Hybrid epidemic spreading - from Internet worms to HIV infection

Changwang Zhang

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computer Science; Department of Security and Crime Science

University College London

May 13, 2015

I, Changwang Zhang, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Epidemic phenomena are ubiquitous, ranging from infectious diseases, computer viruses, to information dissemination. Epidemics have traditionally been studied as a single spreading process, either in a fully mixed population or on a network. Many epidemics, however, are hybrid, employing more than one spreading mechanism. For example, the Internet worm Conficker spreads locally targeting neighbouring computers in local networks as well as globally by randomly probing any computer on the Internet.

This thesis aims to investigate fundamental questions, such as whether a mix of spreading mechanisms gives hybrid epidemics any advantage, and what are the implications for promoting or mitigating such epidemics. We firstly propose a general and simple framework to model hybrid epidemics. Based on theoretical analysis and numerical simulations, we show that properties of hybrid epidemics are critically determined by a hybrid tradeoff, which defines the proportion of resource allocated to local and global spreading mechanisms. We then study two distinct examples of hybrid epidemics: the Internet worm Conficker and the Human Immunodeficiency Virus (HIV) infection within the human body. Using Internet measurement data, we reveal how Conficker combines ineffective spreading mechanisms to produce a serious outbreak on the Internet. We propose a mathematical model that can accurately recapitulate the entire HIV infection course as observed in clinical trials. Our study provides novel insights into the two parallel infection channels of HIV, i.e. cell-free spreading and cell-to-cell spreading, and their joint effect on HIV infection and treatment.

In summary, this thesis has advanced our understanding of hybrid epidemics. It has provided mathematical frameworks for future analysis. It has demonstrated, with two successful case studies, that such research can have a significant impact on important issues such as cyberspace security and human health.

Acknowledgements

First and foremost, I thank both of my supervisors, Dr. Shi Zhou, and Prof. Benjamin Chain. This multi-disciplinary study would not have been possible without their continuous support and immense knowledge input.

My sincere thanks go to Dr. Hervé Borrion, Prof. Nick Tilley, Prof. Gloria Laycock, Prof. Richard Wortley, Mr. Vaseem Khan, and Ms. Kati Carter from the Security and Crime Science Department, for their indispensable support for my study.

I thank the Facilities and Operations Manager of the Computer Science Department, Ms. JJ Giwa, for providing me with the space required to conduct this research.

I would like to thank Prof. Steven Bishop and Dr. Timothy Evans for their valuable time and efforts spent in examining my PhD thesis and viva.

I thank Dr. Clare Jolly from Infection & Immunity Division for informative discussions about the cell-to-cell infection of Human Immunodeficiency Virus (HIV), without which the case study on HIV infection would not have been possible.

I am indebted to my friend Dr. Justin Kurland who help proof the thesis.

Last but not least, I would like to thank my family for always believing in me and helping to keep my spirits up.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Epidemic spreading phenomena are frequently found in the real world, and relevant studies have important practical values. Hybrid epidemics refer to those epidemics that employ more than one spreading mechanism. The motivation of this study is to better understand hybrid epidemics. The study is timely and important as 1) hybrid epidemics pose a great threat to the internet security and human health; but 2) some fundamental problems of hybrid epidemics are still not well understood, including whether and when a mix of spreading mechanisms gives hybrid epidemics any advantage and what are the implications for promoting or mitigating such epidemics.

1.1.1 Ubiquitous epidemic phenomena

Epidemic spreading phenomena exist in a wide range of domains [1–3]. Well-known examples include the spread of infectious diseases, the proliferation of computer viruses, and information propagation. Despite their distinct nature and the environments in which they spread, all these phenomena can be abstracted as packages of information (e.g. a virus) spreading through networks (e.g. computer networks) [1, 4].

Studies on the spread of epidemics not only help understand existing spreading phenomena, but also increasingly provide important insights for designing new systems to solve practical problems. For example, viral marketing promotes the spread of advertisements through social networks [5–7], and epidemic routing aims to maintain and find a path to send information from one device to another [8–10]. While other examples include epidemic-based replicated database which is a distributed system that

stores multiple copies of data in a large number of machines [11–13], and epidemic-based information dissemination [14–16].

