UNIVERSITY OF CALIFORNIA
RIVERSIDE

Spreading Processes on Networks
Theory and Applications

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

by

Nicholas Charles Valler

September 2012

Dissertation Committee:

Professor Michalis Faloutsos, Chairperson
Professor Iulian Neamtiu
Professor Eamonn Keogh

The Dissertation of Nicholas Charles Valler is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First and foremost, I wish to thank my wife and son. Denalee, I dedicate this dissertation to you as a testament to your constant and unwavering patience, support and love. Denalee, you are my reason. And to my son, Oliver. You are my greatest inspiration. Oliver, you are my joy.

To my advisor and committee chairperson, Dr. Michalis Faloutsos. I've describe myself of having 'academic ADHD' and love chasing shiny new problems, without necessarily finishing the last. You always found the shiniest problems–you always kept it interesting–and for that, I offer my deepest thanks. Thank you for the many hours of spirited discussion, your guidance, and sharing your knowledge.

I also wish to thank the members of my committee, Dr.'s Iulian Neamtiu and Eamonn Keogh. I greatly appreciate your time and excellent feedback during this process. Your poignant questions, particularly during my dissertation defense, challenged me to improve.

Finally, thank you to my friends and colleagues. In particular, I would like to thank Mr. Xuetao Wei, you are an excellent researcher with great drive.

To my wife, Denalee.

ABSTRACT OF THE DISSERTATION

Spreading Processes on Networks
Theory and Applications

by

Nicholas Charles Valler

Doctor of Philosophy, Graduate Program in Computer Science
University of California, Riverside, September 2012
Professor Michalis Faloutsos, Chairperson

The interactions between people, technology and modern communication paradigms form large and complex human–machine networks. Complex network theory attempts to address the global and local behavior of such network structures. Of particular interest within the area of network theory is understanding the dynamic behavior of spreading processes on complex networks. In this work, we examine a variety of models covering the intersection of spreading processes and complex network theory, and although we study a large range of problem formulations, we find that–surprisingly–a single parameter effectively summarizes the topology.

We begin by examining the effect that topology has on spreading processes in dynamic networks. Dynamic networks are becoming more common due to our increased reliance on and the functionality of mobile devices, smartphones, etc. Specifically, we ask, given discrete information spread through a proximity-based communication

channel across dynamic network of mobile end-users, what criteria is required such that the information will ultimately die-out; that is, can we determine the "tipping point" between information survival and die-out? We show analytically that yes, such a threshold exists, yet it is computationally infeasible to calculate. To avoid such computationally intensive methods, we go on to provide two approximation methods for determining the tipping point.

Next, we analyze the effect of topology on the propagation of competing information. Using a novel graph structure we refer to as a composite network, we model the intertwined propagation of competing information across a variety of underlying network layers. Through a combination of analytical and empirical methods, we show how the topology affects the competing information, and ultimately, using topology, we predict the winner of competition.

Building on the success of the previous analyses, we formulate a model describing the spread of non-categorical information. Unlike our previous models, the information in this system is represented by a continuous value. We determine the phase transitions of the overall system, relate them to the tipping points in our previous models, and show both analytically and empirically how the structure of the network affects those phase transitions.

Ultimately, for each of these models, a single topological parameter, the largest eigenvalue of the adjacency matrix $\lambda_{\mathbf{A},1}$, is all that is necessary to characterize the effect of topology on the spreading process.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation explores the spread of information across networks. Spreading processes on networks are not a new phenomenon. Perhaps even without realizing it, a variety of spreading processes operate on modern networks everyday, ranging from the malicious propagation of computer viruses, to the consensus algorithms used by Bitcoin; even to the heart of the Internet, where the border gateway protocol propagates network reachability information across autonomous systems. As our use of networked infrastructure continues to evolve, how we use the network has important implications on performance, communication overhead and usability in general.

Spreading processes exist beyond the realm on networked computer systems. Everyday, an individual interacts with hundreds, perhaps thousands, of distinct network structures. As one commutes to work, they interact with transportation networks. When one turns on a light to read, they interact with power grids. When one washes

one's hands, they are interacting with public water networks. And, to complicate the matter further, modern networks need not represent physical infrastructure. Facebook, MySpace and other social networks have no physical counterpart in the real works, yet these online communities exist as networked structures and participate in the spread of information.

Modeling of spreading processes is an old discipline, most commonly associated with epidemic theory and epidemiology. In fact, borrowing from epidemic theory, two computer scientists, Kephart and White [28], proposed the first model of computer virus propagation in 1993. Their assumptions mirrored the assumptions of epidemiologists, namely a concept of homogeneity among agents within the system. That is, every agent within the population is affected by a spreading virus with the same probability. The first heterogeneous model applied to technological networks was introduced by Pastor-Satorras and Vespignani[41]. Their model accounted for non-homogeneous connections among agents in the system, naturally represented by a network structure. Introducing a model that accounted for heterogeneous network connectivity among system agents raised another important area of research, namely, the effect of topology on the spread of information.

Everyday, more and more applications are developed to rely on networked communication across a variety of hosts and servers, ultimately forming a large, decentralized and self-organizing communication paradigm using protocols based on spreading processes to manage communication between participants. Over the past

15 years, complex network theory has attempted to shed light on the interactions of these large, decentralized network structures. Results from complex network theory underly a variety of modern day technology. For instance, Google's PageRank algorithm essentially relies on a back propagation mechanism to accurately determine node importance in a large, complex network, and has changed the way we query the World Wide Web. It is even claimed that network theory, when applied to particular social networks, aided in the capture of Saddam Hussein.

This dissertation builds on the rich history of complex network theory and spreading processes and examines the problem of information spreading processes on complex networks at a theoretical, empirical, and application level. In the following chapters, we will introduce three specific problem areas related to information spreading processes on networks. Each chapter, while related, is self-encapsulated and contains their own brief introductions and background information. The main theme of this work can be summarized as follows:

How does the network structure, or topology, affect the spread of information?

Each chapter that follows presents a self-contained study of a specific spreading process and network formulation. I begin in Chapter 2 and introduce our first area of study, namely epidemic spreading processes on dynamic network structures. We model this spreading process as propagating across short-range, point-to-point communication channels, akin to Bluetooth-equipped mobile devices. This model is

informed by relatively recent attempts of malware to exploit the lax security in many mobile devices. Furthermore, by exploiting proximity-based propagation schemes, many malware will circumvent infrastructure-based detection schemes. The fundamental contributions of this chapter are: (a) we present a framework for analyzing epidemic spreading processes on mobile ad hoc networks, (b) in using our framework, we derive the epidemic threshold for any mobility model operating under the Susceptible–Infected–Susceptible (SIS) compartmental model, and (c) we show that node velocity of the mobility model does node affect the epidemic threshold. Ultimately, we draw an interesting conclusion: the connectivity matrix eigenvalue ($\lambda_{\mathbf{A}}$) determines the epidemic threshold, that is, the point below which the information spreading process is guaranteed to 'die-out'.

Next, in Chapter 3, we pose a new model for the spread of competing information across networks.We motivate this model with the following basic example. Consider a rumor propagating through your social network of friends in Facebook. Assume a false rumor is propagating through an alternate network, such as Twitter. Furthermore, assume that the rumors are so diametrically opposed that upon hearing one, a person will exclude the possibility of the other, i.e., the rumors are mutually exclusive. Given this scenario, we claim these two pieces of information are competing to capture the most hearts and minds, but are there any topological factors that will indicate which will ultimately prevail? Again, we show that topology is strongly connected to the ultimate winner and the winner tends to be the virus with the "strongest" system

matrix eigenvalue ($\lambda_{\mathbf{S}}$), which we show to be closely related to the connectivity matrix.

In Chapter 4 and drawing on both previous results, we propose a discrete-time, continuous value non-linear dynamical information propagation model that captures the essence of all previous models, yet allows one to evaluate the propagation of partial information. This model differs from all previous works, where information was considered to be discrete, binary, or categorical.

Finally, in Chapter 5, I provide a summary and review of the results presented herein.

Chapter 2

Epidemic Spreading Processes on Dynamic Networks

2.1 Overview

Short-range, point-to-point communications for mobile users enjoy increasing popularity, particularly with the rise in Bluetooth-equipped mobile devices. Unfortunately, virus writers have begun exploiting lax security in many mobile devices and subsequently developed malware exploiting proximity-based propagation mechanisms (e.g. Cabir or CommWarrior). So, if given an ad-hoc network of such mobile users, will a proximity-spreading virus survive or die out; that is, can one determine the "tipping point" between survival and die out? What effect does the average user velocity have on such spread? I answer the initial questions and more. The contributions of

this Chapter are:

1. I present a framework for analyzing epidemic spreading processes on mobile ad hoc networks;

2. Using this framework, I derive the epidemic threshold for any mobility model under the SIS model; and

3. I how that the node velocity in mobility models does not affect the epidemic threshold.

Additionally, I introduce a "periodic mobility model" and provide evaluation via the model framework. Finally, I validate the theoretical predictions using a combination of real, synthetic and simulated mobility data, showing ultimately, the predictions accurately estimate the epidemic threshold of such systems.

2.2 Introduction

Mobile phones have experienced an increased prevalence and functionality in recent years, yet, due to the hardware and software limitations and constraints, such devices present a unique opportunity for malicious software (i.e., "malware"). As of 2010, approximately 45.5 million U.S. citizens owned smartphones or comparable mobile devices. So, roughly 15% of the population are walking around, everyday, with a computer more powerful than that of the original Apollo space mission and filled

with a variety of personal and private information. In short, millions of end-users rely on these devices for essential everyday functions.

To compete in the market, mobile device makers have extended their devices with the latest in personal communication technology, such as device–to–device (a.k.a. point–to–point) Bluetooth. Including such point–to–point communication technology allows end-users to bypass the smartphone's service provider's infrastructure and directly pass information to other nearby end-users. There were an estimated 920 million Bluetooth-equipped devices shipped worldwide in 2008, making Bluetooth the most common point–to–point communication protocol in today's smartphones. To capitalize on the prevalence and personal importance of smartphones, at least two worms, *Cabir* and *CommWarrior*, spread using Bluetooth.

In this chapter, I am motivated by the trends described above and examine the effect of network structure on the spread on point–to–point, or proximity, spreading malware.

2.2.1 Problem Statement and Definition

How does the network structure, as described by a system of mobile agents, affect the propagation of a malicious virus[1]? A key question is to identify the tipping point, known as the system's *epidemic threshold*, or *take-off point*, below which a virus is

[1]The focus of this chapter is on the propagation of information through agents in close proximity. I note however that smartphones have a diverse suite of communication methods, such as email, mms, or direct access to the web, yet, as mobility has little affect, if any, on these methods, I elect to exclude them from the following analysis.

guaranteed to "die out." For the epidemic models, I focus on one of the most popular models, the flu-like one susceptible-infected-susceptible - SIS (see Section 2.6 where I handle other models). There, agents maintain no immunity, and become susceptible, immediately after they heal. The key contributions in the chapter are as follows:

1. *Framework and Formula*: We present a framework for estimating the epidemic threshold on *any, arbitrary* mobile ad hoc network model. The idea is to derive a sequence of adjacency matrices, and then compute the first eigenvalue of the so-called *system matrix* (see Theorem 1). There, I show that the epidemic threshold depends only on this first eigenvalue, and nothing else.

2. *Closed Formulas*: I show how to use the proposed framework to derive simple, approximate (but accurate) formulas for several, popular special cases (Random Walk model, Levy flight model).

3. *Insensitivity to Velocity*: The results show that the epidemic threshold does *not* depend on the node velocity ($v > 0$). Experiments confirm the accuracy of the approximations, as well as the 'insensitivity' observation.

Jumping ahead, Figure 2.1 showcases the accuracy of the results (Lemma 4) of the epidemic threshold for the SIS (=flu-like) model on the so-called 'Levy Flight' mobility model. See section 2.5 for more details - but the point is that the prediction for the take-off point (= *epidemic threshold*, indicated with a black arrow) is exactly where all curves take off.

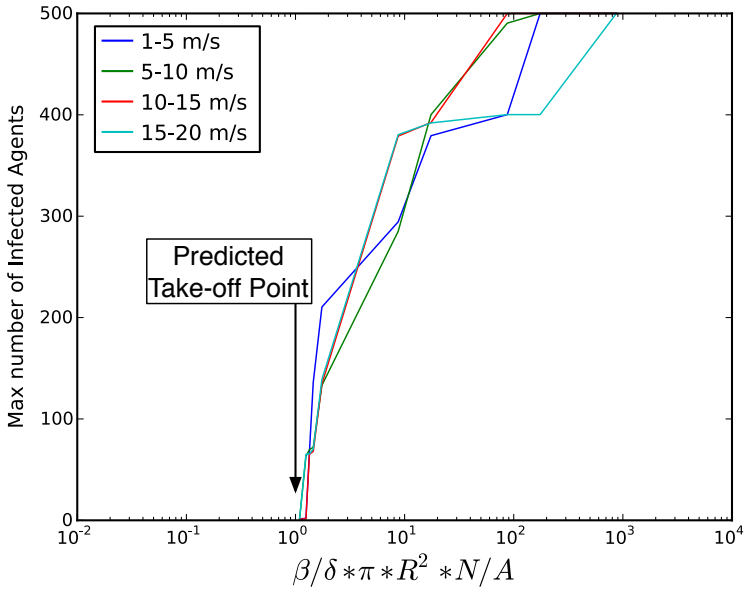


Figure 2.1: Accuracy of results (Lemma 4), for the 'Levy-Flight' model. Take-off plot, plotting the max number of infections vs. strength of the virus. Notice that the predicted take-off point (black arrow) agrees with the simulations, for several node velocities.

I have two additional contributions: through extensive simulation experiments, I show that similar insensitivity results hold for other popular mobility models like Levy flight, Random Waypoint etc.; and, moreover, I introduce the *periodic mobility* model, which is very popular in biological virus epidemiology [2, 9], and show how to use the framework to estimate its epidemic threshold.

The rest of this chapter has the typical organization: background (Section 2.3, proposed framework and theorems (Section 2.4), experiments (Section 2.5), additional observations (Section 2.6), and finally, related work and conclusions. (Sections 2.7 and 2.8).

2.3 Preliminaries

In this section, we present a general background on proximity-based epidemic spreading models and formulate our problem statement.

2.3.1 Epidemic Model: SIS (flu-like)

The SIS epidemic model resembles a flu-like virus, where nodes have no immunity. Healthy ('S' = susceptible) nodes become sick ('I' = infected) stochastically from their infected neighbors with a probability β. Alternatively, a sick node becomes healthy (and open to re-infection), with a probability δ. These two parameters are also referred to as the *birth* rate (β) and *death* rate (δ) of the virus.

The tipping point τ, or *epidemic threshold*, of an SIS epidemic model is the condition under which an infection will die out exponentially quickly irrespective of initial infection, as opposed to spreading out, causing and epidemic (technically, a pandemic). For a survey on SIS and numerous other epidemic models, see Hethcote [24], or [14, 19].

2.3.2 Problem Formulation

Using the background discussed above, we now formulate our problem statement. See Table 2.1 for definitions of various symbols. In this paper, we consider an epidemic on a mobile network, which provides an underlying contact structure for the virus to

Table 2.1: Terminology Specific to Mobility and Epidemic Spreading

Symbol	Definition and Description
General Terms:	
$\mathbf{A}, \mathbf{B}, \ldots$	matrices (bold upper case)
$\mathbf{A}(i,j)$	element at the i^{th} row and j^{th} column of $\mathbf{A}$
$\mathbf{A}(i,:)$	i^{th} row of matrix $\mathbf{A}$
$\mathbf{A}(:,j)$	j^{th} column of matrix $\mathbf{A}$
$\mathbf{I}$	standard $n \times n$ identity matrix
$\mathbf{a}, \mathbf{b}, \ldots$	column vectors
$\mathcal{I}, \mathcal{J}, \ldots$	sets (calligraphic)
$\lambda_{\mathbf{B}}$	first eigenvalue (in absolute value) of a matrix $\mathbf{B}$
Mobility Terms:	
M	mobility model
$P_{i,t}$	position of node i at time t
N	number of nodes
A	simulation area
ρ	node density (N/A)
ΔT	Time step
T	number of different alternating behaviors
$\mathbf{A}_1, \ldots, \mathbf{A}_T$	T corresponding size $(n \times n)$ symmetric alternating adjacency matrices
Epidemic Terms:	
β	virus transmission probability in the SIS model
δ	virus death probability in the SIS model
τ	epidemic threshold
Acronyms and Terms:	
USS	Uniform Steady-State Approximation
EAAM	Eigenvalue of Average Adjacency Matrix Approximation
Take-off Plot	Max number of infected agents vs. Epidemic Threshold τ

use as it propagates. By doing so, at any point in time, the system is non-homogenous, as nodes may only transmit the virus to its neighbors.

Given:

1. Mobile ad hoc network mobility models, described below.

2. The SIS model parameters, i.e., the virus birth and death probabilities β and δ.

Find:

> The epidemic threshold τ or *tipping point* for the system such that for $\tau < 1$ an infection will die out quickly, irrespective of initial conditions.

Our problem naturally leads to other issues like the effect of node velocity in models on the threshold, giving approximations in specific cases etc. We elaborate on them in the upcoming sections.

2.4 Framework

In this section, we detail our framework for analyzing mobility models and then move on to specific approximations and questions arising out of the framework. We will present extensive simulations demonstrating the results later in Section 2.5. Also please see Section 2.5.2 for a description of the mobility models.

Note that node-to-node contacts at a particular time can be represented by an *adjacency matrix* $\mathbf{A}$. We next provide a general theorem expressing the epidemic threshold for mobility models.

2.4.1 Epidemic Thresholds on Mobility Models

Theorem 1 (Mobility model threshold)**.** *If a mobility model can be represented as a sequence of connectivity graphs* $L = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_T\}$*, one adjacency matrix* $\mathbf{A}_t$ *for*

each time step $t \in \{1..T\}$, then the epidemic threshold is:

$$\tau = \lambda_S \tag{2.1}$$

where $\lambda_{\mathbf{S}}$ is the first eigenvalue of matrix $\mathbf{S}$ and $\mathbf{S} = \prod_i \mathbf{S}_i$ and $\forall_i \in \{1..T\}$ $\mathbf{S}_i = (1-\delta)\mathbf{I} + \beta\mathbf{A}_i$ ($\mathbf{I}$ is the standard $N \times N$ identity matrix).

Proof. If the mobility model can be represented as a sequence of graphs, then the epidemic threshold depends on the first eigenvalue of the system matrix [46]. Hence, $\tau = \lambda_{\prod_i((1-\delta)\mathbf{I}+\beta\mathbf{A}_i)}$. □ □

We can now give a simpler closed-form approximation for the threshold in Equation 2.1 in the following lemma:

Lemma 2 (EAAM Approximation for Threshold)**.** *Under the same conditions as in Theorem 1, the following is an approximation for the epidemic threshold:*

$$\tau \approx \frac{\beta}{\delta} \times \lambda_{\mathbf{A}_{avg}} \tag{2.2}$$

where $\mathbf{A}_{avg} = \sum_i \mathbf{A}_i / T$ is the average adjacency matrix.

