms of the normalized *effective strength*, s, of a virus which is a function of the *particular* propagation model and the *particular* underlying contact-network. So we are 'above threshold' when $s > 1$, 'under threshold' when $s < 1$ and the threshold or the tipping point is reached when $s = 1$. The effective strength s can be thought of as the basic reproduction number R_0 frequently used in epidemiology [25; 2]. It (s) is then very roughly, the "net" generalized R_0 for the virus model and an arbitrary graph and is the quantity which determines the tipping point of an infection over a contact-network. Our main result is:

Theorem 1 (G2-threshold theorem) *For any virus propagation model (satisfying our general initial assumptions; see Section 5 for details) operating on an arbitrary undirected graph with adjacency matrix* $\mathbf{A}$ *and largest eigenvalue* λ_1*, the virus will get wiped out if:*

$$\boxed{s < 1} \tag{1}$$

where, s (the effective strength) is:

$$\boxed{s = \lambda_1 \cdot C_{\text{VPM}}} \tag{2}$$

and C_{VPM} *is an explicit constant dependent on the virus propagation model. Hence, the tipping point is reached when* $s = 1$*.*

Proof We give a roadmap in the next section and a detailed proof in the Appendix. □

Firstly, note that our result separates out the effect of the network and the VPM. Secondly, our result subsumes older results on (a) contact-networks, and (b) VPMs as special cases. Results on contact-networks like cliques (everybody contacts everybody else: $\lambda_1 = N - 1$, N is the number of nodes in

Table 3 Threshold Results for Some Models.

Models	Effective Strength (s)	Threshold (tipping point)
SIS, SIR, SIRS, SEIR	$s = \lambda_1 \cdot \left(\frac{\beta}{\delta}\right)$	$s = 1$
SIV, SEIV	$s = \lambda_1 \cdot \left(\frac{\beta\gamma}{\delta(\gamma+\theta)}\right)$	
$SI_1I_2V_1V_2$ ($\sim$ H.I.V.)	$s = \lambda_1 \cdot \left(\frac{\beta_1 v_2 + \beta_2 \epsilon}{v_2(\epsilon+v_1)}\right)$	

SIS (*susceptible/infected/susceptible*) has no immunity (like flu), SIR (*susceptible/infected/recovered*) has permanent immunity (like mumps), SIRS has temporary immunity (like pertussis) while SEIR (*susceptible/exposed/infected/recovered*) has additional virus incubation and $SI_1I_2V_1V_2$ has been used to model some H.I.V. infections [2]. SEIV and SIV are two useful generalizations. β is the attack/transmission probability over a contact link, δ is the healing probability, γ is the immunization-loss probability, $(1-\epsilon)$ is the virus incubation probability and θ is the direct-immunization probability when susceptible (see Figure 2). Our result is a general one and these models just highlight its ready applicability to standard VPMs in use.

the graph), random Erdős-Rényi graphs with expected degree d ($\lambda_1 = d$), 'homogeneous' graphs [30], power-law/scale-free graphs [44], structured hierarchical (near-block-diagonal) topologies [26] (people within a community contact all others in this community, with a few cross-community contacts) etc. are special cases. Likewise, all standard virus propagation models [25; 16] are specific instantiations of the generalized model used in our theorem (see Figure 2; more later).

Table 3 lists a few of our threshold expressions after applying our result on some standard epidemic models. The popular models listed include SIS (no immunity, like flu, Susceptible-Infected-Susceptible), SIR (permanent immunity, like mumps, Susceptible-Infected-Recovered), SIRS (temporary immunity, like pertussis), SEIR (virus incubation in addition to permanent immunity) etc. (note that models like SI inherently don't have an epidemic threshold as all nodes will eventually get infected on any graph - hence our work doesn't apply to them).

Table 3 also lists our SEIV model (Susceptible-Exposed-Infected-Vigilant) which itself generalizes almost all models from [25] (SIS with $\epsilon = 1, \gamma = 1, \theta = 0$; SIR with $\epsilon = 1, \gamma = 0, \theta = 0$; SIRS with $\epsilon = 1, \theta = 0$ and so on). Using our proof, we get that the effective strength for SEIV is $s = \lambda_1 \cdot \frac{\beta\gamma}{\delta(\theta+\gamma)}$ (as before the virus dies out if $s < 1$). Note that this implies that increasing β (the attack probability) strengthens the virus. At the same time, decreasing the healing probability δ also strengthens the virus. Finally, decreasing θ (the direct immunization probability) and increasing γ (the immunization loss probability) also makes the virus stronger. All of these fit with intuition - in fact, the usefulness of our result is partly in enabling us to see these

complex effects on the virus strength very clearly. We discuss some subtler implications later in Section 7. We discuss our terminology, general model and proof sketch next.

5 Proof Overview

We first construct a generalized model ($S^*I^2V^*$- arbitrary number of susceptible and vigilant states, two infectious states) that is powerful enough to generalize all the practical VPMs (and more) and satisfies our very general assumptions, while still being mathematically tractable (Figure 2). We then approximate our general model using a discrete time non-linear dynamical system and transform the tipping point question into a stability problem of the dynamical system at an appropriate equilibrium point. We give the overview and roadmap here. As mentioned before, the full proof can be found in the Appendix.

5.1 Our Terminology

Note that any VPM has some states and the choice of which states to include in a model depends on the particular contagion characteristics. Yet, we can think of every model as having states essentially in any of the following fundamental broad classes:

1. *Susceptible Class*: Nodes in such a state can get infected by any neighboring node (in the contact-network) who is infectious.
2. *Infected Class*: In a state of this class, the node is infectious in the sense that it is capable of transmitting the infection to its neighbors. Note that each such state will have a *transmissibility* parameter (e.g. β in the SIR model for the infectious state I). Thus this can include models with transmissibility parameter = 0 i.e. they are 'exposed' but not infectious (e.g. the E state in the SEIR model is a state which is in the Infected class in the sense that it can potentially cause infections but is not by itself infectious).
3. *Vigilant/Vaccinated Class*: Nodes in any of the states in this class cannot get infected nor can they potentially cause infections. States like R in SIR (the recovered/died state where the node gets permanent immunity/dies and hence does not participate in the epidemic further), M in MSIR (the passive immune state), etc. are conceptually of the Vigilant type.

5.2 Our General Model

Using our terminology above, we can now describe the generalized model we used in Theorem 1: $S^*I^2V^*$ (arbitrary number of susceptible and vigilant states, two infectious states). As our general characterization, $S^*I^2V^*$is powerful enough to seamlessly capture all the practical models (and more)

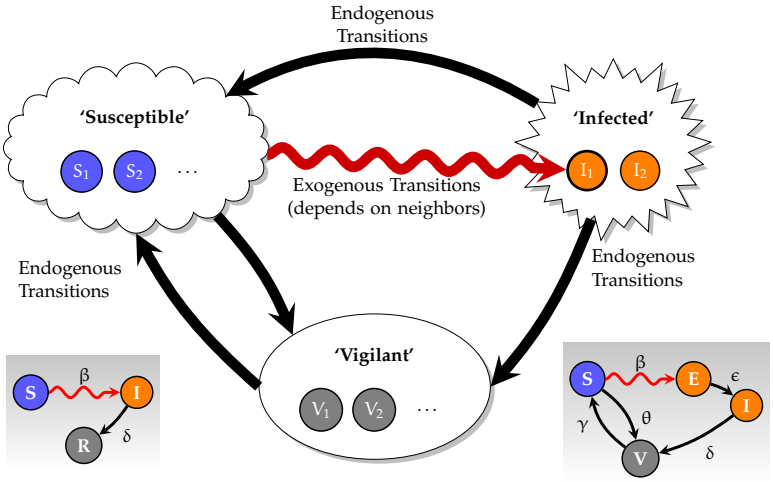


Fig. 2 State Diagram for a node in the graph in our generalized model $S^*I^2V^*$- it is *not* a simple Markov chain. There are three classes (types) of states - Susceptible (healthy but can get infected), Infected (capable of transmission) and Vigilant (healthy and can't get infected). Within-class transitions not shown for clarity. Red-curvy arrow indicates exogenous i.e. *graph-based* transition affected only by the neighbors of the node, all other transitions are *endogenous* (caused by the node itself with some probability at every time step). **(Left Inset)** *Special case*: Transition diagram for the SIR (Susceptible-Infected-Recovered) model. **(Right Inset)** Another special case: Transition diagram for the SEIV (E stands for exposed but not infectious) model. SEIV itself generalizes almost all models from [25] (SIS with $\epsilon = 1, \gamma = 1, \theta = 0$; SIR with $\epsilon = 1, \gamma = 0, \theta = 0$; SIRS with $\epsilon = 1, \theta = 0$ and so on).

like SIS, SIR, SIRS, SEIR, SERIS, MSIR, MSEIR etc. [25; 16], including H.I.V. [2], while being tractable enough to yield simple threshold equations. Figure 2 shows the state diagram under $S^*I^2V^*$for a node in the contact-network together with the assumptions on the transitions. The red-curvy arrow indicates exogenous (*graph-based*) transition caused by infectious neighboring nodes while all other transitions are endogenous, caused by the node itself with some probability. We have shown only cross-class transitions and their types. We make two assumptions:

1. *Infection through Neighbors*: The only way to get infected is through your neighbors i.e. there is no path *to* a state in the Infected class from a state in the Susceptible class composed solely of endogenous transitions.
2. *Starting Infected State*: For the few models that have more than one infectious state, any exogenous (graph-based) transition always results in a transition from a state in the Susceptible class to the I_1 state. Note that this assumption is trivially obeyed for a vast majority of models (with only one infected state).

