Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks

B. Aditya Prakash*, Deepayan Chakrabarti†, Michalis Faloutsos‡, Nicholas Valler‡, Christos Faloutsos*
*Carnegie Mellon University, Email:{badityap, christos}@cs.cmu.edu
†Yahoo! Research, Email:{deepay}@yahoo-inc.com
‡University of California - Riverside, Email:{michalis, nvaller}@cs.ucr.edu

***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.*) [20], [12]. 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 policies. We also demonstrate our result using extensive simulations 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.**

I. 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 [20], [12], on *arbitrary* contact-networks.

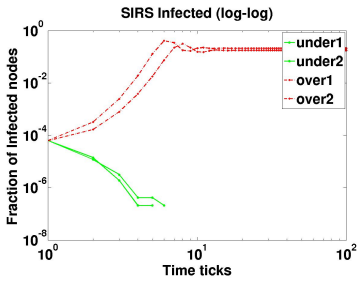


Figure 1. Qualitatively different infection time-series curves (Fraction of Infected population vs Time) for the SIRS model (temporary immunity, like pertussis) on a large contact-network. What is the condition that separates the two regimes - red (epidemic) vs green (extinction)?

Much of previous work focuses on either special types of graphs (typically cliques [25], block-structure and hierarchical graphs [21] and random power-law graphs [34]) or on specific epidemiological models [8]. 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 § VIII) 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 [24] 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 [17]) in sociology are similar in reality to cascade models [11]. 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 [26], [19]. 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 [39] 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 [29].

The rest of the paper is organized as follows: we first give the related work in § II, then formulate the problem (§ III) and state our main result (§ IV), give a proof roadmap and example (§ V) and then show simulation experiments (§ VI) to demonstrate the result. We discuss the broad implications and many applications of the result in § VII and § VIII. We then conclude (§ IX) and finally give a detailed example proof in the Appendix.

II. Related Work

We review related work here, which can be categorized into two parts: epidemic thresholds, and information diffusion. None of these works generalize in two directions: for arbitrary propagation models and arbitrary networks.

Epidemic Thresholds Canonical texts for epidemiology include [2], [20]. The most widely-studied epidemiological models include the so-called *homogeneous models* [30], 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 [25] 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 [14] etc. follow a power law structure instead. Pastor-Satorras and Vespignani [34] 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 [7] 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 [33], [32] 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. [8] and Ganesh et.al [15] 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.

Information diffusion There is a lot of research interest in studying dynamic processes on large graphs, (a) blogs and propagations [18], [27], [24], [37], (b) information cascades [6], [16], [17] and (c) marketing and product penetration [38], [28]. These dynamic processes are all closely related to virus propagation, with many directly based on epidemiological models [5], [24] e.g. the award-winning linear-cascade model [24] is a *special* case of our model : specifically it is essentially a SIR model with $\delta = 1$ and all our results carry through.

III. Problem Formulation

Table I and Table II 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:

Epidemic Threshold Problem *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 I
Common Terminology

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 $\mathbf{A}$

IV. Results

The epidemic threshold is usually defined as the minimum level of virulence to prevent a viral contagion from dying

Table II
SOME VIRUS PROPAGATION MODELS (VPMS)

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

out quickly [2], [20], [3], [26]. 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 [20], [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 V 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:*

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

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

Proof. Full proof in our extended version [35]. We give a roadmap in the next section and a detailed example 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 the graph), random Erdős-Rényi graphs with expected degree d ($\lambda_1 = d$), 'homogeneous' graphs [25], power-law/scale-free graphs [34], structured hierarchical (near-block-diagonal) topologies [21] (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 [20], [12] are specific instantiations of the generalized model used in our theorem (see Figure 2; more later). Table III 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. We discuss our terminology and general model next.

V. PROOF OVERVIEW

We first construct a generalized model ($\mathrm{S}^*\mathrm{I}^2\mathrm{V}^*$- 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 an illustration, we then discuss the result on the SEIV model. Finally, we give a more detailed example proof for the SEIV model in the Appendix: we believe it exemplifies the key aspects of our general proof. As mentioned before, the full proof can be found in our extended version [35].

