Winner Takes All: Competing Viruses or Ideas on fair-play Networks

B. Aditya Prakash Alex Beutel Roni Rosenfeld Christos Faloutsos

Computer Science Department, Carnegie Mellon University, USA
Email: {badityap, abeutel, roni, christos}@cs.cmu.edu

ABSTRACT

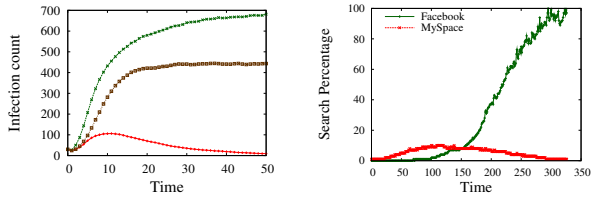
Given two competing products (or memes, or viruses etc.) spreading over a given network, can we predict what will happen at the end, that is, which product will 'win', in terms of highest market share? One may naïvely expect that the better product (stronger virus) will just have a larger footprint, proportional to the quality ratio of the products (or strength ratio of the viruses). However, we prove the surprising result that, under realistic conditions, for *any* graph topology, the stronger virus completely wipes-out the weaker one, thus not merely 'winning' but 'taking it all'. In addition to the proofs, we also demonstrate our result with simulations over diverse, real graph topologies, including the social-contact graph of the city of Portland OR (about *31 million* edges and *1 million* nodes) and internet AS router graphs. Finally, we also provide real data about competing products from Google-Insights, like Facebook-Myspace, and we show again that they agree with our analysis.

1. INTRODUCTION

Given two competing products like iPhone/Android or Blu-ray/HD-DVD, and 'word of mouth' adoption of them, what will happen in the end? This question is of interest in numerous settings. For example, in a biological virus setting, we have the common flu versus avian flu. In a computer virus setting, clever virus authors make sure that their code eliminates most other computer viruses from the victim's disk.

The list continues, with competing scientific theories, competing memes ('coke' vs 'soda' vs 'pop'), and many more.

Our main result is that we answer the above question analytically, and we show that 'winner takes all' (WTA), or, more accurately, the weaker product/virus will soon become extinct. The fate of the stronger virus depends on its strength: below the epidemic threshold (more details, later), it will also become extinct, but above that it has good chances of lingering practically for ever. In more detail, we assume



(a) Simulations (b) Facebook vs Myspace

Figure 1: (a) Number of infected vs time for simulations on the AS-OREGON network for a virus propagating in isolation (brown square) - note that it is above the epidemic threshold and hence doesn't die out, and the same virus competing with an even stronger virus (green) - note that it now dies-out *completely* (red). (b) Winner takes all in search interest data from Google-Insights for Facebook (green) and Myspace (red). Even though Myspace got a head-start, Facebook wiped it out.

(a) an SIS-like model (no immunity, like flu),
(b) perfect mutual immunity (a node can have at most one of the viruses/products, at any given time),
(c) the underlying network is connected (every node can reach every other node)
(d) the network is '*fair-play*', in the sense that all nodes have the same behavior towards the two competing products/viruses: everybody has the same probability β_1 of getting infected with virus-1 by a sick neighbor, and similarly for virus-2, and for the recovery times.

One of the main contributions is that our theoretical analysis holds for *any* graph topology, while earlier work focuses only on specific-topology graphs (cliques, random, etc).

Figure 1(a) gives an illustration of our result: it shows the number of infected nodes vs time for computer simulations on the AS-OREGON network (see Section 5 for details) for a 'above-threshold' virus propagating in isolation (brown square) in one case and the virus competing with an even stronger virus (green) in another case. Clearly it is wiped out during the competition, although it gave a fight (red, note the bump). Note that though both the viruses are above the threshold, the weaker virus is wiped out. We prove this result for *arbitrary* underlying networks in this paper.

Figure 1(b) shows the time evolution of search-interest for a pair of competing products Facebook-Myspace. The data came from Google-Insights. Notice that again, the weaker

competitor is extinct (or close to that). We will give more case-studies later in Section 5.

The outline of the paper is as follows: 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 WTA result in § 4 while we demonstrate it using simulations and real case-studies in § 5. Finally, we discuss some subtle issues in § 6 and conclude in § 7.

2. RELATED WORK

We present the related work in this section, which can be categorized into three parts: epidemic thresholds, information diffusion and ecology. Most of these works either consider only single virus models or typically use only simulation or analyze on very restricted underlying networks.

Epidemic Thresholds Much research in virus propagation has been devoted to studying the so-called epidemic threshold, that is, to determine the condition under which an epidemic will not break out. Widely-studied epidemiological models include the so-called *homogeneous models* [4, 26, 3], which assume that every individual has equal contact to others in the population. While earlier works [20, 28] focus on some specific types of graph structure (e.g., random graphs, power-law graphs, etc), Chakrabarti et al. [9] and Ganesh et al. [11] found that, 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. [30] further discovered that the leading eigenvalue and a model-dependent constant are the only parameters that determine the epidemic threshold for almost all virus propagation models. However, all of these works focus on single virus models.

Information Diffusion There is also a lot of research interest in studying other types of information dissemination processes on large graphs, including (a) information cascades [6, 12], (b) blog propagation [24, 14, 22, 31], and (c) viral marketing and product penetration [23]. Broadly two classes of information cascade models have been proposed (a) independent cascade [19] (essentially a 'SIR' model) and (b) linear threshold [13]. Research work in multiple cascades has looked into extensions of the independent cascade model with the restriction that nodes can't switch from one competitor to the other [5, 21]. One of the few works to consider switching between the competitors is Pathak et. al. [29]. However, their work differs from the current one in several important aspects: (a) they use the linear threshold model, as opposed to the 'flu-like' SIS model (a cascade style model) that we use; (b) they assume that nodes may randomly switch between products; (c) they do *not* find winner-takes-all phenomena; and (d) they give no closed-form results - only an algorithm to compute the steady state.

Ecology In ecology, the principle of '*competitive exclusion*' espouses that two species can not occupy the same ecological niche in the long term. Research has gone into studying this using various propagation models like SIS, SIR, Lotka-Volterra etc. (for example, see [7, 8, 1, 2]). However, they typically did simulations, or they only studied homogenous or very structured topologies like cliques.

Distinguishing features of current work: In short, none of the previous work fulfills all the conditions of this current work: (a) analytical proof of 'WTA' (b) in arbitrary topologies (c) under a SIS-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 ($\mathbf{A}$, $\mathbf{C}$ etc.) or vectors ($\tilde{\mathbf{P}}, \tilde{\mathbf{u}}$ etc.).

Table 1: Terms and Symbols

Symbol	Definition and Description
WTA	Winner-Takes-All
SI_1I_2S	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)
$\mathbf{A}$	adjacency matrix of the underlying graph
$\lambda_{\mathbf{M}}$	set of eigenvalues of the matrix $\mathbf{M}$
$\lambda_1(\mathbf{M})$	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)
$\mathbf{M}^T$	transpose of $\mathbf{M}$
$NE(i)$	set of neighbors of node i in the graph
$\mathbf{I}$	identity matrix of appropriate size
$\mathbf{0}$	all-zeros matrix of appropriate size
diag($\tilde{\mathbf{P}}$)	the diagonal matrix with elements of vector $\tilde{\mathbf{P}}$ in the diagonal

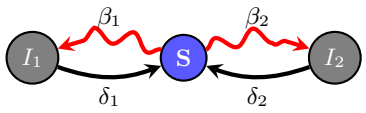


Figure 2: State Diagram for a node in the graph under our SI_1I_2S competing viruses model. The node is in the S state if it doesn't have either competitor (say iPhone or Android). It is in I_1 if it gets the iPhone (virus 1) and is in I_2 if it gets the Android (virus 2). The transitions from S to I_1 or I_2 (red-curvy arrows) depend on the infected neighbors of the node. The remaining transitions, in contrast, are self-transitions, without the aid of any neighbor.

