Interacting Viruses in Networks: Can Both Survive?

Alex Beutel, B. Aditya Prakash, Roni Rosenfeld, and Christos Faloutsos
Computer Science Department, Carnegie Mellon University
Pittsburgh, PA, USA
{abeutel, badityap, roni, christos}@cs.cmu.edu

ABSTRACT

Suppose we have two competing ideas/products/viruses, that propagate over a social or other network. Suppose that they are strong/virulent enough, so that each, if left alone, could lead to an epidemic. What will happen when both operate on the network? Earlier models assume that there is perfect competition: if a user buys product 'A' (or gets infected with virus 'X'), she will never buy product 'B' (or virus 'Y'). This is not always true: for example, a user could install and use both Firefox and Google Chrome as browsers. Similarly, one type of flu may give partial immunity against some other similar disease.

In the case of full competition, it is known that 'winner takes all,' that is the weaker virus/product will become extinct. In the case of no competition, both viruses survive, ignoring each other. What happens in-between these two extremes?

We show that there is a phase transition: if the competition is harsher than a critical level, then 'winner takes all;' otherwise, the weaker virus survives. These are the contributions of this paper (a) the problem definition, which is novel even in epidemiology literature [3, 17, 35] (b) the phase-transition result and (c) experiments on real data, illustrating the suitability of our results.

Categories and Subject Descriptors

H.2.8 [**Database management**]: Database applications—*Data mining*

Keywords

Epidemics, Cascades, Competition, Co-existence

1. INTRODUCTION

Given two partially competing products (like Firefox and Google Chrome; or Android and iPhone), is it possible that they both survive?

KDD'12, August 12–16, 2012, Beijing, China.

The well-known Competitive Exclusion Principle in ecology states that when two species are in complete competition under constant conditions, the more fit one will eventually drive the less fit one into extinction. A more common but less well understood scenario is one where the competing species induce partial immunity against one another. There has been significant work trying to elucidate the conditions under which such partial immunity leads to coexistence [25, 11, 26] but a complete theory has not yet emerged.

Here, we study the general case of two virus strains with partial (and symmetric) cross-immunity spreading over a fixed network topology. In addition to the implications for the evolutionary problem discussed above, our results have direct relevance to the spread of rumors and opinions in social networks and market penetration of products.

The contributions of this work are the following:

- the discovery that there is a phase transition, that is, the weaker virus/product may survive, if the cross-immunity satisfies a threshold condition. This seemed to be an open problem even within the epidemiology community [25]
- experiments on real data, showing that our model fits well

Figure 1 shows the time-plots for partially competing products Hulu vs. Blockbuster (a), and Google Chrome vs. Firefox (b). They plot (normalized) count of Google queries, versus time. We fit our model to the data[1] and plot it as well. Notice that it captures the trends well.

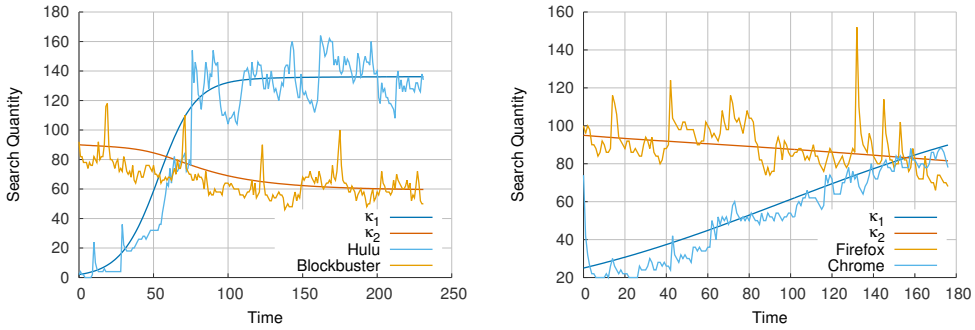
The rest of the paper is organized in the usual way: we review related work in § 2 and formulate the problem giving details of our model in § 3. We give the analysis and proof of our phase-transition and coexistence result in § 4 and demonstrate the validity of the results using simulations and real-world case-studies in § 5. Finally, we discuss other subtle aspects of the model in § 6 and conclude in § 7.

2. RELATED WORK

Here we provide a brief overview of the vast literature on epidemiology, influence propagation and population ecology.

Epidemiology and Epidemic Thresholds There are numerous, well-studied epidemiological models, including the so-called *homogeneous models* [4, 27, 3]. They effectively assume that the network is a full clique, that is, every person has the same probability to contact every other person. A large portion of the literature focuses on a single virus, and

[1]Fitted with `www.alexbeutel.com/jsplot/kdd2012.html`



(a) Hulu and Blockbuster (b) Firefox and Chrome

Figure 1: Plots of real web-search interest vs. time for pairs of competitors with our model fitted to the data.

even then, there are numerous interesting sub-cases. Typical propagation models include SIS (flu-like, with no immunity - Susceptible / Infected / Susceptible), SIR (mumps-like, with lifetime immunity - 'R'= 'recovered'), SIRS (with temporary immunity), and many-many more - see [17] for a detailed survey, or the 700-page book of Anderson and May [3].

A lot of attention focuses on the so-called epidemic threshold: under what conditions will a virus (say, flu-like) become extinct? Earlier works [19, 28] studied specific topologies (e.g., random graphs, power-law graphs, etc). Chakrabarti et al. [10] and Ganesh et al. [13] found that, for arbitrary topologies, and for the flu-like SIS model, the epidemic threshold for any arbitrary graph is determined by the leading eigenvalue of the adjacency matrix of the graph. Prakash et. al. [31] generalized this result to any topology, as well as almost any virus propagation model (SIR, SIRS, etc). However, all of these works focus on single virus models.

Recent work [30] looked at a two-virus SIS model on arbitrary graphs, but focused on the case where there was full mutual immunity between viruses. Their theorem says that, in such a setting, the stronger virus will push the weaker one to extinction ('winner takes all'), even if the weaker one would be able to survive on the network when left alone.

Influence propagation - information diffusion Biological viruses are not the only objects of interest that propagate over the edges of a network. Rumors, ideas, memes, and computer worms behave similarly and have attracted a lot of interest, especially recently thanks to the internet and the web. Research focuses on topics that include (a) information cascades [6, 14], (b) blog propagation [24, 16, 21, 33], and (c) viral marketing and product penetration [23]. Typical cascade models include (a) the independent cascade [18] (essentially an 'SIR' - mumps-like model) and (b) linear threshold [15]. Research work in multiple cascades has examined extensions of the independent cascade model for the case that nodes can not switch from one competitor to the other [5, 20], as well as when they can [29].

