CHAPTER 17

# Influence/Virus Propagation and Immunization

For a given, arbitrary graph, what is the epidemic threshold? That is, what are the conditions under which a virus (or rumor, or new product) will result in an epidemic, taking over the majority of the population as opposed to dying out? Prakash et al. [232] give the *super-model* theorem which generalizes older results in two important, orthogonal dimensions. The theorem shows that

- (a) for a wide range of *virus propagation models* (VPM) that include *all* virus propagation models in standard literature (say, [140][105]), and
- (b) for *any* contact graph, the answer always depends on the first eigenvalue of the connectivity matrix.

Here we give the theorem, arithmetic examples for popular VPMs (=Virus Propagation Models), like flu (SIS), mumps (SIR), SIRS, and more. We also discuss the implications of the theorem: easy (although sometimes *counter-intuitive*) answers to "what-if" questions, easier design and evaluation of immunization policies, and significantly faster agent-based simulations.

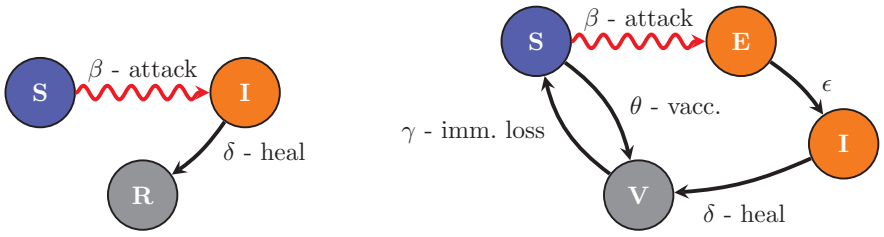
## 17.1 INTRODUCTION—TERMINOLOGY

Given a social or computer network, where the links represent who has the potential to infect whom, can we tell whether a virus will create an epidemic, as opposed to quickly becoming extinct? This is a fundamental question in epidemiology; intuitively, the answer should depend on (a) the graph connectivity, and (b) the virus propagation model (VPM), that is, how virulent is it, how quickly the host recovers (if ever), whether the host obtains (or is born with) immunity, how quickly (if ever) the host loses immunity, etc. This threshold is the level of virulence below which a virus is guaranteed of dying out quickly [168].

The overwhelming majority of earlier work focuses either on full-clique topologies (everybody contacts everybody else), or on "homogeneous" graphs [160, 161], or on power-law graphs [229] or hierarchical (near-block-diagonal) topologies [141] (people within a community contact all others in this community, with a few cross-community contacts). The only exception that examine arbitrary-topology graphs is [276] and its follow-up work [67], [120] which all focused on only a single model, the "SIS" one (flu-like, with no immunity).

The upcoming result shows that, irrespective of the virus propagation model, the effect of the underlying topology can be captured by just one parameter: the first eigenvalue $\lambda_1$ of the adjacency

matrix $\mathbf{A}$. In particular, it covers *all* models given in the standard survey by Hethcote [140], which includes models like SIS (no immunity, like flu – "susceptible, infected, susceptible") and SIR (life-time immunity, like mumps: "susceptible, infected, recovered"). It also includes numerous other cases like SIRS [105] (temporary immunity), as well as some useful generalizations SIV (vigilance/vaccination with temporary immunity) and SEIV (vigilance/vaccination with temporary immunity *and* virus incubation) and many more. A few of these models are shown in Figure 17.4, organized in a lattice: Models in child nodes are special cases of the model in the parent node.




(a) SIR model (like mumps): permanent immunity

(b) SEIV model: temporary immunity, incubation (state 'E'), and vaccination

**Figure 17.1:** Examples of virus propagation models: (a) SIR, like mumps, with life-time immunity. (b) SEIV – more complicated: temporary immunity (lost with prob. $\gamma$), vaccination (with prob. $\theta$), and "Exposed" (but not infectious – yet) state. Both models and many more, follow the first-eigenvalue threshold.