A. 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

Table III

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.)	$s = \lambda_1 \cdot \left(\frac{\beta_1 v_2+\beta_2\epsilon}{v_2(\epsilon+v_1)}\right)$	

Threshold results for some models. 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.

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.

B. 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) like SIS, SIR, SIRS, SEIR, SERIS, MSIR, MSEIR etc. [20], [12], 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) which itself generalizes many known models (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).

C. Proof Sketch

We define the vector $\tilde{\mathbf{P}}_t$ 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 $\tilde{\mathbf{P}}_t \in \Re^{m \cdot N \times 1}$, which captures the probability of each node being in any of m states at a given time t. Specifically:

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

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 $\tilde{\mathbf{P}}_{t+1} = g(\tilde{\mathbf{P}}_t)$ 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 $\tilde{\mathbf{P}}_t$ over time. An equilibrium point (also called a fixed point) of the system is the state vector (i.e. some particular $\tilde{\mathbf{P}}$) which does not change. Thus at the equilibrium point $\tilde{\mathbf{P}}_{t+1} = \tilde{\mathbf{P}}_t = \tilde{\mathbf{x}}$. 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

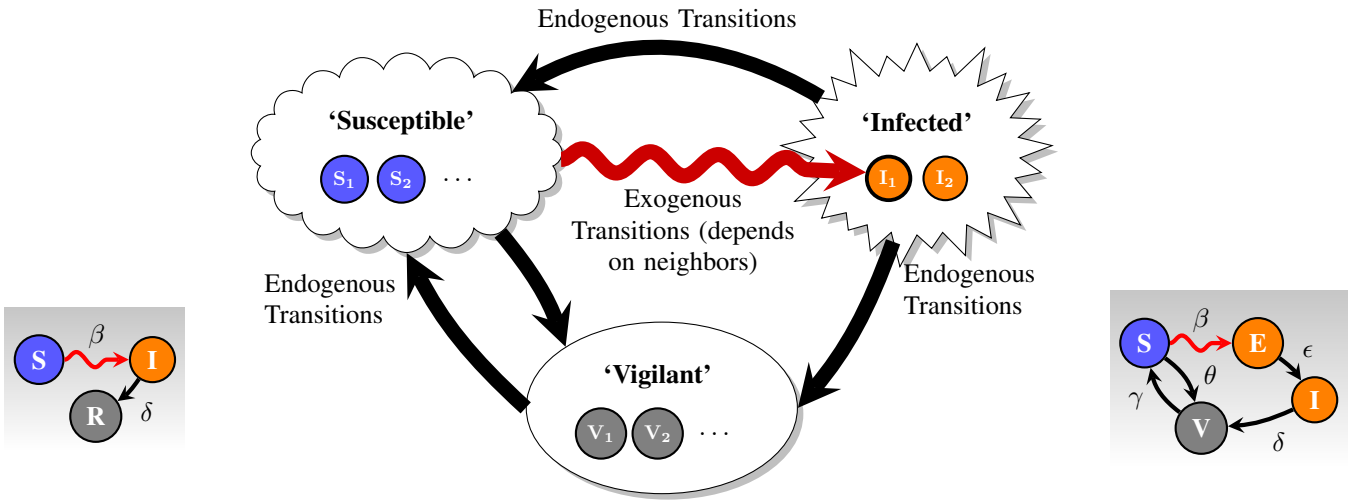


Figure 2. State Diagram for a node in the graph in our generalized model $S^*I^2V^*$- it is *not* a simple Markov chain. There are three classes (types) of states - Susceptible (healthy but can get infected), Infected (capable of transmission) and Vigilant (healthy and can't get infected). Within-class 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 [20] (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).

matrix at that point (i.e. $\bigtriangledown g(\tilde{\mathbf{x}})$). We eventually reduced the requirement on the eigenvalues of $\bigtriangledown g(\tilde{\mathbf{x}})$ 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 $S^*I^2V^*$model described before.