Figure 2 (Left Inset) shows the popular SIR model as an instantiation of our general model $S^*I^2V^*$. Also, Figure 2 (Right Inset) shows an instantiation in the form of our SEIV model (Susceptible-Exposed-Infected-Vigilant).Figure 3 shows the generalization hierarchy for some common epidemic models and

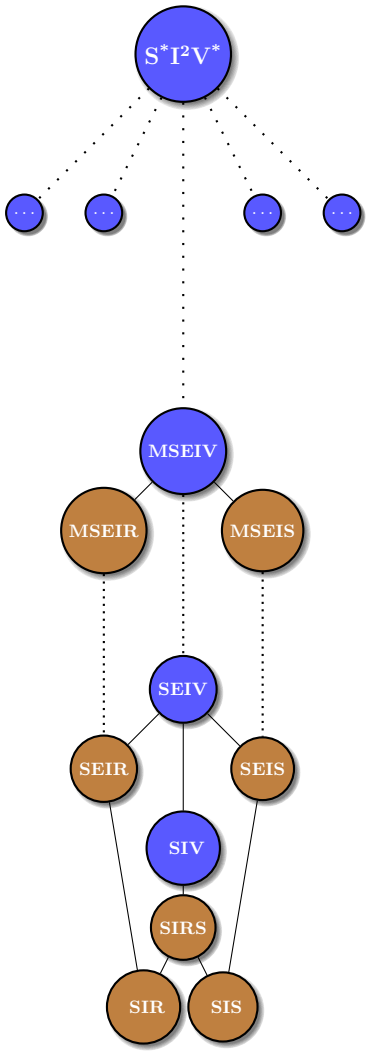


Fig. 3 Virus Propagation model hierarchy (actually, lattice) for some standard models including SIRS (temporary immunity), SIV (vigilance, i.e., pro-active vaccination); SEIV (includes the 'exposed but not infectious' state, and temporary vigilance); MSEIR (with the passive immune state M); and our main generalization $S^*I^2V^*$. The brown colored nodes denote standard VPMs found in literature while the blue colored nodes denote our generalizations. Each VPM is a generalization of all the models below it.

our main generalization $S^*I^2V^*$. The brown colored nodes denote standard VPMs found in literature while the blue colored nodes denote our generalizations. Each VPM is a generalization of all the models below it e.g. SIV is a generalization of SIRS, SIR and SIS.

5.3 Proof Sketch

We define the vector $\mathbf{P}_t$ such that it specifies the state of the system at time t; the exact definition will differ from model to model but it effectively encodes the probability of each node in the graph of being in any given state at time t. Suppose the virus-propagation model has m $(s_1, s_2, \ldots, s_m)$ states (e.g. $m = 3$ for the SIR model with states $s_1 = S$, $s_2 = I$ and $s_3 = R$) and it operates on a graph of N nodes. Consider then a column vector $\mathbf{P}_t \in \Re^{m \cdot N \times 1}$, which captures the probability of each node being in any of m states at a given time t. Specifically:

$$\mathbf{P}_t = [P_{s_1,1,t}, P_{s_1,2,t}, \ldots, P_{s_1,N,t}, P_{s_2,1,t}, \ldots, P_{s_m,N,t}]^{\mathrm{T}} \tag{3}$$

where, $P_{s_i,j,t}$ is the probability that node j is in state s_i at time t. A Non-Linear Dynamical System (NLDS) can be represented by $\mathbf{P}_{t+1} = g(\mathbf{P}_t)$ where g is some non-linear function operating on a vector. The function g in our case is large and complicated. The NLDS equation essentially tracks the evolution of the vector $\mathbf{P}_t$ over time. An equilibrium point (also called a fixed point) of the system is the state vector (i.e. some particular $\mathbf{P}$) which does not change. Thus at the equilibrium point $\mathbf{P}_{t+1} = \mathbf{P}_t = \mathbf{x}$. Intuitively, the tipping point for any model then deals with analyzing the stability of the corresponding NLDS at the point when none of the nodes in the graph are infected, because otherwise the infection can still spread. If the equilibrium is unstable, a small "perturbation" (physically in the form of a few initial nodes getting infected) will push the system further away (which physically means more and more nodes will get infected leading to an epidemic). But if the equilibrium is stable, the system will try to come back to the fixed point without going "too-far" away, in effect, "controlling the damage". At threshold, the tendencies to go further away and come-back will be the same. In other words, the equilibrium is stable below the threshold and is neutral at the tipping point. From dynamical-system literature, we know how to relate the stability of the system at the equilibrium point to the spectrum of the Jacobian matrix at that point (i.e. $\bigtriangledown g(\mathbf{x})$). We eventually reduced the requirement on the eigenvalues of $\bigtriangledown g(\mathbf{x})$ for any virus propagation model to a simple condition on the eigenvalue of the adjacency matrix. This condition translates into the effective strength of the virus under the model. The reason we can reduce the condition to one on the adjacency matrix is due to the special structure of the virus models, which was captured by the $\mathrm{S}^*\mathrm{I}^2\mathrm{V}^*$ model described before. See the Appendix for the full proof.

6 Experiments

We performed computer simulation experiments on two large networks topologies, to demonstrate our result. All the different virus propagation models were implemented as a discrete event simulation in C++. We ran each simulation for 1000 time ticks and took the average of 100 runs. Initially, 10 nodes were infected with the virus and we then let the propagation take over according to the particular model. The datasets we used were:

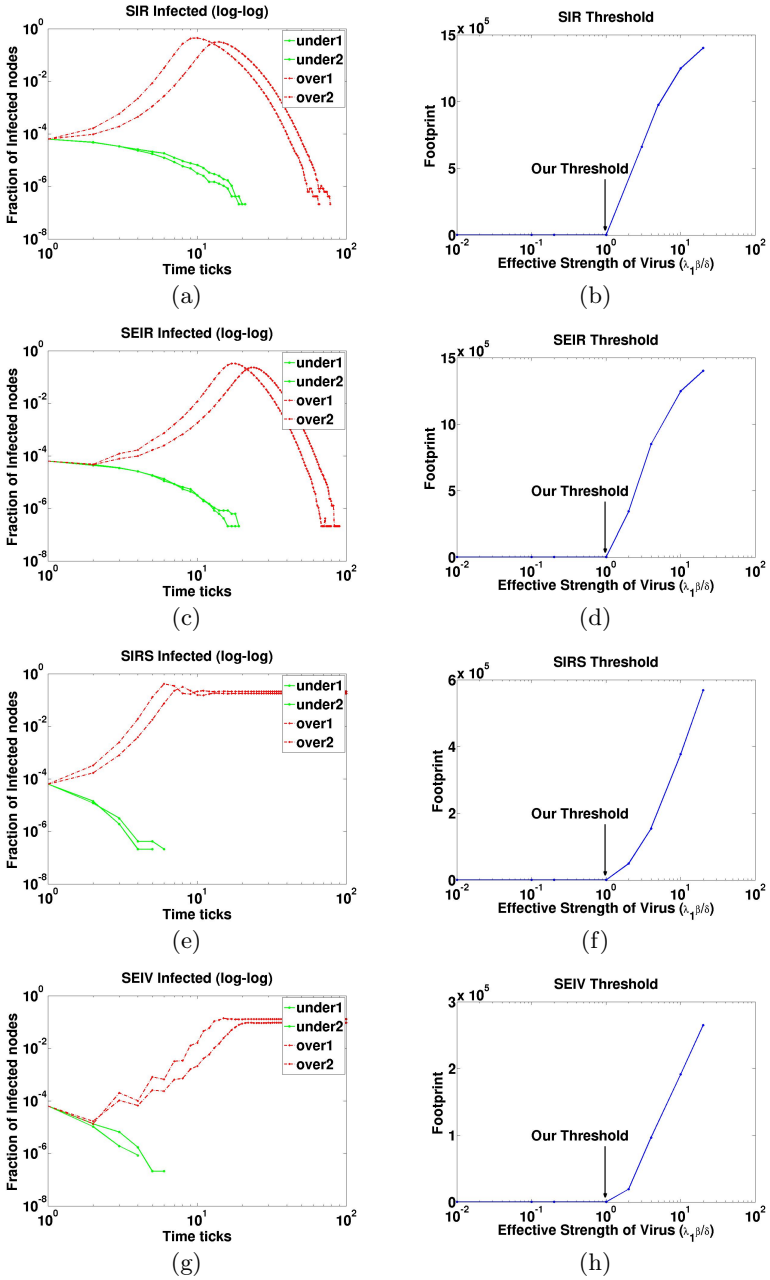


Fig. 4 Simulation Results on the PORTLAND graph, all values averages over 100 runs. **(a),(b),(c)** Plot of Infective Fraction of Population vs Time (log-log) for SIR, SEIR, SIRS and SEIV models. Note the qualitative difference in behavior-two curves *under* (green) the threshold and two curves *above* (red) the threshold. **(d),(e),(f)** "Take-off" plots, *Footprint* (see Section 6) vs Effective Strength (lin-log) for SIR, SEIR, SIRS and SEIV models. The tipping point exactly matches our prediction ($s = 1$) in all cases.