Informally, the main result is:

**Informal Theorem 1** *For any virus propagation model (VPM) in the published literature, operating on an underlying undirected contact-network of any arbitrary topology with adjacency matrix* $\mathbf{A}$*, the epidemic threshold depends only on the first eigenvalue*

$$\lambda_1$$

*of* $\mathbf{A}$ *and some constant* $C_{\text{VPM}}$ *that is determined by the virus propagation model.*

In a nutshell, the typical states are the following, out of which the first three are the most typical ones:

- 'S:' Susceptible/healthy
- 'I:' Infected (and infectious)
- 'R:' Removed/recovered – the node has immunity for life (or is deceased)
- 'V:' Vigilant: the node can not be infected (but may lose it's immunity, depending on the VPM)
- 'E:' Exposed: the node is not infectious, but it is a carrier of the virus, and it will eventually evolve to the "Infected/Infectious" state.

- 'M:' mother-inherited immunity, like a newborn that initially carries the mother's antibodies, but the node will eventually evolve to the "susceptible" state.

Also notice that we have a whole class of infected states (like 'I,' 'E'), where the node has the virus; a whole class of vigilant states (like 'R,' 'V,' 'M'), where the node is healthy and invincible. The up coming model is very general, exactly allowing full classes of susceptible, infected, and vigilant states, as we show in Figure 17.5.

| | |
|---|---|
| VPM | virus-propagation model |
| $\beta$ | attack/transmission probability over a contact-link |
| $\delta$ | healing probability once infected |
| $\gamma$ | immunization-loss probability once recovered (in SIRS) or vigilant (in SIV, SEIV) |
| $\epsilon$ | virus-maturation probability once exposed – hence, $1-\epsilon$ is the virus-incubation probability |
| $\theta$ | direct-immunization probability when susceptible |
| $\mathbf{A}$ | adjacency matrix of the underlying undirected contact-network |
| $N$ | number of nodes in the network |
| $\lambda_1$ | largest (in magnitude) eigenvalue of $\mathbf{A}$ |
| $s$ | effective strength of a epidemic model on a graph with adjacency matrix $\mathbf{A}$ |

**Figure 17.2:** Symbols and definitions

| | |
|---|---|
| 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 |

**Figure 17.3:** Some Virus Propagation Models (VPMs)

Figure 17.4 shows the generalization hierarchy for some common epidemic models. The brown colored nodes denote standard VPMs found in literature while the blue colored nodes denote generalizations introduce in [232]. Each VPM is a generalization of all the models below it, e.g., SIV is a generalization of SIRS, SIR, and SIS. The main generalization, $S^*I^2V^*$, is illustrated in Figure 17.5 and further discussed in [232].

## 17.2 MAIN RESULT AND ITS GENERALITY

The tipping point for each of the models captures a fundamental transition in the behavior of the system between the two extremes: a network-wide epidemic, versus a minor local disturbance that

