CSE 8803 EPI: Data Science for Epidemiology, Fall 2020

Lecturer: B. Aditya Prakash
Scribe: June Cho

Lecture #6
September 14, 2020

1 Summary of Lecture Content

This lecture covered the dynamics of network-based epidemiological models.

We first discussed the dynamics of single-virus models in static networks and examined the threshold at which the behavior of the virus propagation model changes from extinction to invasion. Along with intuitive explanations on how the threshold is derived, we observed its utility through experimental results on synthetic and actual graphs. We also looked at an extension to dynamic graphs in which the topology changes over time.

In the later part of lecture, we studied synchronization of the SIRS model where the infective fraction shows periodic oscillations, an interesting phenomenon that occurs in complex dynamical systems. Lastly, we discussed models with multiple viruses that compete under mutual immunity or cooperate towards survival.

2 Dynamics of the Single-Virus Model on Static Graphs

2.1 Problem Statement

When a virus is introduced, there can be two distinct outcomes intuitively. A strong enough virus will lead to an epidemic (the *above* phase), whereas a weak virus will be extinct after a certain period (the *below* phase). This section tackles the following problem: Can we mathematically distinguish the two phases? Can we find a condition under which the virus will die out quickly regardless of initial infection condition?

Specifically, we consider the threshold phenomenon introduced in Lecture 3 within the context of static graph-based models. Given a graph G and virus specifications such as attack probability β and curing rate δ, we aim to find the condition for virus invasion/extinction.

A solution to this problem can be used to accelerate simulations, as we can avoid simulating the spread of a virus if we already expect it to be extinct. We can also accurately forecast the size of an epidemic, which allows development of effective precautionary measures such as immunization to control the spread.

2.2 Results

The solution is expected to depend on the topology of the graph as well as the parameters of the virus propagation model (VPM). Note that the graph is itself a high dimensional object with many different properties such as the average degree, maximum degree, graph diameter, etc. The VPM also involves multiple parameters such as the attack/cure probabilities. Therefore, we must determine which set of parameters have an effect in the threshold and how are they arranged together mathematically to form the metric.

Exact analysis of this problem is difficult, and thus previous work have imposed different assumptions to simplify the problem towards meaningful results. Wang et al. [12] assumed independence among the events under the SIS model, and Ganesh et al. [2] relaxed the problem setting to continuous time. Van Mieghem et al. [9] also studied the model under a different independence assumption using large Markov chains.

In the case of the SIS model with attack probability β and cure probability δ, the number of infectious individuals approaches zero as time goes to infinity. The theorem due to Ganesh et al. [2] below provides an upper bound on the expectation of the time of extinction.

Theorem 1. *If $\rho(\boldsymbol{A}) < \delta/\beta$, the expected time of extinction τ satisfies the following.*

$$\mathbb{E}(\tau) \leq \frac{\log(n) + 1}{1 - \beta\rho(\boldsymbol{A})} \tag{1}$$

Here $\rho(\boldsymbol{A})$ denotes the spectral radius or the largest-magnitude eigenvalue of the adjacency matrix $\boldsymbol{A}$. This result shows that if $\rho(\boldsymbol{A})$ is small enough, the expected time for the virus to die out is upper-bounded by $\mathcal{O}(\log(n))$.

Generalizing this result to virus propagation models other than the SIS model, the main claim of Prakash et al. [7] states the following under the independence assumption of events:

Theorem 2. *Given any arbitrary graph with adjacency matrix $\boldsymbol{A}$ and any virus propagation model (VPM) satisfying general assumptions, the virus dies out if*

$$s < 1 \tag{2}$$

where the effective strength s is defined as

$$s = \lambda \cdot C_{VPM} \tag{3}$$

λ denotes the largest eigenvalue of $\boldsymbol{A}$, and C_{VPM} is a model-specific constant determined by the virus propagation model. $s = 1$ is referred to as the threshold or the tipping-point.

Table 1 shows a list of effective strengths for different virus propagation models. The V state refers to the Vigilant state, which would correspond to individuals practicing mask-wearing and social distancing in the case of COVID-19. For all models, notice that the topology of the graph only plays a role in the threshold through λ.