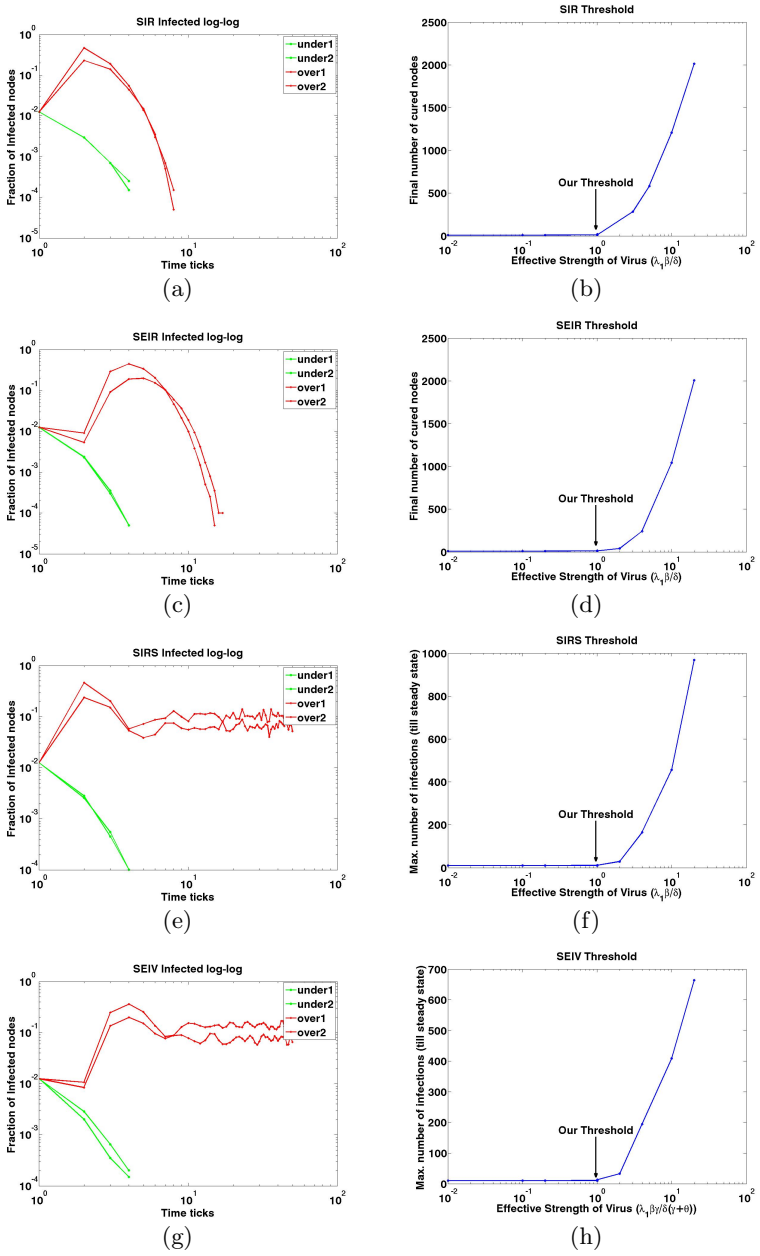


Fig. 5 Simulation Results on the `AS-OREGON` graph, all values averages over 100 runs. **(a),(b),(c)** Plot of Infective Fraction of Population vs Time (log-log) for SIR, SEIR, SIRS and SEIV models. Note the qualitative difference in behavior-two curves *under* (green) the threshold and two curves *above* (red) the threshold. **(d),(e),(f)** "Take-off" plots, *Footprint* (see Section 6) vs Effective Strength (lin-log) for SIR, SEIR, SIRS and SEIV models. The tipping point exactly matches our prediction ($s = 1$) in all cases.

1. `AS-OREGON`: This network represents the Internet's Autonomous System (AS) connectivity derived from public data sets collected by the Oregon Route Views project [11]. It contains 15,420 links among 3,995 AS peers. The Oregon graph is relevant to studying the robustness of router networks to worm attacks [33]. More information can be found from `http://topology.eecs.umich.edu/data.html`.
2. `PORTLAND`: It is one of the biggest available physical contact graphs, representing a synthetic population of the city of Portland, Oregon, USA [41].It is a social-contact graph containing more than *31 mil.* links (interactions) among about *1.6 mil.* nodes (people). The data set is based on detailed microscopic simulation-based modeling and integration techniques and has been used in modeling studies on smallpox outbreaks as well as policy making at the national level [17].

Figure 4 illustrates our result via simulation experiments on `PORTLAND`. Above threshold, note the steady state behavior in SEIV and the initial explosive phase and eventual decay in SIR and SEIR (because the number of susceptible nodes decrease monotonically). Also note the initial "flat" period in the time plots for above threshold for the models having the Exposed (E) state, SEIR and SEIV. This is due to the virus-incubation period because of which there is an initial delay in number of infected nodes. This then results in an initial "silent" period after which the epidemic takes-off. As there is no such incubation period in SIR and SIRS, their plots do not show such silent periods.

In contrast, under threshold, the number of infections aggressively go down to zero in all the models. In addition, as our result predicts, the precise point when the footprint of infection suddenly jumps in all models is at $s = 1$. The footprint measures the extent of infection: For models with a steady-state behavior (SIS/SIRS) it is defined as the maximum number of infections at any instant till we reach steady-state. For models with monotonous decrease of susceptibles (and hence without a steady-state, SIR/SEIR) footprint is the final number of cured/removed nodes from the network at the end of the infection. Figures 4 (d-f) also demonstrate the simplicity and power of our result - the only variable we need for determining the epidemic threshold of the whole system consisting of multiple parameters is the effective strength ($s = \lambda_1 * C_{VPM}$), nothing else.

7 Implications

We first discuss some direct implications of the G2-threshold theorem: the vulnerability of graphs to epidemics and some unexpected results in specific models.

7.1 Vulnerability of Networks–focus on eigenvalues

What exactly does the result mean w.r.t. the graph? Intuitively, λ_1 (also known as the spectral radius) of a graph captures the connectivity of the

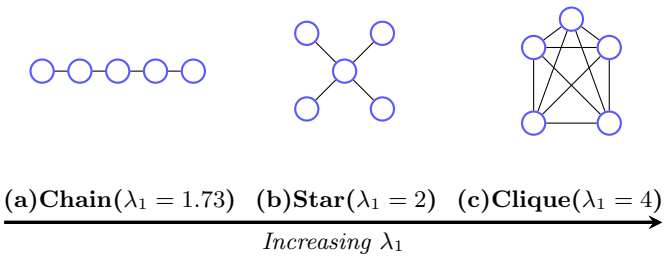
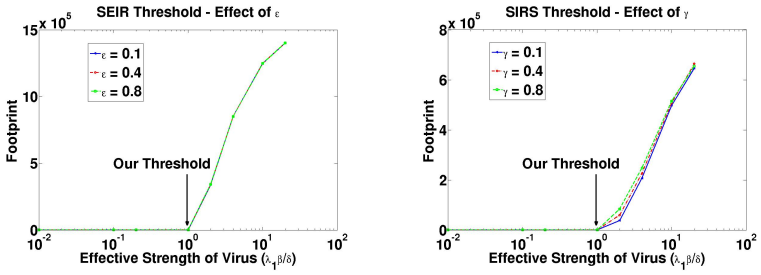


Fig. 6 Why λ_1 matters more than number of edges E: changing connectivity and vulnerability of graphs with changing λ_1. The clique (largest λ_1) is the most vulnerable. Note that E is not enough: star and chain have the same number of edges ($E = 4$) but the star is intuitively more vulnerable, as our result also says (it has a higher λ_1).



(a) (b)

Fig. 7 Counter-intuitive results - neither Incubation rate ϵ or Immunity-loss rate affects the threshold. **(a)** 'Take-off' plot for the SEIR model (a special case of SEIV) on the PORTLAND graph (lin-log scale). All three curves are on top of each other. **(b)** 'Take-off' plot for the SIRS model (a special case of SEIV) on the PORTLAND graph (lin-log scale) - higher means more infections (increasing with the loss of immunization γ). Note that in both the cases it *does not* affect the threshold (the tipping point is still at effective strength $s = 1$). All values are averages over 100 runs.

graph. More connected the graph is, more vulnerable it is to an epidemic by a virus (see Figure 6). Our threshold results suggest that an arbitrary graph behaves in the same way to a λ_1-regular graph (both will have the same λ_1). The entire dynamics of the epidemic may not be captured by λ_1 completely, but the *threshold* is solely dependent on λ_1 (apart from parameters of the VPM). By making the relation between the graph and threshold explicit, our result has many consequences for the vulnerability of real, complex networks as well. For example, our result explains the observed vulnerability of 'small-world' networks [52]: their λ_1 is relatively high compared to a regular graph with the same number of nodes and edges, due to the presence of shortcuts. Also, previous results have shown that the epidemic threshold for the SIS model in case of random scale-free networks like the Internet is vanishingly small as the size N of the network increases [44; 1]. This is a corollary of our result: When a power-law graph grows ($N \to \infty$), the largest eigenvalue

grows with the maximum degree [13], which also grows to infinity, and thus the threshold approaches zero.