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_1I_2S, based on the popular "flu-like" SIS (Susceptible-Infected-Susceptible) model [16]. SI_1I_2S denotes Susceptible - Infected$_1$ - Infected$_2$ - Susceptible. Each node in the graph can be in one of three states: Susceptible (healthy), I_1 (infected by virus 1), or I_2 (infected by 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 state I_1 (or I_2), it recovers on its own with rate δ_1 (or δ_2). This implies that the time taken for each infected node to heal is exponentially distributed with parameter δ_1 (or δ_2). This parameter captures the persistence of the virus in an inverse way: a high δ means low persistence. For example, a very convincing rumor that sticks to one's mind will be modeled with a low δ value.

Attack (virus transmission) rate: β. A healthy node gets infected by infected neighbors, and the virus transmissability is captured by β_1 and β_2. Specifically, an infected node transmits its infection to each of its neighbors *independently* at rate β_1 (or β_2). Hence, the time taken for each infected node to transmit the virus to a neighbor over a link is exponentially distributed with parameter β_1 (or β_2).

This is a novel generalization of the single-virus SIS model to a competing-viruses scenario. Note the competition between the viruses: each virus has to compete with the other for healthy victims. Moreover, note that we assume *full mutual immunity*: while a node is infected by one virus, it cannot be infected by the other.

Fair-play: We assume that the competitors are playing a 'fair game': All nodes in the network have the same model parameters (β's and δ's) 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.

Competing viruses problem

Given: A undirected connected graph G, and the propagation model (SI_1I_2S) parameters (β_1, δ_1 for virus 1, β_2, δ_2 for virus 2)

Find: What will happen at the end i.e. what are the steady-state populations of the two viruses.

4. WTA: RESULTS AND PROOFS

We prove our winner-takes-all (WTA) result on an arbitrary undirected graph in this section. Our main result can be formally stated as follows:

THEOREM 1 (WINNER TAKES ALL). *Given an arbitrary undirected, connected graph with adjacency matrix* $\mathbf{A}$ *and the SI_1I_2S model parameters ($\beta_1, \beta_2, \delta_1, \delta_2$), then virus 1 will dominate and virus 2 will completely die-out in the steady state if virus 1 is above threshold*[1] *and the strength of virus 1 is greater than the strength of virus 2 i.e. if $\sigma_1 > 1$ and $\sigma_1 > \sigma_2$.*

The proof is involved, and we present it in the next few pages. We will first prove it for simpler cases of the underlying network - namely a clique and a barbell before we move on to arbitrary graphs.

4.1 Proof roadmap

In short, the proof has the following steps:

1. **Dynamical System**: construct a suitable dynamical system of differential equations for the propagation process,
2. **Fixed Points**: prove that there are only *three* fixed points and at least one of the viruses has to die out at any fixed point, and
3. **Stability Conditions**: give the precise conditions for each fixed point to be stable (attracting).

Intuitively, the dynamical system generates a field on which we show that only 3 possible fixed points can exist. Moreover the field makes only one of the possible fixed points stable under any given scenario. Figures 3(a-c) shows the field-plots in a simple case - when the underlying graph is a clique[2] of size $N = 1000$. Specifically we show three scenarios (wlog, we assume the first virus is the stronger virus):

BELOW : $1 > \beta_1 N/\delta_1 = 0.6 > \beta_2 N/\delta_2 = 0.2$
(both viruses below the threshold)

MIXED : $\beta_1 N/\delta_1 = 6 > 1 > \beta_2 N/\delta_2 = 0.2$
(one above and one below the threshold)

ABOVE : $\beta_1 N/\delta_1 = 6 > \beta_2 N/\delta_2 = 4 > 1$
(both above the threshold)

The field plots illustrate the fixed points in this setting and their stability. In this case, we have a 2-dimensional field, but for an arbitrary graph it will depend on the number of nodes in the graph. At any point on the field, the direction of the field-arrow tells us where the system will go next. Stable fixed points are marked by bold circles, unstable fixed points by hollow circles, x-axis denotes the # of infected nodes by virus 2 and the y-axis denotes the # infected by virus 1 (the stronger virus). For example, in Figure 3(c), both viruses are above threshold, yet the FP_1 and FP_3 points are unstable while the other fixed point corresponding to the stronger virus winning (FP_2) is stable. The trajectory of the simulation is overlaid on the field plots - we can see that the system follows the field lines and is attracted towards and ends up at the stable fixed point in the steady state. We also show the time-evolution separately in Figures 3(d-f) - especially note part (f) (ABOVE), virus 2 tries to take over, but is over-powered by virus 1 which goes on to dominate. We can similarly observe the BELOW and MIXED scenarios as well.

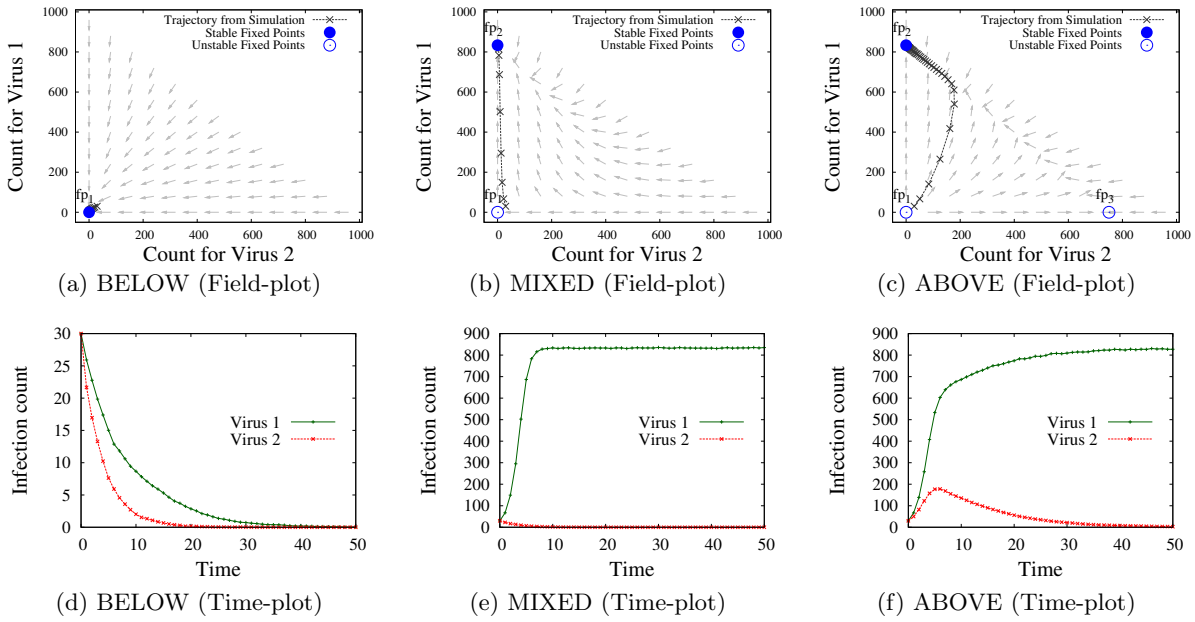
We elaborate a bit more on the steps next. Consider a dynamical system (set of differential equations) of the form $x' = F(x)$, where x' is the (component-wise) time derivative of x, and $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and differentiable. If $F(x_0) = 0$, then x_0 is a stationary point (also called a fixed point). The proof begins by setting up the propagation as a dynamical system of non-linear differential equations and then analyzes the possible fixed points and their stability conditions. In principle, one might expect that there might be several fixed points of the system corresponding to different proportions of the populations of the two virus. But we prove that in fact there are always only three fixed points possible and in each at least one virus gets wiped-out.