D. Proof Example: SEIV

We give a detailed example proof for the SEIV model in the Appendix as we believe it exemplifies the key aspects of our general proof. We can check that SEIV is a special case of our general $S^*I^2V^*$ model satisfying the assumptions. 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 VII.

VI. Experiments

We used one of the biggest available physical contact graphs, PORTLAND, representing a synthetic population of the city of Portland, Oregon, USA [31] for our experiments. It is a social-contact graph containing more than *31 million* links (interactions) among about *1.6 million* 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 [13].

Figure 3 illustrates our result via computer 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 notice the initial "silent" period for above threshold because of virus-incubation (presence of the Exposed state) in SEIV and SEIR. 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 3 (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.

VII. 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.

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 graph. More connected the graph

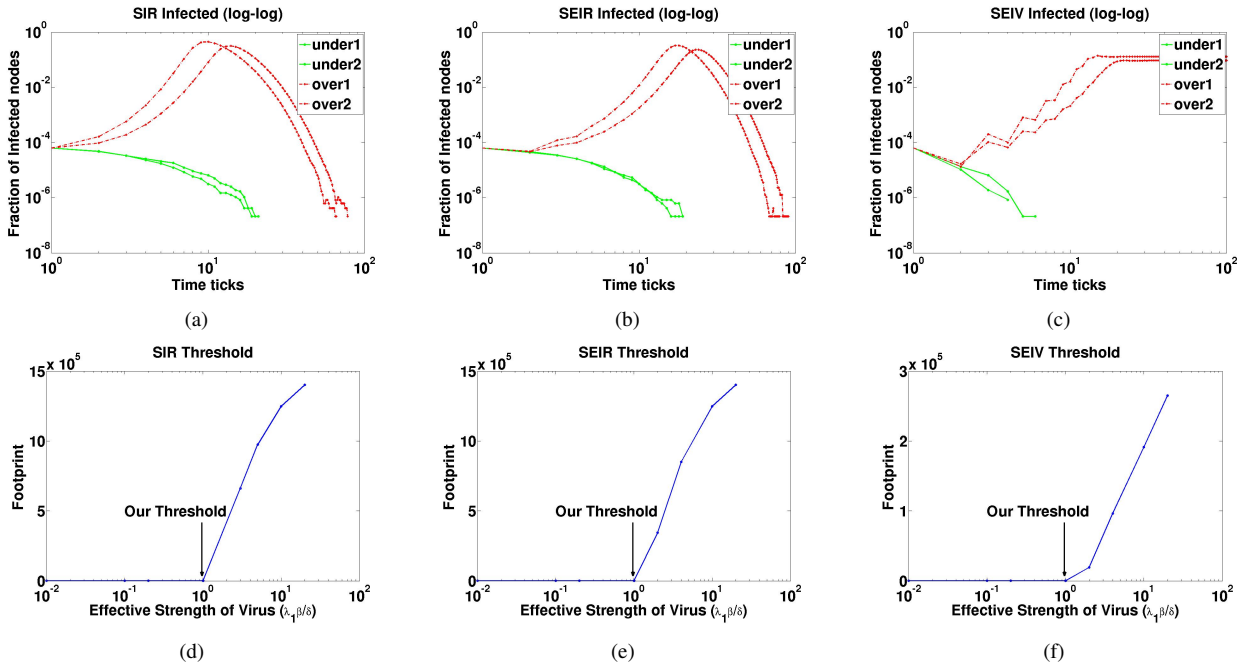


Figure 3. Simulation Results on the PORTLAND graph, all values averages over 100 runs. **(a),(b),(c)** Plot of Infective Fraction of Population vs Time (log-log) for SIR, SEIR, and SEIV models. Note the qualitative difference in behavior- two curves *under* (green) the threshold and two curves *above* (red) the threshold. **(d),(e),(f)** "Take-off" plots, *Footprint* (see Section VI) vs Effective Strength (lin-log) for SIR, SEIR, and SEIV models. The tipping point exactly matches our prediction ($s = 1$) in all cases.