Models	Effective Strength (s)	Threshold (tipping-point)
SIS, SIR, SIRS, SEIR	$s = \lambda \cdot \left(\frac{\beta}{\delta}\right)$	
SIV, SEIV	$s = \lambda \cdot \left(\frac{\beta\gamma}{\delta(\gamma + \theta)}\right)$	$s = 1$
$SI_1I_2V_1V_2$ (HIV)	$s = \lambda \cdot \left(\frac{\beta_1\nu_2 + \beta_2\epsilon}{\nu_2(\epsilon + \nu_1)}\right)$	

Table 1: Effective strengths and thresholds for different VPMs. [7]

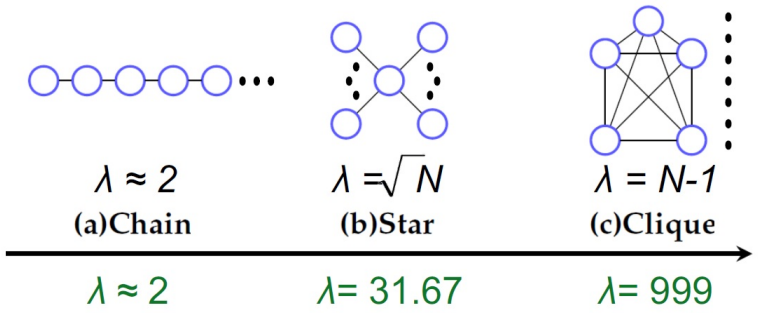


Figure 1: λ for graphs of varying topologies. The arrow going towards the right indicates increasing connectivity of graphs, and below the arrow are λ values for $N = 1000$. [7]

2.3 Intuition

In the realm of linear algebra, λ is defined as the root with the largest magnitude of the characteristic polynomial of adjacency matrix $\boldsymbol{A}$, or $\det(\boldsymbol{A} - x\boldsymbol{I})$ where $\boldsymbol{I}$ is the identity matrix.
A rather more intuitive interpretation of λ is to regard it as an aggregated measure of the connectivity of the graph. This comes from these observations:

- The number of paths within the contact network heavily affects the spread of a disease.
- The (i, j)-th entry of $\boldsymbol{A}^k$ is equal to the number of paths with length k going from node i to node j (there can be repeated nodes within the path).
- The adjacency matrix $\boldsymbol{A}$ can be written as an eigendecomposition $\boldsymbol{A} = \sum_i \lambda_i u_i u_i^T$ [11].
- If we raise the power k of $\boldsymbol{A}^k = \sum_i \lambda_i^k u_i u_i^T$ towards infinity, the largest eigenvalue λ_1 dominates the smaller eigenvalues and thus $\lambda_1^k u_1 u_1^T$ well-approximates $\boldsymbol{A}^k$.

Example topologies in Figure 1 show that λ_1 indeed represents the connectivity of the graph well and can be used to measure how vulnerable the network is against an injected virus. It makes intuitive sense that a star (Figure 1b) is much more likely to spread the disease than a chain (Figure 1a) due to the existence of a central node, but simple network statistics such as the number of edges or the average degree fails to capture such distinction: While a chain and a star both with N number of nodes share approximately the same average degree of 2, $\lambda_1 \approx 2$ for chains and $\lambda_1 = \sqrt{N}$ for stars.

Figure 2 shows simulation results on the popular `PORTLAND` graph from Prakash et al. [7]. The *footprint* is defined as the final number of nodes in state R for the SIR model, or the maximum number of infections at any time step for the SIRS model. We can see that the footprint remains close to zero until the effective strength reaches the tipping-point at $s = 1$, after which the footprint jumps immediately. This threshold phenomenon is also called the *phase transition* from statistical physics due to its analogous behavior to that of freezing water into ice: water remains in liquid form until it reaches 0°C and then starts transitioning to ice.