There are several fascinating, but remotely-related problems: (a) which nodes to immunize, given a finite count of vaccines [32, 36] (b) which are the best nodes to advertize a new product, to maximize product penetration [12] (c) which nodes are the most likely to be the culprits, when we are given a snapshot of infected nodes in a graph. [22, 34]

Ecology In ecology, the principle of '*competitive exclusion*' espouses that two species can not occupy the same ecological niche in the long term. That is, 'winner takes all', using our earlier terminology. Research has gone into studying this using various propagation models like SIS, SIR, Lotka-Volterra etc. (for example, see [8, 9, 1, 2]). Partial immunity models have received much attention in epidemiology. For example, [25] suggests a differential equation based model and analyzes it via simulation. However, for this and most other models of interest, a complete analytical solution has been beyond reach.

Distinguishing features of current work: In short, none of the previous work fulfills all the conditions of this current work: (a) analytical proof of $\epsilon_{\text{critical}}$, the critical value of the competition threshold (b) closed-form steady-state behavior (c) under an SIS (flu-like) model.

3. PROBLEM FORMULATION

In this section, we formulate our problem, giving details about the model used and the assumptions. Table 1 explains the terminology we have used in the paper. Bold letters typically denote matrices (**A**, **M** etc.).

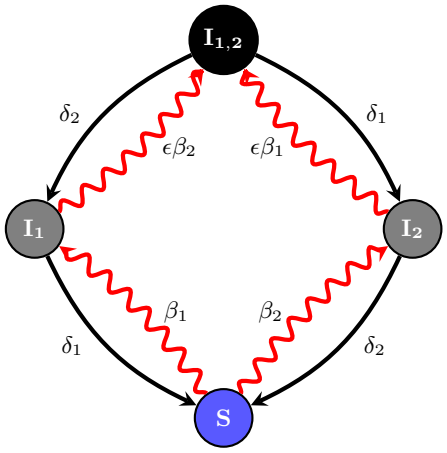


Figure 2: State Diagram for a node in the graph under our partial-competition model.

3.1 The propagation model

We assume that the competing viruses are spreading on the network according to a propagation model, which we describe next. We call our propagation model $SI_{1|2}S$, based on

Table 1: Symbols and Definitions

Symbol	Definition and Description
$SI_{1\|2}S$	our competing viruses model
β_1(or β_2)	attack rate of virus 1 (or virus 2)
δ_1(or δ_2)	cure rate of virus 1 (or virus 2)
ϵ	Interaction factor between virus 1 and 2
$\mathbf{A}$	adjacency matrix of the underlying graph
$\lambda_1(\mathbf{M})$	largest eigenvalue of matrix $\mathbf{M}$
λ	$\lambda_1(\mathbf{A})$
σ_1	$\lambda\beta_1/\delta_1$ (strength of virus 1)
σ_2	$\lambda\beta_2/\delta_2$ (strength of virus 2)
I_1 (or I_2)	The number of nodes infected with only virus 1 (or only virus 2)
$I_{1,2}$	The number of nodes infected with both virus 1 and virus 2
κ_1	Fraction of nodes infected with virus 1 ($(I_1 + I_{1,2})/N$)
κ_2	Fraction of nodes infected with virus 2 ($(I_2 + I_{1,2})/N$)
i_{12}	Fraction of nodes infected with both viruses ($I_{1,2}/N$)
i_1 (or i_2)	Fraction of nodes infected with virus 1: I_1/N (or virus 2: I_2/N)
$\kappa_1^*, \kappa_2^*, i_{12}^*$	The solution at the coexistence equilibrium (if exists)

the popular "flu-like" SIS (Susceptible-Infected-Susceptible) model [17]. $SI_{1|2}S$ denotes Susceptible - Infected$_{1 \text{ or } 2}$ - Susceptible. Each node in the graph can be in one of four states: Susceptible (healthy), I_1 (infected by virus 1), I_2 (infected by virus 2), or $I_{1,2}$ (infected by both virus 1 and virus 2). The state transition diagram as seen from a node in the network is shown in Figure 2. We could have extended other single virus models as well, but we believe that our model is a reasonable starting point, and we leave the analysis of other models as future work.

Healing (virus death) rate: δ. If a node is in an infected state (I_1, I_2, or $I_{1,2}$), it recovers on its own with some rate, δ_1 for virus 1 or δ_2 for virus 2. The healing rate is inversely related with the virus's strength: a high δ means that nodes that are infected heal quickly. For example, a product that lasts a long time such that people using it rarely consider alternatives would be modeled with a low δ value.

Attack (virus transmission) rate: β. An infected node can spread the virus to its neighboring nodes, and the node susceptibility is captured by β_1 and β_2. Specifically, an infected node transmits its infection to each of its healthy neighbors *independently* at rate β_1 (or β_2). The more often an idea or product is shared with friends, frequently referred to as being "viral," the higher the value for β.

Virus interaction factor: ϵ. A node infected with one virus may be more or less susceptible to being infected by the other virus, as determined by the factor ϵ. The transmission rate for a virus becomes $\epsilon\beta_1$ (or $\epsilon\beta_2$) when a node is already infected with one virus. Specifically, if a node is infected with virus 1, each of its neighbors infected with virus 2 have a transmission rate to it of $\epsilon\beta_2$; a node infected with virus 2 can only be infected with virus 1 at a rate of $\epsilon\beta_1$.

This is a novel generalization of the single-virus SIS model to a multiple-virus scenario. The value of ϵ can describe many different virus interactions. If $\epsilon = 0$ then the viruses are fully mutually immune, and $0 < \epsilon \leq 1$ suggests an amount of competition between viruses.

Fair-play: We assume that the competitors are playing a 'fair game': All nodes in the network have the same model parameters (β's, δ's, ϵ) for each of the viruses and behave according to the state-diagram in Figure 2.

3.2 Problem Statement

We are now in a position to state the problem formally. We assume the underlying network is connected - otherwise we just have separate disconnected problems.

Interacting viruses problem

Given: An undirected connected graph G, and the propagation model ($SI_{1|2}S$) parameters (β_1, δ_1 for virus 1, β_2, δ_2 for virus 2, and ϵ)

Find: What are the possible fixed points for the system? In particular, for what values of ϵ is there a fixed point for which both virus 1 and virus 2 survive?