Further, intuitively, if a fixed point is not stable then the system would be repelled whenever it tries to approach that fixed point. Hence, to fully characterize the fixed points, we need to derive the stability conditions, which give us the conditions for each of these fixed points to be stable and attracting.

For characterizing the stability of the fixed points, we use a well-known result from dynamical system theory (c.f. [17]). The fixed points will be a hyperbolic fixed point (i.e. where

[1]It is known that in the *single-virus* SIS model, a virus dies-out unless it is above the epidemic threshold i.e. unless $\beta\lambda/\delta > 1$ [30], where λ is the largest eigenvalue of the adjacency matrix of the underlying graph.

[2]every node is connected to every other node.



(a) BELOW (Field-plot) (b) MIXED (Field-plot) (c) ABOVE (Field-plot)

(d) BELOW (Time-plot) (e) MIXED (Time-plot) (f) ABOVE (Time-plot)

Figure 3: (a-c) Field plots for clique of size 1000 for various cases. Stable fixed points are marked by bold circles, unstable fixed points by hollow circles, while the y-axis denotes the number of infected nodes by virus 1 and the x-axis denotes the number infected by virus 2. The simulation trajectory is overlaid. The field plots illustrate the fixed points in this setting and their stability. (d-f) Corresponding Time evolution plots of the competition. Note that virus 2 (red curve) in (f) could have won in isolation, but lost to virus 1.

the linearized stability analysis can be performed) only when none of the eigenvalues of the corresponding Jacobian[3] has a zero real part. Further, the system will be stable at a hyperbolic fixed point (attractor) only if the real part of the eigenvalues of the Jacobian is negative. So to make a fixed point a hyperbolic stable attractor we need to impose the condition that *the real parts of all the eigenvalues of the corresponding Jacobian should be negative*. We next show this proof scheme for the special case of a clique topology.

4.2 Special case: Clique Topology

In a clique, all nodes are connected to each other with undirected and unweighted edges. Each node is identical to the other and hence our system is a simple continuous time markov chain, due to which we can write down the system equations directly. Let N be the size of the clique and I_1 be the number of nodes infected by virus 1 at some time t. Similarly define I_2.

Dynamical System: Clearly, under our SI_1I_2S model, we have the following system equations:

$$\frac{dI_1}{dt} = \beta_1(N - I_1 - I_2)I_1 - \delta_1 I_1$$
$$\frac{dI_2}{dt} = \beta_2(N - I_1 - I_2)I_2 - \delta_2 I_2$$

Fixed Points There are three fixed points of the system of differential equations above (when the rates of change in I_1 and I_2 are zero):

1. $\{I_1 \to 0, I_2 \to 0\}$ (i.e. the viruses die-out)
2. $\left\{I_1 \to N - \frac{\delta_1}{\beta_1}, I_2 \to 0\right\}$ (i.e. only virus 1 survives)
3. $\left\{I_2 \to N - \frac{\delta_2}{\beta_2}, I_1 \to 0\right\}$ (i.e. only virus 2 survives)

Stability Conditions The corresponding Jacobians at the fixed points are:

1. $J_1 = \begin{bmatrix} N\beta_1 - \delta_1 & 0 \\ 0 & N\beta_2 - \delta_2 \end{bmatrix}$
2. $J_2 = \begin{bmatrix} -\delta_1 + \beta_1\left(N - 2\left(N - \frac{\delta_1}{\beta_1}\right)\right) & -\beta_1\left(N - \frac{\delta_1}{\beta_1}\right) \\ 0 & \frac{\beta_2\delta_1}{\beta_1} - \delta_2 \end{bmatrix}$
3. $J_3 = \begin{bmatrix} -\delta_1 + \frac{\beta_1\delta_2}{\beta_2} & 0 \\ -\beta_2\left(N - \frac{\delta_2}{\beta_2}\right) & -\delta_2 + \beta_2\left(N - 2\left(N - \frac{\delta_2}{\beta_2}\right)\right) \end{bmatrix}$

The eigenvalues of the Jacobians can be seen to be:

1. $\lambda_{\mathbf{J}_1} \equiv \left\{\beta_1(N - \frac{\delta_1}{\beta_1}), \beta_2(N - \frac{\delta_2}{\beta_2})\right\}$
2. $\lambda_{\mathbf{J}_2} \equiv \left\{\beta_1(\frac{\delta_1}{\beta_1} - N), \beta_2(\frac{\delta_1}{\beta_1} - \frac{\delta_2}{\beta_2})\right\}$
3. $\lambda_{\mathbf{J}_3} \equiv \left\{\beta_1(\frac{\delta_2}{\beta_2} - N), \beta_2(\frac{\delta_2}{\beta_2} - \frac{\delta_1}{\beta_1})\right\}$

From our preceding discussion we know that to have stable fixed points we require that the real part of the eigenvalues of the Jacobians should be negative. Clearly the corresponding conditions for the fixed point to be (a) hyperbolic and (b) stable attractor are:

1. $\frac{\beta_1 N}{\delta_1} < 1$ and $\frac{\beta_2 N}{\delta_2} < 1$
   (i.e. both are below threshold)

[3]The Jacobian is the matrix of all component-wise first-order partial derivatives of x' with respect to x evaluated at the fixed point.

2. $\frac{\beta_1 N}{\delta_1} > 1$ and $\frac{\beta_1 N}{\delta_1} > \frac{\beta_2 N}{\delta_2}$
(i.e. virus 1 is above threshold and virus 1 strength is greater than virus 2)
3. $\frac{\beta_2 N}{\delta_2} > 1$ and $\frac{\beta_2 N}{\delta_2} > \frac{\beta_1 N}{\delta_1}$
(i.e. virus 2 is above threshold and virus 2 strength is greater than virus 1)

Firstly note that we recover a result similar to the single-virus SIS model case - that if the viruses are below the epidemic threshold, they both die-out. Secondly, we can conclude that in case of a clique, the stronger virus wipes-out the weaker virus if it is above the epidemic threshold.

Non-hyperbolic fixed points: We can see that the fixed points will be non-hyperbolic if the virus strengths are equal. Hence, in this case, no conclusions can be drawn from the linearized analysis and we take a different route. Note that we always have:

$$\int_{I_1^0}^{I_1} \frac{dI_1}{\beta_1 I_1} + \int_0^t \frac{\delta_1}{\beta_1} dt = \int_{I_2^0}^{I_2} \frac{dI_2}{\beta_2 I_2} + \int_0^t \frac{\delta_2}{\beta_2} dt \quad (1)$$

$$\Rightarrow \frac{I_1^{\beta_2}}{I_2^{\beta_1}} \times \frac{(I_2^0)^{\beta_1}}{(I_1^0)^{\beta_2}} = e^{\beta_1\beta_2(\frac{\delta_2}{\beta_2} - \frac{\delta_1}{\beta_1})t} \quad (2)$$

where I_1^0 and I_2^0 are the initial values of I_1 and I_2. The R.H.S. will evaluate to one in our case (the virus strengths are equal). Hence now, the ratio of virus populations at any given time t will be directly proportional to the initial ratio (up to some exponents). Also, clearly, the maximal ratios are attained at one of the last two fixed points.