Proof. Note that,

$$\begin{aligned} \mathbf{S} &= \prod_i ((1-\delta)\mathbf{I} + \beta \mathbf{A}_i) \\ &= (1-\delta)^T \mathbf{I} + \beta \sum_i \mathbf{A}_i + O(\beta^2) + O(\beta * \delta) + O(\delta^2) \\ &\approx (1 - T\delta)\mathbf{I} + T\beta \mathbf{A}_{avg} \end{aligned} \tag{2.3}$$

where we neglected second or lower order terms involving β and δ. Hence, we find that $\mathbf{B} = (1-T\delta)\mathbf{I} + T\beta\mathbf{A}_{avg}$ is a first order approximation for the $\mathbf{S} = \prod_i \mathbf{S}_i$ matrix. Hence from Theorem 1 we want $\lambda_B < 1$ which implies Equation 2.2. □ □

We will refer to the above approximation as the '**E**igenvalue of the **A**verage **A**djacency **M**atrix' (EAAM) approximation.

2.4.2 Specific Approximations

Lemma 3 (Random-Walk Threshold)**.** *In the random-walk mobility model and under the SIS model, the following is an approximate epidemic threshold:*

$$\tau \approx \beta/\delta \times \pi R^2 \times N/A \tag{2.4}$$

where R is the radius of influence of each node.

Proof. Under the random-walk model, at every point of time, each node is at a random (x, y) position, uniformly distributed on the field of interest. Each node has a radius

of possible connections (like the BlueTooth radius) R. Consequently each node has $d = \pi R^2 \times N/A$ neighbors on average (ignoring boundary effects). The connectivity graph at each time step is roughly a random graph with average degree d. Hence it has first eigenvalue $\lambda_1 = d$ on average. Hence this is approximately equivalent to having a static graph under the SIS model where the epidemic threshold [14] is $\tau = \beta/\delta \times \lambda_1$. We now obtain the lemma after obvious substitutions. □ □

In fact, we can go further and generalize this to *any* mobility model where the geographic steady state distribution is uniform.

Lemma 4 (Uniformly-Distributed Steady State (USS) Threshold)**.** *For any mobility model where the geographic distribution of nodes at the steady state is uniform over the area of interest and under the SIS model, the following is an approximate epidemic threshold:*

$$\tau \approx \beta/\delta \times \pi R^2 \times N/A \tag{2.5}$$

where R is the radius of influence of each node.

Proof. The proof for Lemma 3 goes through even here precisely because of the geographically uniformly distributed nature of the steady state. Each node has roughly the same number of connections and hence the adjacency graph is approximately a homogenous graph with constant first eigenvalue. The result follows as before. □ □

We will refer to the above approximation as the '**U**niform **S**teady-**S**tate' (USS)

approximation. Mobility models like Levy Flight and Random-Walk are examples of models with a geographically uniform distribution of the nodes at the steady state. Lemma 4 allows us to quickly estimate the threshold for these and many other models.

2.4.3 Insensitivity to Node Velocity

As there is no factor depending on the node velocity in Lemma 4, we conclude the following surprising implication:

Corollary 5 (Node velocity and threshold)**.** *The node velocity ($v > 0$) does not affect the epidemic threshold in mobility models where the steady state has a geographically uniform steady state distribution like Random Walk, Levy Flight etc.*

We conjecture that the velocity does not affect the threshold even for non-geographically uniformly distributed steady state mobility models like Random Waypoint. We provide empirical results supporting this claim later in Section 2.5.

Conjecture 6 (Effect of velocity)**.** *The node velocity ($v > 0$) does not affect the epidemic threshold in the Random Waypoint mobility model.*

Does Velocity have an impact at all? The above discussion raises the point whether the node velocity has any effect at all on the dynamics of the epidemic spreading. We expect that the velocity of motion does have an effect, when we are above threshold. Furthermore, simulations resulted in a non-obvious observation. The velocity had an

impact on the steady-state number of infected agents in the system. We elaborate more on these issues in Section 2.6.

2.5 Simulation Methodology and Results

2.5.1 Experimental Setup

To facilitate the simulation, I wrote a custom Python2.6 simulation program using the NumPy/SciPy python libraries. All simulations were conducted on a 4 core Intel(R) Xeon(R) CPU operating at 2.53 GHz and 72 GB of memory running CentOS-5.5 (Linux kernel 2.6). I varied the number of agents (nodes) N between 250 and 1500 within a simulation field of area $A = 40,000m^2$ (200m by 200m). Thus, node density ρ, commonly defined a N/A, was between 0.125 and .125 nodes per m^2. All nodes had a transmission range of 5.0 meters. I did not account for signal attenuation, reflection nor other wireless phenomena. Prior to the beginning of the simulation, nodes were distributed on the simulation field in a uniform fashion. Simulations were generally run for a period of 100s with time intervals of $\Delta T = 0.1$ seconds.

I examined three mobility models common to mobile ad hoc networking: Random Walk, Levy Flight and Random Waypoint. In the following sections, I provide detail on each model as well as simulation results. The position $P_{i,t}$ of each node in the system at time t is a function of mobility model and previous position and time step ΔT, such that $P_{i,t+1} = M(P_{i,t}, \Delta T)$, where M is the mobility model.

The purpose of the simulations was to determine the role of the mobility model in the spread of malware in a point-to-point contact network loosely describing Blue-Tooth communication technology.

2.5.2 Mobility Models

Random Walk. The Random Walk (RW) mobility model (also referred to as *Brownian Motion*) was originally formulated to describe the seemingly random motion of particles. Numerous variations exists, here we describe the implementation.

Each node i in the system is parameterized by *speed (V_i)* and *angle (θ_i)*. Both V_i and θ_i are drawn uniformly from systemwide predefined ranges, $[v_{min}, v_{max}]$ and $[0, 2\pi)$, respectively. Clearly, such a system is memoryless. The model employed varies from the simple RW model by introducing a *flight time* for each node, T_i. Flight time is drawn uniformly from a range $[\tau_{min}, \tau_{max}]$. The spatial distribution of the RW mobility model is uniform over the simulation field. According to my framework, I predict the RW mobility model will follow Lemma 3.

Levy Flight. Levy Flight mobility models have recently attracted attention due to their statistical similarities with human mobility [47]. At the beginning of each flight, each node selects an angle uniformly from within $(0, 2 * \pi]$, a flight time drawn from some distribution, a flight length and a pause time. Flight length and pause time are drawn from Levy distributions $p(t) \propto |t|^{-(1+\alpha)}$ and $\psi(t) \propto t^{-(1+\beta)}$, where time $t > 0$, respectively. When $\alpha = 2$ and $\beta = 2$ the result is a special case of the Levy

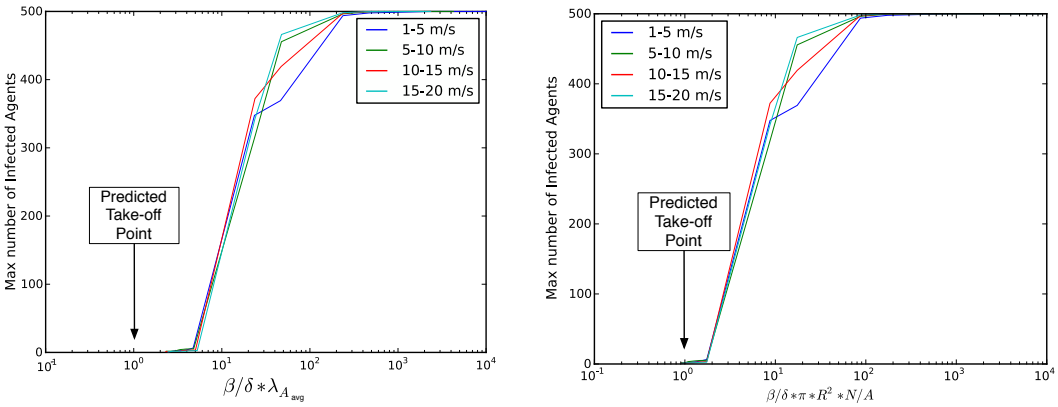
distribution resulting in a Gaussian distribution. As with the Random Walk, the spatial distribution of the Levy Flight mobility model is uniform over the simulation field.

Random Waypoint. The Random Waypoint (RWP) mobility model is often cited as the *de facto* mobility model in ad hoc networks. As originally proposed by Johnson et. al [27], the RWP mobility model each node i is described by three parameters: *current location* (P_i), *speed* (V_i), *waypoint* (W_i), and *pause time* (ρ). In general, the RWP mobility model operates as follows: Initially, a node is stationary. After a pause time ρ, the node selects a waypoint uniformly from the simulation field, then, travels along the shortest path to its waypoint P at a velocity V_i drawn uniformly from $(v_{min}, v_{max}]$. Upon arrival at their waypoint, each node will pause for a time $t = \rho$. After the pause period is done, each node will repeat the process. The spatial distribution of the RWP mobility model is bell-shaped [10].

2.5.3 Summary of Results

Accuracy of Approximations, Random Walk.

Next, I present a series of simulation studies of the Random Walk mobility model in Figure 2.2. In these studies, birth (β) and death (α) parameters were varied of the SIS infection model, and the resulting plots are referred to as "*take-off* plots," which show the maximum number of infected agents seen in the simulation against the approximated epidemic threshold. For each plot, the take-off point is labeled



(a) EAAM (b) USS

Figure 2.2: Accuracy of Framework Approximations, Random Walk Take-off plots for EAAM and USS. Notice that the predictions (black arrows) are accurate.

according to the specific threshold approximation.

Figure 2.2(A), epidemic threshold was approximated using the first eigenvalue of the average adjacency matrix, $\mathbf{A}_{avg}$ (the EAAM threshold approximation of Lemma 2). I indicate the predicted threshold value at $\tau = 1$. As expected, no epidemic was present at values of $\tau < 1$. At values of $\tau > 1$, one observes explosive growth in the max number of infected agents.

In Figure 2.2(B), I plot the the USS approximation of the epidemic threshold (Lemma 4), i.e. $\beta/\delta \times \pi R^2 \times N/A$. As in the EAAM threshold, USS behaves as expected (i.e. no epidemic below $\tau = 1$). In fact, in each of these figures, we note that no infected agents (aside from patient zero) are present for epidemic threshold

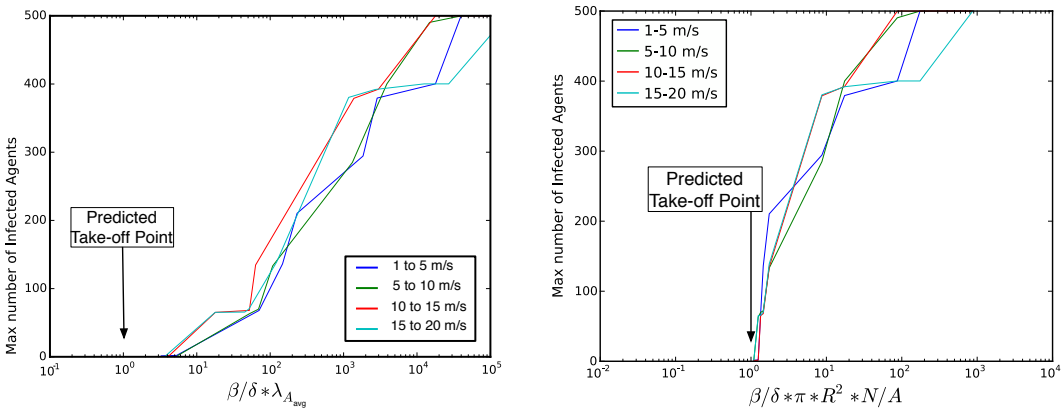
values below 1. Compared to EAAM, this threshold value takes off at values closer to $\tau = 1$.

Accuracy of Approximations, Levy Flight. As with RW, the spatial distribution of nodes following Levy Flight mobility model is uniformly distributed on the simulation field. Thus, we expect Levy flight to perform similar to RW. Figure 2.3 presents a take-off plots for Levy Flight simulations. Again, we note that no infected agents exist below either threshold approximations. Furthermore, we find that USS performs better than EAAM.

Accuracy of Approximations, Random Waypoint. The next series of simulations were conducted using the popular Random Waypoint mobility model. We also selected this model specifically because it does not result in a uniform spatial distribution of nodes, therefore does not fall under USS. The results of these simulations are presented in Figure 2.4. Surprisingly, both threshold approximations perform well against the RWP mobility model. As the earlier mobility models exemplified, USS performs better than EAAM.

Insensitivity to Velocity. In order to illustrate Corollary 5 and validate Conjecture 6, we performed a series of simulation in which we varied the node velocity, the results of which are illustrated in Figures 2.2, 2.3, 2.4.

As expected, for both the RW and Levy Flight mobility models, the take-off points were not greatly affected by increasing the nodes velocity. Furthermore, Figure 2.4 shows that the take-off point for the RWP mobility model was not affected by velocity,



(a) EAAM (b) USS

Figure 2.3: Accuracy of Framework Approximations, Levy Flight Take-off plots for EAAM and USS. Again, predictions (black arrows) prove accurate.

affirming Conjecture 6.

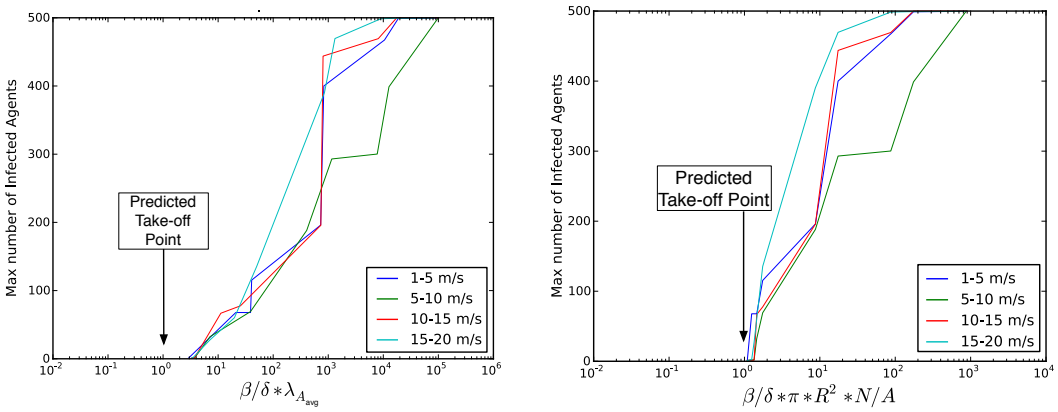
2.6 Discussion

We elaborate here on the effect of node velocity on the dynamics of epidemic spreading. We also introduce the periodic mobility model and present an analysis of it via our framework. In addition, we touch upon other epidemic models as well.

2.6.1 More on impact of node velocity

As discussed previously in Section 2.4, node velocity does not seem to effect the threshold in many models. We now ask whether the velocity affects the epidemic at all?

For an "above threshold" system, two more parameters are of interest: (a) *steady-*

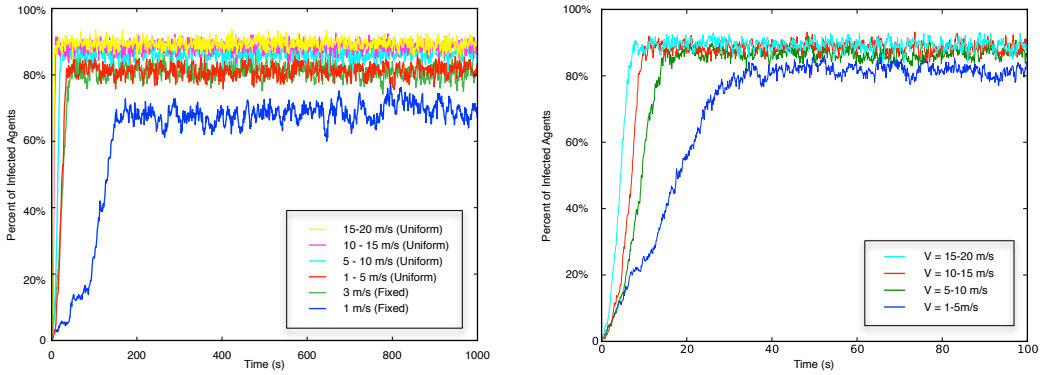


(a) EAAM (b) USS

Figure 2.4: Accuracy of Framework Approximations. Random Waypoint Take-off plots for EAAM and USS.

state maximum, and (b) *warm-up period*. The steady-state maximum is the maximum number of infected agents in the system till steady state, whereas the warm-up period is the time necessary to reach steady state. We expect that the velocity of motion does have an effect, when we are above threshold. Clearly, speed will effect the speed of propagation of the virus and thus the warm-up period. Higher velocity means better mixing of agents, and thus faster convergence to the steady state. This observation is also demonstrated through simulations.

Figure 2.5 (best viewed in color) shows the number of infected agents per unit time (in seconds) for both the Random Walk and Random Waypoint mobility models. The velocity varied between a fixed 1 m/s fixed and an uniform selected $15 - 20$ m/s, as indicated in the legends. We performed a longer simulation in order for the systems to settle in a steady state.



(A) RW, 1000s (B) RWP, 100s

Figure 2.5: Number of Infected Agents vs. Time on the Random Walk and Random Waypoint. Node velocities were varied as indicated in the legends. Steady-State number of infected agents increases with node velocity, while warm-up period shrinks (best viewed in color).

Less intuitive is that velocity appears to affect the number of infected agents at steady-state. For example, in Figure 2.5(A), the line corresponding to 1 m/s appears to reach a steady-state of approximately $65-70\%$ infected agents, whereas, at $15-20$ m/s, the steady-state is roughly 90%. The steady-state for velocities between these two extremes lay in-between. We suspect the degree of mixing, influenced by node velocity, is the root cause of the above observation.

2.6.2 The Periodic Mobility Model

As we indicated in Section 2.4, our framework predicts the epidemic threshold of mobility models that can be represented as a series of adjacency matrices. The *Periodic Mobility Model* is a special case of such a series, where a set of k adjacency matrices $\{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_k\}$ are repeated periodically.

This is a typical model used in biological virus studies [9] to model general movements of a population. As an example, let $\mathbf{A}_1$ be an adjacency matrix of people during the day (say, at the office). Let $\mathbf{A}_2$ be an adjacency matrix representing contacts/interactions during the evening (say, at home). So the series formed by repeating $\{\mathbf{A}_1, \mathbf{A}_2\}$ represents the daily, repeated interactions of our population. The periodic model offers a realistic, yet general model of mobility, capturing general patterns rather than specific movements of the system.