7.2 Counter-intuitive Results

Apart from the dependence of the threshold on λ_1, it is instructive to note unexpected results in some specific models. The SEIR, SIRS and SEIV models serve well to demonstrate the effect of virus-incubation and direct-immunization. See Figure 7. The threshold in SEIR surprisingly does not depend on the virus-incubation probability: the parameter ϵ, in effect, only delays/speeds-up the achievement of the threshold, not what the threshold itself is. Similarly, the threshold in SIRS does not depend on γ. Also, from the threshold equation of SEIV (Table 3), we can infer that lowering the rate of loss of immunity i.e. having a smaller γ (say due to better hygiene) decreases the effective strength s (and makes it harder for the virus to cause an epidemic) only so long as there is a mechanism to give a node direct immunity i.e. having a non-zero θ (say by using a vaccine) *before* an infection (in the Susceptible state) instead of *after* (in the Recovered state). Satisfyingly, this fits well with the old adage 'Prevention is better than Cure'.

8 Impact

Our results can be fundamental to a wide-range of applications. We mentioned broader impact in § 1 before. Here we briefly discuss some immediate applications in epidemiology and cyber-physical infrastructures.

8.1 Effective Immunization

Given the linear dependence on λ_1 of our threshold, we can propose a simple immunization goal. For any virus, remove (immunize) those nodes whose removal will decrease the λ_1 value the most (so that the resultant infection falls below threshold and dies out) e.g. immunize teachers and kindergarten children first to control the epidemic. A lot of work targets immunizing high-degree nodes in scale-free networks [14] which, while a good idea, is not optimal: just concentrating on high-degree nodes will *miss* those low-degree nodes which are good "bridges" and can have an important influence on decreasing λ_1 when immunized. For example, intuitively, the *sole common* friend between two disparate yet internally well-connected groups (like say between scientists and movie celebrities) can have a huge impact in the outbreak of a disease even if (s)he knows only a few people in each community.

8.2 Evaluating 'What-if' Scenarios

Our result can also help quickly determine the result of plausible situations e.g. is there a danger of an epidemic if the virus is twice (or half) as infectious

(virulent)? This can then feed into policy decisions for controlling epidemics, like imposing restrictions on travel so as to not increase the λ_1. Policy makers can assume *any* graph model which best captures the contact behavior of the population and still use our threshold result to guide immunization policies.

8.3 Accelerating Simulations

Similarly, we can considerably simplify expensive epidemiological simulations as well. For example, running a typical simulation with one set of parameters of a flu epidemic on a population of size 33 million ($\sim$ size of the state of California) takes about *2 days* on a cluster of 50 machines [4]. Using our result, we can eliminate parameters which do not affect the effective strength of the contagion and also quickly identify parameter spaces where simulations would be useful (i.e. above threshold). Clearly, the main task of such a testing will be eigenvalue computations. For this purpose, there are already very efficient algorithms like Lanczos for sparse graphs which take *2-5 mins* for networks of millions. Moreover, structured topologies like cliques, block-diagonal matrices lend themselves to even faster eigenvalue computations, making it very easy to apply our result to real world simulations.

8.4 Applications to Computer Networking

As mentioned in Section 2, the epidemic threshold can be applied to a number of network robustness scenarios in cyber-physical infrastructures. For example, the Kademlia DHT [37] is used in a number of P2P networks–such as BitTorrent–to form a decentralized P2P overlay and lookup table. In particular, Mainline BitTorrent [53] implements a version of Kademlia as an alternative to the typical centralized tracker.When a BitTorrent user queries the system for a particular torrent file, it is the DHT's responsibility to return the torrent file (i.e., BitTorrent swarm bootstrap information). Hence, using our result, the network overlay structure of the DHT–in particular, it's eigenvalue–may be used to evaluate the data replication necessary to guarantee the torrent file is reachable.

9 Conclusion

In summary, we studied the problem of determining the epidemic threshold given the virus propagation model and an underlying arbitrary undirected unweighted graph. Intuitively, the answer should depend both on the graph and the propagation model. Earlier results have focused on either special cases of graphs or special models. In this paper, we give a formula for the epidemic threshold which shows:

1. *De-coupling*: The effect of the topology and the propagation model on the threshold is clearly de-coupled,
2. *Arbitrary Topology*: The effect of the undirected underlying topology is determined *only* by λ_1 (the largest eigenvalue of the adjacency matrix),

3. *Arbitrary VPM*: The effect of the virus propagation model is determined by a model dependent constant.

Thus, all previous epidemic threshold results are specific instantiations of our G2-threshold theorem. Our results can be used for forecasting and estimations in 'what-if' scenarios, for control and manipulation of propagation and related dynamical processes (immunization, marketing policies etc.). Moreover, our result can be easily extended to handle even more elaborate settings such as (a) time-varying topologies (extending the SIS-only results of [4; 46]), and (b) multiple competing diseases (extending the random power-law-graphs-only results of [43]).

APPENDIX

We give the full proof of Theorem 1 here.

A Notation

Recall that we are dealing with the $S^*I^2V^*$ generalized model - it has two states I_1 and I_2 in the Infected class. To simplify notation, we refer to state I_1 as E (the 'infection entrance state') and I_2 as the I state in the proofs. The state E has a transmission probability of β_1 and the state I has a transmission probability of β_2. The states E and I here should be thought as to mean *general* infected states of our model and not in the sense of the specific E and I states in epidemic models like SEIR, SEIV etc. We also refer to the exogenous transitions as *graph-based* and endogenous transitions as *internal* interchangeably. Table 4 gives some of the additional notation we will be using in our description of the proof.

Table 4 Additional Notation and Symbols used in the proofs

Symbol	Definition
m	total number of states in the model
q	total number of states in the Susceptible and Vigilant classes of the model; hence $m = q + 2$
w	total number of states in the Susceptible class of the model
$S_1, S_2, \ldots, S_w$	general states in the Susceptible class
E, I	general states in the Infected class
α_{KU}	probability (constant and given) of transition from state K to state U
β_1	transmission probability for state E
β_2	transmission probability for state I
$\zeta_{i,t}(E, I)$	probability that a node i does not receive any infections from E and I at time t
$\mathbf{x}$	the fixed point vector our NLDS corresponding to when no node is in any of the Infected class states
$p^*_{S_y}$	(same for each node) probability of being present in the S_y state at $\mathbf{x}$
$\mathcal{J}$	Jacobian matrix of the NLDS computed at $\mathbf{x}$

B System Equations

We can develop the system equations i.e. explicitly specify the non-linear function g for the NLDS based on the transition diagram of the model. As stated earlier in Section 5.2 we assume that infections are received only from infected neighbors i.e. those in states E and I the Infected class of states. Firstly, let's calculate the probability that a node i does not receive any infections in the next time step (call it $\zeta_{i,t}(E,I)$, E,I denotes that an infection is passed only from a neighbor in the E or I states). No infections are transmitted if:

- Either a neighbor is not any of the infected states E and I
- Or it is in state E and the transmission fails with probability $1-\beta_1$
- Or it is in state I and the transmission fails with probability $1-\beta_2$

Since we assume infinitesimally small time steps ($\Delta t \to 0$), multiple events can be ignored for first-order effects in the time step. Also, assuming the neighbors are *independent*, we get:

$$\begin{aligned}\zeta_{i,t}(E,I) &= \prod_{j \in \mathcal{NE}(i)} (P_{E,j,t}(1-\beta_1) + P_{I,j,t}(1-\beta_2) + (1 - P_{E,j,t} - P_{I,j,t})) \\ &= \prod_{j \in \{1..N\}} (1 - \mathbf{A}_{i,j}(\beta_1 P_{E,j,t} + \beta_2 P_{I,j,t})) \end{aligned} \tag{4}$$

where $\mathcal{NE}(i)$ is the set of neighbors of node i in the graph.

Also, the sum of probabilities of being in all the possible states for each node i should equal 1. Hence,

$$\forall_{i,t} \quad \sum_K P_{K,i,t} = 1 \tag{5}$$

We can now write down the system equations as follows. A node i will be in any particular state S_y of the Susceptible class at time $t+1$ if:

- Either it was in S_y at time t and stayed in state S_y i.e. it did not receive any infections from its neighbors *and* it did not change state internally from S_y to any other state
- Or it was in some other state U and changed state internally from U to S_y

Hence, the probability of node i being in S_y where S_y is any state in the Susceptible class at time $t+1$ is:

$$\forall y = 1,2,\ldots,w \quad P_{S_y,i,t+1} = \sum_{K \neq S_y} \alpha_{K S_y} P_{K,i,t} + P_{S_y,i,t} \left(\zeta_{i,t}(E,I) - \sum_{K \neq E, S_y, I} \alpha_{S_y K} \right) \tag{6}$$

Similarly, for the E state:

$$P_{E,i,t+1} = \sum_{K \neq S_1, S_2, \ldots, S_w} \alpha_{KE} P_{K,i,t} + \sum_{y=1}^{w} P_{S_y,i,t} \left(1 - \zeta_{i,t}(E,I)\right) \tag{7}$$

and for any other state $U \neq \{S_1, S_2, \ldots, S_w, E\}$:

$$P_{U,i,t+1} = \sum_K \alpha_{KU} P_{K,i,t} \tag{8}$$

As discussed earlier (Equation 3), we can now define a probability vector $\mathbf{P}_t$ by "stacking" all these probabilities which will completely describe the system at any time t and evolve according to the above equations. Note that the above equations are non-linear and naturally define the function g for the NLDS $\mathbf{P}_{t+1} = g(\mathbf{P}_t)$.