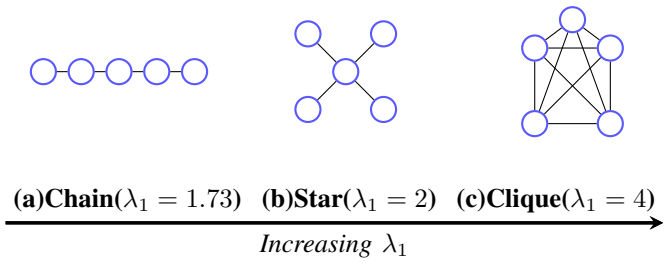


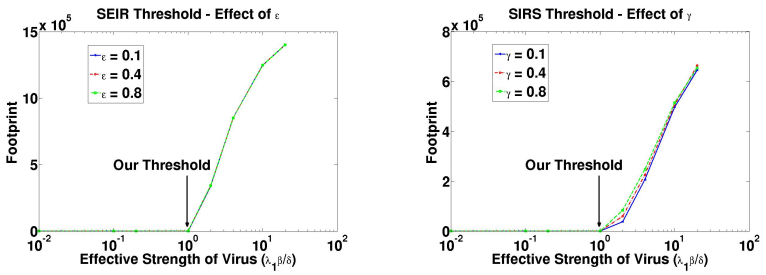
Figure 4. Why λ_1 matters more than number of edges E: changing connectivity and vulnerability of graphs with changing λ_1. The clique (largest λ_1) is the most vulnerable. Note that E is not enough: star and chain have the same number of edges ($E = 4$) but the star is intuitively more vulnerable, as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

is, more vulnerable it is to an epidemic by a virus (see Figure 4). 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 [40]: 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 [34], [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 [9], which also grows to infinity, and thus the threshold approaches zero.

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 5. 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 III), 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'.

VIII. Impact

Our results can be fundamental to a wide-range of applications. We mentioned broader impact in § I before. Here we briefly discuss some immediate applications in



(a) (b)

Figure 5. 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.

epidemiology.

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 [10] 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.

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.

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.

IX. 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], [36]), and (b) multiple competing diseases (extending the random power-law-graphs-only results of [33]).

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.

References

[1] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 407(6794):378–482, July 2000.

[2] R. M. Anderson and R. M. May. *Infectious Diseases of Humans*. Oxford University Press, 1991.

[3] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge, U.K., 2010.

[4] C. L. Barrett, K. R. Bisset, S. G. Eubank, X. Feng, and M. V. Marathe. Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. *ACM/IEEE Conf. on Supercomputing*, 2008.

[5] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[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] C. Castellano and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Phys. Rev. Let.*, 105, Dec. 2010.

[8] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM TISSEC*, 10(4), 2008.

[9] F. Chung, L. Lu, and V. Vu. Eigenvalues of random power law graphs. *Annals of Combinatorics*, 7(1), 2003.

[10] R. Cohen, S. Havlin, and D. ben Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 91(24):247901, December 2003.

[11] P. S. Dodds and D. J. Watts. A generalized model of social and biological contagion. *Journal of Theoretical Biology*, 232:587–604, September 2004.

[12] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[13] S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, May 2004.

[14] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *SIGCOMM*, pages 251–262, Aug-Sept. 1999.

[15] A. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM*, 2005.

[16] 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.

[17] M. Granovetter. Threshold models of collective behavior. *Am. Journal of Sociology*, 83(6):1420–1443, 1978.

[18] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW '04*, 2004.

[19] 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.

[20] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42, 2000.

[21] H. W. Hethcote and J. A. Yorke. Gonorrhea transmission dynamics and control. *Springer Lecture Notes in Biomathematics*, 46, 1984.

[22] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, 1974.

[23] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[24] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.

[25] J. O. Kephart and S. R. White. Measuring and modeling computer virus prevalence. *IEEE Computer Society Symposium on Research in Security and Privacy*, 1993.