Lemma 7 (Periodic Model Threshold)**.** *Under the periodic mobility model with k alternating behaviors repeating periodically, the epidemic threshold is given by $\tau = \lambda_S$ where $\mathbf{S} = \prod_{i=1}^{k} \mathbf{S}_i$ and, as before, $\mathbf{S}_i = (1-\delta)\mathbf{I} + \beta\mathbf{A}_i$.*

Proof. Omitted for brevity, similar to Theorem 1. □

As an example, Figure 2.6 shows the take-off plot of a periodic mobility models where $N = 500$ nodes, $k = 2$ adjacency matrices. As predicted by Lemma 7, the max number of infected agents over the simulation period takes off at $\lambda_S = 1$.

2.6.3 Other Epidemic Models

Given recent results on epidemic thresholds on static networks [46], we believe that our results will carry through for many other epidemic models as well e.g. SIR (mumps-like), SIRS, SEIR, MSEIR etc [23] which capture differences between the way various diseases spread.

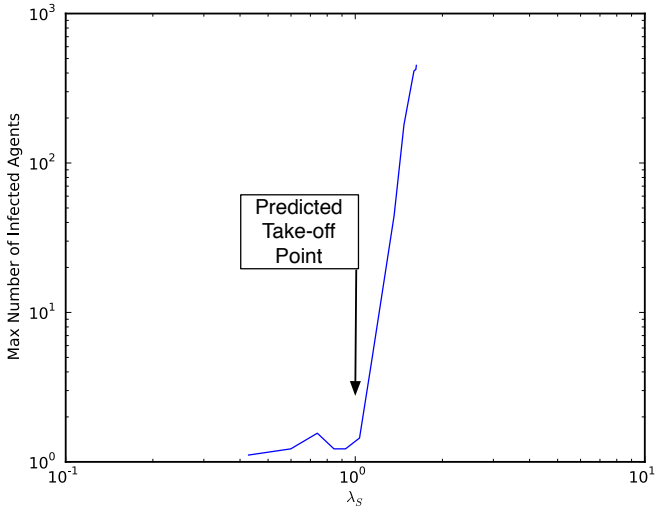


Figure 2.6: Take-off Plot for System Matrix Eigenvalue of a Periodic Mobility Model. A_1 and A_2 had $N = 500$ nodes spread uniformly across a 200×200 simulation field. Notice the accurate our prediction (black arrow).

Conjecture 8 (Other Epidemic models)**.** *Our results for all the mobility models discussed in this chapter hold for all the epidemic models covered in [46] as well.*

2.7 Related Work

Here, I review the related work. I begin by discussing general epidemic models, followed by epidemic spread on static networks. Next, I review epidemic spread in dynamic networks. Finally, I comment on works related to the mobility models described above. It is worth pointing out that while most of existing studies about epidemic spread on mobile networks focus on (1) some particular types of network structures, and/or (2) one specific mobility model; my framework is very general and it applies to *arbitrary* network structure, and all the three popular mobility model.

2.7.1 General Epidemiology

Bailey provides the canonical text on epidemic modeling [6]. A more recent survey is provided by Hethcote in [24].

Kephart and White [30, 31] were among the first to propose epidemiology-based models to analyze the spread of computer viruses. The model they suggest provides a good approximation of virus propagation in networks where contact among individuals is essentially homogeneous. Recent discoveries suggest real networks (including social networks [17], router and AS networks [18], and Gnutella overlay graphs [48]) follow a power-law structure instead, prompting a re-evaluation of the homogeneity assumption common in the works above.

2.7.2 Static Networks

Observation suggests that real networks are not homogeneous, rather, overwhelming evidence suggests real networks follow a power law structure instead. By introducing an underlying structure for a disease to spread, such as a static network, removes the original homogeneous assumption pioneered by those reference above. Newman [39] studied the epidemic thresholds for multiple competing viruses on special, *random* graphs. Pastor-Satorras and Vespignani studied viral propagation for such power-law networks [38, 43, 44]. They developed an analytic model for the Barabási-Albert (BA) power-law topology [7]. However, their derivation depends

on some assumptions which does not hold for many real networks [33, 18]. Pastor-Satorras et al. [44] also proposed an epidemic threshold condition, but this uses the "mean-field" approach, where all graphs with a given degree distribution are considered equal. There is no particular reason why all such graphs should behave similarly in terms of viral propagation. Chakrabarti et. al. [14] observe that epidemic threshold of an arbitrary graph can be captured in a single parameter, the first eigenvalue of the adjacency matrix $\lambda_{1,A}$. Their observation was rigorously confirmed in [15] and independently by [56]. I again leverage the above observations to formulate my solution in Section 2.4

2.7.3 Mobile Networks

Prompted by the emergence of mobile devices, such as Bluetooth-equipped smartphones, researchers introduced mobility to epidemic spread. Mickens et. al. were among the first to examine device-to-device spreading of malicious software in mobile ad hoc networks [37][36]. In their work, they present a *queue*-based technique for the RWP model to overcome the limitations of the earlier homogeneous models of Kephart and White. In a similar work, Yan et. al. extend the observations of Mickens et. al. by examining additional mobility models and their effect on epidemic spreading of a SIS virus [59]. Their work is unique in that it models virus propagation, in detail, a Cabir-like bluetooth worm, including the Bluetooth stack and unique worm properties.

2.7.4 Mobility Models

The mobility models used in Section 2.5 are fairly common, with significant literature devoted to the subject. For an overview on mobility models, I refer readers to the following surveys [13]. The RWP mobility model has been extensively used, despite well know flaws. For a discussion of the merits of RWP, refer to [60]. The Levy mobility model was first described in [51], yet has been used extensively to model human and animal movements [47]

2.8 Summary

To conclude, recent malware in the wild, using device-to-device virus propagation schemes, prompted our study of the epidemic threshold in mobile ad hoc networks. Our contributions in this chapter are:

1. *Framework:* We present a framework for the determining the epidemic threshold (for the SIS model) on *any* mobility model which can be converted into a series of adjacency matrices and give a formula for it (Theorem 1).

2. *Closed Formulas:* We also give a closed-form approximation for the SIS epidemic threshold on general mobility models (Lemma 2).

3. *Insensitivity to Velocity:* We analyze the impact of velocity in popular mobility models like Random walk, Levy Flight, Random waypoint etc. and find that it

unexpectedly does not affect the threshold (Lemmas 3, 4 and Conjecture 6).

In addition, we introduced the "periodic mobility model," popular in other fields like epidemiology [2, 9], to the networking community and solved it using our framework (Lemma 7). Finally we presented extensive simulations to demonstrate our analysis and results.

Future work may concentrate on providing theoretical analysis on the effect of velocities on the steady state behavior of an epidemic on various mobility models.

Chapter 3

Competing Epidemic Spreading Processes

In this chapter, I examine the effect that a network's topology has upon competing information. For example, consider a misinformation campaign through social networks. In such a campaign, false rumors and lies will propagate using the social connectivity between agents within the system. Alternatively, the truth may propagate using an alternate, trusted, social network. Ultimately, the question I address in this chapter *which will win* and *what factors will determine the winner*. To that end, I model competing information using a spreading processes derived from the popular compartmental model and provide network structure using a novel composite network model.

Using this model, I evaluate the intertwined propagation of two distinct pieces

of information given the following scenarios: (1) who will win?; (2) how one or both pieces of information be suppressed efficiently?; and (3) what if competing information was allowed to cross-contaminate alternate composite network layers?. Ultimately, I find that a single topological parameter, the first eigenvalue of the system matrix $\lambda_{\mathbf{S}}$ greatly affects the outcome of each of these scenarios.

3.1 Introduction

Models of spreading processes on networks, like those described in this chapter, cross many scientific disciplines, ranging from computer science, to business marketing, to statistical physics [34, 29, 21, 22, 52, 54, 49]. The datum spread in these models range from social information, rumors, computer viruses, fashion trends, market penetration and product adoption. In fact, the most popular use of spreading models is for disease and epidemic propagation [3, 23], so without loss of generality, I will use the term *virus* to describe the information propagating through a system.

Few of the previous studies examined the intertwined spread of two competing pieces of information. In this chapter, I focus on this problem and consider two competing datum that propagate over distinct, yet partially connected networks.

Figure 3.1 illustrates the two aspects of the competing information spread model. The first component is the *composite network* (Figure 3.1(a)). A composite network is defined as $C = (N, E_1, E_2)$, with a single set of nodes N with two distinct edge sets

Symbol	Definition	Symbol	Definition
V_1, V_2	Virus #1, #2	$\mathbf{A_1}$, $\mathbf{A_2}$	Adjacency matrices
δ_1, δ_2	Meme persistence of V_1, V_2	β_1, β_2	Meme strength of V_1, M_2
S	Susceptible state	I_1, I_2	Infected state for V_1, M_2
$\mathbf{S_1}$, $\mathbf{S_2}$	System matrix for $\mathbf{A_1}$, $\mathbf{A_2}$, where $\mathbf{S} = (1-\delta)\mathbf{I} + \beta\mathbf{A}$	λ_1, λ_2	Largest eigenvalue of $\mathbf{S_1}$, $\mathbf{S_2}$ in absolute value.

Table 3.1: Terminology Specific to Viral Competition on Networks

E_1 and E_2. The second component is the SI_1I_2S compartmental model, described in Figure 3.1(b). As with other compartmental models, SI_1I_2S assumes that an agent may only exist in one state, i.e., the viruses are *mutually exclusive*. So, as the two viruses in our system spread, an agent may only have one virus.

Using this model, I evaluate the intertwined propagation of two distinct pieces of information given the following scenarios: (1) who will win?; (2) how one or both pieces of information be suppressed efficiently?; and (3) what if competing information was allowed to cross-contaminate alternate composite network layers?. Ultimately, I find that a single topological parameter, the first eigenvalue of the system matrix $\lambda_{\mathbf{S}}$ greatly affects the outcome of each of these scenarios.

The problem described above is relevant to a variety of applications and disciplines. For instance, the online social networks provide one domain that competing information is highly relevant. Additionally, anti-virus software companies have experimented with propagating anti-virus definitions using peer-to-peer or social networks. This produces the obvious competition of virus/anti-virus products across distinct networks.

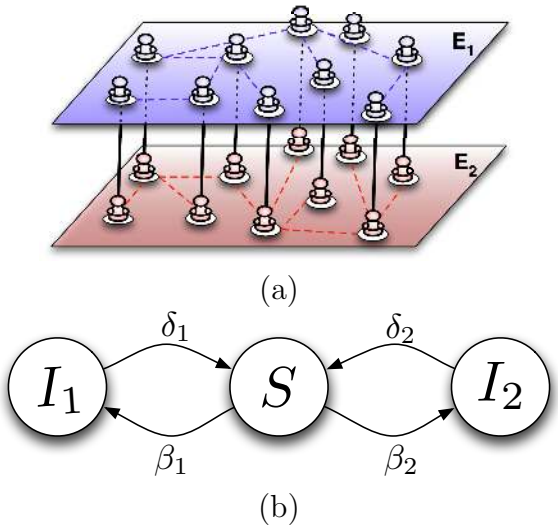


Figure 3.1: (a) Example Composite Network topology: a single set of nodes N with two distinct edge sets E_1 and E_2. (b) The SI_1I_2S State Transition Diagram, where S represents the susceptible state and $I_{\{1,2\}}$ indicate the infected state for viruses V_1 and V_2. The transitions between states are indicated by the directed edges labeled $\beta_{\{1,2\}}$ and $\delta_{\{1,2\}}$.

Contributions

1. Formulate a rigorous model for competing viruses on composite networks using a modified susceptible-infected-susceptible (SIS) propagation mechanism.

2. Demonstrate the phase transition behavior of this system, showing that the first eigenvalue of an appropriately-constructed system matrix for each virus is a critical metric that determines system behavior.

3. Numerous simulation results, modeling the composite network layers after Erdős-Rényi, Barabási-Albert and various synthetic graph models, plus a real-world composite network of 235 mobile phone/text users.

4. Design and evaluate several viral suppression, focusing on two strategies: (a) Unilateral Suppression and (b) Concurrent Suppression.

5. Evaluate the effect of cross-contamination between layers of our composite network, where one virus may, with a certain cross-contamination probability, to "jump" composite network layers and use propagate across both layers simultaneously.

3.2 The SI_1I_2S Propagation Model

Our model is described by two components: (1) a composite network and (2) a propagation mechanism. We define a composite network as $C = (N, E_1, E_2)$, a single set of nodes N with two distinct edge sets E_1 and E_2. Each layer of the composite network corresponds to a single adjacency matrix $\mathbf{A_1}, \mathbf{A_2}$.

The propagation mechanism is based on the popular "flu-like" SIS (Susceptible-Infected-Susceptible) model [23]. We name our model SI_1I_2S (Susceptible – Infected$_1$ – Infected$_2$ – Susceptible). Each node in the composite network is in one of three states: Susceptible (healthy), I_1 (infected by V_1), or I_2 (infected by V_2). The state transitions are shown in Fig. 3.1(b). Note that this is one of the several virus propagation models that one could consider, others being SI, SIR, etc. We believe that our model is a reasonable starting point, and we leave the analysis of other models as future work.

Virus persistence: δ. If a node is in state I_1 (or I_2), it recovers on its own with probability 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.

Note that we assume that while a node is infected by one virus, it cannot be infected by the other. We do not anticipate that allowing a virus to preempt each other (i.e., infect and subsume an infected node) would change the results from a qualitative point of view: this would be equivalent to having a node skip the recovery state and go straight to a new infection. As we will see later on, our metrics and methods consider the δ values explicitly.

Virus strength: β. A healthy node gets infected by infected neighbors, and the virus strength is captured by β_1 and β_2. Specifically, this parameter is the probability that an infected neighbor would pass the infection to a healthy neighbor in the absence of any other interaction. We refer to this potential infection-in-isolation as an **attack**. In the presence of multiple infected neighbors, we need to decide which infection succeeds (infects a susceptible node i) as follows. Let C_1 be the number of attacks (each happening with probability β_1 independently) by node i's neighbors which are in state I_1 (infected by V_1); similarly, let C_2 be the number of neighbors infected by V_2. Then, we have three possible scenarios for a node in the Susceptible state:

- node i remains in the Susceptible state if $C_1 = 0$ and $C_2 = 0$.

- node i gets infected with V_1 with probability $\frac{C_1}{C_1+C_2}$.
- node i gets infected with V_2 with probability $\frac{C_2}{C_1+C_2}$.

It is easy to see that this is a natural generalization of the SIS model to a competing-virus scenario. Moreover, note the competition between the viruses: each virus has to compete with each other for healthy victims.

3.3 The Epidemic Threshold

We want to determine the epidemic threshold (Problem 1) analytically. First, we approximate the infection process by a discrete-time Non-Linear Dynamical System (NLDS) whose general form is $\mathbf{p}_{t+1} = g(\mathbf{p}_t)$. The NLDS gives the evolution of the system with time, as we explain below.

First, we see the probability that node i is infected by neighbor node j with virus V_1 at time t is $\beta_1 P_j^1(t-1)$. This is what we referred to as *attack* earlier or infection by a neighbor in absence of other influences. Then, we have the probability $\zeta_i^1(t)$ that node i does not receive the infection of V_1 from its neighbors (assuming the neighbors are independent) as:

$$\zeta_i^1(t) = \Pi_{j \in i's\, neighbors}(1 - \beta_1 P_j^1(t-1)) \tag{3.1}$$

Thus, the probability that node i receives the infection of V_1 at time t from its

neighbors is:

$$1 - \zeta_i^1(t) = 1 - \Pi_{j \in i's\, neighbors}(1 - \beta_1 P_j^1(t-1)) \tag{3.2}$$

Using the same reasoning, we can obtain the probability of that node i receives the infection of V_2 from its neighbors at time t is:

$$1 - \zeta_i^2(t) = 1 - \Pi_{j \in i's\, neighbors}(1 - \beta_2 P_j^2(t-1)) \tag{3.3}$$

Now, the probability that node i is infected by V_1 from its neighbors at time t is the probability that node i receives the infection of V_1 and does not receive infection of V_2 from its neighbors at time t. Here we assume that the β and δ values are all extremely small (or, equivalently, the time between state transitions is extremely small). This ensures that in any given time step, the probability of having two or more events is vanishingly small. Thus, we get:

$$T_i^1(t) = (1 - \zeta_i^1(t)) \cdot \zeta_i^2(t) \tag{3.4}$$

With the same reasoning, the probability that the node is infected by M_2 at time t is:

$$T_i^2(t) = (1 - \zeta_i^2(t)) \cdot \zeta_i^1(t) \tag{3.5}$$

Hence the probability that node i is in state I_1 is:

$$P_i^1(t) = (1 - \delta_1) \cdot P_i^1(t-1) + T_i^1(t) \cdot S_i(t-1) \tag{3.6}$$

and the probability that it is in state I_2 is:

$$P_i^2(t) = (1 - \delta_2) \cdot P_i^2(t-1) + T_i^2(t) \cdot S_i(t-1) \tag{3.7}$$

and the probability that it is in state S (Susceptible) is:

$$S_i(t) = (1 - T_i^1(t) - T_i^2(t))S_i(t-1) +$$

$$\delta_1 P_i^1(t-1) + \delta_2 P_i^2(t-1)$$

As mentioned before, for $V1$ we define the vector $\vec{P}^1(t) = (P_1^1(t), P_2^1(t), ..., P_N^1(t))'$ where $P_i^1(t)$ is the probability that node i is infected by virus V_1 at time t. Similarly, for V_2, we have $\vec{P}^2(t) = (P_1^2(t), P_2^2(t), ..., P_N^2(t))'$. Let $\vec{V}(t) = (\vec{P}^1(t), \vec{P}^2(t))$ be the concatenation of two vectors. Using the NLDS formulation, we can now describe the whole infection process evolution as $\vec{V}(t) = f(\vec{V}(t-1))$, with:

$$f_i(\vec{V}(t-1)) = \begin{cases} (1 - \delta_1)P_i^1(t-1) + & \\ \quad T_i^1(t)S_i(t-1) & \text{if } i \leq N \\ (1 - \delta_2)P_i^2(t-1) + & \\ \quad T_i^2(t)S_i(t-1) & \text{if } i > N \end{cases} \tag{3.8}$$

Substituting $T_i^1(t)$, $T_i^1(t)$ and $S_i(t-1)$ into equation 3.8, we find that $f_i(\vec{V}(t-1))$

is equal to the following:

$$= \begin{cases} (1-\delta_1)P_i^1(t-1) + (1-\zeta_i^1(t))\zeta_i^2(t) & \\ \quad (1-P_i^1(t-1) - P_i^2(t-1)) & \text{if } i \leq N \\ (1-\delta_2)P_i^2(t-1) + (1-\zeta_i^2(t))\zeta_i^1(t) & \\ \quad (1-P_i^1(t-1) - P_i^2(t-1)) & \text{if } i > N \end{cases}$$

We make use of the following theorem about the asymptotic stability of an NLDS at a fixed point:

Theorem 9 (Hirsch and Smale, 1974 [26])**.** *The system given by* $\mathbf{p}_{t+1} = g(\mathbf{p}_t)$ *is asymptotically stable at an equilibrium point* $\mathbf{p}^*$*, if the eigenvalues of the Jacobian* $J = \bigtriangledown g(\mathbf{p}^*)$ *are less than 1 in absolute value, where*

$$J_{k,l} = [\bigtriangledown g(\mathbf{p}^*)]_{k,l} = \frac{\partial p_{k,t+1}}{\partial p_{l,t}}|_{\mathbf{p}_t = \mathbf{p}^*}$$

The fixed point we are interested in for analyzing the threshold is the point where no node is infected (all nodes are healthy), i.e., $\vec{V}^* = \vec{0}$. Using this, we have:

Theorem 10. *The system is asymptotically stable at* $\vec{V}^* = \vec{0}$ *if the first eigenvalue of the system matrices for both viruses as defined in Table 3.1, are less than 1, i.e.,* $\lambda_1 < 1$ *and* $\lambda_2 < 1$*, where* λ_1 *is the largest eigenvalue of matrix* $S_1 = (1-\delta_1)I + \beta_1 A_1$

(and similarly for λ_2).