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

Theorem 2 (Asymptotic Stability, e.g. see [27]) *The system given by* $\mathbf{P}_{t+1} = g(\mathbf{P}_t)$ *is asymptotically stable at an equilibrium point* $\mathbf{P} = \mathbf{x}$*, if the eigenvalues of* $\mathcal{J} = \bigtriangledown g(\mathbf{x})$ *are less than 1 in absolute value, where,*

$$\mathcal{J}_{i,j} = [\bigtriangledown g(\mathbf{x})]_{i,j} = \frac{\partial g_i}{\partial g_j}|_{\mathbf{P}=\mathbf{x}}$$

Hence, next we compute the fixed point we are interested in and the Jacobian of our NLDS at that point.

C Fixed point

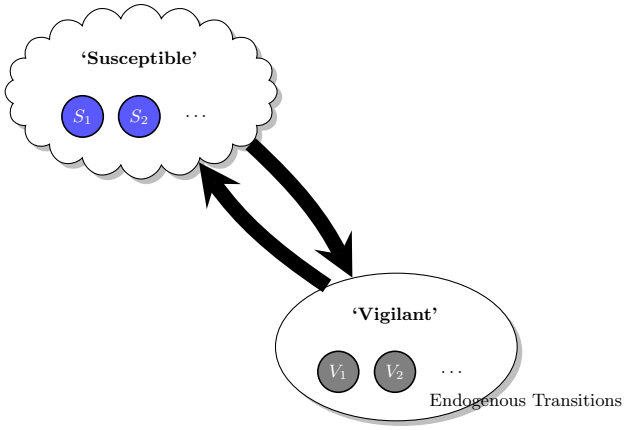


Fig. 8 State Diagram for any node in the graph at the fixed point when *no* node is present in a state in the Infected class. Only cross-class edges are shown. Note that it is now a simple Markov chain with a unique steady state probability.

We are interested in the stability of the equilibrium point (i.e. where $\mathbf{P}_{t+1} = \mathbf{P}_t(= \mathbf{x})$) of the NLDS which corresponds to when no one is infected. Only the transition from the Susceptible class states towards the Infected class states are graph-based (and can happen only when at least one of the nodes is in any of the Infected states), so the state-diagram for each node will be a simple Markov chain (call it MC_{SV}) consisting of the Susceptible and Vigilant states (see Figure 8). Note now there are no graph-based effects, hence each node is independent of others and will converge to steady state probabilities corresponding to the Markov chain. The steady state vector π^* (size $q \times 1$, where q is the number of states in the Susceptible and Vigilant classes) which will be the same for each node can be computed from the following equations from standard Markov chain analysis:

$$\pi^{*T}\ \mathrm{Tran}_{MC_{SV}} = \pi^* \quad \& \quad \sum_{i=1}^{q} \pi_i^* = 1 \tag{9}$$

Hence π^* is a probability vector and is the left eigenvector corresponding to eigenvalue 1 of the stochastic matrix $\mathrm{Tran}_{MC_{SV}}$ of the Markov chain MC_{SV}. The full $(m \times 1)$ probability vector $\mathbf{p^*}$ for each node at this steady state will have the entries in π^* for states in the Susceptible and Vigilant classes and 0 for all states

in the Infected class. The fixed point of the global original NLDS $\mathbf{x}$ can be finally represented as:

$$\mathbf{x} = \begin{bmatrix} \mathbf{p^*} \\ \mathbf{p^*} \\ \vdots \\ \mathbf{p^*} \end{bmatrix} \tag{10}$$

where $\mathbf{p^*}$ is repeated N times (once for each node in the graph). Let $p^*_{S_y}$ be the steady state probability value in the vector $\mathbf{p^*}$ corresponding to the S_y state. In other words, each node will have a probability of p^*_S of being present in the S_y state at the fixed point. Also define,

$$p^*_S = \sum_{y=1}^{w} p^*_{S_y} \tag{11}$$

i.e. p^*_S is the total probability of each node at the fixed point of being present in any of the states of the Susceptible class.

D The Jacobian

We know from Theorem 2 that $\mathbf{x}$ is stable if the eigenvalues of $\mathcal{J} = \bigtriangledown g(\mathbf{x})$ are less than 1 in absolute value. From the definition of $\mathcal{J}$ we can see that it is a $m \cdot N \times m \cdot N$ matrix with m (for each state) square blocks of size $N \times N$ each (corresponding to every node in the graph). We can calculate $\mathcal{J}$ to be (states have been mentioned on the top and side for ease of exposition and $\mathbb{I}$ is the identity matrix of size $N \times N$):

	S_y	K	$\dots$	E	I
S_y	$(1 - \sum_{K \neq S_y, E} \alpha_{S_y K})\mathbb{I}$	$\alpha_{K S_y}\mathbb{I}$	$\dots$	$\alpha_{E S_y}\mathbb{I} - p^*_{S_y}\beta_1\mathbf{A}$	$\alpha_{I S_y}\mathbb{I} - p^*_S\beta_2\mathbf{A}$
$\vdots$			$\ddots$		
U	$\alpha_{S_y U}\mathbb{I}$	$\alpha_{KU}\mathbb{I}$	$\dots$	$\alpha_{EU}\mathbb{I}$	$\alpha_{IU}\mathbb{I}$
$\vdots$			$\ddots$		
E	$\alpha_{S_y E}\mathbb{I}$	$\alpha_{KE}\mathbb{I}$	$\dots$	$\alpha_{EE}\mathbb{I} + p^*_S\beta_1\mathbf{A}$	$\alpha_{IE}\mathbb{I} + p^*_S\beta_2\mathbf{A}$
I	$\alpha_{S_y I}\mathbb{I}$	$\alpha_{KI}\mathbb{I}$	$\dots$	$\alpha_{EI}\mathbb{I}$	$\alpha_{II}\mathbb{I}$

where K is any state $\neq \{E, I\}$ and U is any state $\neq \{S_1, S_2, \dots, S_w, E, I\}$.

Recall the properties we are assuming for the epidemic models discussed in Section 5.2 (also see Figure 2). Crucially, they imply $\forall_{K \neq E,I}\ \alpha_{KE} = 0$ and $\forall_{K \neq E,I}\ \alpha_{KI} = 0$. Hence $\mathcal{J}$ reduces to:

	S_y	K	$\dots$	E	I
S_y	$(1 - \sum_{K \neq S_y, E} \alpha_{S_y K})\mathbb{I}$	$\alpha_{K S_y}\mathbb{I}$	$\dots$	$\alpha_{E S_y}\mathbb{I} - p^*_{S_y}\beta_1\mathbf{A}$	$\alpha_{I S_y}\mathbb{I} - p^*_S\beta_2\mathbf{A}$
$\vdots$			$\ddots$		
U	$\alpha_{S_y U}\mathbb{I}$	$\alpha_{KU}\mathbb{I}$	$\dots$	$\alpha_{EU}\mathbb{I}$	$\alpha_{IU}\mathbb{I}$
$\vdots$			$\ddots$		
E	$\mathbf{0}_{N,N}$	$\mathbf{0}_{N,N}$	$\dots$	$\alpha_{EE}\mathbb{I} + p^*_S\beta_1\mathbf{A}$	$\alpha_{IE}\mathbb{I} + p^*_S\beta_2\mathbf{A}$
I	$\mathbf{0}_{N,N}$	$\mathbf{0}_{N,N}$	$\dots$	$\alpha_{EI}\mathbb{I}$	$\alpha_{II}\mathbb{I}$

where $\mathbf{0}_{N,N}$ is a $N \times N$ matrix with all zeros.

E Eigenvalues of the Jacobian

Note that $\mathcal{J}$ is very structured and can be written as:

$$\mathcal{J} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{0}_{2N,(m-2)N} & \mathbf{B}_3 \end{bmatrix} \tag{12}$$

where $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ are matrices of size $(m-2)N \times (m-2)N$, $(m-2)N \times 2N$ and $2N \times 2N$ respectively. $\mathbf{B}_3$ corresponds to the E and I rows and columns of $\mathcal{J}$ i.e.:

$$\mathbf{B}_3 = \begin{bmatrix} \alpha_{EE}\mathbb{I} + p_S^*\beta_1\mathbf{A} & \alpha_{IE}\mathbb{I} + p_S^*\beta_2\mathbf{A} \\ \alpha_{EI}\mathbb{I} & \alpha_{II}\mathbb{I} \end{bmatrix} \tag{13}$$

$\mathbf{B}_1$ and $\mathbf{B}_2$ are defined similarly. Consider any eigenvector $\mathbf{v}$ (size $mN \times 1$) and corresponding eigenvalue $\lambda_{\mathcal{J}}$ of $\mathcal{J}$. We can write $\mathbf{v}$ as being composed of vector $\mathbf{v}_1$ of size $(m-2)N \times 1$ and vector $\mathbf{v}_2$ of size $2N \times 1$ i.e:

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \tag{14}$$

Also $\mathbf{x}$ and $\lambda_{\mathcal{J}}$ satisfy the eigenvalue equation:

$$\mathcal{J}\mathbf{v} = \lambda_{\mathcal{J}}\mathbf{v} \tag{15}$$

Substituting from Equations 12 and 14 we get:

$$\begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{0} & \mathbf{B}_3 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = \lambda_{\mathcal{J}} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \tag{16}$$