4.3 Special Case: Barbell Graph

A barbell graph G has two cliques (say clique C_1 and C_2 of size N each) connected through weak edges. Specifically, we assume that all nodes in C_1 are connected to all nodes in C_2 (and vice versa) with edges of weight ϵ, whereas they are connected with nodes within the same clique with edges of weight 1. In this case, by symmetry we can see that the virus populations in both the cliques should remain the same at the steady state. If we follow the steps in the case of a single clique, we get at steady state:

$$\beta_1(N - I_1 - I_2)I_1(1+\epsilon) = \delta_1 I_1$$
$$\beta_2(N - I_1 - I_2)I_2(1+\epsilon) = \delta_2 I_2$$

Hence, the only possible fixed points are:

1. $\{I_1 \to 0, I_2 \to 0\}$ (i.e. the viruses die-out)
2. $\left\{I_1 \to N - \frac{\delta_1}{\beta_1 * (1+\epsilon)}, I_2 \to 0\right\}$ (i.e. only virus 1 survives)
3. $\left\{I_2 \to N - \frac{\delta_2}{\beta_2 * (1+\epsilon)}, I_1 \to 0\right\}$ (i.e. only virus 2 survives)

Moreover, continuing similarly to the single clique case, we can see that the stronger virus again wipes-out the weaker virus as long as it is above the epidemic threshold (note that in this case $\lambda = (1+\epsilon)N$, hence the threshold condition for a single virus is $(1+\epsilon)\beta N/\delta > 1$).

4.4 General Arbitrary Graph

Let $\mathbf{A}$ be the adjacency matrix of the arbitrary graph of N nodes. Let $p_{i,1}$ be the probability of node i to be in the I_1 state. Similarly define $p_{i,2}$ and s_i is the probability of node i being in the susceptible state. Clearly, $s_i + p_{i,1} + p_{i,2} = 1$.

Dynamical System: As we have a continuous time process, we can write the following system equations, for each node i:

$$\frac{dp_{i,1}}{dt} = -\delta_1 p_{i,1} + \beta_1(1 - p_{i,1} - p_{i,2}) \sum_j (\mathbf{A}_{ij}\mathbf{1}_{j,1})$$

$$\frac{dp_{i,2}}{dt} = -\delta_2 p_{i,2} + \beta_2(1 - p_{i,1} - p_{i,2}) \sum_j (\mathbf{A}_{ij}\mathbf{1}_{j,2})$$

where $\mathbf{1}_{j,k}$ (for $k = 1, 2$) is the indicator random variable denoting if node j is infected with virus k. Our system is *not* a markov chain due to the presence of random variables $\mathbf{1}_{j,k}$ in the rate equations. But after making a mean-field approximation ($\mathbf{1}_{j,1} \approx E[\mathbf{1}_{j,1}] = p_{j,1}$ and $\mathbf{1}_{j,2} \approx p_{j,2}$, where $E[X]$ is the expected value of the random variable X), we get the following dynamical system:

$$\frac{dp_{i,1}}{dt} = -\delta_1 p_{i,1} + \beta_1(1 - p_{i,1} - p_{i,2}) \sum_j (\mathbf{A}_{ij}p_{j,1}) \quad (3)$$

$$\frac{dp_{i,2}}{dt} = -\delta_2 p_{i,2} + \beta_2(1 - p_{i,1} - p_{i,2}) \sum_j (\mathbf{A}_{ij}p_{j,2}) \quad (4)$$

(for each node i).

Fixed Points: At the steady state i.e. at fixed points where the change in probabilities will be zero, we get (for each node i):

$$\delta_1 p_{i,1} = \beta_1(1 - p_{i,1} - p_{i,2}) \sum_j (\mathbf{A}_{ij}p_{j,1}) \quad (5)$$

$$\delta_2 p_{i,2} = \beta_2(1 - p_{i,1} - p_{i,2}) \sum_j (\mathbf{A}_{ij}p_{j,2}) \quad (6)$$

which can be written in vector-form as:

$$\beta_1 \mathbf{S}\mathbf{A}\tilde{\mathbf{P}}_1 = \delta_1 \tilde{\mathbf{P}}_1 \quad (7)$$

$$\beta_2 \mathbf{S}\mathbf{A}\tilde{\mathbf{P}}_2 = \delta_2 \tilde{\mathbf{P}}_2 \quad (8)$$

where $\tilde{\mathbf{P}}_1 = [p_{1,1}, p_{2,1}, \ldots, p_{N,1}]^T$, $\tilde{\mathbf{P}}_2 = [p_{1,2}, p_{2,2}, \ldots, p_{N,2}]^T$ and $\mathbf{S} = \text{diag}(s_i) = \mathbf{I} - \text{diag}(\tilde{\mathbf{P}}_1 + \tilde{\mathbf{P}}_2)$.

In all of the following analysis, we assume we are operating at fixed point unless stated otherwise, i.e. Equations 5 and 6 or equivalently Equations 7 and 8 hold. Additionally, we assume that $\mathbf{A}$ is connected. First we have the following series of lemmas.

LEMMA 1. *$\forall\ i$ we have that $s_i \neq 0$.*

PROOF. If $s_i = 0$ for any i, then Equations 5 and 6 immediately imply that $p_{i,1} = p_{i,2} = 0$ which contradicts $s_i + p_{i,1} + p_{i,2} = 1$. □

LEMMA 2. *If $\exists\ i\ p_{i,1} = 0 \Rightarrow \forall\ i\ p_{i,1} = 0$. Similarly $\exists\ i\ p_{i,2} = 0 \Rightarrow \forall\ i\ p_{i,2} = 0$.*

PROOF. If $\exists\ i\ p_{i,1} = 0$, then from Equation 5 we have $\sum_j (\mathbf{A}_{ij}p_{j,1}) = 0$ (as $s_i \neq 0$ from Lemma 1). Clearly, $\mathbf{A}_{ij}$'s are positive only for those nodes j which are neighbors of node i, i.e. for $j \in NE(i)$ (and there is at least one such j as the graph is connected). For these j, as $p_{j,1}$ can not be negative (they are probabilities), they have to be zero so that the above is true. Now we can apply the same argument we applied for node i in turn for all the neighbors $j \in NE(i)$ and so on. Finally we get that $\forall\ i\ p_{i,1} = 0$ as the graph is connected. We can prove similarly for $p_{i,2}$. □

LEMMA 3. *The matrix* $\mathbf{SA}$ *is non-negative and irreducible.*

PROOF. $\mathbf{A}$ is symmetric and irreducible as it is connected. From Lemma 1 we have that $\mathbf{S}$ is a diagonal positive matrix. Clearly, it follows that $\mathbf{S} \cdot \mathbf{A}$ maybe asymmetric but it is a non-negative and irreducible matrix (intuitively, multiplying by $\mathbf{S}$ preserves the original edges in $\mathbf{A}$). □