Proof. Recall that we are interested in the stability of the fixed point $\vec{V}^* = \vec{0}$. Let the Jacobian at this point be $\nabla(f)$ (a $2N$ x $2N$ matrix). Then

$$[\nabla(f)]_{ij} = \frac{\partial f_i(\vec{V}(t-1))}{\partial \vec{V}_j(t-1)}$$

We can write it into a block matrix composed of the system matrices:

$$\nabla(f) = \left[\begin{array}{c|c} S_1 & S_3 \\ \hline S_4 & S_2 \end{array}\right]$$

In order to find the first eigenvalue of $\nabla(f)|_{\vec{V_f}}$, we define $\vec{X}$ as 2N elements vector:

$$\vec{X} = \left[\begin{array}{c} \vec{X}_1 \\ \hline \vec{X}_2 \end{array}\right]$$

where $\vec{X}_1$ and $\vec{X}_2$ have N elements respectively. We then have:

$$\bigtriangledown(f)|_{\vec{v_f}}\vec{X} = \left[\begin{array}{c|c} S_1 & S_3 \\ \hline S_4 & S_2 \end{array}\right] \cdot \left[\begin{array}{c} \vec{X}_1 \\ \\ \vec{X}_2 \end{array}\right] = \lambda_{\bigtriangledown(f)}|_{\vec{v_f}} \left[\begin{array}{c} \vec{X}_1 \\ \\ \vec{X}_2 \end{array}\right]$$

Doing the matrix multiplications, we get:

$$S_1\vec{X}_1 + S_3\vec{X}_2 = \lambda_{\bigtriangledown(f)}|_{\vec{v_f}}\vec{X}_1$$

$$S_4\vec{X_1} + S_2\vec{X_2} = \lambda_{\bigtriangledown(f)}|_{\vec{v_f}}\vec{X_2}$$

with $S_1 = (1-\delta_1)I + \beta_1 A_1$, $S_2 = (1-\delta_2)I + \beta_2 A_2$ and $S_3 = S_4 = 0$ (where I is the N x N identity matrix), as we show in Table 3.1 and as discussed below.

Hence, the Jacobian $\nabla(f)$ is a block diagonal matrix and its eigenvalues are the same as the eigenvalues of S_1 and S_2. So the largest eigenvalue of $\nabla(f)$ can be either λ_1 or λ_2. □

Discussion: adjacency versus system matrix. We can understand how the eigenvalue of the system matrix is the key parameter, if we consider the definition of the system matrix. At the same time, it is useful to stress the difference between the adjacency matrix, **A**, and the system matrix, **S**. One such matrix exists for each virus but here we drop the virus subscript.

The key point that we make is the following: the system matrix for a virus and thus the related eigenvalue are a function of the topology *and* the properties of the virus. The eigenvalues of the adjacency matrix $\lambda_{\mathbf{A}}$ are related to the eigenvalues of the system matrix $\lambda_{\mathbf{S}}$. Recall that the system matrix is defined as $\mathbf{S} = (1-\delta)\mathbf{I} + \beta\mathbf{A}$, where **A** is the adjacency matrix. Therefore, if we consider an eigenvector for **A**, that would also be an eigenvector for **S** and the following will hold for the eigenvalues:

$$\lambda_{\mathbf{S}} = 1 - \delta + \beta\ \lambda_{\mathbf{A}} \tag{3.9}$$

In conclusion, the system eigenvalue $\lambda_{\mathbf{S}}$ increases with the virus strength, β and the adjacency eigenvalue. Naturally, $\lambda_{\mathbf{S}}$ decreases as the virus persistence δ increases.

3.4 Simulation Study

We use a discrete-time simulation of our system that simulates the stochastic behavior of competing virus on several different synthetic and real composite networks.

3.4.1 Small-scale data sets ($N < 1,000$)

Real-world enterprise composite network (ENT). We have obtained a composite network dataset that represents the phone call and SMS text message communications within an urban branch of a large Chinese corporation [57]. Each node is an employee ($|N| = 235$), the edges in E_1 correspond to SMS messages exchanged between employees, and edges in E_2 correspond to phone calls made between employees. The data was captured over the course of six months. Among all communicating pairs of users, 31% communicate via calls alone, 28% via SMS alone, and 41% via both calls and SMS.

Synthetic composite networks. We have created two synthetic graphs with 1,000 nodes: the first one is an Erdős-Rényi graph, whereas the second one is a scale-free graph; we use the Barabási-Albert model [8]. We have experimented with several

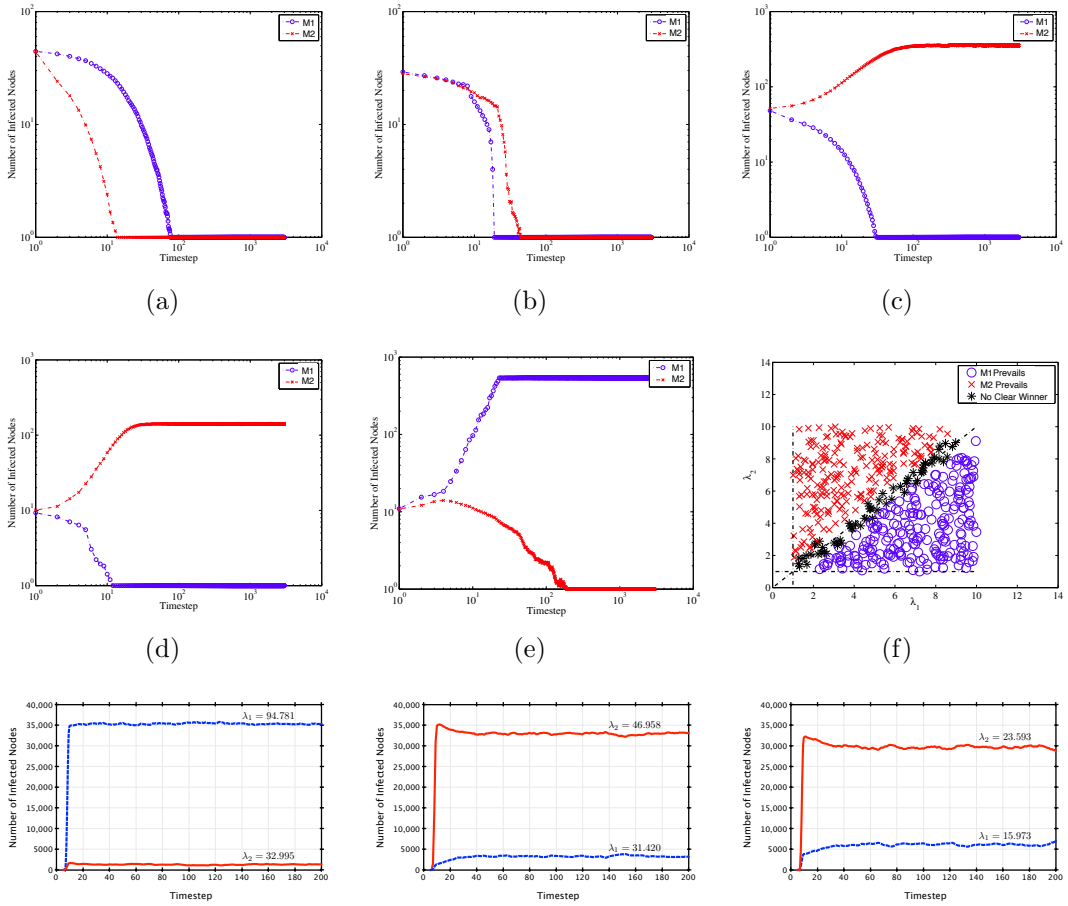


Figure 3.2: Simulation Results: Infection plot over time (log-log) in Figure(a)-(e). 3.2(a): Synthetic Composite Networks: $\lambda_1 = 0.97, \lambda_2 = 0.96$; 3.2(b): Real Composite Networks: $\lambda_1 = 0.9, \lambda_2 = 0.94$; 3.2(c): Synthetic Composite Networks: $\lambda_1 = 0.91, \lambda_2 = 1.63$; 3.2(d): Real Composite Networks: $\lambda_1 = 0.99, \lambda_2 = 1.4$; 3.2(e): $\lambda_1 = 4.5, \lambda_2 = 1.7$; 3.2(f): The outcomes for different combinations of system eigenvalues: $1 < \lambda_1 < 10$ and $1 < \lambda_2 < 10$; black dotted lines represent three lines λ_1=1, λ_2=1, and λ_1=λ_2. When the eigenvalues are roughly equal there is no clear winner.

different combinations of topologies. Here, we focus on these two, because: (a) we would like to have significantly different topologies, in order to show that our methods are not tailored to a particular family of graphs, and (b) scale-free graphs are known

to emerge in complex human and communication networks [8].

3.4.2 Large-scale data Sets ($1,000 < N < 50,000$)

To further stress-test the accuracy of our model, we conducted experiments on synthetic social networks with $1,000 < N < 50,000$ nodes using the `forestFire`, `randomWalk`, and `nearestNeighbor` graph generation models provided by Sala et al. [50]. These synthetic models are informed by real world measurements of social networks(e.g, Facebook) and provide graph structures that resemble such networks.

3.4.3 Simulation experiments

All experiments on real and synthetic composite networks were conducted using a combination of Matlab and Python. All values are averaged over 100 simulation runs. In each experiment, each virus infects a unique set of nodes Ini_1 and Ini_2, each with the same size, selected uniformly at random from N, subject to the constraint $Ini_1 \cap Ini_2 = \emptyset$ (i.e., mutually exclusive). We run each simulation until it reaches a relatively stable state as we discussed in Section 3.2, at which point, we determine the average number of nodes infected by M_1 and M_2 and report the outcome, which then gets averaged across 100 runs.

3.4.4 Analysis of results

From Section 3.3, we know that if the system matrix's first eigenvalue of one virus is less than 1, the corresponding virus will die-out eventually. Therefore, in this scenario, we can predict which virus prevails eventually using the following three rules:

(i) if $\lambda_1 < 1$ and $\lambda_2 > 1$, then M_2 tends to prevail eventually in the composite networks;

(ii) if $\lambda_1 > 1$ and $\lambda_2 < 1$, then M_1 tends to prevail eventually in the composite networks;

(iii) if $\lambda_1 < 1$ and $\lambda_2 < 1$, then both viruses will die out and none of them can be said to prevail.

Figures 3.2(a)-(e) demonstrate the proposed rules on both synthetic and real composite networks. The infection starts by infecting 30 nodes for each virus in Figure 3.2(a), Figure 3.2(b) and Figure 3.2(c), and 10 nodes for each virus in both Figure 3.2(d) and Figure 3.2(e). The outcomes of below- and above-threshold from these rules can be distinctly seen in these figures. These results show that, though simple, our proposed rules are very effective for predicting which virus tends to prevail eventually in the composite networks.

This is the more interesting case in terms of competition: each virus in isolation would not die-out, so it is a "fight for dominance." As shown in Figure 3.2(f), we

find again that the system eigenvalues play a critical role: the virus whose first eigenvalue is larger tends to prevail eventually in the composite networks. In addition, Figures 3.2(g)-(i) show experimental results of large scale epidemic simulations using a ForestFire and Nearest Neighbor synthetic graph models. Results are for $N = 40,000$ nodes, but are consistent for results of $10,000$ to $50,000$ node experiments. Unlike smaller-scale experiments, these results show that the weaker virus may retain some endemic population, yet the virus with the larger eigenvalue clearly dominates the simulation.

3.5 Suppression and Control

In this section, we design and evaluate suppression methods based on two distinct strategies.

1. *Unilateral Suppression.* The goal of this strategy is to suppress one meme, while leaving the other unscathed, thus free to spread unimpeded. Using the five techniques described below, we intend to suppress V_1 by removing ability to spread the memes from a subset of carefully selected nodes. The intention of unilateral suppression is to reduce λ_1 to below λ_2 (i.e., $\lambda_1 < \lambda_2$), thus affecting the outcome.

2. *Concurrent Suppression.* The goal of this strategy is to suppress the spread of both viruses by removing a set of nodes from both graphs in the composite

Method Name	Unilateral	Concurrent	Intuition
`Random`	$rand(node_{G_1})$	$rand(node_{G_1\|G_2})$	Randomly select a node and remove it from G_1 ($G_1\|G_2$)
`Acquaintance`	$rand(neighbor)$ of $rand(node)$	$rand(neighbor)$ of $rand(node)$ of $rand(G_1\|G_2)$	Acquaintance immunization, remove a random neighbor of a randomly selected node in G_1 ($G_1\|G_2$).
`Max Degree`	$max(deg(G_1))$	$max(deg(G_1\|G_2))$	Remove node with the maximum degree in G_1 ($G_1\|G_2$).
`Social Hierarchy`	$max(rank(node))$	$max(rank(node))$	Remove node with the highest rank.
`Greedy`	$max(\lambda_1)$	$max(\lambda_1\|\lambda_2)$	Remove the node that causes the largest eigenvalue drop in either λ_1 or λ_2

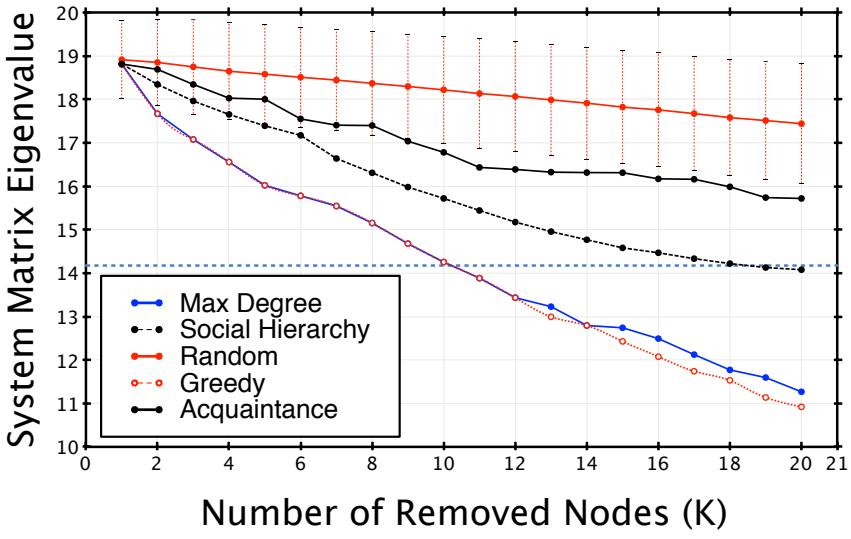
Table 3.2: Suppression Methods

network. Ultimately, we want to reduce λ_1 and λ_2 to below 1 (i.e., $\lambda_1, \lambda_2 < 1$), thus stopping the spread of both viruses.

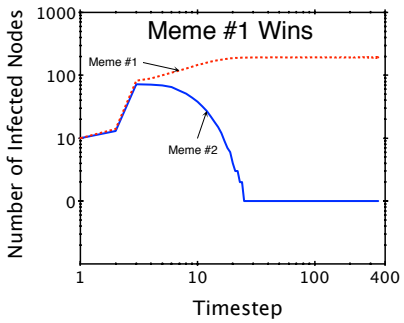
Given these two main strategies, we propose 5 methods, described in Table 3.2, partially motivated by the methods used in single virus/disease propagation on a single network [16, 55]. We evaluate each method's effect on the system matrix eigenvalues for each subgraph in the composite network (λ_1,λ_2). The proposed methods are: (a) `Random`, (b) `Acquaintance`, (c) `Max Degree`, (d) `Social Hierarchy` and (e) `Greedy`.

3.5.1 Unilateral Suppression

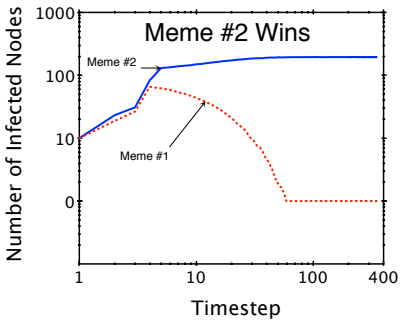
As mentioned above, the objective of unilateral suppression is to reduce λ_1 to less than λ_2, thus reversing the prediction of our EigenPredictor. That is, we seek to answer: *What set of nodes should we suppress in order to reduce the spread of*



(a)



(b) Before Suppression



(c) After Suppression

Figure 3.3: Example of Unilateral Suppression on the enterprise data set. The methods `Greedy`, `Max Degree` and `Social Hierarchy` drop the system matrix eigenvalue λ_1 below λ_2 (thus reversing the prediction of the EigenPredictor); (b) shows the original competition results without removing nodes; note that V_1 wins, while V_2 dies out; (c) shows the competition results after removing $k = 20$ nodes using the `Max Degree` method; the result is reversed, with V_2 winning and V_1 dying out.

one virus, ultimately resulting in the dominance of the other, unsuppressed, virus?

We present the results of using Unilateral Suppression on the enterprise data set in Figure 3.3. Note that V_1 will eventually prevail in the composite network prior to

applying any unilateral suppression strategies. Then, observe that the value of λ_1 decreases as nodes are removed from the system. At $k = 10$, λ_1 is reduced to below λ_2 (thus reversing the prediction of the *EigenPredictor*).