3.3 Model Formulation for a Clique

For a clique, the following differential equations fully describe the transitions of the system, seen in Figure 2. Here I_1, I_2, and $I_{1,2}$ are the number of nodes infected with only virus 1, only virus 2, and both virus 1 and 2 respectively. N is the total number of nodes, and S is the number of susceptible nodes ($S = N - I_1 - I_2 - I_{12}$).

$$\frac{dI_1}{dt} = \beta_1 S(I_1 + I_{12}) + \delta_2 I_{12} - \delta_1 I_1 - \epsilon\beta_2 I_1(I_2 + I_{12}) \quad (1)$$

$$\frac{dI_2}{dt} = \beta_2 S(I_2 + I_{12}) + \delta_1 I_{12} - \delta_2 I_2 - \epsilon\beta_1 I_2(I_1 + I_{12}) \quad (2)$$

$$\frac{dI_{12}}{dt} = \epsilon\beta_1 I_2(I_1 + I_{12}) + \epsilon\beta_2 I_1(I_2 + I_{12}) - (\delta_1 + \delta_2) I_{12} \quad (3)$$

4. RESULTS AND PROOFS

The goal of our analysis is to find for what values of ϵ is there an equilibrium point for which both virus 1 and virus 2 survive. We find that there is an $\epsilon_{\text{critical}}$ such that if $\epsilon > \epsilon_{\text{critical}}$ then an equilibrium point for which the viruses coexist.

4.1 Formulating the problem

At an equilibrium point, all derivatives are zero. Thus, we can find a simple equation for I_{12}

$$\epsilon(\beta_1 + \beta_2) I_1 I_2 \quad = \quad (\delta_1 + \delta_2 - \epsilon(\beta_1 I_2 + \beta_2 I_1)) I_{12}$$

LEMMA 1. *The number of people infected by both virus 1 and virus 2 will obey the following equation:*

$$I_{12} \quad = \quad I_1 I_2 \epsilon(\beta_1 + \beta_2)/(\delta_1 + \delta_2 - \epsilon(\beta_1 I_2 + \beta_2 I_1))$$

PROOF. Trivial, given the above. □

Thus we have the expected three equilibrium points

- $I_1 = I_2 = I_{12} = 0$
- $I_1 = I_{12} = 0$, $I_2 = N - \frac{\delta_2}{\beta_2}$
- $I_2 = I_{12} = 0$, $I_1 = N - \frac{\delta_1}{\beta_1}$

and possibly one for which $I_1, I_2 > 0$ and obeys the differential equations outlined:

$$0 = \beta_1 S(I_1 + I_{12}) + \delta_2 I_{12} - \delta_1 I_1 - \epsilon\beta_2 I_1(I_2 + I_{12}) \quad (4)$$
$$0 = \beta_2 S(I_2 + I_{12}) + \delta_1 I_{12} - \delta_2 I_2 - \epsilon\beta_1 I_2(I_1 + I_{12}) \quad (5)$$
$$0 = \epsilon\beta_1 I_2(I_1 + I_{12}) + \epsilon\beta_2 I_1(I_2 + I_{12}) - (\delta_1 + \delta_2) I_{12} \quad (6)$$

We rework these equations to be primarily in terms of κ_1, κ_2, i_{12}, where $\kappa_1 = (I_1 + I_{12})/N$, $\kappa_2 = (I_2 + I_{12})/N$, $i_{12} = I_{12}/N$. As such, each of these terms represent a fraction of the population that is infected. We first convert the constraints to

$$N\kappa_1\beta_1[1 - \kappa_1 - (1-\epsilon)i_2] = \delta_1\kappa_1 \quad (7)$$
$$N\kappa_2\beta_2[1 - \kappa_2 - (1-\epsilon)i_1] = \delta_2\kappa_2 \quad (8)$$
$$\epsilon N(\beta_1\kappa_1 i_2 + \beta_2\kappa_2 i_1) = (\delta_1 + \delta_2) i_{12} \quad (9)$$

where $i_1 = I_1/N$ and $i_2 = I_2/N$.

Manipulating (9) to remove i_1 and i_2, we find

$$\epsilon\kappa_1\kappa_2[\sigma_1\delta_1 + \sigma_2\delta_2] = i_{12}[\delta_1 + \delta_2 + \epsilon\sigma_1\delta_1\kappa_1 + \epsilon\sigma_2\delta_2\kappa_2] \quad (10)$$

Remember, because we are working with a clique the virus strengths are $\sigma_1 = N\beta_1/\delta_1$ and $\sigma_2 = N\beta_2/\delta_2$.

4.2 Results

From these constraints, we look to find a lower bound on ϵ, such that for any less competition there can be coexistence.

THEOREM 1 (EPSILON THRESHOLD THEOREM). *Given a fully connected graph with the $SI_{1|2}S$ model parameters $\sigma_1 \geq \sigma_2$, an equilibrium point for which $\kappa_1, \kappa_2 > 0$ exists if $\epsilon > \epsilon_{\text{critical}}$, where*

$$\epsilon_{\text{critical}} = \begin{cases} \frac{\sigma_1 - \sigma_2}{\sigma_2(\sigma_1 - 1)} & \text{if } \sigma_1 + \sigma_2 \geq 2 \\ \frac{2(1+\sqrt{1-\sigma_1\sigma_2})}{\sigma_1\sigma_2} & \text{if } \sigma_1 + \sigma_2 < 2 \end{cases} \quad (11)$$

PROOF. In Lemma 3 we give the possible fixed point for coexistence. In Lemma 4 we show the constraints for the fixed points to be real, which contribute to the bounds in (11). In Lemmas 5 through 9 we give the proofs for the constraints on the fixed points being positive, and in Lemma 10 we give the proof that the fixed points are less than one. □

Next we describe all of the Lemmas, which contribute to the proof.

LEMMA 2. *If a fourth equilibrium point exists, then it should satisfy the follow equation:*

$$\epsilon(\kappa_2 - \kappa_1) = 1/\sigma_1 - 1/\sigma_2 \quad (12)$$

PROOF. Since we are only looking for non-zero solutions for κ_1 and κ_2, we can eliminate them in (7) and (8).

$$1 - \kappa_1 - (1-\epsilon)i_2 = 1/\sigma_1 \quad (13)$$
$$1 - \kappa_2 - (1-\epsilon)i_1 = 1/\sigma_2 \quad (14)$$

Subtracting, we get the lemma. □

LEMMA 3 (COEXISTENCE LEMMA). *If an equilibrium point exists for which both viruses coexist in the network, $\kappa_1, \kappa_2 > 0$, it will be at:*

$$i_{12} = \epsilon\kappa_1\kappa_2\left[\frac{\sigma_1\delta_1 + \sigma_2\delta_2}{\delta_1 + \delta_2 + \epsilon\sigma_1\delta_1\kappa_1 + \epsilon\sigma_2\delta_2\kappa_2}\right] \quad (15)$$

$$\kappa_1 = \kappa_2 + \frac{1}{\epsilon}\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right) \quad (16)$$

$$\kappa_2 = \frac{-2\epsilon\sigma_1\sigma_2 + \epsilon^2\sigma_1\sigma_2^2 \pm \epsilon\sqrt{\sigma_1}\sigma_2^{3/2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2}}{2\epsilon^2\sigma_1\sigma_2^2} \quad (17)$$

We will denote the solution to these three equations for fixed-points as i_{12}^*, κ_1^*, and κ_2^* respectively.