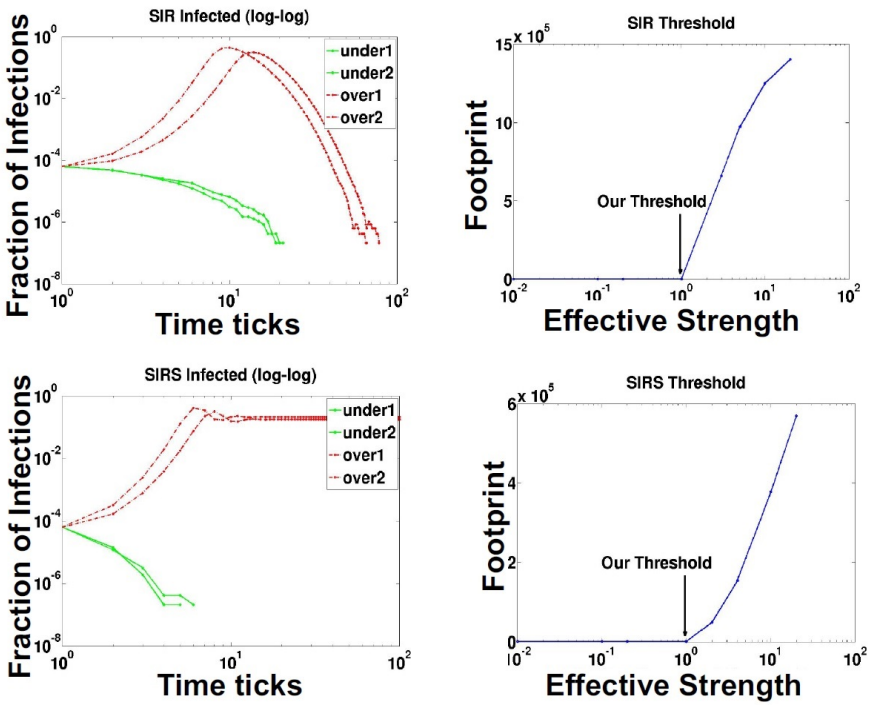


Figure 2: Simulation results of SIR and SIRS models on the PORTLAND graph with 6 million nodes and 31 million edges. Left column: infection profiles with time vs. infective fraction. Right column: "take-off" plots with effective strength s vs. footprint [7]

2.4 Proof Ideas

The proof of Theorem 2 has two main ingredients: 1) λ comes from the graph-based analysis of the topology and stability, and 2) C_{VPM} comes from the generalized VPM structure. Here we only discuss the sketch of the proof, but a more detailed analysis can be found in [7].

2.4.1 A Generalized Model

The case of Typhoid Mary, the first person in the US identified as a *healthy carrier* of typhoid fever, illustrates the need of diverse models that can capture complex propagations present in real life. One example would be the SICR model with two infected states I and C (for Carrier), each representing symptomatic and asymptomatic states. To generalize the argument across different propagation models, Prakash et al. [7] first constructed a generalized model $S^*I^2V^*$ (Figure 3) that can capture most practical variations. One could use $S^*I^*V^*$, but we consider having only 2 infected states for the sake of simplicity. Models such as SIR and $SI_1I_2V_1V_2$ (HIV) can easily be seen as special cases of this meta-model.

Roughly speaking, all states are divided into three different classes: There are susceptible states with healthy nodes that can be infected, there are infected states where nodes can actively infect neighbors, and there are vigilant states where nodes have immunity against the virus. While nodes can go back and forth between susceptible and vigilant states, vigilant nodes cannot be infected by itself. Transitions from susceptible to infected states are called *exogenous* transitions, as those transitions depend on other nodes attacking. All other transitions are *endogenous*, since they can happen on their own at each timestep.

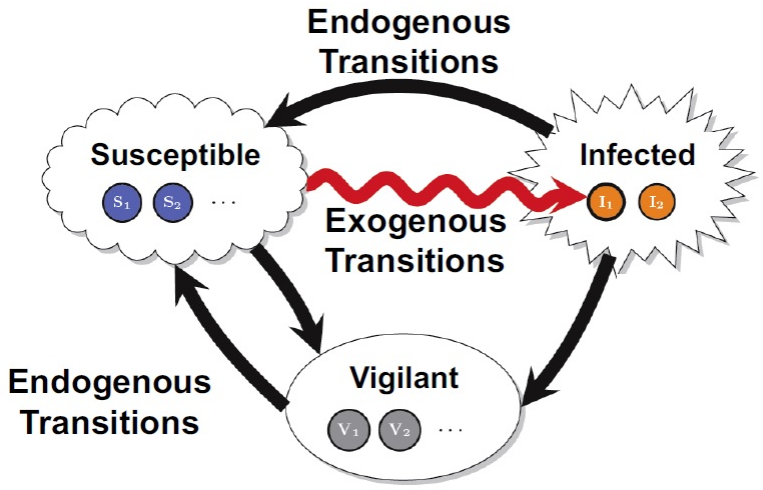


Figure 3: Generalized model $S^*I^2V^*$ [7]