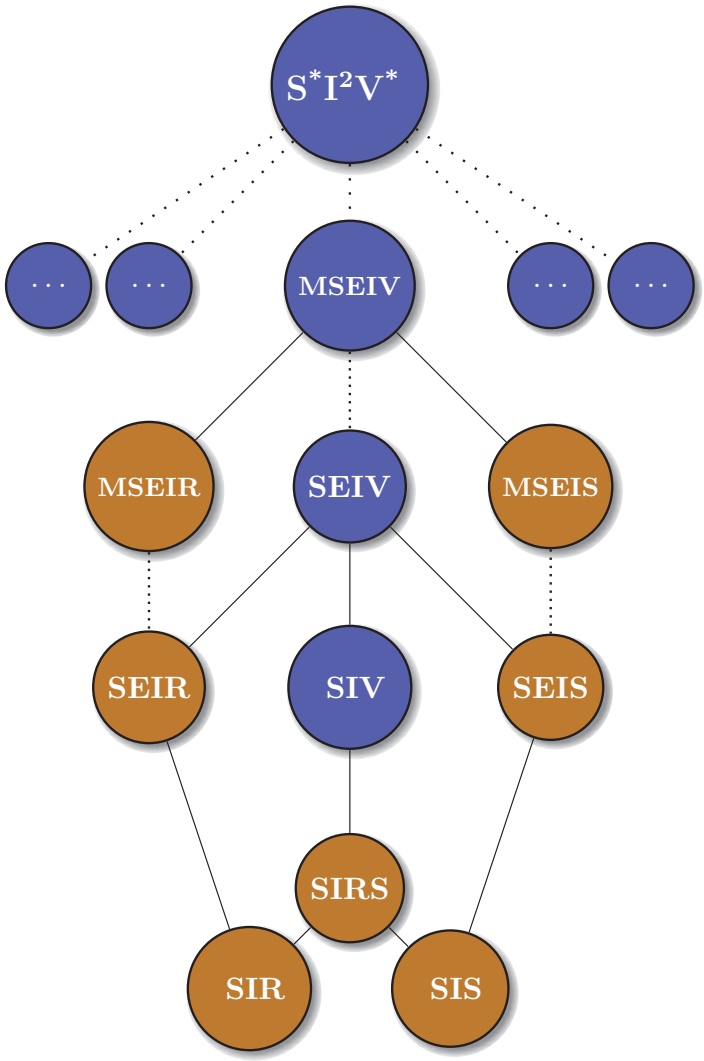



**Figure 17.4:** Lattice of Virus Propagation models, including SIS (flu); SIR (mumps); SIRS (temporary immunity); SIV (vigilance, i.e., pro-active vaccination); SEIV (like SIV, with virus incubation, i.e., the "exposed but not infectious" state); MSEIR (with the passive immune state $M$, like newborns that inherit mother's immunity); and the main generalization $S^*I^2V^*$. Each VPM is a generalization of all the models below it.

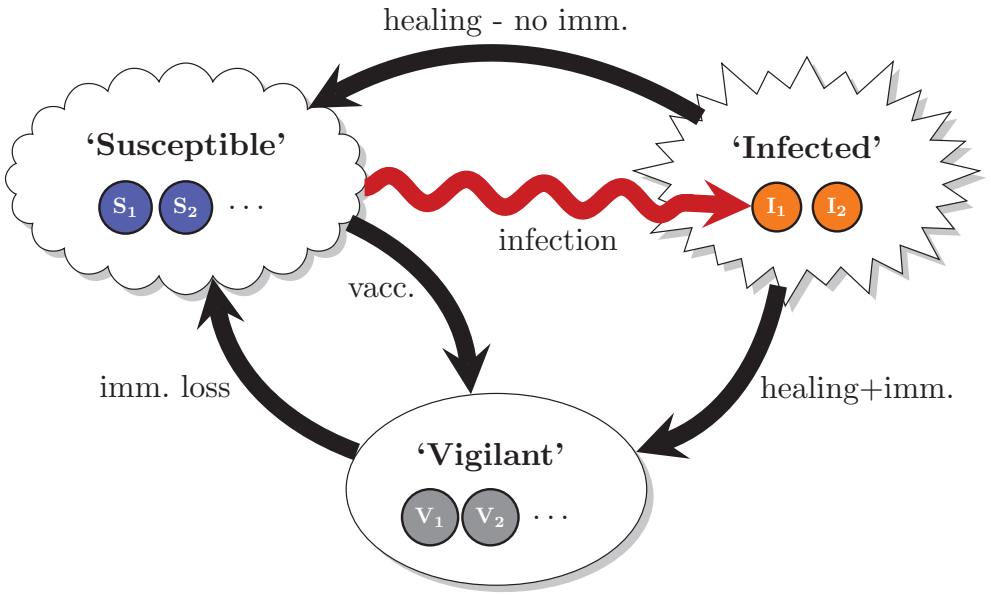



**Figure 17.5:** Pseudo-State-Diagram for a node in the graph in the generalized model $S^*I^2V^*$. Three types of states – Susceptible (healthy but can get infected), Infected (capable of transmission), and Vigilant (healthy and can't get infected). Within-class transitions not shown for clarity. Red-curvy arrow indicates *graph-based* transition, due to neighbors of the node; all other transitions are *endogenous* (caused by the node itself with some probability at every time step).

fizzles out quickly. We use the typical definition of the threshold used in literature [40, 67, 120, 140]. Intuitively, below threshold the virus has no chance of spreading the infection while, above threshold, the virus may take over and create an epidemic. For the SIS model (=no immunity), the tipping point describes some maximum *strength* of a virus, that will guarantee no epidemic [67, 140]. We define *strength* later (see Equation 17.2). Similarly, for the SIR model, the tipping point relates the explosiveness of the infection phase with respect to the virus strength, since in this model the virus will become extinct.