PROOF. Equation (15) is a simple rearrangement of equation (10), and equation (16) is a rearrangement of equation (12). Plugging (15) and (16) into (13) allows us to solve for κ_2 resulting in (17).

For κ_2^* (and by extension κ_1^* and i_{12}^*) to be a valid fixed-point, κ_2^* must be: (a) real, (b) $\kappa_2^* \geq 0$, (c) $\kappa_2^* \leq 1$.

LEMMA 4. *In order for fixed-point solution κ_2^*, and by extension κ_1^* and i_{12}^*, to be real valued, either $\sigma_1\sigma_2 > 1$ or*

$$\epsilon < \frac{2(1-\sqrt{1-\sigma_1\sigma_2})}{\sigma_1\sigma_2} \quad \text{or} \quad \epsilon > \frac{2(1+\sqrt{1-\sigma_1\sigma_2})}{\sigma_1\sigma_2}.$$

PROOF. This constraint comes from the square root in equation (17) for κ_2^*. We analyze the quadratic equation $4 - 4\epsilon + \epsilon^2\sigma_1\sigma_2$ (in terms of ϵ) from inside the square root. It is a simple, upward-facing parabola. Solving for the roots of the quadratic equation in terms of ϵ we find

$$\epsilon = \frac{2(1 \pm \sqrt{1-\sigma_1\sigma_2})}{\sigma_1\sigma_2}.$$

For $\sigma_1\sigma_2 > 1$ there is no solution because the equation is positive for all values of ϵ. Thus, if $\sigma_1\sigma_2 > 1$ then κ_2^* must be real valued. For $\sigma_1\sigma_2 < 1$ a portion of the parabola is negative. Therefore, we require that ϵ be in the positive region of the quadratic equation, where ϵ is less than the lower root or greater than the upper root. □

To find when $\kappa_2^* \geq 0$, we consider the cases above for which it is real. As we explained before, we will focus on the lower bound for ϵ.

LEMMA 5. *For strengths $\sigma_1\sigma_2 > 1$, fixed-point κ_2^* is monotonically increasing as a function of ϵ.*

PROOF. Taking the derivative of (17) we get

$$\frac{\pm(-2+\epsilon)\sqrt{\sigma_2} + \sqrt{\sigma_1}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2}}{\epsilon^2\sqrt{\sigma_1}\sigma_2\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2}}.$$

Because $\sigma_1\sigma_2 > 1$, all of the square roots are real valued. The denominator is clearly positive, so to prove that κ_2^* is monotonically increasing, we must show that the numerator is positive. To show that the numerator is always positive we would like to show that

$$\pm(-2+\epsilon)\sqrt{\sigma_2} < \sqrt{\sigma_1}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2}$$

or alternatively

$$1 < \frac{\sigma_1}{\sigma_2}\frac{4-4\epsilon+\epsilon^2\sigma_1\sigma_2}{4-4\epsilon+\epsilon^2}.$$

Because $\sigma_1 \geq \sigma_2$ the first term is clearly > 1. For $\sigma_1\sigma_2 > 1$ (and of course $\epsilon > 0$) this is trivially true. □

LEMMA 6. *Fixed-point solution κ_2^-, defined by*

$$\kappa_2^- = \frac{-2\epsilon\sigma_1\sigma_2 + \epsilon^2\sigma_1\sigma_2^2 - \epsilon\sqrt{\sigma_1}\sigma_2^{3/2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2}}{2\epsilon^2\sigma_1\sigma_2^2}, \quad (18)$$

can only be positive when κ_2^+, defined by

$$\kappa_2^+ = \frac{-2\epsilon\sigma_1\sigma_2 + \epsilon^2\sigma_1\sigma_2^2 + \epsilon\sqrt{\sigma_1}\sigma_2^{3/2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2}}{2\epsilon^2\sigma_1\sigma_2^2}, \quad (19)$$

is positive.

PROOF. As a simple case, for $\sigma_1\sigma_2 > 1$, $\kappa_2^- < 0$ and thus invalid for all $\epsilon > 0$. As ϵ approaches 0, it is clear that $\kappa_2^- \to -\infty$, and as $\epsilon \to \infty$, we see that κ_2^- approaches 0. Since from the previous lemma we know that it is monotonically increasing, $\kappa_2^- < 0$ for $\sigma_1\sigma_2 > 1$.

If we do not restrict σ_1 and σ_2, it is still clear that $\kappa_2^- < \kappa_2^+$ for all $\epsilon \geq 0$, since the last term is always positive. We will see later that $\kappa_2^+ < 1$ for all $\epsilon > 0$. Therefore, the range for which κ_2^- is valid is a strict subset of that for which κ_2^+ is valid. □

Because κ_2^- is only valid when κ_2^+ is valid, it has no impact on the phase transition claimed in Theorem 1. As a result, we will focus on κ_2^+ for the remainder of the proof and, with a slight abuse of notation, use κ_2^* to denote κ_2^+.

LEMMA 7. *When strengths $\sigma_1\sigma_2 \geq 1$, the fixed-point for the population infected by virus 2 is positive, $\kappa_2^* > 0$, if and only if*

$$\epsilon > \frac{\sigma_1 - \sigma_2}{\sigma_2(\sigma_1 - 1)}.$$

PROOF. Solving equation (19) for $\kappa_2^* = 0$ produces $\epsilon = \frac{\sigma_1-\sigma_2}{\sigma_2(\sigma_1-1)}$. Because κ_2^* is monotonically increasing in this region ($\sigma_1\sigma_2 > 1$), for all ϵ greater than this solution, $\kappa_2^* > 0$, and for all ϵ less than this solution $\kappa_2^* \leq 0$. □