2.4.2 Non-Linear Dynamical System + Stability

The next important piece is to view the generalized epidemiological model as a discrete time non-linear dynamical system (NLDS). In other words, we assume that time increases in a step-wise fashion and that the system follows the recurrence $\vec{P}_{t+1} = \mathcal{G}(\vec{P}_t)$. Given a graph of N nodes and m different states, $\vec{P}_t \in \mathbb{R}^{mN}$ is a probability vector which specifies the probability of each node being in each state at time t, and $\mathcal{G} : \mathbb{R}^{mN} \to \mathbb{R}^{mN}$ is a non-linear function that is applied to $\vec{P}_t$ to give the probability vector of the next time step $\vec{P}_{t+1}$. For instance, the special case of SIR model can be written as

$$\vec{P}_t = [P_{S,1,t}, P_{S,2,t}, \ldots, P_{S,N,t}, P_{I,1,t}, \ldots, P_{R,N,t}]^T \in \mathbb{R}^{3N} \tag{4}$$

$$\mathcal{G} : \mathbb{R}^{3N} \to \mathbb{R}^{3N},\ \mathcal{G}(\vec{P}_t) = \begin{cases} P_{S,i,t+1} & = P_{S,i,t}\zeta_{i,t}(I) \\ P_{I,i,t+1} & = P_{S,i,t}(1-\zeta_{i,t}(I)) + (1-\delta)P_{I,i,t} \\ P_{R,i,t+1} & = \delta P_{I,i,t} + P_{R,i,t} \end{cases} \tag{5}$$

where $\zeta_{i,t}(I)$ is the probability that node i is not attacked by any of its infected neighbors. Just as a sidenote, we can instead represent $\vec{P}_t$ as a smaller length-$(2N)$ vector since $P_{R,i,t} = 1 - P_{S,i,t} + P_{I,i,t}$ for all i, t.

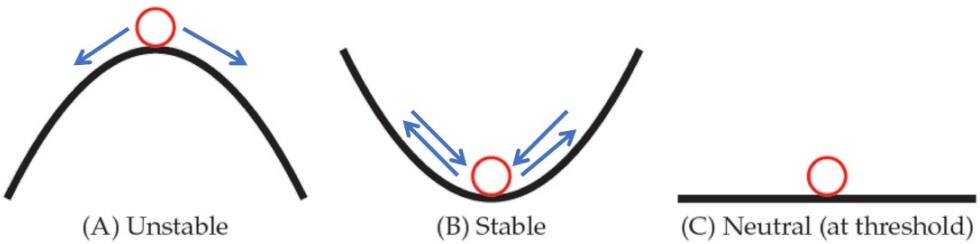


Figure 4: Illustrative example of dynamical systems under gravity. (A) Slight perturbations on an unstable system will cause the ball to fall, (B) A stable system will return to its original state after perturbation (C) A neutral system has no tendency to be stable or unstable. [7]

With this formulation, we can view the threshold problem as a NLDS stability problem, in which we examine whether the system is stable or unstable at the initial fixed state with no infections (Figure 4). In an epidemiological viewpoint, a perturbation would be equivalent to injecting a virus into several nodes, after which the system can respond in different ways. The number of infections will rise in an unstable system, diverging away from the initial fixed state. In a stable system, on the other hand, the tendency to go back to the fixed point will reduce the number of infections towards zero.