LEMMA 4. *The matrix* $\mathbf{SA}$ *has a unique positive real number (say* λ*) as its largest eigenvalue (in magnitude). Further the algebraic multiplicity of* λ *is 1 and it has a positive eigenvector (say* $\tilde{\mathbf{v}}$*: then all components of* $\tilde{\mathbf{v}}$ *are positive).*

PROOF. As $\mathbf{SA}$ is non-negative and irreducible (Lemma 3), we can apply the Perron-Frobenius theorem [25]. The lemma follows directly then. □

LEMMA 5. *There are no positive eigenvectors of* $\mathbf{SA}$ *other than* $\tilde{\mathbf{v}}$ *(the Perron eigenvector of* $\mathbf{SA}$ *corresponding to the largest eigenvalue).*

PROOF. From Lemma 3, it follows that $(\mathbf{SA})^T$ is non-negative and irreducible as well. Moreover, note that the eigenvalues of any matrix $\mathbf{M}$ and $\mathbf{M}^T$ are the same. Hence, again applying the Perron-Forbenius theorem to $(\mathbf{SA})^T$, we have the largest eigenvalue as λ and the corresponding positive eigenvector as say $\tilde{\mathbf{u}}$. From the eigenvalue equation, we know that $\tilde{\mathbf{u}}^T\mathbf{SA} = \lambda\tilde{\mathbf{u}}^T$.

Now suppose we have another positive eigenvector, say $\tilde{\mathbf{w}}$, corresponding to eigenvalue t of $\mathbf{SA}$ (so, $\mathbf{SA}\tilde{\mathbf{w}} = t\tilde{\mathbf{w}}$). Then:

$$\lambda\tilde{\mathbf{u}}^T\tilde{\mathbf{w}} = \tilde{\mathbf{u}}^T\mathbf{SA}\tilde{\mathbf{w}} = t\tilde{\mathbf{u}}^T\tilde{\mathbf{w}}$$

Hence $(\lambda - t)\tilde{\mathbf{u}}^T\tilde{\mathbf{w}} = 0$. But $\tilde{\mathbf{u}}^T\tilde{\mathbf{w}} \neq 0$ as both $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{w}}$ are positive. Hence $t = \lambda$. But λ has multiplicity 1 (Lemma 4) and hence $\tilde{\mathbf{w}} = \tilde{\mathbf{v}}$. □

LEMMA 6. *At fixed point,* $\tilde{\mathbf{P}}_1 > 0$ *and* $\tilde{\mathbf{P}}_2 > 0$ *both can not hold unless* $\frac{\delta_1}{\beta_1} = \frac{\delta_2}{\beta_2}$.

PROOF. Together with Lemma 2, Equation 7 implies that either $\tilde{\mathbf{P}}_1 = 0$ or it is a positive eigenvector of $\mathbf{SA}$ with eigenvalue δ_1/β_1. Similarly from Equation 8 (and Lemma 2) we get that either $\tilde{\mathbf{P}}_2 = 0$ or it is a positive eigenvector of $\mathbf{SA}$ with eigenvalue δ_2/β_2. From Lemma 5, the only positive eigenvector of $\mathbf{SA}$ is the one corresponding to the largest eigenvalue. Hence both $\tilde{\mathbf{P}}_1 > 0$ and $\tilde{\mathbf{P}}_2 > 0$ can hold only if $\frac{\delta_1}{\beta_1} = \frac{\delta_2}{\beta_2}$. Otherwise at least one of them is zero. □

Assuming the virus strengths are not equal, Lemma 6 implies the following theorem:

THEOREM 2. *Assuming the virus strengths are not equal, the system has only the following possible fixed points:*

1. $\left\{\tilde{\mathbf{P}}_1 \to 0, \tilde{\mathbf{P}}_2 \to 0\right\}$ *(i.e. the viruses die-out)*
2. $\left\{\tilde{\mathbf{P}}_1 \to$ perron eigenvector of $\mathbf{SA}, \tilde{\mathbf{P}}_2 \to 0\right\}$ *(i.e. only virus 1 survives)*
3. $\left\{\tilde{\mathbf{P}}_2 \to$ perron eigenvector of $\mathbf{SA}, \tilde{\mathbf{P}}_1 \to 0\right\}$ *(i.e. only virus 2 survives)*

We can assert the next lemma immediately:

LEMMA 7. *The second and third fixed points in Theorem 2 require* $\sigma_1 > 1$ *and* $\sigma_2 > 1$ *respectively.*

PROOF. In the second fixed point, virus 2 dies-out and only virus 1 survives. Hence the system now is equivalent to a single virus operating on the whole graph under the standard flu-like SIS model. For this we already know that the virus should be above the 'epidemic threshold' if it has to survive (and not die-out exponentially quickly) [9, 30]. Hence $\lambda\beta_1/\delta_1 = \sigma_1 > 1$ is necessary for the second fixed point. Similarly we can prove the case for when virus 2 survives. □

Stability Conditions: We first compute the Jacobian at each of the fixed points.

LEMMA 8. *The Jacobians at the three fixed points can be written as below. (each Jacobian is a* $2N \times 2N$ *matrix, each sub-matrix block below is a matrix of size* $N \times N$*).*

1. $\mathbf{J}_1 = \begin{bmatrix} \beta_1\mathbf{A} - \delta_1\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \beta_2\mathbf{A} - \delta_2\mathbf{I} \end{bmatrix}$
2. $\mathbf{J}_2 = \begin{bmatrix} \beta_1\mathbf{SA} - \delta_1\mathbf{I} - \beta_1\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1) & -\beta_1\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1) \\ \mathbf{0} & \beta_2\mathbf{SA} - \delta_2\mathbf{I} \end{bmatrix}$
3. $\mathbf{J}_3 = \begin{bmatrix} \beta_1\mathbf{SA} - \delta_1\mathbf{I} & \mathbf{0} \\ -\beta_2\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_2) & \beta_2\mathbf{SA} - \delta_2\mathbf{I} - \beta_2\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_2) \end{bmatrix}$

Here, in $\mathbf{J}_2$*,* $\tilde{\mathbf{P}}_1$ *is the Perron eigenvector of* $\mathbf{SA}$ *with eigenvalue* δ_1/β_1 *(i.e. it satisfies Equation 7 and is non-zero). Similarly* $\tilde{\mathbf{P}}_2$ *in* $\mathbf{J}_3$.

PROOF. Can be computed using standard differentiation. Details omitted for brevity. □

Given the discussion before, we can analyze the corresponding conditions for the fixed point to be hyperbolic stable attractor.

LEMMA 9. *The conditions for the fixed points to be hyperbolic and stable attractor are:*

1. $\sigma_1 < 1$ *and* $\sigma_1 < 1$
2. $\sigma_1 > \sigma_2$
3. $\sigma_2 > \sigma_1$

PROOF. We prove the conditions for each fixed point separately below (we omit some details for brevity):

1. The eigenvalues of matrix $\mathbf{J}_1$ are simply the eigenvalues of the matrices $\mathbf{M}_1 = \beta_1\mathbf{A} - \delta_1\mathbf{I}$ and $\mathbf{M}_2 = \beta_2\mathbf{A} - \delta_2\mathbf{I}$. The real part of all the eigenvalues of these matrices will be negative if the real part of the largest eigenvalue is negative (as $\mathbf{M}_1$ and $\mathbf{M}_2$ are real and symmetric, all their eigenvalues are real). Hence the conditions for this are $\beta_1\lambda/\delta_1 < 1$ and $\beta_2\lambda/\delta_2 < 1$, where λ is the largest eigenvalue of $\mathbf{A}$.