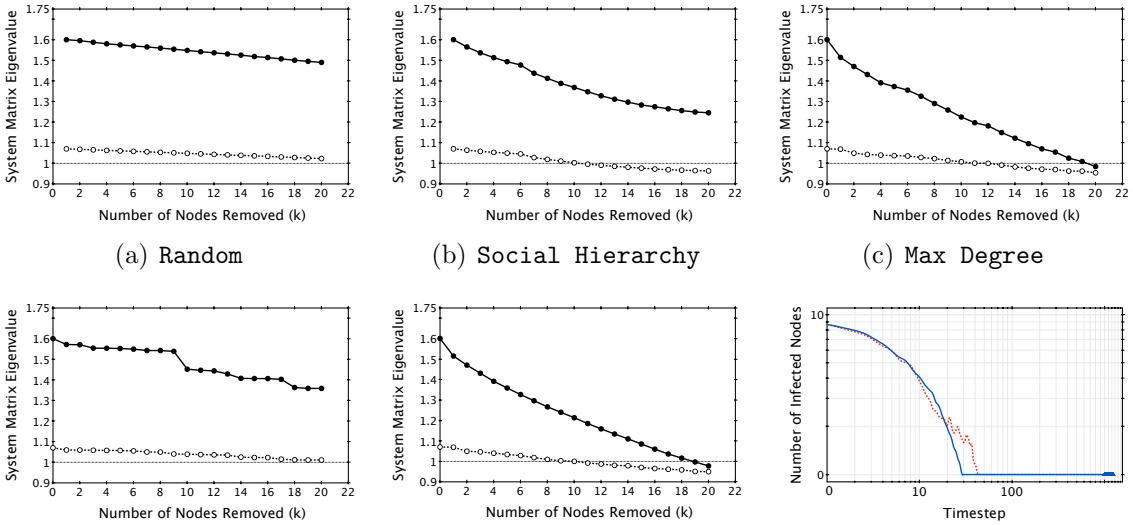
As expected, the two methods that rely on randomness (i.e., `Random` and `Acquaintance`) have the worst performance compared to the other methods. In contrast, `Greedy` performs better than the others, yet is the most expensive computationally. `Max Degree` performs surprisingly well, within 1% of `Greedy` at much lower computational cost.

Interestingly, when we remove nodes based on their social status (e.g., remove "bosses" before "managers," and so on), the method `Social Hierarchy` performs better than the random methods, yet not as well as the topologically-informed models, and eventually (at $k = 20$) crosses the value of λ_2. Though not as effective, in situations where we lack topological information, we could potentially rely on easily observable social hierarchy information to inform our suppression process.

3.5.2 Concurrent Suppression

Under the concurrent suppression scheme, the goal is to reduce the effective spreading power of two viruses spreading through different modes of communication (i.e., edge sets E_1, E_2 in a composite network). Simply put, we ask: *What set of shared nodes should we inoculate in order to reduce the spread of both viruses the most?*

We present the results of our suppression methods in Figure 3.4. As before, we



(a) `Random` (b) `Social Hierarchy` (c) `Max Degree`

(d) `Acquaintance` (e) `Greedy` (f) Outcome of Concurrent Suppression

Figure 3.4: Example of Concurrent Suppression on the enterprise data set, using each method. The epidemic threshold is marked in each plot at 1. Again, both the `Greedy` and `Max Degree` methods drop λ_1 and λ_2 below the epidemic threshold. Subplot 3.4(f) shows suppression results after removing $k = 20$ nodes selected using the `Max Degree` method—both viruses die out.

observe that `Max Degree` and `Greedy` reduce both λ_1 and λ_2 to below the epidemic threshold (indicate by the horizontal line at 1) at approximately $k = 19$.

`Social Hierarchy` provides mixed results. As indicated by the lower plotted line (representing the SMS portion of the enterprise network), `Social Hierarchy` is nearly as effective as `Max Degree` and `Greedy`. Yet, on the upper line (call graph of the enterprise network), the `Social Hierarchy` method is not as effective and does not reduce the spreading power to below the epidemic threshold.

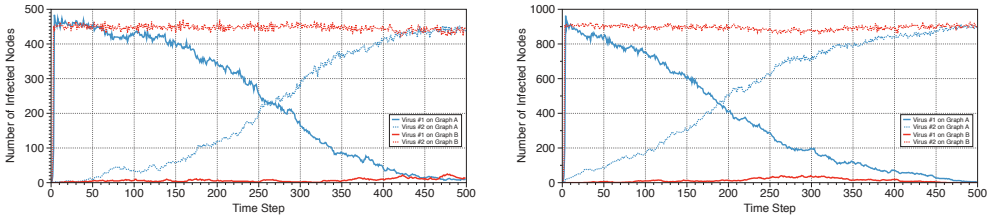
3.5.3 Summary

In summary, we have designed and evaluated several techniques for unilateral and concurrent suppressions, which are based on randomness, topological information and social hierarchy. The results from both suppressions show that the topological properties-based method (i.e., `Max Degree`) is very effective in controlling virus propagation compared to other methods. Put another way, removing the highest-connected node is a very effective suppression strategy. In situations where we lack topological information, we could potentially rely on the explicit information of social hierarchy to design our suppression scheme (e.g., `Social Hierarchy`), though not as effective as the topological properties-based method.

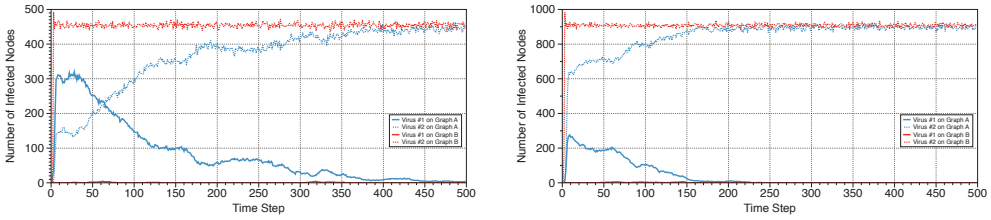
3.6 Cross-Contamination

3.6.1 Cross-Contamination Experiments

Until this point, I have considered viruses spreading on a composite network to be mutually exclusive, i.e., a virus will spread using only the edges associated with their own network. In this section, I evaluate the effect of cross-contamination across various synthetic graphs. Here, I change the model slightly to allow the virus from one network to eventually "transform" into a virus that can propagate on the edges of the other network. For example, consider a rumor propagating in Facebook being



(a) $N = 500$, Erdős-Rényi vs. Erdős-Rényi (b) $N = 1000$, Erdős-Rényi vs. Erdős-Rényi

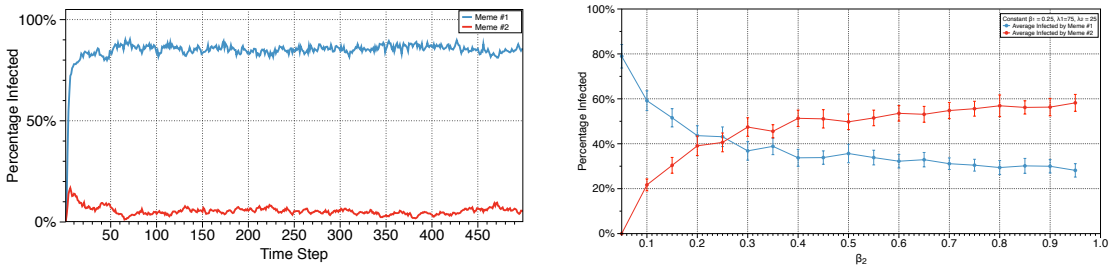


(c) $N = 500$, Erdős-Rényi vs. Barabási-Albert (d) $N = 1000$, Erdős-Rényi vs. Barabási-Albert

Figure 3.5: Simulation Results of Cross Contaminations.

transformed into a rumor spread on twitter by an individual user. The user creates a new meme and releases it on twitter, which spreads the same information, but now this meme is spreading across twitter edges. Our simulation model emulates this exact scenario, by carefully following the propagation of these "cross-over" virus. To account for the new ability of a virus to jump composite network layers, each virus is assigned a cross-contamination parameter, denoted $0 < X_{A\to B}, X_{B\to A} \leq 1.0$. Specifically, $X_{A\to B}$ describes the ability of a virus propagating on composite network layer A to cross to layer B. $X_{A\to B}$ is similarly defined.

In the shown scenarios, if simulated in isolation, each of the viruses would propagate and capture the graph. While in competition, in each of these simulations,



(a) Example of Cross-Contamination

(b) Cross-Contamination vs. β_2

Figure 3.6: Cross-Contamination Simulation Examples

ultimately one virus dominates. I have established that the virus propagating with the largest eigenvalue will eventually dominate the graph in the absence of cross-over. However, the crossing over allows the virus with higher such likelihood to spill over and propagate on the other topology, thus giving an advantage to the virus with higher such likelihood.

In Figure 3.5(a), I demonstrate the results of cross contamination on a composite network of 500 nodes. In this example, $\lambda_{S,1} = 26.43$, and $\lambda_{S,2} = 13.24$ with cross-contamination probabilities of $X_{A\to B} = .15$ and $X_{B\to A} = 0.05$. Observing that $\lambda_{S,1} > \lambda_{S_2}$, I see that virus #1 ultimately captures the greatest number of nodes. Interestingly, virus #2 does not completely die out, possible due to the ability of virus #2 to cross over to both layers of the composite network.

In Figure 3.5(b), I show the average steady-state percentage of infected nodes versus the propagations strength of virus #2 (β_2). All other parameters are constant, in particular $\beta_1 = 0.25$, $X_{A\to B} = 0.10$, and $X_{B\to A} = 0.10$. I observe, that as β_2

increase, eventually, virus #2 will eventually dominate the simulation.

3.7 Discussion

In this section, we discuss the limitations of our work and possible future directions.

Choice of epidemic model. The flu-like SIS (Susceptible-Infected-Susceptible) epidemiological model is simple, yet illustrative, and has been extensively studied in past literature in a single-virus setting (cf. [3, 34, 29]). Therefore, we chose to extend SIS in order to gain fundamental insights into the dynamics of competing viruses. We leave the investigation of other epidemic models as future work.

Real composite networks. Finding real data for any networked system or communication is non-trivial due to privacy concerns, infrastructure limitations, and measurement biases. Finding real data sets of *composite networks* is even more challenging. Obtaining the enterprise dataset used in this paper was instrumental in modeling and understanding how real composite network operate, but the dataset comes with dissemination restrictions. I believe that the research community could greatly benefit from the creation of an open repository of real composite networks.

3.8 Related Work

Single-virus propagation. Many works focus on single meme propagation on one single topology. Compartmental models like SIS, SIR, etc., have been well-studied in many epidemiological texts [3, 34]. The evolution of blogs and the maximization of influence propagation are studied in [22]. Information cascades models are proposed to study the meme propagation in word-of-mouth communications [21]. Numerous studies exist on virus propagation on the Internet based on the basic epidemic models of infection [52, 54]. A fundamental question in epidemiology is the presence of a threshold, under which an epidemic is guaranteed not to happen. Pastor-Satorras et al. [42] proposed an epidemic threshold condition for random power law networks, which uses the "mean-field" approach. Ganesh et al. [20] and Yang et al. [58] provided epidemic threshold for the single-virus on single topology. Prakash et al. [46] gave the epidemic threshold condition for almost all single-virus epidemic models on a single static network.

Multiple viruses and interdependent networks. Newman [40] studied multiple viruses on a single, special random graph and provided the epidemic threshold for the case when the second virus propagates over the residual network after the propagation of the first virus has completed. This scenario is close to the dynamics of propagation of a single virus—one virus passed over the network, the second virus starts to pass over the residual network. Prakash et al. [45] only considered multiple

ideas spreading on a simple and fair-play network. Models for multiple cascades have been studied as extensions of the independent cascade model, where once a node is infected with a cascade, it never change its state [11]. The effects of cascades in inter-dependent networks (e.g., Internet router and power electricity networks) were investigated by Buldyrev et al. [12]. However, all of these works are completely different from our problem as we consider the more realistic and challenging scenario of competing viruses propagating simultaneously on composite networks.

Game theory. Meier et al. [35] studied inoculation games in social networks, where each node selfishly decides whether or not to protect itself. The game between a virus and an alert over a network was investigated by Aspnes et al. [5]. Kostka et al. [32] studied competing campaigns as a game-theoretical problem and showed that being the first player was not always advantageous. However, these works using game theory are different from our problem where we assume that all nodes are passive and follow the same propagation model.

3.9 Summary

In this paper, we have formulated the scenario of competing viruses on composite networks as an SI_1I_2S model and showed that the epidemic threshold depends on the largest eigenvalues of the system matrices of both viruses. Extensive simulations on different datasets demonstrate the effectiveness of the epidemic threshold and the

largest eigenvalue.

Chapter 4

Non-binary Information Propagation

This chapter examines the qualitative behavior and phase transitions points of a non-linear, continuous value spreading process on a discrete-time, multi-agent network. I begin with by presenting background information on the classic logistic map, $f(x(t)) = x(t+1) = r_0 x(t)(1-x(t))$, which inspires my two networked logistic map models, eNLM and iNLM. Using these two models, I demonstrate empirically similarities with the classic logistic map. Ultimately, through empirical analysis, I find that the network topology, summarized by the first eigenvalue of its adjacency matrix, is critical to the phase transition points and equilibrium behaviors of the eNLM and iNLM models. Despite the simplicity of our system, we show that these results include a variety of dynamical systems and apply to diverse scientific fields, from

biology to network routing protocols.

4.1 What is non-binary propagation?

In the previous chapters, I presented binary propagation models for a variety of mobility and competition scenarios. In each of those models, I assumed that the information state propagating through the system was represented by discrete state value, e.g., "infected" or "not infected." In this chapter, I re-evaluate the discrete information assumption and present models for propagating non-binary information.

This change in the nature of the propagating information reflects my observations of real-world information. To clarify, consider the following scenario. Google's Android-based smartphones are gaining popularity and market share through word-of-mouth campaigns and advertisements. In a word-of-mouth campaign, an individual is told by a friend that Android phones are great, and at some point, the individual will be persuaded to purchase the product and share their opinions with their friends. Yet, one's excitement about the smartphone is not strictly "excited" or "not excited" (i.e., binary), rather one experiences a range of excited states, and one's strength of excitement influences family, friends and others proportionally. The strength of one's excitement can be measured as a continuous real value that spreads influence across one's social network of acquaintances.

This example summarizes the key insight that motivates this work, which–simply

put–is that *certain phenomena are not adequately described by a single binary digit (i.e., infected/not infected), but better described as a spectrum.* Other examples of phenomena that may benefit when modeled using a continuous value include market penetration, product adoption, gossip/rumors and public opinion.

In the study that follows, I focus on developing the simplest possible model that characterizes the propagation of real value properties across a networked system. I modify the classic logistic map to account for non-binary, continuous value state propagating across networked agents. The selection of the logistic map simplifies analysis and simulation implementation.

This chapter is separated into three parts. First, I will begin by reviewing the canonical results on fixed points, phase transitions and chaos in the classic logistic map function. Next, I introduce our two models for propagating non-binary information across networks and qualitatively evaluate the results of a series of simulation experiments on a variety of random graph models. Finally, using the insights gained from the qualitative study, I present a series of analytical results that, like the classic logistic map, demonstrate the existence of fixed points and phase transitions in our networked logistic map models.

To the best of my knowledge, this is the first work that examines such a network wide propagation problem and sets the stage for further theoretical analysis of these systems. From a theoretical point of view, our work may be seen as a first attempt to explore the outcome of two competing phenomena: *attrition* at every propagation

hop and *amplification* by propagation of properties from adjacent nodes (neighbors).

4.1.1 Contributions

I present the following contributions.

1. **Existence Threshold,** r_e. There exists an existence threshold r_e, such that, for all $0 < r_0 < r_e$, the system will converge to $\vec{x^*} = \vec{0}$ as $t \to \infty$. Furthermore, we show that $r_e = \frac{1}{\lambda_{\mathbf{A},1}}$, where $\lambda_{\mathbf{A},1}$ is the largest absolute value eigenvalue of the adjacency matrix for all graph models. Numerical simulation results are presented in Chapter 4.5.3 followed by analysis in Chapter 4.6.1.

2. **Dampening Threshold,** r_d. I demonstrate the dampening threshold r_d, such that, for any value of r_0 in the range $r_e < r_0 < r_d$, the system stabilizes to a fixed point x^* and at no time during the stabilization process does the system state exceed x^*. Furthermore, I demonstrate that for $r_d < r_0 < r_p$, the system converges to x^*, yet during the stabilization process, the system state exceeds x^*. These results are show empirically that $r_d \approx \frac{2}{\lambda}$ (Chapter 4.5.4).

3. **Periodic Threshold,** r_p. I show there exists a minimum value for the periodic threshold r_p such that, for values $r_p < r_0$, the system will exhibit a periodic steady state behavior, i.e., there will exist a period p such that $\vec{x}(t) = \vec{x}(t-p)$. Thus, the steady state behavior is a set $\vec{x^*} \in \{\vec{x}(t-p+1), \vec{x}(t-p+2), \cdots, \vec{x}(t)\}$.

The periodic threshold will occur at $r_p \leq \frac{3}{\lambda_{\mathbf{A},1}}$. I show this results empirically for for random graphs in Chapter 4.5.5)

4. **Initial Condition Insensitivity**. Given a graph G, a parameter r_0, and any appropriately bounded set of initial conditions $\vec{x}(0) = [x_0(0), x_1(0), \cdots, x_{N-1}(0)]^\intercal$, the system will reach equilibrium at the fixed point $\vec{x}^*$ for all $\vec{x}(0)$. This results is demonstrated in Chapter 4.5.6

4.2 What are the challenges?

Most work on spreading processes falls into one of two categories: (1) continuous-time, continuous-state systems or (2) discrete-time, discrete-state systems. Most classic models, such as those presented in [53], are continuous-time, continuous-state systems and fall into the first category. Such systems typically assume homogeneity of agents across the system, yet, with by adding a discrete structure, e.g., a network or graph, this assumption is invalidated and new analytical approaches are required. In order for these systems to account for discrete network structures, such structures are often reduced to a statistical model representing the mean-value behavior of all agents in the system and ignoring or losing localized conditions.

The models presented in Chapters 2 and 3 fall into the second category and do not assume agent homogeneity. As a result, analysis of these models include discrete network structures and have produced a number of important results that quantify

the effect of network structure on the spread of information.

The challenges we attempt to address by using a discrete-time, continuous-value models are as follows.

1. Can we retain localized information?

2. Does the epidemic threshold have any meaning in such systems? If so, how does it relate to classic models?

3. Does our model lend itself to analysis not available to the classic models?

4.3 Preliminaries

Before we discuss our proposed network spreading models, we begin with a discussion the function that forms the heart of our models, namely the classic logistic map. With the logistic map as the basis of our analysis, in this section we introduce the concepts of phase transition, fixed (or equilibrium) points, and chaos.

4.3.1 The Classic Logistic Map

Throughout the remainder of this chapter, the classic logistic map, shown below in Equation 4.1, forms the basis of our networked propagation models. Expressed as a difference equation, the logistic map is

$$f(x(t)) = x(t+1) = r_0 x(t)\,(1 - x(t)) \tag{4.1}$$

Symbol	Definition
N	Number of agents in the system (vertices)
M	Number of connections in the system (edges)
$V = \{v_0, v_1, \cdots, v_{N-1}\}$	Set of vertices v_i in G, where $\lvert V\rvert = N$
$E = \{e_0, e_1, \cdots, e_{M-1}\}$	Set of edges in G, where $e_i = (v_k, v_j)$ and $\lvert E\rvert = M$
$G = (V, E)$	Graph topology
$\mathbf{A} = [a_{ij}]$	The adjacency matrix representation of G
$\vec{x}(t) = [x_0(t), x_1(t), \cdots, x_{N-1}(t)]^\intercal$	The vector of states for each agent v_i, where $\vec{x} \in \mathbb{R}^N$
x^*	Fixed point (with period $p = 1$)
x^{*p}	Periodic point with period p
$x^* = \sum_i \vec{x}_i^*$	Sum of $\vec{x}^*$ values
$f_i(\vec{x}(t+1)) = f(x_i(t))$	Node-level dynamical function

Table 4.1: Terminology Specific to Non-binary Propagation

and is subject to the constraints

1. $x(0) = x_0$,

2. $x_0 \in [0, 1]$, and

3. $0 < r_0 < 4$.