[26] J. Kleinberg. The wireless epidemic. *Nature, Vol. 449*, Sep 2007.

[27] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. On the bursty evolution of blogspace. In *WWW*, 2003.

[28] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC*, pages 228–237, New York, NY, USA, 2006. ACM Press.

[29] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. *ACM SIGKDD*, 2009.

[30] A. G. McKendrick. Applications of mathematics to medical problems. In *Proceedings of Edin. Math. Society*, volume 14, pages 98–130, 1926.

[31] NDSSL. Synthetic Data Products for Societal Infrastructures and Protopopulations: Data Set 2.0. *NDSSL-TR-07-003*, 2007.

[32] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1):016128, Jul 2002.

[33] M. E. J. Newman. Threshold effects for two pathogens spreading on a network. *Physical Review Letters*, 95(10):108701, September 2005.

[34] R. Pastor-Santorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters 86*, 14, 2001.

[35] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos. Got the flu (or mumps)? check the eigenvalue! *arXiv:1004.0060v1 [physics.soc-ph]*, 2010.

[36] B. A. Prakash, H. Tong, N. Valler, M. Faloutsos, and C. Faloutsos. Virus propagation on time-varying networks: Theory and immunization algorithms. *ECML-PKDD*, 2010.

[37] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. *SIGKDD*, 2002.

[38] E. M. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, August 2003.

[39] M. Vojnovic, V. Gupta, T. Karagiannis, and C. Gkantsidis. Sampling strategies for epidemic-style information dissemination. *IEEE INFOCOM*, 2008.

[40] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

APPENDIX

We give a detailed example proof for the SEIV model as an illustration here. As noted before, we can check that SEIV is a special case of our general $S^*I^2V^*$ model satisfying the assumptions. To demonstrate generality in the proof, we assume that a node in the E state infects its neighbors with probability β_1 while a node in the I state infects a neighbor with β_2 probability (in SEIV, $\beta_1 = 0$ and $\beta_2 = \beta$). Due to this, much of the sample proof carries over without major changes to our general model. For sake of standard notation, we refer to the constant and given probability of transition from any state K to any state U as α_{KU} (hence, $\alpha_{SV} = \theta$, $\alpha_{VS} = \gamma$ etc.).

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 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) a neighbor is not in any of the infected states E and I; or (b) it is in state E and the transmission fails with probability $1-\beta_1$; or (c) 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:

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

where $\mathcal{NE}(i)$ is the set of neighbors of node i in the graph. A node i will be in S at time $t+1$ if: either (a) it was in S at time t and it did not receive any infections from its neighbors *and* it did not change state internally from S to V; or (b) it was in V and changed state internally from V to S. Hence, the probability of node i being in S at time $t+1$ is:

$$P_{S,i,t+1} = \alpha_{VS} P_{V,i,t} + P_{S,i,t}\,(\zeta_{i,t}(E,I) - \alpha_{SV}) \quad (5)$$

Similarly, for the E and I states:

$$\begin{aligned}P_{E,i,t+1} &= P_{S,i,t}\,(1 - \zeta_{i,t}(E,I)) \quad (6)\\ P_{I,i,t+1} &= \alpha_{EI} P_{E,i,t} + P_{S,i,t}\,(1 - \zeta_{i,t}(I)) \quad (7)\end{aligned}$$

Also, we can compute $P_{V,i,t}$ using the relation $\forall_{i,t}\ P_{S,i,t} + P_{E,i,t} + P_{I,i,t} + P_{V,i,t} = 1$. As discussed earlier (Equation 3), we can now define a probability vector $\tilde{\mathbf{P}}_t$ 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 $\tilde{\mathbf{P}}_{t+1} = g(\tilde{\mathbf{P}}_t)$. We have the following theorem about NLDS stability at a fixed point:

Theorem 2 (Asymptotic Stability, e.g. see [22])**.** *The system given by $\tilde{\mathbf{P}}_{t+1} = g(\tilde{\mathbf{P}}_t)$ is asymptotically stable at an equilibrium point $\tilde{\mathbf{P}} = \tilde{\mathbf{x}}$, if the eigenvalues of $\mathcal{J} = \bigtriangledown g(\tilde{\mathbf{x}})$ are less than 1 in absolute value, where, $\mathcal{J}_{i,j} = [\bigtriangledown g(\tilde{\mathbf{x}})]_{i,j} = \frac{\partial g_i}{\partial p_j}|_{\tilde{\mathbf{P}}=\tilde{\mathbf{x}}}$.*

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

Fixed point We are interested in the equilibrium point (i.e. where $\tilde{\mathbf{P}}_{t+1} = \tilde{\mathbf{P}}_t (= \tilde{\mathbf{x}})$) of the NLDS which corresponds to when no one is infected. Just the transition from S towards E is graph-based and can happen only when at least one of the nodes is in either E or I, so the state-diagram for each node will be a simple Markov chain consisting of S and V (call it MC_{SV}, see Figure 6). As there are no graph-based effects, each node is independent of others and will converge to the same steady state probabilities. The steady state vector π^* can be computed using Markov chain analysis: it will be a probability vector such that it is the left eigenvector corresponding to eigenvalue 1 of the MC_{SV} transition matrix:

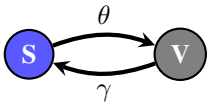


Figure 6. State Diagram at the fixed point for SEIV, when *no* node is present in E or I. Note that it is now a simple Markov chain with a unique steady state probability.

$$\pi^{*T} \begin{bmatrix} 1-\theta & \theta \\ \gamma & 1-\gamma \end{bmatrix} = \pi^* \ \ \& \ \ \sum_{i=1}^{2} \pi_i^* = 1$$
$$\Rightarrow \pi^{*T} = [\frac{\gamma}{\theta+\gamma}, \frac{\theta}{\theta+\gamma}]^T \quad (8)$$

Hence, the probability of being present in S and V for any node at the the fixed point is $x_S = \frac{\gamma}{\theta+\gamma}, x_I = \frac{\theta}{\theta+\gamma}$. Fixed point $\tilde{\mathbf{x}}$ of the global original NLDS can be constructed appropriately now.

The Jacobian We know from Theorem 2 that $\tilde{\mathbf{x}}$ is stable if the eigenvalues of $\mathcal{J} = \bigtriangledown g(\tilde{\mathbf{x}})$ are less than 1 in absolute value. From the definition of $\mathcal{J}$ we can see that it is a $4 \cdot$

$N \times 4 \cdot N$ matrix with 4 (for each state) square blocks of size $N \times N$ each (corresponding to every node in the graph). We can calculate $\mathcal{J}$ as below ($\mathbb{I}$ is the identity matrix of size $N \times N$, $\mathbf{0}_{N,N}$ is a $N \times N$ matrix with all zeros):

$$\begin{bmatrix} (1-\alpha_{SV})\mathbb{I} & \alpha_{VS}\mathbb{I} & -x_S\beta_1\mathbf{A} & -x_S\beta_2\mathbf{A} \\ \alpha_{SV}\mathbb{I} & 1-\alpha_{VS}\mathbb{I} & \mathbf{0}_{N,N} & \alpha_{IV}\mathbb{I} \\ \mathbf{0}_{N,N} & \mathbf{0}_{N,N} & \alpha_{EE}\mathbb{I}+x_S\beta_1\mathbf{A} & x_S\beta_2\mathbf{A} \\ \mathbf{0}_{N,N} & \mathbf{0}_{N,N} & \alpha_{EI}\mathbb{I} & \alpha_{II}\mathbb{I} \end{bmatrix} \quad (9)$$