LEMMA 8. *If virus strengths $\sigma_1 + \sigma_2 < 2$, then the fixed-point for the population infected by virus 2 is positive, $\kappa_2^* > 0$, for*

$$\epsilon > \frac{2(1+\sqrt{1-\sigma_1\sigma_2})}{\sigma_1\sigma_2}.$$

PROOF. For κ_2^* to be positive, the numerator of (19) must be positive. We can reduce this as follows:

$$\begin{aligned}
&-2\epsilon\sigma_1\sigma_2 + \epsilon\sigma_1\sigma_2^2 + \epsilon\sqrt{\sigma_1}\sigma_2^{3/2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2} \\
&= \epsilon\sigma_2(\sqrt{\sigma_1\sigma_2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2} - 2\sigma_1 + \epsilon\sigma_1\sigma_2) \\
&\geq \epsilon\sigma_2(0 - 2\sigma_1 + 2(1+\sqrt{1-\sigma_1\sigma_2})) \\
&= 2\epsilon\sigma_2(-\sigma_1 + 1 + \sqrt{1-\sigma_1\sigma_2})
\end{aligned}$$

For this to be positive we must have $\sqrt{1-\sigma_1\sigma_2} > \sigma_1 - 1$, which is true for $\sigma_1 + \sigma_2 < 2$. □

LEMMA 9. *If virus strengths $\sigma_1 + \sigma_2 \geq 2$ then the fixed-point for the population infected by virus 2 is positive, $\kappa_2^* > 0$, for*

$$\epsilon > \frac{\sigma_1 - \sigma_2}{\sigma_2(\sigma_1 - 1)}.$$

PROOF. Again, for κ_2^* to be positive, the numerator of (19) must be positive. We can reduce this as follows:

$$\begin{aligned}
&-2\epsilon\sigma_1\sigma_2 + \epsilon\sigma_1\sigma_2^2 + \epsilon\sqrt{\sigma_1}\sigma_2^{3/2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2} \\
&= \epsilon\sigma_2(\sqrt{\sigma_1\sigma_2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2} - 2\sigma_1 + \epsilon\sigma_1\sigma_2) \\
&\geq \epsilon\sigma_2\left(\sqrt{\sigma_1\sigma_2}\sqrt{\frac{\sigma_1(-2+\sigma_1+\sigma_2)^2}{\sigma_2(-1+\sigma_1)^2}} - 2\sigma_1 + \sigma_1\sigma_2\left(\frac{\sigma_1-\sigma_2}{\sigma_2(1-\sigma_1)}\right)\right) \\
&= \epsilon\sigma_1\sigma_2\left(\frac{2-\sigma_1-\sigma_2}{1-\sigma_1} - 2 + \frac{\sigma_1-\sigma_2}{1-\sigma_1}\right) = 0
\end{aligned}$$

□

LEMMA 10. *The fixed-point for the population infected by virus 2 is valid, $\kappa_2^* \leq 1$, for $\sigma_1 \geq \sigma_2$ and $\epsilon \geq 0$.*

PROOF. The constraint $\kappa_2^* \leq 1$ is equivalent to

$$-2\epsilon\sigma_1\sigma_2 - \epsilon^2\sigma_1\sigma_2^2 + \epsilon\sqrt{\sigma_1}\sigma_2^{3/2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2} < 0.$$

This can be simplified as follows:

$$\sqrt{\sigma_1\sigma_2}\sqrt{4-4\epsilon+\epsilon^2\sigma_1\sigma_2} < 2\sigma_1 + \epsilon\sigma_1\sigma_2 \quad (20)$$

$$\sigma_1\sigma_2(4-4\epsilon+\epsilon^2\sigma_1\sigma_2) < 4\sigma_1^2 + \epsilon^2\sigma_1^2\sigma_2^2 + 4\epsilon\sigma_1^2\sigma_2 \quad (21)$$

$$\sigma_1\sigma_2 - \epsilon\sigma_1\sigma_2 < \sigma_1^2 + \sigma_1^2\sigma_2 \quad (22)$$

$$\frac{\sigma_2}{\sigma_1}\frac{1-\epsilon}{1+\epsilon\sigma_2} < 1 \quad (23)$$

The simplification to (23) makes it clear that the lemma is true for $\sigma_1 \geq \sigma_2 > 0$. □

As such, for any interaction factor $\epsilon > \epsilon_{\text{critical}}$, we have proved that κ_1^* and κ_2^* are valid equilibrium points for which the population infected by each virus $\kappa_1, \kappa_2 > 0$. **QED**

5. EXPERIMENTS

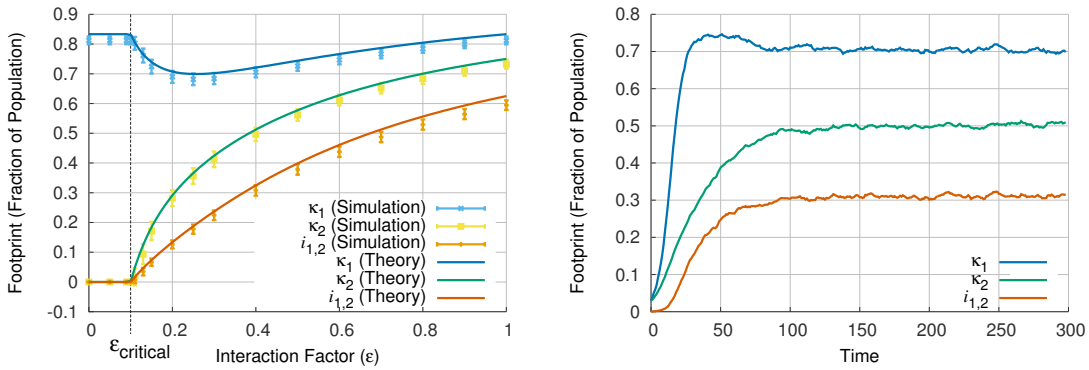
We demonstrate our result using (a) simulation experiments and (b) case studies using real data in this section.

5.1 Setup

Without loss of generality, in our experiments we assumed that the first virus is the stronger virus. We primarily focus on the case where $\sigma_1 > \sigma_2 > 1$. For our simulations we use $\sigma_1 = 6$ and $\sigma_2 = 4$.

We run a simulation on a fully-connected clique of 1000 nodes. We vary ϵ around our expected threshold and for each value of ϵ perform 10 runs over 4000 time steps. On each run we begin by infecting 30 nodes at random with each virus.

We analyze the results in two ways. First, we create a *steady-state plot* of mean values and standard deviations for κ_1, κ_2, and i_{12} at steady-state over a range of values for ϵ. Over the results of the simulation we draw the behavior predicted by our results. Second, for one $\epsilon > \epsilon_{\text{critical}}$ we track each virus's development over time with a *time-plot*. The time-plot takes the average number of nodes infected (κ_1, κ_2, and i_{12}) at each time step and plots this against time. Although the simulations were run for 4000 time steps, the plots are truncated to give more detail to the initial fluctuations of the virus counts.