To standardize the discussion of threshold results, we cast the threshold problem as expressing the normalized *effective strength* of a virus as a function of the *particular* propagation model and the *particular* underlying contact-network. So we are "above threshold" when the effective strength $s > 1$, "under threshold" when $s < 1$ and the threshold or the tipping point is reached when $s = 1$.

Formally, the main result is:

**Theorem 17.1 Super-model theorem – sufficient condition for stability** *For virus propagation models that follow the $S^*I^2V^*$ model (see [232]) and for any arbitrary undirected graph with adjacency*

*matrix* $\mathbf{A}$ *the sufficient condition for stability is given by:*

$$\boxed{s < 1} \tag{17.1}$$

*where,* $s$ *(the effective strength) is:*

$$\boxed{s = \lambda_1 \cdot C_{\text{VPM}}} \tag{17.2}$$

with $\lambda_1$ being the largest eigenvalue of the adjacency matrix, $C_{\text{VPM}}$ is a constant that depends on the virus propagation model (see Figure 17.6). Hence, the tipping point is reached when $s = 1$.

The result generalizes along two different, difficult directions: (a) arbitrary contact-network topologies, and (b) several virus propagation models (VPMs).

***General Topologies*** Much of previous work [25] has concentrated on the analysis of VPMs on *specific* types of contact-networks, typically cliques or homogeneous graphs. We include them all, as *special* cases. Specifically

- Cliques, where every node contacts every other node. In that case, our result gives $\lambda_1 = N$, where $N$ is the number of nodes in the graph.
- Homogeneous graphs, with fixed degree $d$ and random Erdös-Rényi graphs with expected degree $d$ (e.g., see [160, 161]). In all these cases we have $\lambda_1 = d$, and our theorem includes the previous results.
- Hierarchical (i.e., near-block-diagonal), e.g. [141].
- Power-law random graphs (e.g. [229]).

Theorem 17.1 provides a simple and natural generalization of these results to arbitrary graphs. For example, previous results [229] have shown that the epidemic threshold in case of scale-free (power-law) networks is vanishingly small as the size $N$ of the network increases. This is a corollary of our theorem: When a power-law graph grows ($N \to \infty$), the largest eigenvalue grows with the highest degree, which also grows infinity, and thus the threshold approaches zero.

***General VPMs*** The result generalizes the threshold results for *any* VPM that is a special case of the $\text{S}^*\text{I}^2\text{V}^*$ model. We refer to this generalized model as $\text{S}^*\text{I}^2\text{V}^*$, because it has an arbitrary number of susceptible states, two infectious/infected states, and an arbitrary number of vigilant/vaccinated (= recovered) states. All the standard models (like see [140], [105]) are simply *special* cases of $\text{S}^*\text{I}^2\text{V}^*$:

- the typical flu model, SIS, is a special case;
- the typical mumps model, SIR, which corresponds to permanent immunity, is a special case, with one state for each class – $S$ belongs to the Susceptible class, $I$ belongs to the Infected class, and $R$ belongs to the Vigilant class:
- the SIRS model (temporary immunity), similar to the SIR model;
- the SEIRS model ([140], page 601) where the virus has an incubation period (state 'E:' exposed, but not infectious), and all other ingredients of the SIRS model (temporary immunity).