The parameter r_0 is often referred to as the bifurcation parameter[1] and may be considered akin to a combination of the birth rate β and death rate δ parameters discussed in previous chapters. The term $x(t)$ represents the value of the system at time t, and x_0 represents the initial condition value.

4.3.2 Definitions

With Equation 4.1 in mind, we introduce the following definitions. For more information, refer to [53]. Please note, the definitions presented below are well-known,

[1]Not to be confused with r_0 ("r-not") previously discussed as the epidemic threshold, yet, as we show below, the two parameters perform a similar function.

yet reproduced here to add clarity and consistency for the work that follows.

Definition 11. (Fixed Point[2])

A fixed point, denoted x^*, is a point such that $f(x^*) = x^*$.

Definition 12. (Fixed Point Stability)

A fixed point x^* is called stable if and only if $|f'(x^*)| < 1$ and unstable if and only if $|f'(x^*)| > 1$.

Definition 13. (Periodic stability)

A point x^{*n} that satisfies the fold[3] $f^n(x^{*n}) = f(f(\ldots(x^{*n}))) = x^{*n}$. The period is the smallest value of n for which $f^n(x^{*n}) = x^{*n}$ holds. Note that the fixed point x^* is a special case of the periodic fixed point with a period $n = 1$, i.e., $x^* = x^{*1}$.

4.3.3 The Fixed point, x^*

To determine the fixed point x^* for the classic logistic map, one must first solve $f(x^*) - x^* = 0$, i.e., the point at which $f(x^*) = x^*$. Next, one must determine the stability criteria for which x^* is stable. In Lemma 14, I determine x^* for the classic logistic map, following the two steps outlined above.

Lemma 14 (Logistic Map Fixed Point x^*)**.** *The logistic map has fixed points x^* (with*

[2]May also referred to as *equilibrium point*.

[3]i.e., the function f applied to x^{*n} n-times, so if $f^3(x^{*3}) = x^{*3}$, then the periodic fixed point is x^{*3} and the period is $n = 3$.

period $p = 1$*)*

$$x^* = 0 \ \textit{and} \ x^* = \frac{r_0 - 1}{r_0}$$

Proof. Recall Definition 11, for the logistic map[4] the fixed points are given by $f(x^*) = x^*$. Given $r_0 x^*(1-x^*) = x^*$, we solve for $f(x^*) - x^* = 0$ as follows:

$$\begin{aligned} 0 &= f(x^*) - x^* \\ &= r_0 x^*(1 - x^*) - x^* \\ &= r_0 x^* - r_0 x^{*2} - x^* \\ &= (x^*)(r_0 - r_0 x^* - 1) \end{aligned}$$

Now, ignoring the trivial solution $x^* = 0$, I concentrate on the term $(r_0 - x^* - 1) = 0$.

$$\begin{aligned} 0 &= (r_0 - r_0 x^* - 1) \\ r_0 x^* &= r_0 - 1 \\ x^* &= \frac{r_0 - 1}{r_0} \end{aligned}$$

Thus, $x^* = 0$ and $x^* = \frac{r_0 - 1}{r_0}$. □

The two fixed points are stable only for a subrange of r_0 values, which we can find by determining what range of r_0 satisfies Definition 12, that is, where $|f'(x^*)| < 1$.

[4]Or any general difference equation

Solving for $|f'(x^*)| = |f'(\frac{r_0-1}{r_0})|$, we find that x^* is stable in the range $1 < r_0 < 3$.

4.3.4 The Periodic Fixed Point, x^{*n}

What happens at values of $r_0 > 3$ beyond which the x^* identified above are unstable? At $r_0 > 3$, $|f'(\frac{3-1}{3})| > -1$, meaning for values of $r_0 > 3$, the solution $x^* = \frac{r_0-1}{r_0}$ in unstable. Let's begin by assuming that there exists a fixed point x^{*2} solves the fold f^2 with period $n = 2$, where f^2 is the composite function $f^2(x) \equiv f(f(x))$.

Lemma 15. *Two stable fixed points exists for the system $f^2 \equiv f(f(x))$ if and only if $3 < r_0 < (1 + \sqrt{6})$.*

Proof. Begin by determining $[f^2(x)]'$.

$$\begin{aligned}\left[f^2(x)\right]' &= f'(f(x)) \times f'(x) \text{ (by chain rule)} \\ &= r_0\left[1 - 2(r_0x(1-x)\right] \times r_0(1-2x) \\ &= r_0^2(1-2x)(1-2r_0x+2r_0x^2)\end{aligned}$$

Now, evaluating f^2 for stable fixed points $x^{*2} = \frac{(r_0+1)\pm\sqrt{(r_0+1)(r_0-3)}}{2r_0}$, we find that $|\,[f^2(x^{*2})]'\,| < 1$ for all values of $3 < r_0 < (1 + \sqrt{6})$. □

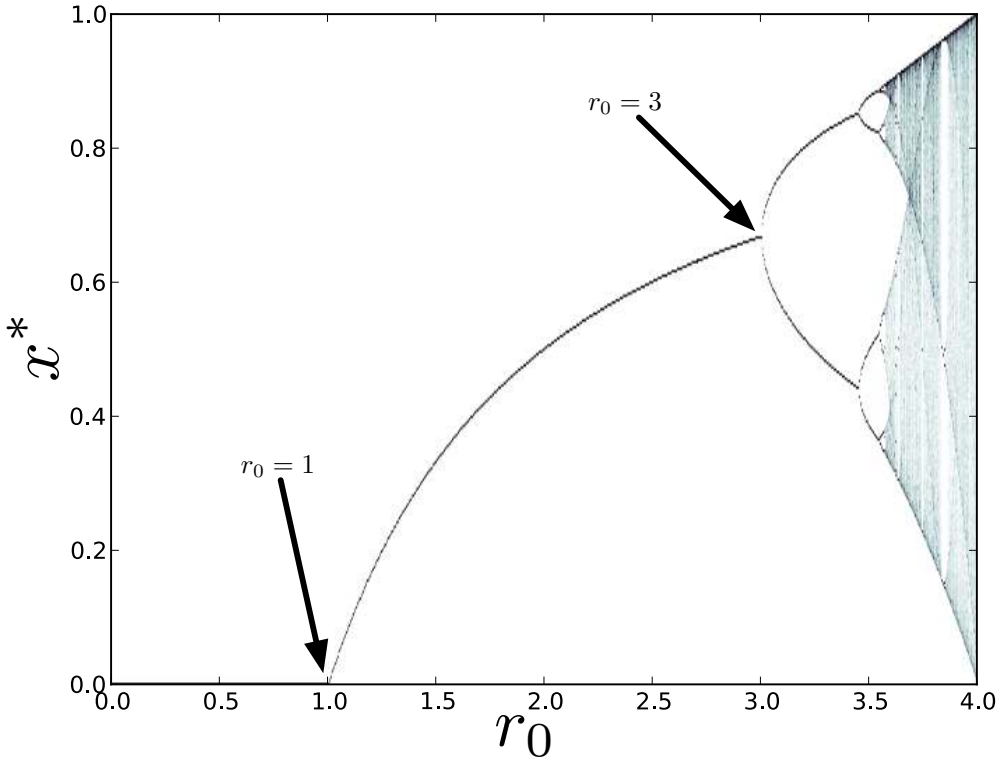


Figure 4.1: The bifurcation diagram of the classic logistic map for the r_0 in the range $[0, 4]$.

4.3.5 Interpretation Fixed Points and Phase Transitions

Thus far, we have identified three regions of stability for the stable fixed point solutions presented above. The trivial x^*=0 is stable in the range bounded by $r_0 \in (0, 1)$, $x^* = \frac{r_0 - 1}{r_0}$ is stable in the range bounded by $r_0 \in (1, 3)$. And finally, the third fixed point x^{*2} is delineated by the region bounded by $r_0 \in (3, 1+\sqrt{6})$. As we increase the values of r_0, the system experiences a variety of phase changes.

Definition 16. (Phase change)

Given the logistic map as described in 4.1, a phase change occurs when the behavior of $f(x)$ experiences a bifurcation.

In Figure 4.1, we present a plot of the r_0-value against the value of the fixed point x^* and shows the phase transition behavior of the logistic map for values of r_0 in the range $[0, 4]$. Indicated on the plot are the points corresponding to $r_0 = 1$ and $r_0 = 3$, i.e., the r_0-values that delineate the boundaries between the three fixed points regions. We observe that the behavior of x^* changes as each threshold is crossed. For instance, as the r_0-value increases crossing $r_0 = 1.0$, x^* achieves a non-negative value. Furthermore, as the r_0-value crosses $r_0 = 3.0$, the fixed point behavior splits into two distinct values, corresponding to x^{*2}. These phase changes are referred to as bifurcations, and this type of plot is commonly referred to as a *bifurcation diagram*. For values of $0 < r_0 < 1$, the system converges to a single state $x^* = 0$.

Beginning at $r_0 > 1$, the logistic map converges to a single, non-zero value. This behavior changes phase again at $r_0 > 3$, at which, we observe the system bifurcate into a periodic fixed point $x^* = x^{*2}$, with a period of $n = 2$. Simply put, when the logistic map system settles into an equilibrium x^*, the system oscillates between values in the sequence defined by $x^{*2} = \{x_0, x_1\}$.

For convenience, we will refer to $r_0 = 1$, i.e., the point at which x^* transitions to a non-zero value, as the *existence threshold* (r_e). The point $r_0 = 3$ will be referred to as the *periodic threshold* (r_p), i.e., the point at which x^* transitions from having a single solution to having multiple solutions. These thresholds signify a change in the

behavior of the logistic map, also known as a phase change.

4.3.6 Other fixed points

For brevity, the discussion above omits one other fixed point/phase transition that is important to the rest of this chapter, namely, the phase transition that occurs at $r_0 = 2$. We refer to this point as the *dampening threshold* (r_d). The dampening threshold occurs at $r_0 = 2$ and delineates two distinct phase region. In the first region, bounded by r_e and r_d, i.e., $1 < r_0 < 2$, the system converges to a single solution such that for all values $x(t)$, subsequent values of $x(t+1)$ are strictly non-decreasing, i.e., $x(t+1) \geq x(t)$. The second region is bounded by r_d and r_p, i.e., $2 < r_0 < 3$. In this region, the condition strictly non-decreasing condition does not hold.

For a complete discussion on the logistic map, fixed points and phase transitions, refer to Strogatz [53].

4.4 The NLM Propagation Model

In our networked models described below, we modify the logistic map from a scalar function f that maps $f : \mathbb{R} \to \mathbb{R}$ to a matrix function $\mathbf{F} : \mathbb{R}^{\mathbb{N}\times\mathbb{N}} \to \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$, where the dimension N is defined by the number of nodes in our system.

Consider a graph $G = (V, E)$, where V is a set of vertices (nodes) and E is a set of edges. Each node $v_i \in V$ influences its neighbors according to a logistic map

function, where $0 \leq i \leq N-1$. Furthermore, G is represented as an adjacency matrix $\mathbf{A} = [a_{ij}]$, where $a_{ij} = 1$ when node v_i connects to v_j and 0 otherwise. Let $x_i(t)$ denote the state of node v_i at time t and $\vec{x}(t) = [x_0(t), x_1(t), \cdots, x_{N-1}(t)]^\intercal$ denote the vector state of the nodes in the system.

Now, for our system, we must determine at what point do we accumulate the neighbor's input $x_j(t)$. In Figure 4.2, we describe the accumulation options. Ultimately, these two options resolve to taking the *inner product* versus the *outer product* during matrix multiplication.

The accumulation options depicted in Figs 4.2(a) and 4.2(b) are formalized as follows.

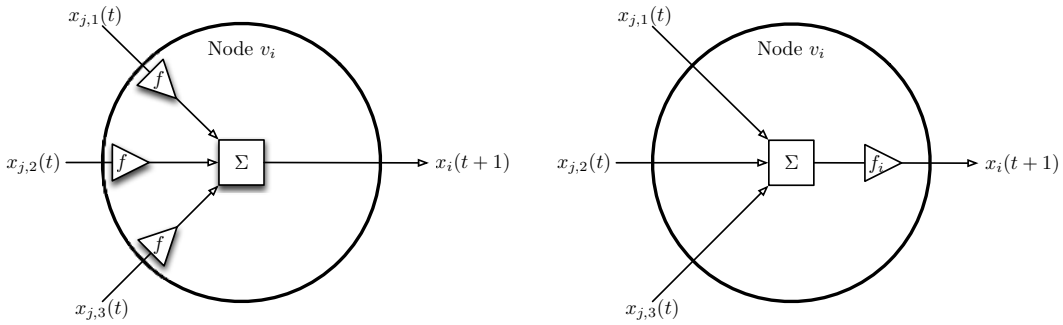
1. **Fig 4.2(a).** For each time t, we accumulate each of node i's neighbor's input, $\vec{a}_i^\intercal \times \vec{x}(t)$, then apply the logistic map such that $F_{LM}(\vec{a}_i^\intercal \vec{x}(t))$; or

2. **Fig 4.2(b).** For each time t, we apply the logistic map to each of i's neighbors $a_{ji} \times \vec{x}(t)$, such that $\sum_{j=0}^{N-1} F_{LM}(a_{ji} x_j)$.

where F_{LM} the function for the logistic map kernel, $F_{LM}(x) = r_0 \times x \times (1-x)$ (i.e., Equation 4.1).

Given the scenarios above, we find that at time $t+1$, the amount of influence spread $x_i(t+1)$ by node v_i is, respectively,

$$f_i(\vec{x}(t)) = x_i(t+1) = r_0 \times \vec{a}_i \vec{x}(t)^\intercal \times (1 - \vec{a}_i \vec{x}(t)^\intercal) \tag{4.2}$$



(a) Internal NLM (b) External NLM

Figure 4.2: Accumulation Options.

$$f_i(\vec{x}(t)) = x_i(t+1) = \sum_{j=0}^{N-1} [r_0 \times a_{ji} x_j(t) \times (1 - a_{ji} x_j(t))] \tag{4.3}$$

where the node v_j is a neighbor of v_i (i.e., edge $(v_j, v_i) \in E$ or $a_{ji} = 1$). For convenience, we refer to Equation 4.2 as the **Internally Summed Networked Logistic Map (iNLM)** and Equation 4.3 as the **Externally Summed Networked Logistic Map (eNLM)**.

Finally, we define $\mathbf{F_A}$ as

$$\mathbf{F_A}(\vec{x}(t)) = \vec{x}(t+1) = \begin{bmatrix} f_0(\vec{x}(t)) \\ f_1(\vec{x}(t)) \\ \vdots \\ f_{N-1}(\vec{x}(t)) \end{bmatrix} \tag{4.4}$$

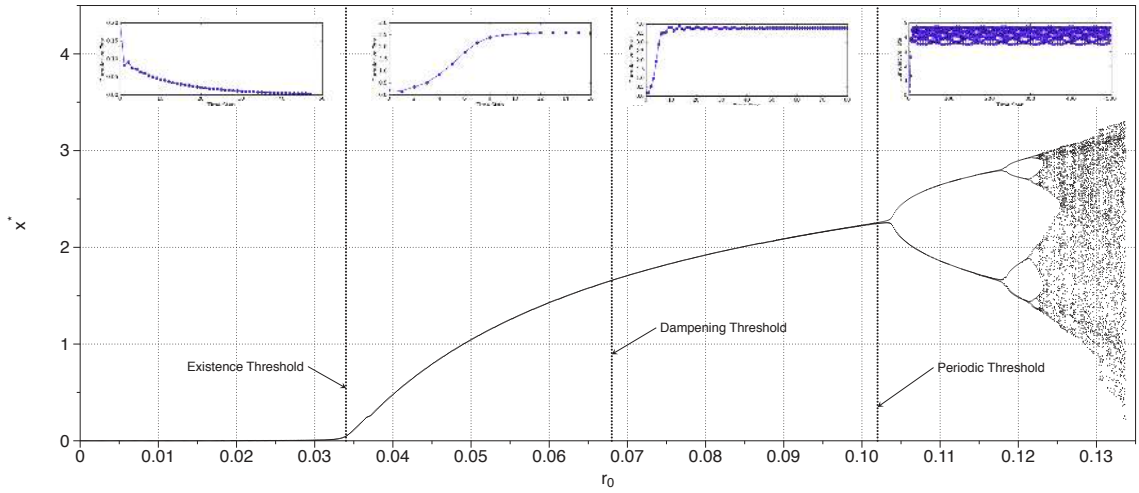


Figure 4.3: Example of eNLM thresholds simulated on an Erdős-Rényi random graph with $N = 100$, $p = 0.3$. Each vertical line indicates a particular threshold condition: (1) Existence Threshold, (2) Dampened Behavior Threshold, (3) Periodicity Threshold. Insets show typical behavior for each phase region.

4.4.1 Initial Observations of NLMs

In this chapter, we examine the effect of graph topology on the phase transitions of the systems described above. We begin by defining NLM analogues to the thresholds observed in the classic logistic mapping (i.e., §4.3.5 and §4.3.6), specifically, the *existence threshold* (r_e), *dampening threshold* (r_d), and *periodic threshold* (r_p).

In the following, we will refer to Figure 4.3 to illustrate the qualitative behaviors of each threshold under consideration. Figure 4.3 shows a bifurcation diagram for an eNLM spreading model on an Erdős-Rényi random graph. We are encouraged by the following qualitative observations.

- The general shape of the bifurcation diagram shown in Figure 4.3 resembles that shown in Figure 4.1.

- The behavior of each phase region (depicted in insets of Figure 4.3), behaves similar to the classic logistic map. For instance, in the region bounded by $0.034 < r_0 < 0.068$, all $r(t) \leq r(t+1)$, i.e., satisfy the strictly non-decreasing condition.

4.4.2 Limitations of our model

As with the classic logistic map, our model suffers from the problem of invalid or divergent initial conditions. We say the system diverges when $\vec{x^*} \to \infty$ as $t \to \infty$ and never settles on a solution. That is, there exists a vector $\vec{x_0}$ that when $\vec{x}(0) = \vec{x_0}$, the system diverges and will never settle to a fixed point $\vec{x^*}$. We must take care not to initialize simulations with values of $\vec{x_0}$ that cause the system to diverge.

4.5 Simulation Results and Implications

In the following sections, we perform a qualitative comparison of the eNLM/iNLM models versus the classic logistic mapping as exemplified in Figure 4.1.

4.5.1 Methodology

For each trial, we instantiated a specific graph based on classic random network models, namely the Erdős-Rényi random graph and 2. Barabási-Albert random graph models. Once instantiated, we selected a node seeding methodology, either *individual,*

uniform (iu), *subset, uniform (su)*, or *all, uniform (au)*, described below.