2.5 Extension towards Dynamic Graphs

The result from static graphs can also be extended to dynamic graphs in which the contact patterns change across discrete time steps. For example, the contact network can differ between during the day when people go to work or school and during the night when families gather back together at home. This can be represented as a set of T adjacency matrices $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_T\}$. One obvious result for the SIS model with dynamic graphs is that the virus dies out if

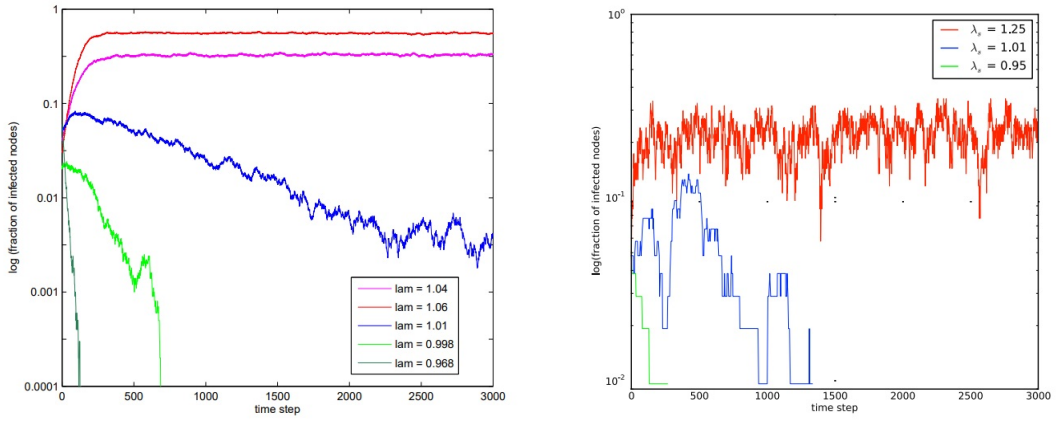
$$\frac{\lambda_{max}\beta}{\delta} < 1 \tag{6}$$

where λ_{max} is the largest eigenvalue of all T adjacency matrices. This natural result induced from the static-graph case says that if the virus goes extinct in the worst-case graph with the largest connectivity, it will go extinct in other graphs as well. However, this result is too pessimistic as it only captures monotonic decrease in infection counts, and is thus not so practical to be applied towards real life study. Work from Prakash et al. [8] provides a more realistic analysis of the threshold problem in time-varying graphs.

Theorem 3. *Let* $\boldsymbol{S}_i = (1-\delta)\boldsymbol{I} + \beta\boldsymbol{A}_i$ *for all* $i \in \{1, 2, \ldots, T\}$. *Then, the virus dies out if*

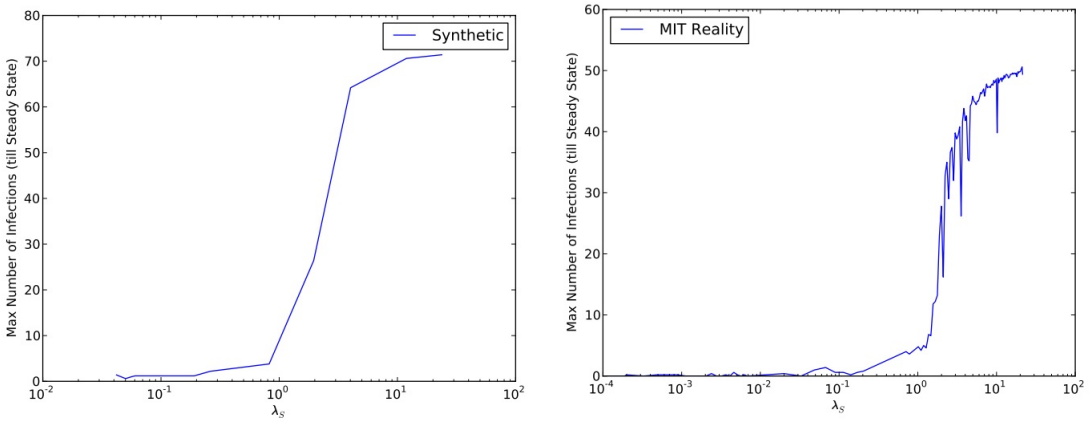
$$\lambda_{\boldsymbol{S}} < 1 \tag{7}$$

where $\lambda_{\boldsymbol{S}}$ *is the largest eigenvalue of the system matrix* $\boldsymbol{S} = \prod_i \boldsymbol{S}_i$.