We now give a brief summary of the threshold results (Figure 17.6) by applying Theorem 17.1 on some standard epidemic models. Note the effect of the contact-network in effective strength for *each* model is captured solely by one parameter, $\lambda_1$, the first eigenvalue of the adjacency matrix of the network. Again, our result is a general one and these models just highlight the ready applicability of the result to standard VPMs in use.

| **Model** | **Effective Strength ($s$)** | **Threshold** |
|---|---|---|
| SIS (e.g., flu) | $\lambda_1 \cdot \left(\frac{\beta}{\delta}\right)$ | $s = 1$ |
| SIR (e.g., mumps) | $\lambda_1 \cdot \left(\frac{\beta}{\delta}\right)$ | |
| SIRS (e.g., pertussis) | $\lambda_1 \cdot \left(\frac{\beta}{\delta}\right)$ | |
| SIV | $\lambda_1 \cdot \left(\frac{\beta\gamma}{\delta(\gamma+\theta)}\right)$ | |
| SEIR | $\lambda_1 \cdot \left(\frac{\beta}{\delta}\right)$ | |
| SEIV | $\lambda_1 \cdot \left(\frac{\beta\gamma}{\delta(\gamma+\theta)}\right)$ | |
| $SI_1I_2V_1V_2$ (*used to model the H.I.V. virus, e.g., see [25]*) | $\lambda_1 \cdot \left(\frac{\beta_1 v_2+\beta_2\epsilon}{v_2(\epsilon+v_1)}\right)$ | |

**Figure 17.6:** Threshold results for some models. $\beta$, $\delta$, $\gamma$ are the probabilities to attack, heal, and lose-immunity.

## 17.3 APPLICATIONS

The results in this chapter can be fundamental to numerous applications. We describe a few important ones next.

***Fast answers to "what-if" questions and guiding policy*** The threshold results can help quickly determine the result of plausible situations. For example, what happens if the virus is twice as infectious (virulent)? Similarly, what happens when there is a weaker strain of the virus? Our results

will help in determining whether there is a danger of the infection taking off or not. Naturally then this can feed into policy decisions for controlling epidemics. Assuming some models for the underlying contact network (like scale-free, small-world, hierarchical, etc.) we can estimate which nodes/classes should be quarantined or immunized first. Given the linear dependence on $\lambda_1$, we want to immunize nodes (and hence remove them from the contact graph) which will drop the $\lambda_1$ value the most so that the resultant infection becomes below threshold and dies out. For example, they may decide to immunize teachers and kindergarten children first to control the epidemic. In addition, they can impose restrictions on travel so as to not increase the $\lambda_1$ and hence the effective strength for the virus. The above discussion also illustrates the generality of our result. Policy makers can assume *any* graph model which captures the contact behavior of the population the best and still use our threshold result to guide policy.

A lot of work has been done to show that immunizing high-degree nodes in scale-free networks is a good idea because of the vanishing threshold result [229]. But significantly, just concentrating on high-degree nodes will *miss* those low-degree nodes which are good "bridges" and hence can have an important influence on decreasing $\lambda_1$ when immunized. Intuitively, how *disparate* the groups are to which node connects is also important in addition to how *many* groups one is connected to. For example, a single common friend of only some sportspeople and movie stars can have a huge impact on the outbreak of a disease even if he/she knows only a few sportspeople and movie stars (while sportspeople and movie stars are themselves very tightly connected).

We have been concentrating on biological viruses only, but various biological virus models have been used to model computer viruses as well [168], e.g., [138] introduced the SHIR model ("susceptible," "hidden," "infected," "recoverable") to model computers under email attack. In these cases, more so than the biological ones, it is easier to get the entire underlying network. Hence, our threshold results can be precisely used to make the network more robust to malware and computer viruses. This can be done by selectively "removing" nodes from the contact-network by immunizing them like installing a firewall on them, etc.

***Immunization policies*** Another related topic has been on finding the right immunization policy. Pastor-Satorras and Vespignani [230] find that randomly selecting nodes for immunization performs much worse than "targeted" immunization, which selects the nodes with the highest connectivity. This is as expected; removing the highest-degree nodes quickly disconnects the graph [20, 50, 220], preventing the spread of infection.