2. We can see that the eigenvalues of the matrix $\mathbf{J}_2$ are either the eigenvalues of matrix $\mathbf{M}_1 = \beta_1\mathbf{SA} - \delta_1\mathbf{I} - \beta_1\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1)$ or the eigenvalues of the matrix $\mathbf{M}_2 = \beta_2\mathbf{SA} - \delta_2\mathbf{I}$.

   The eigenvalues of $\mathbf{M}_2$ are just the eigenvalues of $\beta_2\mathbf{SA}$ subtracted by δ_2. From Lemma 4 and Equation 7 we know that under this fixed point, the largest eigenvalue of $\mathbf{SA}$ is δ_1/β_1. This implies that $\mathbb{R}(\lambda_{\mathbf{SA}}) < \delta_1/\beta_1$ for any eigenvalue $\lambda_{\mathbf{SA}}$ of $\mathbf{SA}$[4]. Thus,

$$\mathbb{R}(\lambda_{\mathbf{M}_2}) = \beta_2\mathbb{R}(\lambda_{\mathbf{SA}}) - \delta_2 < \beta_2\delta_1/\beta_1 - \delta_2 < 0$$

[4] $\mathbb{R}(x)$ denotes the real part of x

where the last step follows if $\beta_2/\delta_2 < \beta_1/\delta_1$. Hence, if $\sigma_2 < \sigma_1$, the real part of all the eigenvalues of $\mathbf{M}_2$ are negative.

Consider the matrix $\mathbf{D} = \mathbf{M}_1 + \mathbf{M}_1^T = \beta_1(\mathbf{SA} + \mathbf{AS}) - 2\delta_1\mathbf{I} - 2\beta_1\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1)$. Matrix $\mathbf{D}$ is clearly real and symmetric and so has all real eigenvalues. Due to Lemma 4 we can apply the Perron-Frobenius theorem to $\mathbf{SA} + \mathbf{AS}$ as well and deduce that its largest eigenvalue $\lambda_1(\mathbf{SA} + \mathbf{AS})$ is positive. Further, from matrix theory [18, 15], we know that for any real non-negative matrix $\mathbf{C}$, $\lambda_1(\mathbf{C} + \mathbf{C}^T) \leq 2\lambda_1(\mathbf{C})$. Hence $0 < \lambda_1(\mathbf{SA} + \mathbf{AS}) \leq 2\lambda_1(\mathbf{SA}) = 2\delta_1/\beta_1$. Again we know from standard linear algebra [18], that $\lambda_1(\mathbf{X} + \mathbf{Y}) \leq \lambda_1(\mathbf{X}) + \lambda_1(\mathbf{Y})$ if $\mathbf{X}$ and $\mathbf{Y}$ are symmetric. Hence,

$$\begin{aligned}\lambda_1(\mathbf{D}) &\leq \beta_1\lambda_1(\mathbf{SA} + \mathbf{AS}) - 2\delta_1 - 2\beta_1\lambda_1(\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1)) \\ &\leq 2\beta_1\delta_1/\beta_1 - 2\delta_1 - 2\beta_1\lambda_1(\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1)) \\ &\leq -2\beta_1\lambda_1(\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1)) \\ &< 0\end{aligned}$$

as under this fixed point, $\text{diag}(\mathbf{A}\tilde{\mathbf{P}}_1)$ is a diagonal matrix with positive entries and hence has all positive eigenvalues. As $\mathbf{D}$ has all real eigenvalues, $\lambda_1(\mathbf{D}) < 0$ implies that it has all negative eigenvalues. The Lyapunov theorem [17] states that a matrix $\mathbf{C}$ is stable (has $\mathbb{R}(\lambda_\mathbf{C}) < 0$) if $\mathbf{C}^T + \mathbf{C}$ has all negative eigenvalues. Applying it to our case, we can see that matrix $\mathbf{D}$ having all negative eigenvalues implies that $\mathbf{M}_1$ is stable unconditionally under this fixed point.

Finally, as $\mathbf{M}_1$ and $\mathbf{M}_2$ both (and so $\mathbf{J}_2$ as well) have the real part of their eigenvalues negative under the condition $\sigma_2 < \sigma_1$, the fixed point is a hyperbolic stable attractor if $\sigma_2 < \sigma_1$.

3. Analogous to the case of the fixed point above.

Proved. □

Lemma 9 combined with Lemma 7 allows us to conclude the following:

THEOREM 3. *The corresponding conditions for each of the fixed points to (a) exist, and (b) have stability (i.e. be a hyperbolic and stable attractor) are:*

1. $\sigma_1 < 1$ *and* $\sigma_2 < 1$
   (i.e. both are below threshold)
2. $\sigma_1 > 1$ *and* $\sigma_1 > \sigma_2$
   (i.e. virus 1 is above threshold and virus 1 strength is greater than virus 2)
3. $\sigma_2 > 1$ *and* $\sigma_2 > \sigma_1$
   (i.e. virus 2 is above threshold and virus 2 strength is greater than virus 1)

Combining Theorem 2 and Theorem 3, we again have a result similar to the single virus epidemic threshold - that viruses die-out if they are below the individual epidemic threshold (i.e. if $\beta\lambda/\delta < 1$). Finally, they also imply our WTA result (Theorem 1).

5. EXPERIMENTS

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

5.1 Setup

We first briefly describe our experimental setup for the simulations as well as the case studies.

Simulations: WLOG, in our experiments, we assumed that the first virus is the stronger virus. We then considered the following three cases:

BELOW : $1 > \beta_1\lambda/\delta_1 = 0.6 > \beta_2\lambda/\delta_2 = 0.2$
(both viruses below the threshold)
MIXED : $\beta_1\lambda/\delta_1 = 6 > 1 > \beta_2\lambda/\delta_2 = 0.2$
(one above and one below the threshold)
ABOVE : $\beta_1\lambda/\delta_1 = 6 > \beta_2\lambda/\delta_2 = 4 > 1$
(both above the threshold)

We used the following real-world and synthetic network datasets for the simulations:

1. AS-OREGON: The Oregon AS router graph which is a network graph collected from the Oregon router views. It contains 15,420 links among 3,995 AS peers. More information can be found from `http://topology.eecs.umich.edu/data.html`.
2. PORTLAND: One of the biggest available physical contact graphs, representing a synthetic population of the city of Portland, Oregon, USA [27], and has been used in smallpox modeling studies [10]. It is a social-contact graph containing more than *31 mil.* links (interactions) among about *1.6 mil.* nodes (people).
3. Clique: A fully connected clique of 1000 nodes.
4. Barbell: Two cliques of 500 nodes joined together by weak edges of weight $\epsilon = 0.01$ (see Section 4.3 for a description).