1. Individual, Uniform: A single node in the graph is initialized with a seed value drawn uniformly from (0,1).

2. Subset, Uniform: A random subset of k nodes are initialized with a seed value drawn uniformly from (0,1).

3. All, Uniform: A nodes in the system are initialized with a seed value drawn uniformly from (0,1).

Within each trial, all nodes were instantiated with either the iNLM or eNLM model described in Section §4.3. Trials ran until convergence on a fixed point or at least $t = 250$ time steps, whichever came first.

4.5.2 Graph Models

Erdős-Rényi Random Graphs

The Erdős-Rényi random graph models is defined as follows. Given N nodes and edge probability p, an edge exists between nodes i and j as follows:

$$e_{ij} = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } (1-p) \end{cases} \tag{4.5}$$

Barabási-Albert Random Graphs

This model is often referred to as the *Preferential Attachment Model*. Barabási-Albert grows a power-law network (with $\gamma = 3$) as follows: Begin with an initial network of $m_0 \geq 2$ nodes with degree of each node at least 1. New nodes are added until the desired graph size of N is reached. Each new node is connected to m existing nodes with a probability that is proportional to the number of edges that the existing nodes already have, more formally:

$$p_i = \frac{k_i}{\sum_j k_j}, \tag{4.6}$$

where k_i is the degree of node i.

4.5.3 Evaluation of Existence Threshold, r_e

I begin with an examination of the Existence Threshold on various network models. Recall that the existence threshold (r_e) is the point below which the system converges to $\vec{x}^*=\vec{0}$. The existence threshold has a number of implications related to information survivability and epidemic threshold.

Figures 4.4 and 4.4 show a series of plots for the iNLM on Erdős-Rényi graphs. The existence threshold is indicated in the first figure and shows that r_e is approximately 0.034. Curiously, we find this is approximately $\frac{1}{\lambda_{\mathbf{A},1}}$.

Figures 4.4 and 4.4 also demonstrate that r_e does not change for any given set

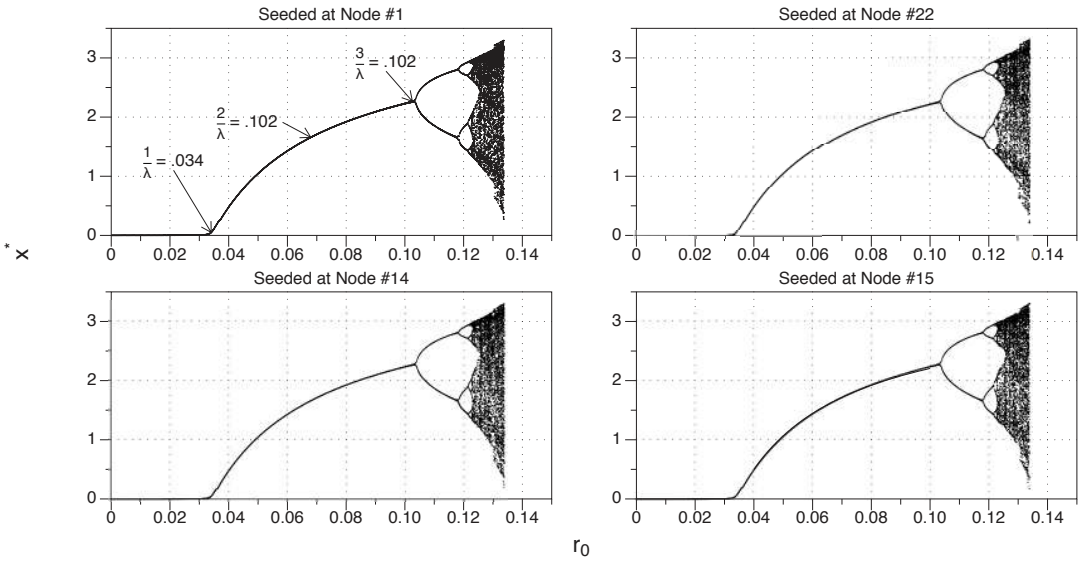


Figure 4.4: Bifurcation diagram of iNLM applied to an Erdős-Rényi Graph with $N = 100$ and $p = 0.3$ (and graph seed value of 0). For each figure, nodes were initialized using the individual, uniform method. $\lambda_{\mathbf{A},1} = 29.37$

of initial conditions, that is, despite changing the initial seeding method and values, the existence threshold always occurs at $r_e = 0.34$. When qualitatively compared to When compared to the classic logistic map in Figure 4.1, we note that all plots share the same general features, e.g., similar bifurcation patterns. The primary qualitative difference between the classic and the iNLM on Erdős-Rényi is the scale. For iNLM on Erdős-Rényi, the magnitude of $r^* = \sum \vec{r}_i^*$ is greater than the classic logistic map. Furthermore, as mentioned above, r_e is scaled by $\frac{1}{\lambda_{\mathbf{A},1}}$.

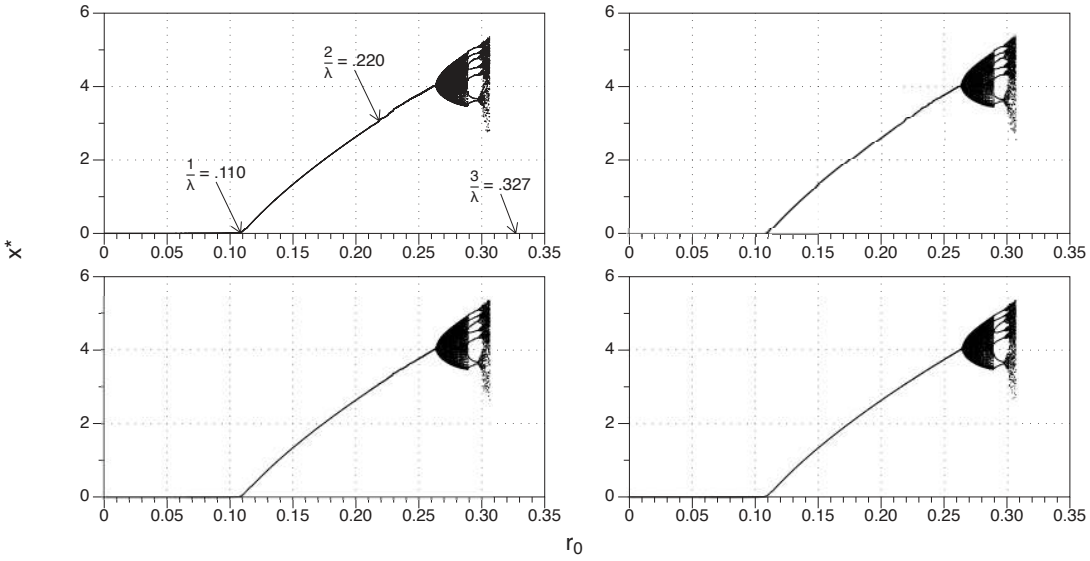
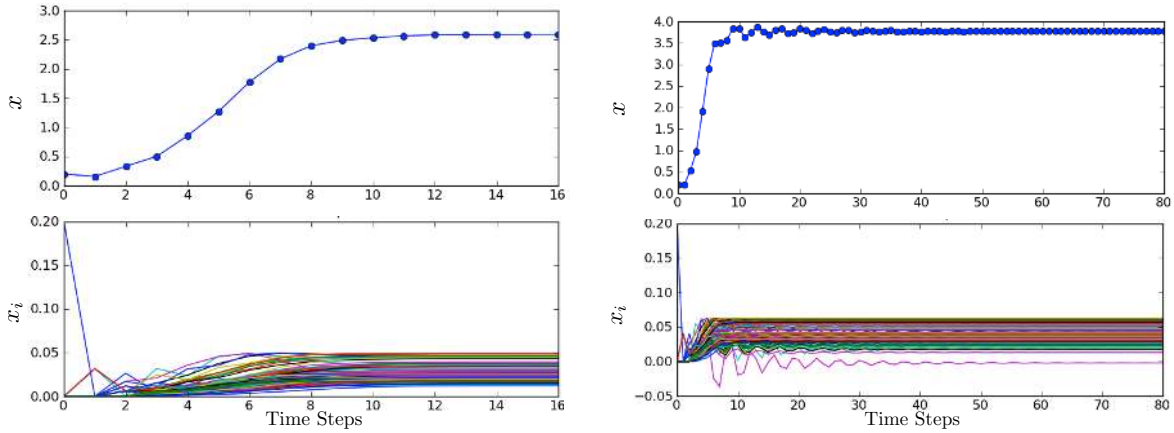


Figure 4.5: Bifurcation diagram of iNLM applied to a Barabási-Albert Graph with $N = 100$ and $m = 3$ (and graph seed value of 0). For each figure, nodes were initialized using the all, uniform method. $\lambda_{\mathbf{A},1} = 9.16$.

4.5.4 Evaluation of Dampening Threshold, r_d

A set of results demonstrating the effect of the dampening regions is shown in Figure 4.6. The non-dampening results are presented in Figure 4.6(a) and dampening results shown in Figure 4.6(b). For each plot, the top subgraph shows the cumulative x, whereas the bottom subplot shows individual node x_i, where $x = \sum_i x_i$.

Focusing on the top subplots, we see that for models exhibiting dampening behavior, the cumulative sum x fluctuates until finally settling to the fixed point of x^*. In contrast, whereas in the non-dampening subplot, the system settles to the fixed point x^* such that $x(t+1) \geq x(t)$, i.e., in a non-decreasing manner. For the examples shown in Figure 4.6, I empirically determined $r_d = 0.1986$.



(a) Non-Dampening, $r_0 = .195$

(b) Dampening, $r_0 = .250$

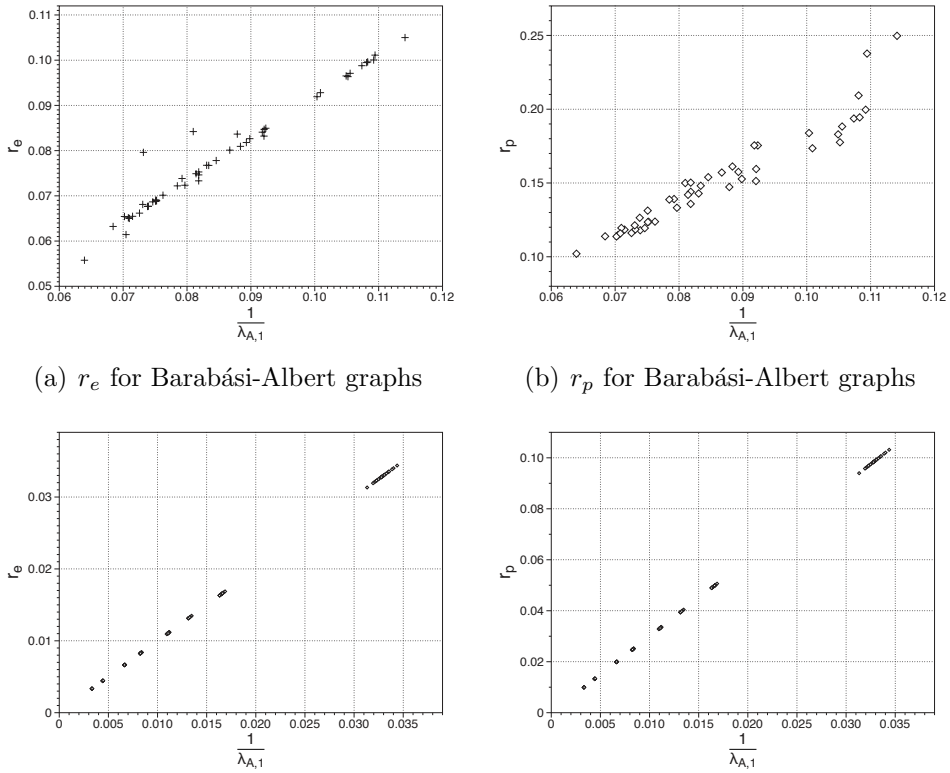
Figure 4.6: Non-dampening versus dampening. These figures were produced using a Barabási-Albert random graph with $N = 100$ and $m = 3$ and graph seed of 0. For each plot, the top subgraph shows the cumulative x, whereas the bottom subplot shows individual node x_i, where $x = \sum_i x_i$.

For all graphs and graph models examined, I determined that $r_d = \frac{3}{\lambda_{\mathbf{A},1}} \pm 0.01205N$, where N is the number of nodes in the system.

4.5.5 Evaluation of Periodic Threshold, r_p

In Figure 4.7, I summarize the values of r_p for both Erdős-Rényi and Barabási-Albert random graphs. These results are consistent for both iNLM and eNLM models.

In Figures 4.7(c) and 4.7(d) the results for iNLM and eNLM are shown against the inverse of $\lambda_{\mathbf{A},1}$. There is a clear linear relationship observed for both r_e and r_e, corresponding to $r_e = \frac{1}{\lambda_{\mathbf{A},1}}$ and $r_p = \frac{3}{\lambda_{\mathbf{A},1}}$. This relationship exists for Barabási-Albert graphs as well, indicated in Figures 4.7(a) and 4.7(b), yet there is a greater degree of variance.



(a) r_e for Barabási-Albert graphs

(b) r_p for Barabási-Albert graphs

(c) r_e for Erdős-Rényi graphs

(d) r_p for Erdős-Rényi graphs

Figure 4.7: These figures show the experimentally determined values of r_e and r_p for the Erdős-Rényi and Barabási-Albert graph on both eNLM and iNLM models against $\frac{1}{\lambda_{\mathbf{A},1}}$. For both thresholds, Erdős-Rényi have linear relation equivalent $r_e = \frac{1}{\lambda_{\mathbf{A},1}}$ and $r_p = \frac{2}{\lambda_{\mathbf{A},1}}$. Alternatively, Barabási-Albert graphs show more variation, particularly in r_p.

4.5.6 Evaluation of Initial Condition Sensitivity

Finally, each of the plots presented in Figures 4.5 and 4.4 were seeded at random from a different seed node. Yet, despite the seeding method or criteria, for a given graph, the system will converge to the same r^* value. Furthermore, the thresholds are consistent across each set of initial conditions.

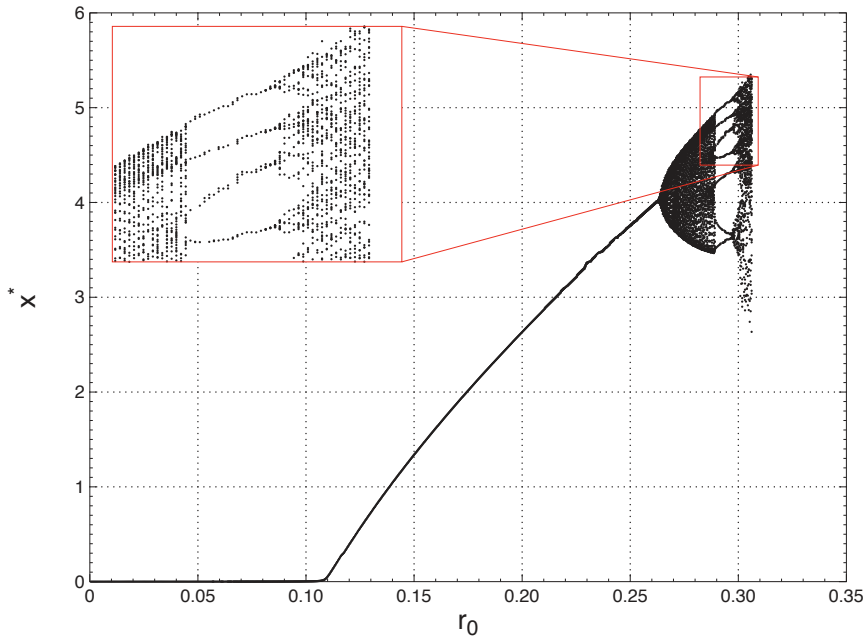


Figure 4.8: Interesting results from Barabási-Albert plot. Inset shows that, like the classic logistic map, at certain r_0 values the fixed point values x^* will "re-converge" to a periodic fixed point.

4.6 Analysis

The system is represented as a vector of node states $\vec{x}(t) = [x_0(t), x_1(t), \cdots, x_{N-1}(t)]^\intercal$ at time t. Our system-level propagation function is encoded as the iterate map $F_{\mathbf{A}} : \mathbb{R}^N \to \mathbb{R}^N$. That is, $F_{\mathbf{A}}$ maps the vector of node states $\vec{x}(t)$ to $\vec{x}(t+1)$.

In this section, we examine the aggregate behavior of system states, that is $s(t)$ as calculated by the aggregation function $H : \mathbb{R}^N \to \mathbb{R}$. Recall from Section 4.3, we define the propagation function $F_{\mathbf{A}}$ as the networked logistic map and the aggregation function H as $\sum_i x_i$.

Before we begin analysis of the existence threshold on the two NLM models, we

must first review some relevant mathematical concepts. The following proofs rely on element-wise multiplication of matrices, referred to as the Hadamard product of matrices.

Definition 17. Hadamard Matrix Multiplication

Given two matrices $\mathbf{A}, \mathbf{B}$*, the Hadamard product, denoted* $\mathbf{A} \circ \mathbf{B}$*, is equal to* $[a_{ij} b_{ij}]$*, where* a_{ij}, b_{ij} *denote the* i, j^{th} *element of the matrices* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively.*

4.6.1 Existence Threshold

Existence Threshold Condition

In this section, I determine the Existence Threshold, r_e, of the iNLM model as described in Equation 4.2. To simplify the determining the iNLM fixed points and stability criteria, I formulate the following system-wide iNLM model, $F_{\mathbf{A}}$, as follows.

$$\mathbf{F_A}(\vec{x}) = \begin{bmatrix} f_0(\vec{x}) \\ f_1(\vec{x}) \\ \vdots \\ f_{N-1}(\vec{x}) \end{bmatrix} = \vec{x}(t+1) = \left(\left(r_0 \mathbf{A} \vec{x}^* \vec{1}^\intercal \right) \circ \left(\mathbf{I} - \mathbf{A} \vec{x}^* \vec{1}^\intercal \right) \right) \vec{1} \qquad (4.7)$$

where $f_i(\vec{x})$ is the node-level iNLM function.

Now, with Equation 4.7 in mind, I will determine r_e for iNLM.

Theorem 18. *Given the global propagation function for the iNLM system* $\mathbf{F}_{\mathbf{A}}(\vec{x}(t))$ *(Eq. 4.7), there exists two stable fixed point* $\vec{x}^*$*, such that*

$$\vec{x}^* - \frac{r - \mathbf{A}^{-1}\vec{1}}{r} \text{ and } \vec{x}^* = \vec{0} \tag{4.8}$$

Proof.