Theorem 17.1 dictates an optimal immunization policy, namely, try to minimize the eigenvalue of the matrix after immunization. If eigen-drop is the difference in eigenvalue before and after immunization, our goal is to maximize the eigen-drop, *no matter* what is the behavior of the virus is (mumps-like, flu-like, HIV-like, etc.). And, as intuitively expected, immunizing nodes with a high degree usually (but not always) lead to best eigen-drop. A systematic way of choosing nodes to immunize is given in [264] and [267]: The optimal solution is infeasible, and they propose carefully designed heuristics to maximize the eigen-drop, given a fixed budget $k$ of vaccines.

***Time-varying graphs*** An orthogonal extension is the case that the contact/connectivity matrix change over time, say, with a daily periodicity: During the day, we come in contact with our colleagues, teachers, school-mates; during the night we are only in contact with family members; and similarly, we have different contacts during the weekend (friends, extended family, etc.). It turns out that, at least for the SIS model, we can compute the epidemic threshold [233], and it is again related to the eigenvalue of a product of matrices.

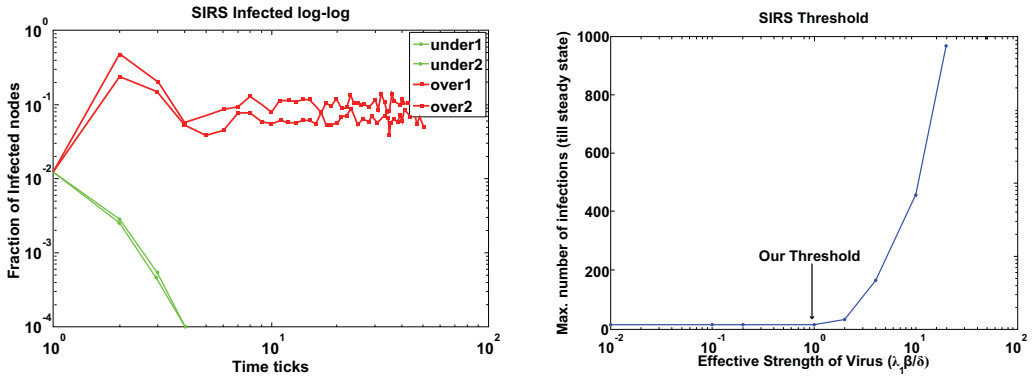
The problem of time-varying graphs is especially of interest for the propagation of computer viruses on mobile phones. The eigenvalue arguments provide solutions to such cases, too [272].

## 17.4 DISCUSSION

We discuss some simulation examples for the threshold result in some models and a few direct implications of the super-model theorem in this section. We also illustrate what the result implies for the "vulnerability" of the underlying contact graph for epidemics. Apart from the dependence of the threshold on $\lambda_1$, it is instructive to note some unexpected results in specific models as well.

### 17.4.1 SIMULATION EXAMPLES

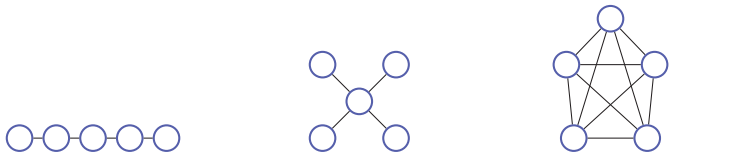
The work in [232] reports computer simulation experiments on the *Oregon* dataset (see Figure 1.2, p. 4): This is the so-called Oregon autonomous system (AS) router graph, a real network collected




(A) SIRS Infective Fraction Time Plot (log-log)

(B) SIRS Max. Infections till steady state vs Strength (lin-log)

**Figure 17.7:** SIRS (all values averages over several runs): (A) Plot of Infective Fraction of Population vs Time (log-log). Note the qualitative difference in behavior *under* (green) and *above* (red) the threshold. (B) Plot of Max. number of infected nodes till steady state vs Effective Strength (lin-log). Note the tipping point is exactly when the effective strength $s = 1$.