(a) $\sigma_1 = 6$, $\sigma_2 = 4$ (b) $\epsilon = 0.4 > \epsilon_{\text{critical}} = 1$, $\sigma_1 = 6$, $\sigma_2 = 4$

Figure 3: Coexistence is possible: Results from simulations on clique of size 1000 with theoretical fixed points overlayed. (a) shows the steady-state population values, κ_1, κ_2, and i_{12}, for each value of ϵ, with the theoretical $\epsilon_{\text{critical}}$ marked. (b) shows the development of the two viruses over time for $\epsilon > \epsilon_{\text{critical}}$. Notice that both viruses survive as expected.

5.2 Simulation Results

Figure 3 displays our results. In short, the plots agree exactly with our result, as expected. Figure 3(a) shows the steady-state plot for the two viruses, and the theoretical predictions closely match the simulation results. Similarly, the viruses' growth as shown in the time-plot in Figure 3(b) matches what is expected.

For the steady-state plot, we expect the steady-state value to be one of the other fixed points where at least one virus dies out for $\epsilon \leq \epsilon_{\text{critical}}$ and then a co-existence for $\epsilon > \epsilon_{\text{critical}}$. In Figure 3(a) we see for $\sigma_1 > \sigma_2 > 1$ and $\epsilon = 0$, winner takes all, as was proven in [30]. However, for $\epsilon > \epsilon_{\text{critical}}$ we see a coexistence between the viruses as expected; this is true even when the viruses are competing ($\epsilon < 1$). For $\epsilon = 0.4 > \epsilon_{\text{critical}}$ we have the time-plot, Figure 3(b), showing the growth of both viruses to steady-state from a small infection in the system.

5.3 Case-Studies using Real Data

We collected historical data for 'web-search interest' for various competing products from the Google-Insights website[2], which aims to "provide insights into broad search patterns." This allows us to use the data as a proxy for sales/interest for each product. We used the following pairs of rival products:

1. Hulu[3] and Blockbuster[4]: Although not direct competitors, both offer video entertainment services, though under very different models.
2. Firefox[5] and Google Chrome[6]: Two rival web browsers.

We consider both pairs of products to be examples of cases where there is partial mutual immunity; people can use both products, but the use of one we expect would detract from the use of the other. While our model does not describe the situations perfectly, we believe it is a good approximation.

[2] www.google.com/insights/search/
[3] www.hulu.com
[4] www.blockbuster.com
[5] www.mozilla.org/en-US/firefox/new/
[6] www.google.com/chrome

In Figure 4 we show plots of the web-interest vs. time for both pairs of products, along with our model fitted to the data[7]. In Figure 4(a), we used a virus interaction factor of $\epsilon = 0.7$ (along with virus parameters $\delta_{\text{Hulu}} = 0.04$, $\beta_{\text{Hulu}} = 0.0007$, $\delta_{\text{Blockbuster}} = 0.05$, $\beta_{\text{Blockbuster}} = 0.00045$). In Figure 4(b), we used a virus interaction factor of $\epsilon = 0.6$ (along with virus parameters $\delta_{\text{Firefox}} = 0.01$, $\beta_{\text{Firefox}} = 0.000095$, $\delta_{\text{Chrome}} = 0.01$, $\beta_{\text{Chrome}} = 0.00015$). In Figure 4(c) we use the same model as (b) but let the model continue to see the projected steady state behavior. We note that the plots begin when Hulu and Chrome are first introduced and with Blockbuster and Firefox at a previous steady-state behavior. In each of these fittings we see that our model fits the data well. The fact that the model fits the data well demonstrates the suitability of our $SI_{1|2}S$ model.

6. DISCUSSION

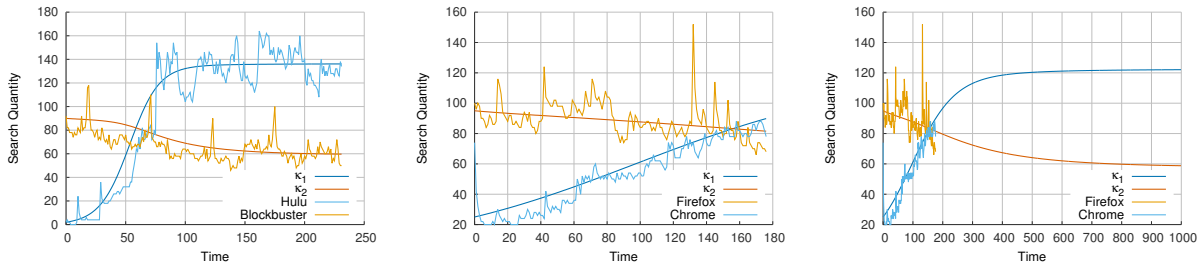
6.1 A general upper bound

Conjecture 1 (Epsilon Threshold Upper Bound). *Given an arbitrary graph with the $SI_{1|2}S$ model parameters $\sigma_1 \geq \sigma_2 \geq 1$, an equilibrium point for which both virus 1 and virus 2 survive exists if $\epsilon > \epsilon_{\text{critical}}$, where*

$$\epsilon_{\text{critical}} \leq \frac{1}{\sigma_2} \tag{24}$$

Justification: Since $\sigma_1 \geq \sigma_2 \geq 1$ and $0 < \epsilon < 1$, we know that both virus 1 and virus 2 would be strong enough to survive independently but there is some competition between them. Because of the competition, as virus 1 spreads to more nodes, virus 2's attack rate on average decreases and thus its strength decreases. Therefore, if we overestimate the strength or number of people with virus 1, this only makes it more difficult for virus 2 survive and thus decreases the maximum amount of competition virus 2 can handle, increasing $\epsilon_{\text{critical}}$. To simplify the problem, we assume that *every person* is infected with virus 1 (as if virus 1 was infinitely strong). In this case, a node can only be in state I_1

[7] Fitted with www.alexbeutel.com/jsplot/kdd2012.html



(a) Hulu and Blockbuster (b) Firefox and Chrome (c) Firefox and Chrome (Prediction)

Figure 4: Real web-search interest vs. time plots for pairs of competitors with our model fitted to it. (c) Predicts steady state values based on our model. Data acquired from Google Trends.

or $I_{1,2}$, and the probability of a node with virus 2 infecting a neighbor is always $\epsilon\beta_2$. This is now equivalent to a one virus model where the strength of virus 2 is $\sigma_2' = \epsilon\sigma_2$. Therefore, as shown in previous research [10, 13, 31], if $\sigma_2' > 1$ then virus 2 will survive. In this case, $\epsilon_{\text{critical}} = \frac{1}{\sigma_2}$. However, because this is a relaxation of the original problem, we know that this is an upper bound and in fact $\epsilon_{\text{critical}} \leq \frac{1}{\sigma_2}$.