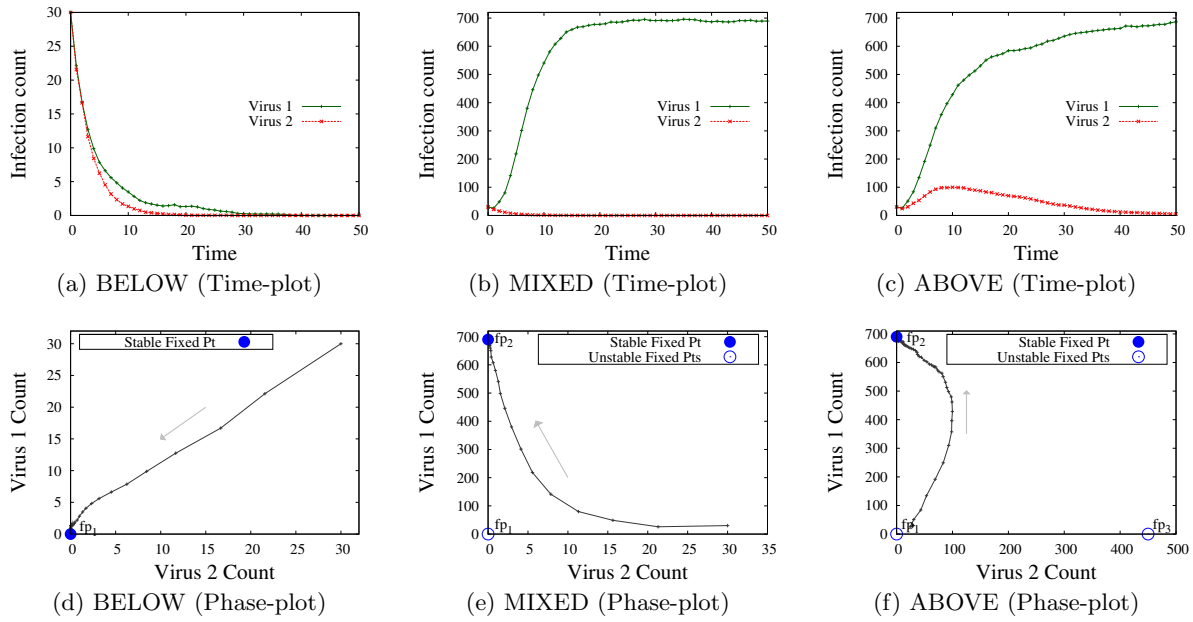
We implemented our competing viruses model SI_1I_2S as an event based discrete simulation in C++. We randomly infect 30 nodes for each of the viruses at the start of any simulation. All simulations were run over 1000 time steps and the plots show averaged results from 100 runs.

Case-studies: We collected historical data for 'web-search interest' for various competing products from the Google-Insights website[5] which aims to 'provide insights into broad search patterns'. This allows us to use the data as a proxy for product sales/adoption for each product. We used the following pairs of rival products:

1. Reddit and Digg: Two social news websites, where users post links to interesting memes/news articles.[6]
2. Facebook and Myspace: Two social network websites, where users add their friends and share posts, pictures etc.[7]
3. Blu-ray and HD-DVD: Two rival competing standards of high-density optical media.

The full mutual immunity model doesn't describe all the above situations perfectly, but it is a very good approximation. We understand that not all of the pairs are mutually exclusive in the strict sense e.g., people can go and put links on both Digg and Reddit, however, people are unlikely to be part of both communities as they have to choose a site while sharing content.

[5] `www.google.com/insights/search/`
[6] `www.reddit.com`, `www.digg.com`
[7] `www.facebook.com`, `www.myspace.com`



(a) BELOW (Time-plot) (b) MIXED (Time-plot) (c) ABOVE (Time-plot)

(d) BELOW (Phase-plot) (e) MIXED (Phase-plot) (f) ABOVE (Phase-plot)

Figure 4: (a-c) Number of infected vs time plots for simulations on the AS-OREGON network for different scenarios. (d-f) Corresponding Phase plots (scatter plot of number of infected nodes by virus 1 (y-axis) and number of infected nodes by virus 2). Stable fixed points are marked by bold circles, unstable by hollow circles. Clearly, the stronger virus wins (as long as it is above threshold) and the weaker dies-out completely as our result predicted.

5.2 Simulation Results

Figures 4 and 5 demonstrate our results. In short, the plots agree exactly with our result, as expected.

Figure 4 shows the *Time-plots* and *Phase-plots* for simulations on the AS-OREGON graph for our three scenarios as discussed before in the setup. The time-plots show the Number of nodes Infected vs Time for each of the viruses (red for the weaker virus, green for the stronger one). The phase plot is the scatter plot of number of infected nodes by the stronger virus on the y-axis and number of infected nodes by the weaker virus on the x-axis. Thus a phase plot shows the trajectory of the simulation in the 2-d plane. The stable points in each scenario are marked with solid circles.

In the BELOW case, we expect that both of them die-out. This is borne out by both the time and phase plots (Figures 4(a) and (d)). Point $FP_1(0,0)$ is the only stable fixed point in this case and hence the system converges to it very quickly (see the phase plot). On the other hand, when the stronger virus is above threshold (MIXED) we can see that it takes-over and the other virus dies-out (Figures 4(b) and (e)). In this case, point FP_2 is stable and attracting while FP_1 becomes unstable. As a result, we converge to the steady state where only the stronger survives. Finally, in case ABOVE, when each could have dominated in isolation, the stronger virus clearly wins and wipes-out the weaker virus (Figures 4(c) and (f)). Here, FP_2 is again stable while the other fixed points are unstable. Moreover, note that the stronger virus reaches the same steady-state as in MIXED. This agrees with our analysis as well (see Lemma 7): in both scenarios, the stronger virus will reach the same fixed-point as it would have if operating in isolation, without the presence of a competitor.

Similarly, Figure 5 shows the phase-plots for simulations on the other graph datasets - PORTLAND, Clique and Barbell. For lack of space, we just show the plots for case ABOVE. As before, the stronger virus wins and the weaker virus dies-out completely, no matter the network, in perfect agreement with our result (Theorem 1).

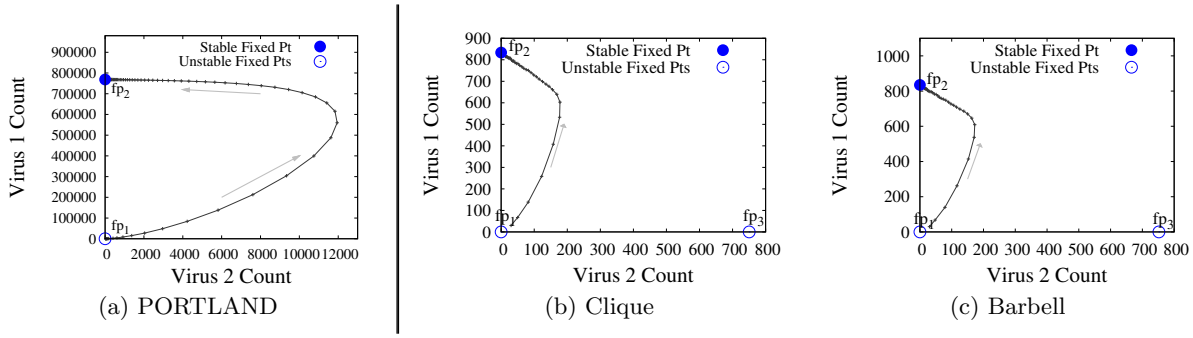
5.3 Case-Studies using Real Data

Figure 6 shows the historical data we collected from Google-Insights. In short, they provide corroboration to the WTA phenomenon in real-world as well.