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

$$\mathcal{J} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{0}_{2N,2N} & \mathbf{B}_3 \end{bmatrix}, \quad (10)$$

$$\text{with } \mathbf{B}_1 = T \otimes \mathbb{I} = \begin{bmatrix} 1-\alpha_{SV} & \alpha_{VS} \\ \alpha_{SV} & 1-\alpha_{VS} \end{bmatrix} \otimes \mathbb{I} \quad (11)$$

$$\mathbf{B}_3 = \begin{bmatrix} \alpha_{EE}\mathbb{I}+x_S\beta_1\mathbf{A} & x_S\beta_2\mathbf{A} \\ \alpha_{EI}\mathbb{I} & \alpha_{II}\mathbb{I} \end{bmatrix}$$

where $\otimes$ is the Kronecker product and $\mathbf{B}_2$ is defined similarly. Consider any eigenvector $\tilde{\mathbf{v}}$ (size $4N \times 1$) and corresponding eigenvalue $\lambda_{\mathcal{J}}$ of $\mathcal{J}$. We can write $\tilde{\mathbf{v}}$ as being composed of vectors $\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ of sizes $2N \times 1$ each i.e: $\tilde{\mathbf{v}}^T = [\tilde{\mathbf{v}}_1^T, \tilde{\mathbf{v}}_2^T]$. From $\mathcal{J}\tilde{\mathbf{v}} = \lambda_{\mathcal{J}}\tilde{\mathbf{v}}$ we get:

$$\begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{0} & \mathbf{B}_3 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{v}}_1 \\ \tilde{\mathbf{v}}_2 \end{bmatrix} = \lambda_{\mathcal{J}} \begin{bmatrix} \tilde{\mathbf{v}}_1 \\ \tilde{\mathbf{v}}_2 \end{bmatrix} \quad (12)$$

Equation 12 implies the following two relations:

$$\mathbf{B}_1\tilde{\mathbf{v}}_1 + \mathbf{B}_2\tilde{\mathbf{v}}_2 = \lambda_{\mathcal{J}}\tilde{\mathbf{v}}_1 \quad (13)$$

$$\mathbf{B}_3\tilde{\mathbf{v}}_2 = \lambda_{\mathcal{J}}\tilde{\mathbf{v}}_2 \quad (14)$$

From Equation 14 we can infer that precisely *one* of the following holds: (a) $\tilde{\mathbf{v}}_2 = \tilde{\mathbf{0}}$; or (b) $\tilde{\mathbf{v}}_2$ is the eigenvector of $\mathbf{B}_3$ (and consequently $\lambda_{\mathcal{J}}$ is the matching eigenvalue of $\mathbf{B}_3$). If $\tilde{\mathbf{v}}_2 = \tilde{\mathbf{0}}$, Equation 13 reduces to $\mathbf{B}_1\tilde{\mathbf{v}}_1 = \lambda_{\mathcal{J}}\tilde{\mathbf{v}}_1$ wherein again, either $\tilde{\mathbf{v}}_1 = \tilde{\mathbf{0}}$ or $\lambda_{\mathcal{J}}$ is an eigenvalue of $\mathbf{B}_1$. The condition $\tilde{\mathbf{v}}_1 = \tilde{\mathbf{0}}$ is not meaningful as then $\tilde{\mathbf{v}} = \tilde{\mathbf{0}}$ ($\tilde{\mathbf{v}}$ is an eigenvector of $\mathcal{J}$ implies $\tilde{\mathbf{v}}$ is non-zero). Therefore the eigenvalues of $\mathcal{J}$ are given by the eigenvalues of $\mathbf{B}_1$ (with $\tilde{\mathbf{v}}_2 = \tilde{\mathbf{0}}$) and the eigenvalues of $\mathbf{B}_3$ (note that $\mathbf{B}_2$ doesn't matter). We know from matrix algebra [23] 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 from Equation 11 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.