6.2 Case-Study: Qualitative Analysis

We also consider the example of educational ideas, and specifically sex education, as a virus. Sociology literature analyzing the success of sex education programs notes the impact of network effects and social structure on sex education success [7]. We match education policy to our $SI_{1|2}S$ model and analyze the implications.

Policy 1: Abstinence-Only Education Abstinence-only education teaches abstinence until marriage as the only way to live a healthy life, with students often taking an abstinence pledge. Under our model, virus 1 is believing in abstinence (through education or pledge) and virus 2 is sexual activity. Therefore, those who are in I_1 have taken a pledge of abstinence, those who are in I_2 are sexually active, and those people who are in S do not believe in abstinence but are not sexually active either. It is obviously impossible to both be following an abstinence pledge and to be sexually active so nobody can be in state $I_{1,2}$. Equivalently in this case there is full mutual immunity or $\epsilon = 0$.

Model 1 Predictions and Results Based on this fit, because $\epsilon = 0$, our model predicts 'winner takes all:' the weaker virus dies out and the stronger virus survives. Sociology research [7], studying over 11,000 people over 7 years, notes that of the 2399 people claiming to have taken an abstinence pledge, 1622 (67%) over time forgot. This suggests that $\sigma_{\text{Abstinence}} < \sigma_{\text{Sexual Activity}}$, and as a result, in the long run sexual activity will win over abstinence.

Policy 2: Comprehensive Sex Education Comprehensive sex education teaches numerous methods to have a safe, healthy sex life, discussing both contraception and abstinence. Matching this to our model, virus 1 is being educated in safe-sex practices and valuing their importance and virus 2 is sexual activity. Therefore, those who are in I_1 have been educated about safe-sex practices and believe they are important but are not sexually active, those in I_2 are sexually active but do not practice safe sex, those in $I_{1,2}$ practice safe-sex, and those in S are neither educated on safe-sex practices nor sexually active. Here we expect little to no competition between the two viruses and thus have an ϵ value close to, if not equal to, 1.

Model 2 Predictions and Results Because ϵ is close to 1, we expect that $\epsilon > \epsilon_{\text{critical}}$. As a result, it is possible for there to be coexistence of the two viruses, such that there can be a steady-state in which people are sexually active and practice safe-sex. This appears to match sociology literature claiming that those who initially use condoms will keep using condoms [7].

In summary, our model qualitatively agrees with sociology research and offers a plausible explanation for the results of the study. Additionally, these two cases demonstrate the value of a phase transition. In the first case, the model suggests winner takes all and the ineffectiveness of abstinence-only education. On the contrary, for policy 2, the model predicts coexistence, which agrees with the findings, and is better for society.

6.3 Subtle Points

There are several subtle points, that we deferred until now, for clarity of exposition. Specifically, here we discuss the following issues:

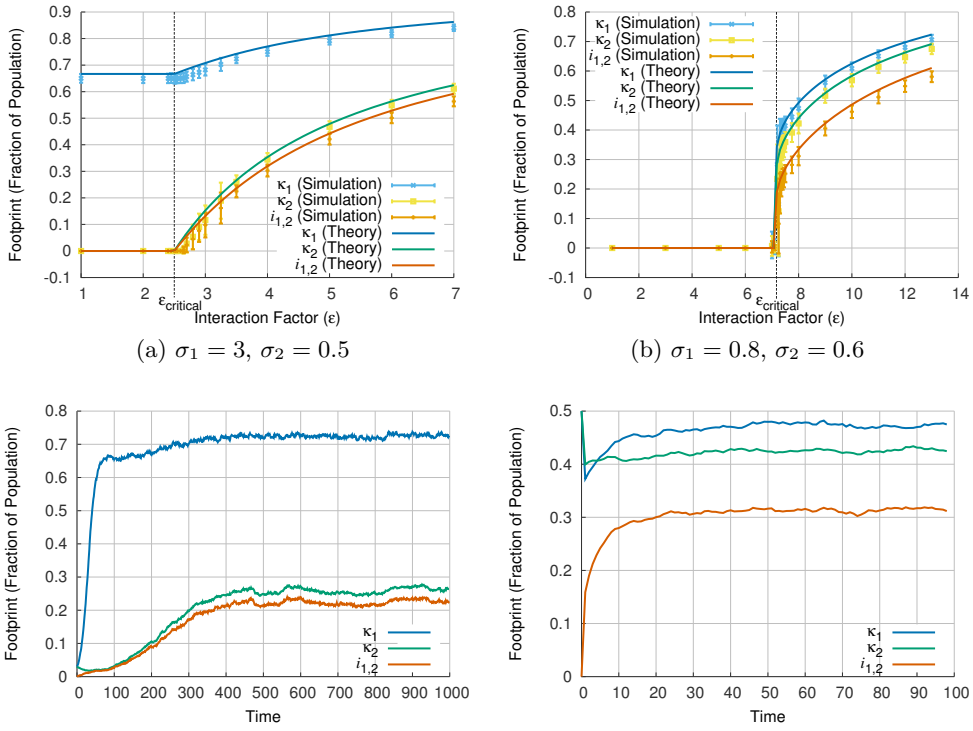
6.3.1 *What does it mean for $\epsilon > 1$?*

As before, the virus 2 transmission rate for a virus infected with virus 1 becomes $\epsilon\beta_2$ and the virus 1 transmission rate for a virus infected with virus 2 becomes $\epsilon\beta_1$. However, because $\epsilon > 1$ the transmission rate for each virus increases for neighbors that are already infected. We consider this to be a form of cooperation between the viruses (products, ideas, etc.).

This pattern of cooperation between products is common in product ecosystems. An example of this is that people who have an iPod are more likely to buy music and videos through Apple's iTunes. Making use of such cooperation can be seen in 'freebie marketing' or the 'razor and blades business model,' in which the company producing razor blades sells the razors at an artificially low price creating a market for the blades. This method of tightly integrating products is common in a variety of industries.

6.3.2 *What happens if σ_2 or $\sigma_1 \leq 1$?*

Because $\sigma_1 \geq \sigma_2$ there are two cases we can analyze. The first is when $\sigma_1 \geq 1 > \sigma_2$, in which the second virus is too weak to survive on its own. We will refer to this as the "piggyback setting," because virus 2 can only survive with the help of the first. The second condition is when $1 > \sigma_1 \geq \sigma_2$, where both viruses are independently too weak to survive. We will refer to this as the "teamwork setting," because only through cooperation can both viruses survive.