(A) Chain ($\lambda_1 = 1.73$) ($\lambda_1 \approx 2$) (B) Star ($\lambda_1 = 2$) $\lambda_1 = \sqrt{N-1}$ (C) Clique ($\lambda_1 = 4$) $\lambda_1 = N - 1$

**Figure 17.8:** Changing connectivity and vulnerability of graphs with changing $\lambda_1$. A chain has the least connectivity, with $\lambda_1 \approx 2$, even for a chain of $N$ nodes; an $N$-node star comes next, with $\lambda_1 = \sqrt{N-1}$; and finally an $N$-node clique, with $\lambda_1 = N - 1$. Notice that the average degree labels the $N$-node star as equivalent to the $N$-node chain, focusing only on the one-step-away paths, while the eigenvalue takes into account paths of all lengths.

from the Oregon router views. It has 15,420 links and 3,995 nodes (AS peers)[1]. We use this dataset to illustrate the super-model theorem.

Figure 17.7 gives an overview of the simulations for the SIRS (whooping cough, limited-duration immunity) model. All values are average over several runs of the simulations. In short, as expected from the theorem, there is a qualitative difference of behavior when we are above (red), below (green), and at threshold.

Figure 17.7(A) shows a time-evolution plot of the fraction of infected nodes (*footprint*) in the graph, for different values of the effective strength of the virus. Specifically it gives results for *above threshold* (in red) and *under threshold* (in green).

We also give a "take-off" plot (Figure 17.7(B)). It shows the number of infections at steady state ('footprint'), versus the different strengths of the virus. If the infection resulted in an epidemic then the footprint will be large. As the theorem predicted, the plot shows that the tipping point is at the point when the effective strength $s = 1$.

### 17.4.2 $\lambda_1$: MEASURE OF CONNECTIVITY

What does the result mean exactly? Clearly, a graph that is better connected should be better for the virus. The most natural measure of connectivity is the average degree – why is it not enough? Why is it that the first eigenvalue is the determining factor?

Intuitively, $\lambda_1$ (also known as the spectral radius) of a graph captures the connectivity of the graph. It is superior to the average degree, for the following intuitive reasons:

- for "homogeneous"/"regular" graphs (i.e., where every node has the same degree $d$, reaching $d$ random nodes), the eigenvalue is $\lambda_1 \approx d$, completely agreeing with intuition.
- for graphs with skewed degree distributions, like, say a "star," the average degree tells us how many nodes can reach other nodes, within just one step; the $\lambda_1$ takes more steps into account.

[1] See http://topology.eecs.umich.edu/data.html.

The intuition is best illustrated through the toy graphs of Figure 17.8. Although an $N$-node chain (A) and an $N$-node star (B) have the same count of edges and thus the same average degree, intuitively, the star has a better connectivity. The eigenvalue captures exactly that. In a chain, we have $\approx 2N$ pairs within one-step, and $\approx 4N$ pairs within two-steps; for a star, we still have $\approx 2N$ pairs within one-step, but $\approx N^2$ pairs within two steps. This is why the first eigenvalue is larger for the star. Very informally, here $\lambda_1$ can be seen as the square root of two step-away paths. Not surprisingly, for a clique of $N$ nodes, the eigenvalue is higher than the chain and the star ($\lambda_1 = N - 1$), correctly reflecting the excellent connectivity of the clique.

In short, the summary of our intuitive discussion is:

**Observation 17.2** As a measure of connectivity, the eigenvalue $\lambda_1$ is like the average degree, but averaged properly, to take into account all path-lengths.

And, of course, the better connected a graph is (high $\lambda_1$), the better it is for the virus.

## 17.5 CONCLUSION

The work in [232] provided two orthogonal generalizations of earlier epidemic threshold results.

- In the first direction, the result gives the threshold of the generalized $S^*I^2V^*$ model, which encompasses *any* epidemic model in published literature ([140], [105], etc.).
- In the second direction, topology, it showed that for any, arbitrary, undirected contact-network, the effect of the topology can be captured solely by $\lambda_1$, the first eigenvalue of the adjacency matrix.

Moreover, we discussed some important applications and implications of the result for policy makers, scientists, marketers, and specifically:

- Fast answers to "what-if" questions.
- Guiding immunization policies: immunize those nodes that drop the eigenvalue as much as possible.
- Simplifying epidemiological simulations: if we are below threshold, we don't need to do a simulation at all, since the virus will die out.