Let $\tilde{\mathbf{u}}^T = [\tilde{\mathbf{u}}_1^T, \tilde{\mathbf{u}}_2^T]$ be a corresponding eigenvector of $\mathbf{B}_3$ ($\tilde{\mathbf{u}}_1$ and $\tilde{\mathbf{u}}_2$ 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 $\mathbf{B}_3\tilde{\mathbf{u}} = \lambda_{\mathcal{J}}\tilde{\mathbf{u}}$ requires the following equations to be satisfied:

$$(\alpha_{EE}\mathbb{I}+x_S\beta_1\mathbf{A})\tilde{\mathbf{u}}_1 + (\alpha_{IE}\mathbb{I}+x_S\beta_2\mathbf{A})\tilde{\mathbf{u}}_2 = \lambda_{\mathcal{J}}\tilde{\mathbf{u}}_1$$

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

Eliminating $\tilde{\mathbf{u}}_1$, we get:

$$\mathbf{A}\tilde{\mathbf{u}}_2 = \left(\frac{\lambda_{\mathcal{J}}^2 - (\alpha_{II}+\alpha_{EE})\lambda_{\mathcal{J}} + \alpha_{II}\alpha_{EE}}{x_S\beta_1(\lambda_{\mathcal{J}} - \alpha_{II}) + x_S\beta_2\alpha_{EI}} \right) \tilde{\mathbf{u}}_2 \quad (15)$$

Again, Equation 15 tells us that either $\tilde{\mathbf{u}}_2 = \tilde{\mathbf{0}}$ or it is an eigenvector for $\mathbf{A}$. But $\tilde{\mathbf{u}}_2 = \tilde{\mathbf{0}} \Rightarrow \tilde{\mathbf{u}}_1 = \tilde{\mathbf{0}} \Rightarrow \tilde{\mathbf{u}} = \tilde{\mathbf{0}}$ which is not possible. Thus Equation 15 is an eigenvalue equation for the adjacency matrix $\mathbf{A}$ and we are looking for solutions $\lambda_{\mathcal{J}}$ and $\tilde{\mathbf{u}}_2$ such that they satisfy it. Hence,

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

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

$$\lambda_{\mathcal{J}}^2 - \lambda_{\mathcal{J}}(\alpha_{EE}+\alpha_{II}+x_S\beta_1\lambda_{\mathbf{A}}) + (\alpha_{II}\alpha_{EE} + x_S\lambda_{\mathbf{A}}(\beta_1\alpha_{II} - \beta_2\alpha_{EI})) = 0 \quad (16)$$

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 $\mathbf{T}$ (Equation 11) and the roots of the Q.Es given by Equation 16 for each eigenvalue $\lambda_{\mathbf{A}}$ of $\mathbf{A}$.*

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 need to handle two cases. We state the following lemmas without proof (omitted for lack of space).

Lemma 2 (Stability Case 1)**.** *All eigenvalues of the matrix $\mathbf{T}$ (given by Equation 11) are less than 1 in absolute value.*

Lemma 3 (Stability Case 2)**.** *All the roots of all the Q.Es given by Equation 16 for each eigenvalue $\lambda_{\mathbf{A}}$ of $\mathbf{A}$ are less than 1 in absolute value if:*

$$\lambda_1 x_S \left(\frac{\beta_1(1-\alpha_{II}) + \beta_2\alpha_{EI}}{(1-\alpha_{II})(1-\alpha_{EE})} \right) < 1 \quad (17)$$

Effective Strength for SEIV Lemma 2 and Lemma 3 together with Lemma 1 imply that the eigenvalues of the Jacobian $\mathcal{J}$ of our NLDS computed at the fixed point $\tilde{\mathbf{x}}$ are less than 1 in magnitude if Equation 17 is true. From Theorem 2, our NLDS is stable at its fixed point $\tilde{\mathbf{x}}$ if Equation 17 holds. Recall that $\tilde{\mathbf{x}}$ is the point when there are no infected nodes in the system and that this is the fixed point whose stability conditions determine the epidemic threshold. Thus, we get:

$$s = \lambda_1 \cdot x_S \left(\frac{\beta_1(1-\alpha_{II}) + \beta_2\alpha_{EI}}{(1-\alpha_{II})(1-\alpha_{EE})} \right) = \lambda_1 \cdot \frac{\beta\gamma}{\delta(\theta+\gamma)}$$

where we used the actual parameters of the SEIV model. □