Equation 16 implies the following the two relations:

$$\mathbf{B}_1\mathbf{v}_1 + \mathbf{B}_2\mathbf{v}_2 = \lambda_{\mathcal{J}}\mathbf{v}_1 \tag{17}$$

$$\mathbf{B}_3\mathbf{v}_2 = \lambda_{\mathcal{J}}\mathbf{v}_2 \tag{18}$$

From Equation 18 we can infer that precisely *one* of the following holds:

1. $\mathbf{v}_2 = \mathbf{0}$
2. $\mathbf{v}_2$ is the eigenvector of $\mathbf{B}_3$ (and consequently $\lambda_{\mathcal{J}}$ is the matching eigenvalue of $\mathbf{B}_3$)

If $\mathbf{v}_2 = \mathbf{0}$, Equation 17 reduces to

$$\mathbf{B}_1\mathbf{v}_1 = \lambda_{\mathcal{J}}\mathbf{v}_1$$

wherein again, either $\mathbf{v}_1 = \mathbf{0}$ or $\lambda_{\mathcal{J}}$ is an eigenvalue of $\mathbf{B}_1$. The condition $\mathbf{v}_1 = \mathbf{0}$ is not meaningful as then $\mathbf{v} = \mathbf{0}$ ($\mathbf{v}$ is an eigenvector of $\mathcal{J}$ implies $\mathbf{v}$ is non-zero). Therefore the eigenvalues of $\mathcal{J}$ are given by the eigenvalues of $\mathbf{B}_1$ (with $\mathbf{v}_2 = \mathbf{0}$) and the eigenvalues of $\mathbf{B}_3$.

E.1 Eigenvalues of $\mathbf{B}_1$

From the expression for $\mathcal{J}$ derived in Section D, note that:

$$\mathbf{B}_1 = \mathbf{T} \otimes \mathbb{I} \tag{19}$$

where $\otimes$ is the Kronecker product of two matrices and

$$\mathbf{T} = \begin{bmatrix} (1 - \sum_{K \neq S_y, E} \alpha_{S_y K}) & \alpha_{KS_y} & \dots \\ \vdots & \vdots & \vdots \\ \alpha_{S_y U} & \alpha_{KU} & \dots \\ \vdots & \ddots & \vdots \end{bmatrix} \tag{20}$$

We know from matrix algebra [28] that if $\mathbf{C} = \mathbf{D} \otimes \mathbf{E}$ then $\mathbf{C}_\lambda = \mathbf{D}_\lambda \otimes \mathbf{E}_\lambda$, where $\mathbf{C}_\lambda$ denotes a diagonal matrix with eigenvalues of the matrix $\mathbf{C}$ on the diagonal. But $\mathbb{I}_\lambda = \mathbb{I}$, hence the eigenvalues of $\mathbf{B}_1$ are the same as the eigenvalues of $\mathbf{T}$ (although with repetition). In other words, eigenvalues of $\mathbf{T}$ are eigenvalues of $\mathcal{J}$ as well.

E.2 Eigenvalues of $\mathbf{B}_3$

Let $\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$ be a corresponding eigenvector of $\mathbf{B}_3$ ($\mathbf{u}_1$ and $\mathbf{u}_2$ are of size $N \times 1$ each and as the eigenvalues of $\mathbf{B}_3$ are also eigenvalues of $\mathcal{J}$, we use $\lambda_\mathcal{J}$ for an eigenvalue of $\mathbf{B}_3$). Hence, the standard eigenvalue relation $\mathbf{B}_3\mathbf{u} = \lambda_\mathcal{J}\mathbf{u}$ requires the following equations to be satisfied:

$$(\alpha_{EE}\mathbb{I} + p_S^*\beta_1\mathbf{A})\mathbf{u}_1 + (\alpha_{IE}\mathbb{I} + p_S^*\beta_2\mathbf{A})\mathbf{u}_2 = \lambda_\mathcal{J}\mathbf{u}_1 \tag{21}$$

$$\alpha_{EI}\mathbf{u}_1 + \alpha_{II}\mathbf{u}_2 = \lambda_\mathcal{J}\mathbf{u}_2 \tag{22}$$

Using Equation 22, we can compute $\mathbf{u}_1$ in terms of $\mathbf{u}_2$ as:

$$\mathbf{u}_1 = \left(\frac{\lambda_\mathcal{J} - \alpha_{II}}{\alpha_{EI}}\right)\mathbf{u}_2 \tag{23}$$

Substituting it back into Equation 21 we get:

$$\left((\alpha_{EE}\mathbb{I} + p_S^*\beta_1\mathbf{A})\left(\frac{\lambda_\mathcal{J} - \alpha_{II}}{\alpha_{EI}}\right) + \alpha_{IE}\mathbb{I} + p_S^*\beta_2\mathbf{A}\right)\mathbf{u}_2 = \lambda_\mathcal{J}\left(\frac{\lambda_\mathcal{J} - \alpha_{II}}{\alpha_{EI}}\right)\mathbf{u}_2$$

$$\Rightarrow (\alpha_{EE}(\lambda_\mathcal{J} - \alpha_{II})\mathbb{I} + \alpha_{IE}\alpha_{EI}\mathbb{I} + (p_S^*\beta_1(\lambda_\mathcal{J} - \alpha_{II}) + p_S^*\beta_2\alpha_{EI})\mathbf{A})\,\mathbf{u}_2 = \lambda_\mathcal{J}(\lambda_\mathcal{J} - \alpha_{II})\mathbf{u}_2$$

which finally gives,

$$\mathbf{A}\mathbf{u}_2 = \left(\frac{\lambda_\mathcal{J}^2 - (\alpha_{II} + \alpha_{EE})\lambda_\mathcal{J} + \alpha_{II}\alpha_{EE} - \alpha_{IE}\alpha_{EI}}{p_S^*\beta_1(\lambda_\mathcal{J} - \alpha_{II}) + p_S^*\beta_2\alpha_{EI}}\right)\mathbf{u}_2 \tag{24}$$

Again, Equation 24 tells us that either $\mathbf{u}_2 = \mathbf{0}$ or it is an eigenvector for $\mathbf{A}$. But $\mathbf{u}_2 = \mathbf{0} \Rightarrow \mathbf{u}_1 = \mathbf{0} \Rightarrow \mathbf{u} = \mathbf{0}$ which is not possible. Thus Equation 24 is an eigenvalue equation for the adjacency matrix $\mathbf{A}$ and we are looking for solutions $\lambda_\mathcal{J}$ and $\mathbf{u}_2$ such that they satisfy it. Hence,

$$\lambda_\mathbf{A} = \frac{\lambda_\mathcal{J}^2 - (\alpha_{II} + \alpha_{EE})\lambda_\mathcal{J} + \alpha_{II}\alpha_{EE} - \alpha_{IE}\alpha_{EI}}{p_S^*\beta_1(\lambda_\mathcal{J} - \alpha_{II}) + p_S^*\beta_2\alpha_{EI}}$$

where $\lambda_\mathbf{A}$ is an eigenvalue of $\mathbf{A}$. This finally gives

$$\lambda_\mathcal{J}^2 - \lambda_\mathcal{J}(\alpha_{EE} + \alpha_{II} + p_S^*\beta_1\lambda_\mathbf{A}) + (\alpha_{II}\alpha_{EE} - \alpha_{IE}\alpha_{EI} + p_S^*\lambda_\mathbf{A}(\beta_1\alpha_{II} - \beta_2\alpha_{EI})) = 0 \tag{25}$$

Thus we have a different quadratic equation (*Q.E.*) for each eigenvalue $\lambda_\mathbf{A}$ of $\mathbf{A}$. Each *Q.E.* gives us two eigenvalues (possibly repeated) of $\mathcal{J}$.

So, finally, we can conclude the following lemma:

Lemma 1 (Eigenvalues of $\mathcal{J}$) *Eigenvalues of $\mathcal{J}$ are given by the eigenvalues of $\mathbf{T}$ (Equations 19 and 20) and the roots of the Q.Es given by Equation 25 for each eigenvalue $\lambda_\mathbf{A}$ of $\mathbf{A}$.*

F Stability

We require that all the eigenvalues of $\mathcal{J}$ to be less than 1 in absolute value (according to Theorem 2). From Lemma 1, we have two cases to handle in enforcing this:

(C1) All the eigenvalues of $\mathbf{T}$ should be less than 1 in absolute value
(C2) All the roots of the *Q.Es* given by Equation 25 for each eigenvalue $\lambda_{\mathbf{A}}$ of $\mathbf{A}$ should be less than 1 in absolute value

F.1 Case C1

Note that this case depends *only* on the model as the matrix $\mathbf{T}$ is *independent* of the adjacency matrix $\mathbf{A}$. But $\mathbf{T}$ is a stochastic matrix i.e. all the column sums are equal to 1 - consequently all its eigenvalues are less than 1 in absolute value.