Figures 6(a-c) show the web-search interest vs time for the three pairs of competitors we discussed before in the setup. Figures 6(d-f) show the corresponding phase-plots (the final data-point is marked by a diamond). Firstly, as it is real data, due to various reasons they do show significant deviations over the smooth steady states observed from our models (e.g., the spikes in Figure 6(c) denote Christmas shopping sales). Nevertheless, they broadly give positive evidence for the WTA result e.g., in (a-b) and (d-e), even though Digg and MySpace had a head-start and even dominate for a while, the stronger product (Reddit and Facebook) eventually takes-over. The phase plots also show the trajectories in effect similar to the ones found in our simulations. Clearly, in all the plots we can see that the eventual winner and dominant competitor (Reddit, Facebook, Blu-ray) almost completely wipes-out the weaker competitor, just as our result predicts.

6. DISCUSSION

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



(a) PORTLAND (b) Clique (c) Barbell

Figure 5: Phase plots (scatter plot of number of infected nodes by virus 2 and number of infected nodes by virus 1) plots for simulations on (a) the PORTLAND network, (b) a clique and (c) a barbell graph for scenario ABOVE. Again, stable fixed points are marked by bold circles and unstable fixed points by hollow circles (FP_3 not shown in (a) for sake of clarity of the trajectory). The weaker virus tries to dominate (note the bulge), but it dies-out completely and the stronger virus wins, as our result predicted.

following issues:

Question: Explain the counter-examples, of 'winner takes all'. If 'winner takes all', how come there are competing products where the weaker one still has a non-trivial market share, like 'Windows', 'Mac-os' (and 'Linux')?

Answer: Not 'level-fields'; or not enough time. There are indeed numerous cases where two (or more) competing products or ideas, co-exist. For example, in the OS 'wars', MS-windows has a large market share, with mac-OS having a smaller, but near-constant market share. There are several settings that could cause such deviations.

- One is that we are violating our assumption of 'fair-play', e.g., some nodes (like enthusiastic Apple^{TM} fans) exhibit much lower infection probability β, or even zero for one of the viruses. Thus by catering to just that niche where it is much stronger, the competitor can survive.
- A second cause is weak connectivity, like a bar-bell graph with a narrow bridge, and not enough time to reach steady-state.
- A third cause is viruses of near-equal strength. We omit the simulation results here, but similar-strength viruses take too long to reach 'WTA'. This is analogous to the case of two near-equal-strength tennis players, that need several games, and several tie-breakers, before a winner emerges.

Question: Has this WTA phenomenon appeared in other settings?

Answer: Yes, with simulation results. In epidemiology studies, WTA is referred to as '*competitive exclusion*' e.g. see [7, 8, 1, 2]. However, they typically did simulations, or they only studied homogenous or very structured topologies like cliques.

Question: How about other propagation models (SIR etc)? Will 'WTA', then?

Answer: We conjecture that the answer is 'yes'. The full analysis for SIR (= life-long immunity, like mumps) SIRS (= long, but not permanent, immunity) and more, are the focus of our ongoing research. We conjecture that similar results may hold, too, extrapolating from the results of (Prakash et al. [30]): that work showed that, for a single virus, the epidemic threshold has the same form, for almost any virus propagation model.

Question: Will 'WTA' hold, under *partial* mutual immunity?

Answer: Future work - no conjectures. In this work, we assume full mutual exclusion, that is a given node will have at most one of the two viruses/products (iPhone/Android), at any given point in time, but not both. There are marketing, and biological settings that a person may have both products/viruses. Will 'WTA' hold, then? This seems like a difficult question, and left for future work. We suspect that the answer will not be a simple 'yes' or 'no'.

7. CONCLUSIONS

In summary, we tackled the setting of two competing products (or viruses or ideas etc.) spreading over a network and studied the problem of what happens in the end (i.e. in the steady state). In addition to problem formulation and getting ecological concepts to web-like phenomena, the main contributions of our work are as follows:

1. *WTA Result and Proof*: We provided a theoretical analysis of the propagation model for *arbitrary* graph topology, proving that the 'winner-takes-all' i.e. the stronger virus dominates and wipes-out the weaker virus (if it is above threshold). See Theorem 1.
2. *Experiments and Case-studies*: We also demonstrated our result using extensive simulations on real and synthetic networks showing that they match exactly with our predictions. Moreover, using case-studies of historical data of competing products (Blu-ray/HD-DVD, Facebook/MySpace, Reddit/Digg), we provided positive evidence of WTA in real-life.

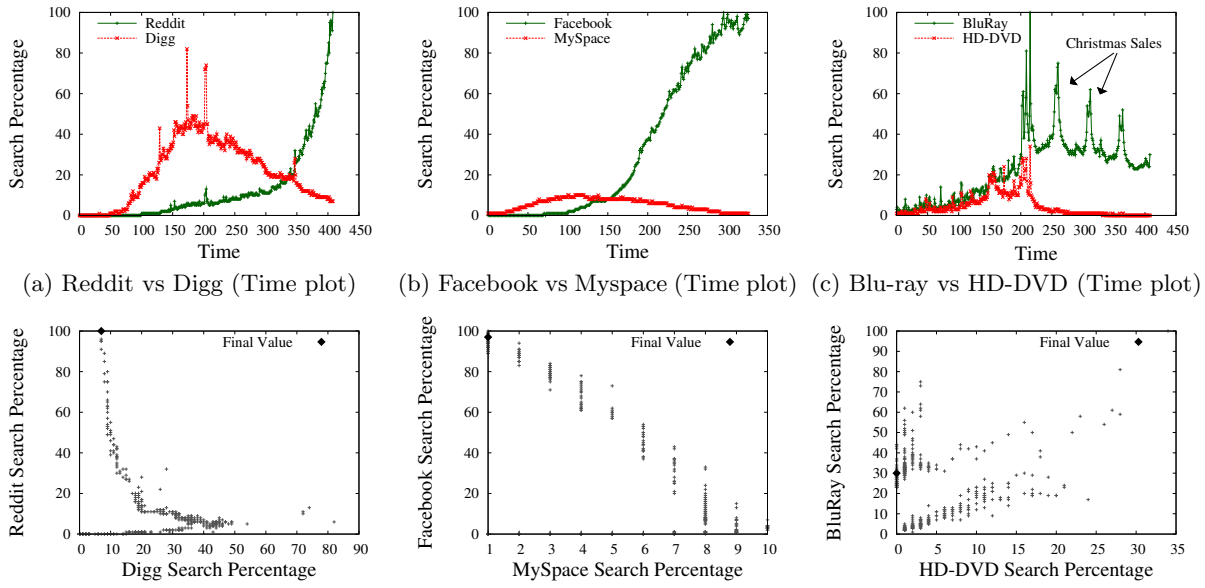
8. REFERENCES

[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.



(a) Reddit vs Digg (Time plot) (b) Facebook vs Myspace (Time plot) (c) Blu-ray vs HD-DVD (Time plot)

(d) Reddit vs Digg (Phase plot) (e) Facebook vs Myspace (Phase plot) (f) Blu-ray vs HD-DVD (Phase plot)

Figure 6: (a-c) Real web-search interest vs time plots for pair of competitors (see Section 5.1 for more details) (d-f) Corresponding Phase plots. As expected from our WTA result, note that the stronger rival dominates and weaker product almost dies-out.

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

[8] 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.

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

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

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

[12] 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.

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

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

[15] C. He and G. A. Watson. An algorithm for computing the numerical radius. *IMA Journal of Numerical Analysis*, 17, 1997.

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

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

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

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

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

[21] 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.

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

[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 *SDM*, 2007.

[25] C. R. McCuler. The many proofs and applications of perron's theorem. *SIAM Review*, 42, 2000.

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

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

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

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