(a) $\sigma_1 = 3$, $\sigma_2 = 0.5$

(b) $\sigma_1 = 0.8$, $\sigma_2 = 0.6$

(c) $\epsilon = 3.5$, $\sigma_1 = 3$, $\sigma_2 = 0.5$

(d) $\epsilon = 8$, $\sigma_1 = 0.8$, $\sigma_2 = 0.6$

Figure 5: With enough collaboration, even weak viruses can survive. Results from simulations on population of $N = 1000$ with theoretical values overlayed, and $1 > \sigma_2$. The first column, (a) and (c), is the "piggyback" case where just virus 2 is too weak to survive. The second column, (b) and (d), is the "teamwork" case where neither virus is on its own strong enough to survive. The top row gives the steady-state plots, showing the steady-state footprint vs. the interaction factor ϵ. The second row gives the time-plots, showing the infection footprint developing over time.

In each of these cases, Theorem 1 still holds - plugging in σ_1 and σ_2 we find an $\epsilon_{\text{critical}}$ for which a fixed point in which $\kappa_1, \kappa_2 > 0$ exists. This suggests that even if both viruses are independently too weak to survive on their own, with enough cooperation they can.

To test our theorem, we ran similar simulations where either one or both viruses were too weak to independently survive. In Figure 5(a) we show the steady-state plot for the piggyback case of $\sigma_1 > 1 > \sigma_2$. As expected, for $\epsilon = 1$ when the viruses are independent, virus 1 survives but virus 2 is not strong enough and dies out. For a sufficient amount of cooperation, $\epsilon > \epsilon_{\text{critical}}$, we find that virus 2 can survive as well. We see in Figure 5(c), the corresponding time-plot where $\epsilon = 3.5 > \epsilon_{\text{critical}}$, that once virus 1 grows, virus 2 is able to survive as well.

We also simulated the teamwork case, where neither virus is independently strong enough to survive, $1 > \sigma_1 \geq \sigma_2$. In Figure 5(b) we show the steady-state plot for this case. Again, the theoretical result and predicted phase-transition match the simulation results. For $\epsilon = 1$, the two viruses are independent and, since they cannot survive on their own, die out. In this case, the phase transition is based on the second part of Theorem 1 where $\sigma_1 + \sigma_2 < 2$, and as such the bound is a result on the restriction of κ_2 being real. As such, at $\epsilon_{\text{critical}}$ both $\kappa_1, \kappa_2 > 0$ rather than equal to 0. We see here at the threshold a large amount of uncertainty in the simulation but as we move away from the threshold the simulation follows this new fixed point. Interestingly, we must initially infect a large portion of the graph for the system to go to this fixed point, and not die out. For $\epsilon = 7.75 > \epsilon_{\text{critical}}$, Figure 5(d) shows the time-plot, demonstrating that both viruses quickly reach steady-state with a high amount of overlap.

7. CONCLUSIONS

We defined and studied the problem of partial competition, where two viruses/products provide partial immunity against each other.

The main contributions of our work are as follows:

1. *Problem Definition*: The problem is novel, in the data mining and web mining communities, and even in the epidemiology literature [3, 17].
2. *Threshold Result and Proof*: We showed that there is a phase transition: 'winner takes all,' until the competition level drops below a critical value. Above this critical value we find a closed-form steady-state solution with coexistence.
3. *Experiments and Case-studies*: We showed results from real settings (like browsers - Firefox vs Google Chrome), which agree with our model.

Acknowledgments: 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 Grant No. IIS-1017415, the National Institute of General

Medical Sciences under MIDAS grant 1U54GM088491-0109, and a National Science Foundation Graduate Research Fellowship.

8. REFERENCES

[1] A. Ackleh and L. Allen. Competitive exclusion in sis and sir epidemic models with total cross immunity and density-dependent host mortality. *Discrete and Continuous Dynamical Systems-Series B*, 5, 2005.

[2] R. M. Anderson and R. M. May. Coevolution of hosts and parasites. *Parasitology*, 85, 1982.

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

[4] N. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, London, 1975.

[5] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. *WINE*, 2007.

[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] H. Brückner and P. Bearman. After the promise: the std consequences of adolescent virginity pledges. *Journal of Adolescent Health*, 2005.

[8] C. Castillo-Chavez, W. Huang, and J. Li. Competitive exclusion in gonorrhea models and other sexually transmitted diseases. *SIAM J. Appl. Math*, 56, 1996.

[9] C. Castillo-Chavez, W. Huang, and J. Li. Competitive exclusion and coexistence of multiple strains in an sis std model. *SIAM J. Appl. Math*, 59, 1999.

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

[11] C. Colijn, T. Cohen, C. Fraser, W. Hanage, E. Goldstein, N. Givon-Lavi, R. Dagan, and M. Lipsitch. What is the mechanism for persistent coexistence of drug-susceptible and drug-resistant strains of streptococcus pneumoniae? *Journal of The Royal Society Interface*, 7(47):905–919, 2010.

[12] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66, New York, NY, USA, 2001. ACM Press.

[13] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology in spread of epidemics. *IEEE INFOCOM*, 2005.

[14] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 2001.

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

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

[17] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42:599–653, 2000.

[18] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03*, 2003.

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

[20] J. Kosta, Y. A. Oswald, and R. Wattenhofer. Word of mouth: Rumor dissemination in social networks. *15 Intl. Coll. on Struct. Inform. and Comm. Complexity SIROCO*, 2008.

[21] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. On the bursty evolution of blogspace. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 568–576, New York, NY, USA, 2003. ACM Press.

[22] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila. Finding effectors in social networks. In *KDD*, pages 1059–1068, 2010.

[23] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC*, 2006.

[24] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. Cascading behavior in large blog graphs: Patterns and a model. In *Society of Applied and Industrial Mathematics: Data Mining*, 2007.

[25] M. Lipsitch, C. Colijn, T. Cohen, W. P. Hanage, and C. Fraser. No coexistence for free: Neutral null models for multistrain pathogens. *Epidemics*, 1(1):2 – 13, 2009.

[26] M. Lipsitch, S. Siller, and M. A. Nowak. The evolution of virulence in pathogens with vertical and horizontal transmission. *Evolution*, 50, 1996.

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

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

[29] N. Pathak, A. Banerjee, and J. Srivastava. A generalized linear threshold model for multiple cascades. *ICDM*, 2010.

[30] B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos. Winner takes all: Competiting viruses or ideas on fair-play networks. *WWW*, 2012.

[31] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. In *ICDM*, 2011.

[32] 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.

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

[34] D. Shah and T. Zaman. Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.

[35] M. H. H. Stevens. *A Primer of Ecology with R*. Springer Press, 2009.

[36] H. Tong, B. A. Prakash, C. E. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, and D. H. Chau. On the vulnerability of large graphs. In *ICDM*, 2010.