(A) Synthetic (B) MIT Reality

Figure 5: Infection profiles (time vs. infected fraction) from (A) synthesized graphs - one clique and one chain, (B) MIT reality graphs constructed by tracking mobile devices. We see varying behaviors depending on $\lambda_{\boldsymbol{S}}$: below 1 (green), at 1 (blue), and above 1(red). [8]



(A) Synthetic (B) MIT Reality

Figure 6: "Take-off" plots ($\lambda_{\boldsymbol{S}}$ vs. maximum number of infections) of same dynamic graphs. As shown in Theorem 3, the sudden jump occurs at $\lambda_{\boldsymbol{S}} = 1$ for both datasets. [8]

3 Other Interesting Dynamics

Representing more states and interactions within the virus propagation model often leads to unexpected behaviors. In this section, we discuss interesting dynamics that occur in complex systems under increasing disorder of networks or when multiple viruses are introduced.

3.1 Synchronization

A complex model like SIRS can show interesting phenomenon such as synchronization. In our context, synchronization refers to the dynamical system showing oscillations repeating at a certain frequency. Through numerical simulations of the SIRS model on random networks, Kuperman and Abramson [5] have verified formation of sinusoidal oscillations in the fraction of infected nodes (Figure 7). Empirical studies on Measles cases in the UK [4] and Syphilis infections [3] have confirmed that synchrony occurs in real life as well. One open research topic is to pinpoint the parameter c at which synchronization happens.

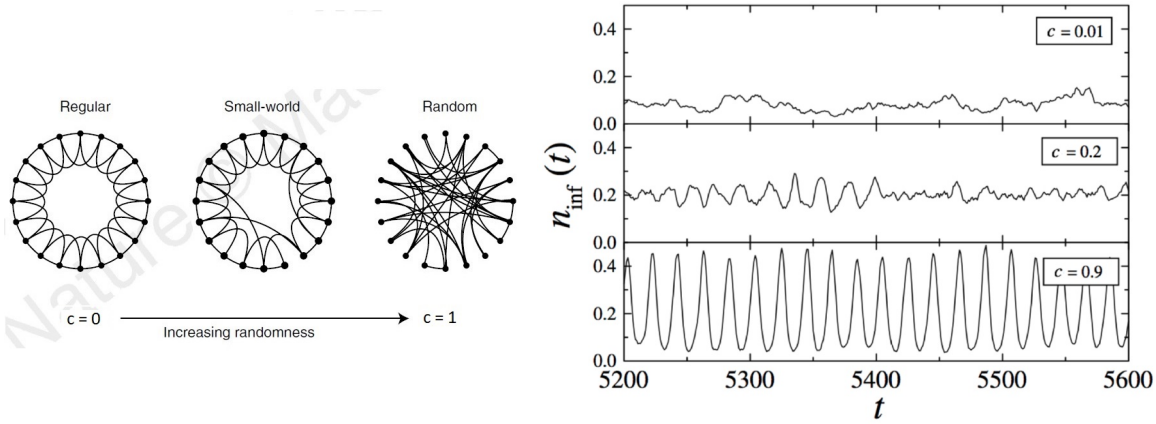
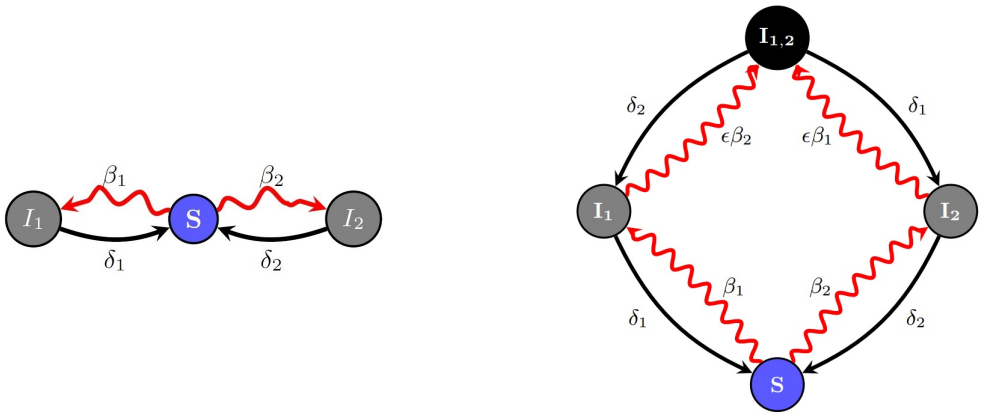


Figure 7: Left: Random graphs can be formed by rewiring edges in a regular graph randomly with probability c. [10] Right: The SIRS model exhibits an endemic state when c is small, but as we increase the randomness we observe synchronization as with $c = 0.9$. [5]



(A) SI_1I_2S [6] (B) $SI_{1|2}S$ [1]

Figure 8: State diagrams of models with two viruses.