Lemma 2 (Stability C1) *All eigenvalues of the matrix* $\mathbf{T}$ *(given by Equation 20) are less than 1 in absolute value.*

F.2 Case C2

As **C1** is always true, we need to only ensure case **C2**. We can prove here the following:

Lemma 3 (Stability C2) *All the roots of the Q.Es given by Equation 25 for each eigenvalue* $\lambda_{\mathbf{A}}$ *of* $\mathbf{A}$ *are less than 1 in absolute value if:*

$$\lambda_1 p_S^* \left(\frac{\beta_1(1-\alpha_{II}) + \beta_2\alpha_{EI}}{(1-\alpha_{II})(1-\alpha_{EE}) - \alpha_{IE}\alpha_{EI}} \right) < 1$$

Proof Let r_1 and r_2 be the roots of Equation 25 (r_1 and r_2 can be real or complex depending on $\lambda_{\mathbf{A}}$). Then we want

$$|r_1| < 1 \text{ and } |r_2| < 1$$

r_1 and r_2 are real As the roots are real, $\lambda_{\mathbf{A}}$ is such that the discriminant $\mathcal{D}$ of the quadratic equation is greater than zero. In this situation:

$$|r_1| < 1 \text{ and } |r_2| < 1 \Rightarrow r_1 \in (-1,1) \text{ and } r_2 \in (-1,1) \tag{26}$$

From the theory of quadratic equations, it is well known (see e.g. [40]) that for real roots x_1 and x_2 of a *Q.E.* $f(x) = ax^2 + bx + c$ (with $a > 0$) to lie in the interval $(-1, 1)$ the following conditions must be true:

$$\begin{aligned} a - c &> 0, \\ a - b + c &> 0, \\ a + b + c &> 0. \end{aligned}$$

Intuitively, the first condition forces the product of the roots to be less than 1 while the last two conditions state that value of $f(x)$ at -1 and 1 should not be "too small". In our case, these then translate into:

$$\alpha_{II}\alpha_{EE} - \alpha_{IE}\alpha_{EI} + p_S^*\lambda_{\mathbf{A}}(\beta_1\alpha_{II} - \beta_2\alpha_{EI}) < 1 \tag{27a}$$

$$1 + \alpha_{EE} + \alpha_{II} + p_S^*\beta_1\lambda_{\mathbf{A}} + \alpha_{II}\alpha_{EE} - \alpha_{IE}\alpha_{EI} + p_S^*\lambda_{\mathbf{A}}(\beta_1\alpha_{II} - \beta_2\alpha_{EI}) > 0 \tag{27b}$$

$$1 - \alpha_{EE} - \alpha_{II} - p_S^*\beta_1\lambda_{\mathbf{A}} + \alpha_{II}\alpha_{EE} - \alpha_{IE}\alpha_{EI} + p_S^*\lambda_{\mathbf{A}}(\beta_1\alpha_{II} - \beta_2\alpha_{EI}) > 0 \tag{27c}$$

Equations 27b and 27c can be written as:

$$\lambda_{\mathbf{A}} p_S^* \left(\frac{-\beta_1(1+\alpha_{II}) + \beta_2\alpha_{EI}}{(1+\alpha_{II})(1+\alpha_{EE}) - \alpha_{IE}\alpha_{EI}} \right) < 1 \tag{28a}$$

$$\lambda_{\mathbf{A}} p_S^* \left(\frac{\beta_1(1-\alpha_{II}) + \beta_2\alpha_{EI}}{(1-\alpha_{II})(1-\alpha_{EE}) - \alpha_{IE}\alpha_{EI}} \right) < 1 \tag{28b}$$

respectively. The above equations should be true for *any* eigenvalue $\lambda_{\mathbf{A}}$ of $\mathbf{A}$ which makes $\mathcal{D} > 0$. Recall that we are considering only undirected graphs, hence $\mathbf{A}$ is a symmetric binary (0/1) square irreducible matrix. As a result firstly, all its eigenvalues are real. Secondly, from the Perron-Frobenius theorem [38] the algebraically largest eigenvalue λ_1 of $\mathbf{A}$ is a positive real number and also has the largest magnitude among all eigenvalues. Hence if the above equations are true for $\lambda_{\mathbf{A}} = \lambda_1$ we are done. Now note that

$$(1+\alpha_{II})(1+\alpha_{EE}) - \alpha_{IE}\alpha_{EI} > (1-\alpha_{II})(1-\alpha_{EE}) - \alpha_{IE}\alpha_{EI}$$

and that

$$\beta_1(1-\alpha_{II}) + \beta_2\alpha_{EI} > -\beta_1(1+\alpha_{II}) + \beta_2\alpha_{EI}$$

In addition the L.H.S in both equations is positive under $\lambda_{\mathbf{A}} = \lambda_1$. So Equation 28a is always true if Equation 28b holds (under $\lambda_{\mathbf{A}} = \lambda_1$) i.e.

$$\lambda_1 p_S^* \left(\frac{\beta_1(1-\alpha_{II}) + \beta_2\alpha_{EI}}{(1-\alpha_{II})(1-\alpha_{EE}) - \alpha_{IE}\alpha_{EI}} \right) < 1 \tag{29}$$

As λ_1 is the largest eigenvalue both algebraically and in magnitude, under Equation 29,

$$\begin{aligned}
&1 - \alpha_{II}\alpha_{EE} + \alpha_{IE}\alpha_{EI} - p_S^*\lambda_{\mathbf{A}}(\beta_1\alpha_{II} - \beta_2\alpha_{EI}) \\
&> 1 - \alpha_{II}\alpha_{EE} + \alpha_{IE}\alpha_{EI} - \left(\frac{(1-\alpha_{II})(1-\alpha_{EE}) - \alpha_{IE}\alpha_{EI}}{\beta_1(1-\alpha_{II}) + \beta_2\alpha_{EI}} \right) (\beta_1\alpha_{II} - \beta_2\alpha_{EI}) \\
&= \frac{\left((1-\alpha_{II})^2 + \alpha_{IE}\alpha_{EI}\right)\beta_1 + \alpha_{EI}(2 - \alpha_{II} - \alpha_{EE})\beta_2}{(1-\alpha_{II})\beta_1 + \alpha_{EI}\beta_2} \\
&> 0
\end{aligned}$$

$\therefore$ Equation 27a is also true if Equation 29 holds. Thus the condition for the roots to be in $(-1, 1)$ when they are real is given simply by Equation 29.

r_1 and r_2 are complex In this case $\lambda_{\mathbf{A}}$ is such that $\mathcal{D} < 0$. Also as Equation 25 has real co-efficients, r_1 and r_2 are complex conjugate of each other and so $|r_1| = |r_2| = \sqrt{r_1 \cdot r_2}$. But the product of roots x_1 and x_2 of the equation $ax^2 + bx + c = 0$ is equal to c/a. Hence we want to enforce $c/a < 1$. In our case it is

$$\alpha_{II}\alpha_{EE} - \alpha_{IE}\alpha_{EI} + p_S^*\lambda_{\mathbf{A}}(\beta_1\alpha_{II} - \beta_2\alpha_{EI}) < 1$$

which is exactly Equation 27a. From the above analysis, we already know that it is true if Equation 29 holds. So, for any eigenvalue $\lambda_{\mathbf{A}}$ for which $\mathcal{D} < 0$, the roots have magnitude less than 1 given Equation 29 is true.

Thus in both cases, whether roots are real or complex, Equation 29 is a sufficient condition for the roots to have magnitude less than 1. □

To re-cap we state our result and then give its proof:

Theorem 3 (G2 theorem) *For virus propagation models which satisfy our general initial assumptions and for any arbitrary undirected graph with adjacency matrix* $\mathbf{A}$ *and largest eigenvalue* λ_1*, the sufficient condition for stability is given by:*

$$s < 1$$

where, s (the effective strength) is:

$$s = \lambda_1 \cdot C$$

and C is a constant dependent on the model (given by Equation 30). Hence, the tipping point is reached when $s = 1$.

Proof Lemma 2 and Lemma 3 ensure cases C1 and C2 and hence together with Lemma 1 imply that the eigenvalues of the Jacobian $\mathcal{J}$ of our general NLDS computed at the fixed point $\mathbf{x}$ are less than 1 in magnitude if Equation 29 is true.

$\therefore$ using Theorem 2, our general NLDS is stable at its fixed point $\mathbf{x}$ if Equation 29 holds. Recall that $\mathbf{x}$ is the point when there no infected nodes in the system (Appendix C) and that this is the fixed point whose stability conditions determine the epidemic threshold (Section 5).

$\therefore$ finally we can conclude the theorem with

$$C_{\text{VPM}} = p_S^* \left(\frac{\beta_1(1-\alpha_{II}) + \beta_2 \alpha_{EI}}{(1-\alpha_{II})(1-\alpha_{EE}) - \alpha_{IE}\alpha_{EI}} \right) \tag{30}$$

and the effective strength $s = \lambda_1 \cdot C_{\text{VPM}}$. The parameter C_{VPM} is a constant for a given propagation model while the only parameter involved from the underlying contact-network is λ_1, the first eigenvalue of the adjacency matrix. □

Acknowledgements 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, CNS-0721889, CNS-0832069 and IIS-1017415 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.