There is a long history of research on the spread of epidemics. Various models have been proposed to analyse the phenomenon [3, 17]. Findings from this stream of research help us better understand the dynamics involved with the spread of an epidemic in the real-world. The results of the studies have important practical values for controlling epidemics, either for mitigating negative spreading, such as computer worms, or promoting positive spreading, such as useful information propagation.

1.1.2 Hybrid and critically hybrid epidemics

Various spreading mechanisms have been studied [3,18]. The two most common mechanisms are *local spreading*, where infected nodes only infect a limited subset of target nodes [19]; and *global spreading*, where nodes are fully-mixed such that an infected node can infect any other node [3,20]. For example, local spreading of infectious organisms occurs by physical contact between infected and susceptible hosts. The pattern of epidemic spread is therefore determined by the pattern of physical interaction between susceptible and infected individuals. But infections can also spread globally, for example via infectious agents that can travel great distances, and then randomly infect any individual within a large target population.

In reality, many epidemics use *hybrid spreading*, which involves a combination of two or more spreading mechanisms. For example the computer worms Conficker [21] and Code-Red [22] try to infect both computers in a local network (*local spreading*) and any randomly chosen computers on the Internet (*global spreading*).

Most previous studies on hybrid epidemics have focussed on what we call *non-critically* hybrid epidemics where a combination of multiple mechanisms is not a necessary condition for an outbreak. In this case, using a fixed total spreading power, a hybrid epidemic will always be less infectious than an epidemic using only the more infectious one of the two spreading mechanisms [23, 24]. However, many real world examples suggest the existence of *critically* hybrid epidemics where a combination of spreading mechanisms may be more infectious than using only one mechanism.

We aim to tackle fundamental but hitherto little investigated questions about hybrid epidemics: whether a mix of spreading mechanisms gives hybrid epidemics any

advantage and what are the implications for promoting or mitigating such epidemics. This will provide insight into designing new and more efficient information dissemination systems.

1.1.3 Two specific examples of epidemic threats to the Internet and health

Large-scale epidemics can cause huge amounts of societal and economic damage. Many hybrid epidemics are especially dangerous. In this thesis we focus on two examples of such epidemic processes including the computer worm Conficker spreading on the Internet and HIV infection within human body, both of which are very difficult to throttle and can last for years and cause enduring damage [21, 25]. There is an urgent need to understand why these hybrid epidemics can be so infectious and so resistant to treatments.

A further aim of this thesis is to investigate how hybrid spreading contributes to an epidemic's infectiousness and resistance to mitigation strategies. So that, evidence gleaned from this thesis can be utilised practically to inform future strategies and interventions that help thwart damaging hybrid epidemics.

1.1.3.1 Computer worm Conficker spread on the Internet

Computer worms are malicious applications that automatically propagate themselves through computer networks. Conficker is one of the most contagious and "smart" computer worms on record [26]. According to analysis from the computer security company - Symantec, the worm employs various spreading mechanisms including local, neighbourhood, and global spreading [26]. Conficker exploded into a full-scale Internet epidemic on 21 November 2008 and has since infected over seven million computers [21]. It was believed at the time that the worm had infected so many computers that it could threaten the critical infrastructures of the Internet [27].

To thwart the worm, the Conficker Working Group [27] coordinated an unprecedented collaboration that consisted of major organisations in the cybersecurity community and beyond, including Microsoft, Internet Corporation for Assigned Names and Numbers (ICANN), Facebook, Cisco, McAfee, Symantec, F-Secure, and so on. The working group successfully prevented the worm infected computers from being exploited for causing damage, but still failed to eliminate the worm [27]. The worm is

still active on the Internet [28], six years after its outbreak, continuing to threaten the security of the Internet.