Outline of Proof. First, determine x^* by solving $\mathbf{F}_{\mathbf{A}}(x^*) - x^* = 0$, similar to the proof for Lemma 14.

$$\begin{aligned}
\vec{0} &= \mathbf{F}_{\mathbf{A}}(\vec{x}^*) - \vec{x}^* \\
&= \left(\left(r_0 \mathbf{A}\vec{x}^*\vec{1}^\intercal\right) \circ \left(\mathbf{I} - \mathbf{A}\vec{x}^*\vec{1}^\intercal\right)\right)\vec{1} - \vec{x}^* \\
&= \left(\left(r_0 \mathbf{A}\vec{x}^*\vec{1}^\intercal\right) \circ \left(\mathbf{I} - \mathbf{A}\vec{x}^*\vec{1}^\intercal\right)\right)\vec{1} - (\vec{x}^* \circ \mathbf{I})\vec{1} \\
\vec{0}\vec{1}^\intercal = \mathbf{0} &= \left(\left(r_0 \mathbf{A}\vec{x}^*\vec{1}^\intercal\right) \circ \left(\mathbf{I} - \mathbf{A}\vec{x}^*\vec{1}^\intercal\right)\right) - (\vec{x}^* \circ \mathbf{I}) \\
&= r_0 \mathbf{A}\vec{x}^*\vec{1}^\intercal - r_0(\mathbf{A}\vec{x}^*\vec{1}^\intercal)^{\circ 2} - (\vec{x}^* \circ \mathbf{I}) \\
&= r_0 \mathbf{A}\vec{x}^*\vec{1}^\intercal - r_0 \mathbf{A}(\vec{x}^*)^{\circ 2}\vec{1}^\intercal - (\vec{x}^* \circ \mathbf{I}) \\
&= r_0 \mathbf{A}\vec{x}^*\vec{1}^\intercal - r_0 \mathbf{A}(\mathbf{I} \circ \vec{x}^*)^2\vec{1}\vec{1}^\intercal - (\vec{x}^* \circ \mathbf{I})
\end{aligned}$$

$$
\begin{aligned}
&= r_0\mathbf{A}\vec{x^*}\vec{1}^\intercal - r_0\mathbf{A}(\mathbf{I}\circ\vec{x^*})^2\mathbf{I} - (\vec{x^*}\circ\mathbf{I}) \\
&= r_0\mathbf{A}\vec{x^*}\vec{1}^\intercal - r_0\mathbf{A}(\mathbf{I}\circ\vec{x^*})^2\mathbf{I} - (\vec{x^*}\circ\mathbf{I}) \\
&= r_0\mathbf{A}(\vec{x}\circ\mathbf{I})\vec{1}\vec{1}^\intercal - r_0\mathbf{A}(\mathbf{I}\circ\vec{x^*})^2\mathbf{I} - (\vec{x^*}\circ\mathbf{I}) \\
&= r_0\mathbf{A}(\vec{x}\circ\mathbf{I})\mathbf{I} - r_0\mathbf{A}(\mathbf{I}\circ\vec{x^*})^2\mathbf{I} - (\vec{x^*}\circ\mathbf{I}) \\
&= r_0\mathbf{A}(\vec{x}\circ\mathbf{I}) - r_0\mathbf{A}(\mathbf{I}\circ\vec{x^*})^2 - (\vec{x^*}\circ\mathbf{I}) \\
\mathbf{0} &= (\vec{x^*}\circ\mathbf{I})\left(r_0\mathbf{A} - r_0\mathbf{A}(\vec{x^*}\circ\mathbf{I}) - \mathbf{I}\right)
\end{aligned}
$$

At this point, two solutions present themselves, namely, $\vec{x^*}\circ\mathbf{I} = \mathbf{0}$ and $r_0\mathbf{A} - r_0\mathbf{A}(\vec{x^*} - \mathbf{I}) - \mathbf{I} = \mathbf{0}$. Solving the trivial solution, one finds $\vec{x^*} = \vec{0}$. Now, let's focus on the second solution.

$$
\begin{aligned}
\mathbf{0} &= r_0\mathbf{A} - r_0\mathbf{A}(\vec{x^*}\circ\mathbf{I}) - \mathbf{I} \\
\mathbf{0}\vec{1} &= \left(r_0\mathbf{A} - r_0\mathbf{A}(\vec{x^*}\circ\mathbf{I}) - \mathbf{I}\right)\vec{1} \\
\vec{0} &= r_0\mathbf{A}\vec{1} - r_0\mathbf{A}(\vec{x^*}\circ\mathbf{I})\vec{1} - \mathbf{I}\vec{1} \\
&= r_0\mathbf{A}\vec{1} - r_0\mathbf{A}\vec{x^*} - \vec{1}
\end{aligned}
$$

Immediately, we find that we can rearrange this equation, yielding

$$
\vec{x^*} = \left(\frac{r_0 - \mathbf{A}^{-1}}{r_0}\right)\vec{1} \tag{4.9}
$$

Note that the term $\frac{r_0 - \mathbf{A}^{-1}}{r_0}$ is similar to the solution to the classic logistic map given in §4.3.3.

The next step is to determine the range of r_0 for which the solutions above are valid. Recall from the Lemma 14 that to determine whether or not a fixed point was stable one simply determine if $f'(x^*) < 1$. A complimentary theorem exists for multivariate systems.

Theorem 19. *A fixed point x^* is stable if and only if $|\rho(\mathbf{J}(\mathbf{F}_\mathbf{A}(\tilde{\mathbf{x}}^*)))| = |\rho(\nabla \mathbf{F}_\mathbf{A}(\vec{x}^*))| < 1$, where the ρ is the eigenvalue function and $\mathbf{J}$ is the Jacobian matrix function.*

The Jacobian matrix is the multivariate analog of the derivative f', where $\mathbf{J}$ is defined as follows:

$$\mathbf{J} = \nabla \mathbf{F}_\mathbf{A}(\vec{x}) = \begin{bmatrix} \frac{\partial x_0(t+1)}{\partial x_0(t)} & \frac{\partial x_1(t+1)}{\partial x_0(t)} & \cdots & \frac{\partial x_{N-1}(t+1)}{\partial x_0(t)} \\ \frac{\partial x_0(t+1)}{\partial x_1(t)} & \frac{\partial x_1(t+1)}{\partial x_1(t)} & \cdots & \frac{\partial x_{N-1}(t+1)}{\partial x_1(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_0(t+1)}{\partial x_{N-1}(t)} & \frac{\partial x_1(t+1)}{\partial x_{N-1}(t)} & \cdots & \frac{\partial x_{N-1}(t+1)}{\partial x_{N-1}(t)} \end{bmatrix} \tag{4.10}$$

Let ρ[5] for more detail. be a function that returns all eigenvalues of a particular matrix. In order for $\vec{x}^*$ to be stable, all eigenvalues of $|\rho(\nabla \mathbf{F}_\mathbf{A}(\vec{x}^*))| < 1$. So, we must determine what values of r_0 for a given adjacency matrix $\mathbf{A}$ satisfy $|\rho(\nabla \mathbf{F}_\mathbf{A}(\vec{x}^*))| < 1$.

Now, by substituting the fixed point solution expressed in Eq. 4.9 into $|\rho(\nabla \mathbf{F}_\mathbf{A}(\vec{x}^*))| <$

[5]Note that the Perron-Frobenius theorem guarantees $\lambda_{\mathbf{A},1} > \lambda_{\mathbf{A},2} > \ldots \geq \lambda_{\mathbf{A},N}$. See [53]

1, we find

$$|\rho(\nabla \mathbf{F}_{\mathbf{A}}(\vec{x}^*))| < 1$$

$$|\rho(\mathbf{AC})| < 1$$

where $\mathbf{C}$ is convenient substitution where $\mathbf{C}_{ii} = \mathbf{J}_{ii}$ and $\mathbf{C}_{ij} = 0$ for all i, j. Ultimately, due to this decomposition, each non-zero element of $\mathbf{C} = 2 - r_0$, thus the eigenvalue of $\mathbf{C} = 2 - r_0$. So, finally, we find So solving for the equation above, we find

$$r_0 \times \lambda_{\mathbf{A},1} > 1$$

and

$$r_0 \times \lambda_{\mathbf{A},1} < 3$$

4.7 Related Works

In this section, we briefly cover works related to our own.

4.7.1 Information Propagation

At its heart, the networked logistic map is a means of information propagation, analogous to disease (or epidemic) spreading processes. For a gentle introduction to this subject, refer to the survey by Hethcote [25]. Early propagation models

assumed *homogeneous* populations; that is, a population with no social or spatial structure. Newer models apply underlying structure to the population, thus creating a *heterogeneous* population. We too apply a network structure to our model through the use of generative graph models, such as Erdős-Rényi or Barabási-Albert . The key difference between our work and such disease spreading processes is our introduction of a continuous valued measure of node state.

4.7.2 Applications

The study of dynamical processes on networks applies to a wide variety of diverse scientific fields, ranging from computer science and engineering to neuroscience to finance. For a comprehensive survey, refer to [1] or [4].

The area of distributed systems benefits from deeper understanding of the behavior of dynamical systems operating on networks. Consensus algorithms often require the distributed system to achieve a global state (signifying a consensus among the population). Understanding the effect of networks on such dynamical processes will result in better performance of such systems. Global coordination algorithms (such as routing algorithms) will also benefit from a deeper understanding of dynamical processes on networks.

The efficiency gains promised by a deeper understanding will fall into one of two general categories: 1) Modifying dynamical system to exploit a fixed structure, or 2) Modifying structure to benefit a dynamical system.

The area of data mining often exploits relations between data to effect efficient extraction of knowledge from large data sets. Thus, it is another natural area that will benefit from the understanding of dynamical systems and their relation to graph structures.

4.8 Summary

In the preceding study, we have effectively de-coupled propagation model from connectivity structure for a deterministic, non-binary propagation function.

Using a variety of graph models, I demonstrated the following series of results:

1. **Novel Propagation Models** Based on the classic logistic map function, I developed two networked propagation models, eNLM and iNLM.

2. **Existence Threshold,** r_e. For each of the NLM models, I demonstrated through empirical analysis that the existence threshold $r_e = \frac{1}{\lambda_{\mathbf{A},1}}$. Furthermore, I analyzed the models and confirmed the empirical results. These results are analyzed in §4.6.1 and numerical simulation results are given in §4.5.3.

3. **Dampening Threshold,** r_d. Empirically, I showed that the NLM models exhibit a dampening behavior similar to the classic logistic map. Additionally, I demonstrated through simulations that $r_d \approx \frac{2}{\lambda_{\mathbf{A},1}}$ (§4.5.4).

4. **Periodic Threshold,** r_p. I showed that there exists a periodic threshold for

the two NLMs presented in this study. Extending the proof for the existence threshold in §4.5.3, I showed that $r_p < \frac{3}{\lambda_{\mathbf{A},1}}$. Furthermore, I demonstrated empirically that the more uniform the degree distribution of $\mathbf{A}$, the closer r_p will be to $\frac{3}{\lambda_{\mathbf{A},1}}$.

5. **Initial Condition Insensitivity**. Finally, through a series of empirical results, I demonstrated that, given an network topology $\mathbf{A}$, the system will converge on the same set of x^*'s for all valid initial conditions.

Chapter 5

Conclusion

Each chapter of this dissertation addresses a specific problem in the area of spreading processes on complex network structures. First, I introduced epidemic spreading processes on dynamic network structures, modeling short-range, point-to-point communication between mobile devices. The key finding was that the tipping point–a.k.a. the epidemic threshold–is intimately related to the first eigenvalue of the connectivity matrix $\lambda_{\mathbf{A}}$. This was followed by an examination competing information (or viruses) on composite networks, and again, $\lambda_{\mathbf{A}}$ predicted not only the epidemic threshold, but also the winner of a given competition. Finally, I introduced a model of non-binary information propagation based on the classic logistic map. Through a rigorous analysis and numerical results, I again found an close relationship between various phase transitions and $\lambda_{\mathbf{A}}$.

At the beginning of this thesis, I asked what affect does topology have on spreading

processes on networks. The results presented herein, as well as other results from different authors, clearly show that $\lambda_{\mathbf{A}}$ is the key measure of topology affecting the spreading processes evaluated herein.

Bibliography

[1] Linda J.S. Allen and P. van den Driessche. The basic reproduction number in some discrete-time epidemic models. *Journal of Difference Equations and Applications*, 2008.

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

[3] Roy M. Anderson and Robert M. May. *Infectious diseases of humans: Dynamics and control*. Oxford Press, 2002.

[4] Alex Arenas, Albert Díaz-Guilera, Jurgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93 – 153, 2008.

[5] James Aspnes, Navin Rustagi, and Jared Saia. Worm versus alert: Who wins in a battle for control of a large-scale network? In *OPODIS*. 2007.

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

[7] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 15 October 1999.

[8] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 1999.

[9] Christopher L. Barrett, Keith R. Bisset, Stephen G. Eubank, Xizhou Feng, and Madhav V. Marathe. Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. *ACM/IEEE SC 2008*.

[10] Christian Betsetter. The Node Distribution of the Random Waypoint Mobility model for Wireless Ad Hoc Networks. *IEEE Transactions on Mobile Computingt*, 2:257–269, 2003.

[11] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the Spread of Misinformation in Social Networks. In *WWW*, 2011.

[12] Sergey V. Buldyrev, Roni Parshani, Gerald Paul, H. Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 2010.

[13] T. Camp, J. Boleng, , and V. Davies. A Survey of Mobility Models for Ad Hoc Network Research. In *Wireless Communication and Mobile Computing Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, pages 483–502, 2002.

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

[15] Deepayan Chakrabarti, Jure Leskovec, Christos Faloutsos, Samuel Madden, Carlos Guestrin, and Michalis Faloutsos. Information survival threshold in sensor and p2p networks. In *Proceedings of 26th Annual IEEE ICC*, 2007.

[16] Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 91(24), December 2003.

[17] Pedro Domingos and Matthew Richardson. Mining the network value of customers. In *KDD*, pages 57–66, 2001.

[18] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *SIGCOMM*, pages 251–262, Aug-Sept. 1999.

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

[20] A. Ganesh, L. Massoulie, and D. Towsley. The Effect of Network Topology in Spread of Epidemics. In *IEEE INFOCOM*, 2005.

[21] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 2001.

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

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

[24] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Rev.*, 42(4):599–653, 2000.

[25] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Rev.*, 42(4):599–653, 2000. http://www.math.rutgers.edu/ leenheer/hethcote.pdf.

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

[27] David B. Johnson and David A. Maltz. Dynamic Source Routing in Ad Hoc Wireless Networks. In *Mobile Computing*, 1996.

[28] J. O. Kephart and S. R. White. Directed-Graph Epidemiological Models of Computer Viruses. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359, May 1991.

[29] J. O. Kephart and S. R. White. Directed-Graph Epidemiological Models of Computer Viruses. *IEEE S&P*, 1991.

[30] Jeffrey O Kephart and Steve R White. Directed-graph epidemiological models of computer viruses. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359, May 1991.

[31] Jeffrey O Kephart and Steve R White. Measuring and modeling computer virus prevalence. In *Proceedings of the 1993 IEEE Computer Society Symposium on Research in Security and Privacy*, pages 2–15, May 1993.

[32] Jan Kostka, Yvonne Oswald, and Roger Wattenhofer. Word of mouth: Rumor dissemination in social networks. In *SIROCCO*, 2008.

[33] S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Trawling the web for emerging cyber-communities. *Computer Networks*, 31(11-16):1481–1493, 1999.

[34] A. G. McKendrick. Applications of mathematics to medical problems. *Edin. Math. Society*, 1926.

[35] Dominic Meier, Yvonne Anne Oswald, Stefan Schmid, and Roger Wattenhofer. On the windfall of friendship: inoculation strategies on social networks. In *ACM EC*, 2008.

[36] James W. Mickens and Brian D. Noble. Modeling epidemic spreading in mobile environments. In *WiSe '05: Proceedings of the 4th ACM workshop on Wireless security*, pages 77–86, New York, NY, USA, 2005. ACM.

[37] James W. Mickens and Brian D. Noble. Analytical models for epidemics in mobile networks. *Wireless and Mobile Computing, Networking and Communication, IEEE International Conference on*, 0:77, 2007.

[38] Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B*, 26:521–529, 4 February 2002.

[39] M. E. J. Newman. Threshold effects for two pathogens spreading on a network. *Physical Review Letters*, 2005.

[40] M. E. J. Newman. Threshold effects for two pathogens spreading on a network. *Physical Review Letters*, 2005.

[41] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Lettes*, 86(14), 2001.

[42] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E*, 65, 2002.

[43] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.

[44] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65:035108, 2002.

[45] B. A. Prakash, A Beutel, R Rosenfeld, and C Faloutsos. Winner-takes-all: Competing Viruses on Fair-play Networks. In *WWW*, 2012.

[46] B. Aditya Prakash, Deepayan Chakrabarti, Michalis Faloutsos, Nicholas Valler, and Christos Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. *IEEE ICDM*, 2011.

[47] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, and Song Chong. On the levy-walk nature of human mobility. In *INFOCOM*, pages 924–932. IEEE, 2008.

[48] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal*, 6(1), 2002.

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

[50] Alessandra Sala, Lili Cao, Christo Wilson, Robert Zablit, Haitao Zheng, and Ben Y. Zhao. Measurement-calibrated graph models for social network experiments. In *WWW*, 2010.

[51] M. F. Shlesinger, J. Klafter, and Y. M. Wong. Random walks with infinite spatial and temporal moments. In *J. Stat. Phys*, volume 27, pages 499– 512, 1982.

[52] S. Staniford, V. Paxson, and N. Weaver. How to 0wn the Internet in Your Spare Time. In *USENIX Security Symposium*, 2002.

[53] Steven H. Strogatz. *Nonlinear Dynamics and Chaos*. Westview Press, 1994.

[54] R. W Thommes and M. J. Coates. Epidemiological Modelling of Peer-to-Peer Viruses and Pollution. In *IEEE INFOCOM*, 2006.

[55] Hanghang Tong, B. Aditya Prakash, Charalampos E. Tsourakakis, Tina Eliassi-Rad, Christos Faloutsos, and Duen Horng Chau. On the vulnerability of large graphs. In *IEEE ICDM*, 2010.

[56] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. Virus spread in networks. *IEEE/ACM Trans. Netw.*, 17(1):1–14, 2009.

[57] Y. Wang, M. Iliofotou, M. Faloutsos, and B. Wu. Analyzing Interaction Communication Networks in Enterprises and Identifying Hierarchies. In *Int. Wksp on Network Science*, 2011.

[58] Yang Wang, D. Chakrabarti, Chenxi Wang, and C. Faloutsos. Epidemic spreading in real networks: an eigenvalue viewpoint. In *IEEE SRDS*, 2003.

[59] Guanhua Yan, Hector D. Flores, Leticia Cuellar, Nicolas Hengartner, Stephan Eidenbenz, and Vincent Vu. Bluetooth worm propagation: mobility pattern matters! In *ASIACCS '07: Proceedings of the 2nd ACM symposium on Information, computer and communications security*, pages 32–44, New York, NY, USA, 2007. ACM.

[60] J. Yoon, M. Liu, and B. Noble. Random waypoint considered harmful. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 2, pages 1312 – 1321 vol.2, 30 2003.