References

1. Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. Nature 407(6794):378–482
2. Anderson RM, May RM (1991) Infectious Diseases of Humans. Oxford University Press
3. Barrat A, Barthélemy M, Vespignani A (2010) Dynamical Processes on Complex Networks. Cambridge University Press, Cambridge, U.K.
4. Barrett CL, Bisset KR, Eubank SG, Feng X, Marathe MV (2008) Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. ACM/IEEE Conf on Supercomputing
5. Bass FM (1969) A new product growth for model consumer durables. Management Science 15(5):215–227
6. Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change in informational cascades. Journal of Political Economy 100(5):992–1026
7. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic Cascade of Failures in Interdependent Networks. Nature 464(7291):1025–1028
8. Castellano C, Pastor-Satorras R (2010) Thresholds for epidemic spreading in networks. Phys Rev Let 105
9. Chakrabarti D, Leskovec J, Faloutsos C, Madden S, Guestrin C, Faloutsos M (2007) Information Survival Threshold in Sensor and P2P Networks. In: IEEE INFOCOM

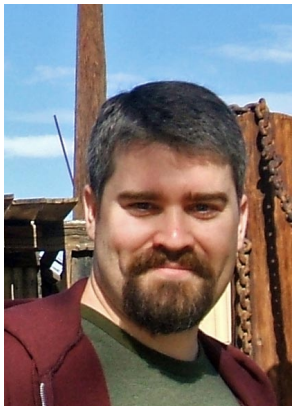
10. Chakrabarti D, Wang Y, Wang C, Leskovec J, Faloutsos C (2008) Epidemic thresholds in real networks. ACM TISSEC 10(4)
11. Chen Q, Chang H, Govindan R, Jamin S, Shenker SJ, Willinger W (2002) The origin of power laws in internet topologies revisited. IEEE INFOCOMM URL http://topology.eecs.umich.edu/data.html
12. Chen W, Wang C, Wang Y (2010) Scalable influence maximization for prevalent viral marketing in large-scale social networks. KDD
13. Chung F, Lu L, Vu V (2003) Eigenvalues of random power law graphs. Annals of Combinatorics 7(1)
14. Cohen R, Havlin S, ben Avraham D (2003) Efficient immunization strategies for computer networks and populations. Physical Review Letters 91(24):247,901
15. Dodds PS, Watts DJ (2004) A generalized model of social and biological contagion. Journal of Theoretical Biology 232:587–604
16. Easley D, Kleinberg J (2010) Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press
17. Eubank S, Guclu H, Anil Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. Nature 429(6988):180–184
18. Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the internet topology. SIGCOMM pp 251–262
19. Ganesh A, Massoulié L, Towsley D (2005) The effect of network topology on the spread of epidemics. In: INFOCOM
20. Goldenberg J, Libai B, Muller E (2001) Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters
21. Gomez Rodriguez M, Leskovec J, Krause A (2010) Inferring networks of diffusion and influence. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, NY, USA, KDD '10, pp 1019–1028
22. Granovetter M (1978) Threshold models of collective behavior. Am Journal of Sociology 83(6):1420–1443
23. Gruhl D, Guha R, Liben-Nowell D, Tomkins A (2004) Information diffusion through blogspace. In: WWW '04, URL www.www2004.org/proceedings/docs/1p491.pdf
24. Hayashi Y, Minoura M, Matsukubo J (2003) Recoverable prevalence in growing scale-free networks and the effective immunization. arXiv:cond-mat/0305549 v2
25. Hethcote HW (2000) The mathematics of infectious diseases. SIAM Review 42
26. Hethcote HW, Yorke JA (1984) Gonorrhea transmission dynamics and control. Springer Lecture Notes in Biomathematics 46
27. Hirsch MW, Smale S (1974) Differential Equations, Dynamical Systems and Linear Algebra. Academic Press
28. Horn RA, Johnson CR (1991) Topics in Matrix Analysis. Cambridge University Press
29. Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: KDD
30. Kephart JO, White SR (1993) Measuring and modeling computer virus prevalence. IEEE Computer Society Symposium on Research in Security and Privacy
31. Kleinberg J (Sep 2007) The wireless epidemic. Nature, Vol 449
32. Kumar R, Novak J, Raghavan P, Tomkins A (2003) On the bursty evolution of blogspace. In: WWW
33. Lad M, Zhao X, Zhang B, Massey D, Zhang L (2003) Analysis of BGP Update Surge during Slammer Worm Attack. In: 5th International Workshop on Distributed Computing (IWDC), URL citeseer.ist.psu.edu/lad03analysis.html
34. Lappas T, Terzi E, Gunopoulos D, Mannila H (2010) Finding effectors in social networks. SIGKDD
35. Leskovec J, Adamic LA, Huberman BA (2006) The dynamics of viral marketing. In: EC, ACM Press, New York, NY, USA, pp 228–237
36. Leskovec J, Backstrom L, Kleinberg J (2009) Meme-tracking and the dynamics of the news cycle. ACM SIGKDD

37. Maymounkov P, Mazières D (2002) Kademlia: A Peer-to-Peer Information System Based on the XOR Metric. In: Revised Papers from the First International Workshop on Peer-to-Peer Systems, Springer-Verlag, London, UK, IPTPS '01, pp 53–65, URL http://dl.acm.org/citation.cfm?id=646334.687801
38. McCuler CR (2000) The many proofs and applications of perron's theorem. SIAM Review 42
39. McKendrick AG (1926) Applications of mathematics to medical problems. In: Proceedings of Edin. Math. Society, vol 14, pp 98–130
40. Milnes HW (Aug. - Sept. 1963) Conditions that the zeros of a polynomial lie in the interval [-1, 1] when all zeros are real. The American Mathematical Monthly 70, No. 7
41. NDSSL (2007) Synthetic Data Products for Societal Infrastructures and Protopopulations: Data Set 2.0. NDSSL-TR-07-003 URL http://ndssl.vbi.vt.edu/Publications/ndssl-tr-07-003.pdf
42. Newman MEJ (2002) Spread of epidemic disease on networks. Phys Rev E 66(1):016,128, DOI 10.1103/PhysRevE.66.016128
43. Newman MEJ (2005) Threshold effects for two pathogens spreading on a network. Physical Review Letters 95(10):108,701
44. Pastor-Santorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Physical Review Letters 86 14
45. Pathak N, Banerjee A, Srivastava J (2010) A generalized linear threshold model for multiple cascades. ICDM
46. Prakash BA, Tong H, Valler N, Faloutsos M, Faloutsos C (2010) Virus propagation on time-varying networks: Theory and immunization algorithms. ECML-PKDD
47. Prakash BA, Beutel A, Rosenfeld R, Faloutsos C (2012) Winner takes all: Competiting viruses or ideas on fair-play networks. WWW
48. Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. SIGKDD
49. Rogers EM (2003) Diffusion of Innovations, 5th Edition. Free Press, URL http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20\&path=ASIN/0743222091
50. Saito K, Kimura M, Ohara K, Motoda H (2012) Efficient discovery of influential nodes for sis models in social networks. Knowledge and Information Systems (KAIS) 30(3):613–635
51. Vojnovic M, Gupta V, Karagiannis T, Gkantsidis C (2008) Sampling strategies for epidemic-style information dissemination. IEEE INFOCOM
52. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442
53. (Website) (November 15, 2011) Mainline bittorrent website URL http://www.bittorrent.com/
54. Zhao J, Wu J, Feng X, Xiong H, Xu K (2011) Information propagation in online social networks: a tie-strength perspective. Knowledge and Information Systems pp 1–20, URL http://dx.doi.org/10.1007/s10115-011-0445-x

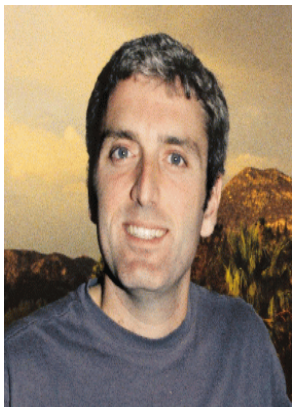
Author Biographies

B. Aditya Prakash is a Ph.D. student in the Computer Science Department at Carnegie Mellon University. He got his B.Tech (in CS) from the Indian Institute of Technology (IIT) - Bombay. He has published 16 refereed papers in major venues and holds two U.S. patents. His interests include Data Mining, Applied Machine Learning and Databases, with emphasis on large real-world networks and streams.

Deepayan Chakrabarti obtained his B.Tech. from IIT Kanpur, and his Ph.D. from Carnegie Mellon University. He has worked on graph mining problems, including clustering and community detection on graphs, and building realistic graph generators. He is interested in applying such techniques on large graph datasets such as the Web and large social networks. Recently, he has also been working on bandit formulations for online learning.



Nicholas C. Valler is a Ph.D. candidate in the Computer Science and Engineering Department at the University of California, Riverside. He received his bachelor's degree from the California State University, Long Beach in 2007. His research interests include understanding the affect of network topology on dynamical systems, network routing protocols, P2P networks and network security.



Michalis Faloutsos is a faculty member at the Computer Science Department in University of California, Riverside. He got his bachelor's degree at the National Technical University of Athens and his M.Sc and Ph.D. at the University of Toronto. His interests include, Internet protocols and measurements, peer-to-peer networks, network security, BGP routing, and ad-hoc networks. With his two brothers, he co-authored the paper on powerlaws of the Internet topology (SIGCOMM'99, which is one of the top ten most cited papers of 1999). His most recent work on peer-to-peer measurements have been widely cited in popular printed and electronic press such as slashdot, ACM Electronic News, USA Today, and Wired. He also works in the area of Internet routing (BGP), and ad hoc networks routing, and network security, with emphasis on routing.

Christos Faloutsos is a Professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, the SIGKDD Innovations Award (2010), seventeen "best paper" awards, (including two "test of time") and four teaching awards. He has served as a member of the executive committee of SIGKDD; he is an ACM Fellow; he has published over 200 refereed articles, 11 book chapters and one monograph. He holds five patents and he has given over 30 tutorials and over 10 invited distinguished lectures. His research interests include data mining for graphs and streams, fractals, database performance, and indexing for multimedia and bioinformatics data.