3.2 Competing viruses

While we mostly covered single-virus models, there are also models that involve multiple viruses. One example is the SI_1I_2S model (Figure 8A) with viruses that compete under mutual immunity. If both viruses are weak by themselves, no epidemic is expected to occur at all. A more interesting setting is when both viruses are strong but one virus is stronger than the other, and the following theorem due to Prakash et al. [6] shows that the *winner takes all* in such cases.

Theorem 4. *Under the SI_1I_2S model with parameters $(\beta_1, \beta_2, \delta_1, \delta_2)$, virus 1 dominates and virus 2 completely dies out in the steady state if the strength of virus 1 is above the threshold $\left(\lambda\frac{\beta_1}{\delta_1} > 1\right)$ and is larger than the strength of virus 2 $\left(\lambda\frac{\beta_1}{\delta_1} > \lambda\frac{\beta_2}{\delta_2}\right)$.*

3.3 Interacting viruses

We can also have a model with cooperation among viruses. Figure 8B shows the $SI_{1|2}S$ model, a modified flu-like (SIS) model that introduces an interaction factor ϵ to allow classification of different interactions between viruses. The dynamics of the model varies depending on ϵ as follows.

- $\epsilon = 0$: The model has full mutual immunity and the stronger virus dominates (given that its strength is above the threshold) as in the SI_1I_2S model.
- $0 < \epsilon < 1$: There is partial competition and strong viruses show footprints lower than when injected independently.
- $\epsilon = 1$: Model exhibits independent co-existence and each virus propagates as if it was the only virus present.
- $1 < \epsilon$: The viruses cooperate and start to show footprints higher than when injected independently. Beutel et al. [1] showed that viruses too weak to survive by itself can also survive when provided enough cooperation.

The main result from Beutel et al. [1] states the following.

Theorem 5. *Under the $SI_{1|2}S$ model with parameters $(\beta_1, \beta_2, \delta_1, \delta_2, \epsilon)$, there exists an $\epsilon_{critical}$ such that if $\epsilon > \epsilon_{critical}$, there is a fixed point where both viruses survive.*

References

[1] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos. Interacting viruses in networks: Can both survive? In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, page 426–434, New York, NY, USA, 2012. Association for Computing Machinery.

[2] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, pages 1455–1466 vol. 2, 2005.

[3] N. C. Grassly, C. Fraser, and G. P. Garnett. Host immunity and synchronized epidemics of syphilis across the united states. *Nature*, 433(7024):417–421, 2005.

[4] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.

[5] M. Kuperman and G. Abramson. Small world effect in an epidemiological model. *Physical Review Letters*, 86(13):2909–2912, Mar 2001.

[6] B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos. Winner takes all: Competing viruses or ideas on fair-play networks. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, page 1037–1046, New York, NY, USA, 2012. Association for Computing Machinery.

[7] B. A. Prakash, D. Chakrabarti, N. C. Valler, M. Faloutsos, and C. Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. *Knowledge and information systems*, 33(3):549–575, 2012.

[8] B. A. Prakash, H. Tong, N. Valler, M. Faloutsos, and C. Faloutsos. Virus propagation on time-varying networks: Theory and immunization algorithms. In *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part III*, ECML PKDD'10, page 99–114, Berlin, Heidelberg, 2010. Springer-Verlag.

[9] P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Transactions on Networking*, 17(1):1–14, 2009.

[10] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440–442, 1998.

[11] Wikipedia, the free encyclopedia. "Eigendecomposition of a matrix", 2020. [Online; accessed September 25, 2020].

[12] Yang Wang, D. Chakrabarti, Chenxi Wang, and C. Faloutsos. Epidemic spreading in real networks: an eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34, 2003.