1.1.3.2 HIV infection within human body

Human Immunodeficiency Virus (HIV) attacks the human immune system that protects the body against diseases. The HIV infection within human body is similar to the spread of computer worms in that it is a hybrid epidemic. It spreads among T cells through two distinctive mechanisms: cell-free (*global*) spreading and cell-to-cell (*local*) spreading.

Tremendous efforts have been spent over the course of the past three decades to better understand and treat the virus. Indeed, presently state-of-the-art HIV treatments can manage and control the virus's infection, but there is still no permanent cure for the disease to date [25]. The sheer volume of people infected with the disease globally - 34 million [25] - is a compelling call for further understanding and better treatments of the disease.

We note that the HIV infection *within* human body is different from the infectious disease spreading in a population [1, 29] although they are both studied as epidemics.

1.2 Challenges

To study hybrid epidemics, especially critically hybrid epidemics, we conduct mathematical modelling, numerical simulations, and real data set analysis. There are remarkable research challenges corresponding to each of the three methods. An additional complexity of the study is the requirement to understand two real epidemics from two distinctive disciplines: computer worms and HIV infection.

1.2.1 Mathematical modelling of hybrid epidemics

Mathematical modelling can play an important role in helping understand the dynamics of epidemic spreading [17]. Most existing epidemic models only consider one spreading mechanism, e.g global spreading in a fully mixed population [30–33] or local spreading among connected individuals [19, 34]. Hybrid epidemics, however, consist of multiple spreading mechanisms. The modelling of hybrid epidemics therefore needs to also address the multiplicative effect of different spreading mechanisms.

Another challenge to modelling is best determining a mathematical model that captures the major dynamics of hybrid epidemics yet remains simple enough to provide

analytical insight and make fast predictions. It is possible that more complex mathematical models can capture every detail of a hybrid epidemic. However, a drawback of such an approach is that they can be overspecified, i.e. requiring so many parameters that some of them cannot be estimated from measurement data. In addition, mathematical models that are highly detailed can require tremendous computational power to analyse epidemic dynamics that is often accomplished using intensive calculations or simulations.

At the same time, a good model does need a certain level of complexity to faithfully capture the major dynamics of hybrid epidemics. Overly simplistic models often require strong assumptions that are violated in reality. A consequence of such models is that they may fail to show the true dynamics of an epidemic. For example, reference [23] found that hybrid spreading can never result in a larger scale outbreak than at least one of the two single-mechanism based cases - either completely local spreading or completely global spreading - when using the same spreading power. A potential problem with this finding is that it implicitly assumed a well connected network for the hybrid epidemic to spread on. In reality, hybrid epidemics can often spread on many weakly connected network patches, i.e. a metapopulation. We test this explicitly, finding that, given the same spreading power, a mixture of spreading mechanisms can be notably more infectious than using only one mechanism. In order to theoretically explain and reproduce this finding, we need to propose a new mathematical model that relies on fewer assumptions and incorporates enough details of hybrid spreading mechanisms.

1.2.2 Processing large, sparse, and incomplete data

Epidemic spreading data can be very large. For example, the spreading dataset for computer worm Conficker [35] collected by the network telescope [36] of the Center for Applied Internet Data Analysis (CAIDA) includes 68 files and has a total size of 164 Giga Bytes (GB), i.e. over $16k$ times larger than this thesis. It includes $430k$ infected nodes located in $92k$ sub-networks. To efficiently analyse such a large dataset, we need to fully utilise computing clusters to conduct data analysis in parallel. Chapter 6 analyses this dataset.

Due to practical issues related to measurements, epidemic spreading data often

includes sparse and incomplete records. HIV measurements, which are analysed in Chapter 7, for example, often start months after the initial HIV infection in many patients. Consequently, the initial spreading period of the epidemic is not included in the data. In addition, the data measurement stops and medical treatments start for a patient when he/she shows symptoms of disease progression, i.e. the late period of HIV spreading is also absent. During the period with measurements, the data points are not only sparse for many patients but also are unevenly distributed temporally for some patients. These shortcomings in the data present a challenge to the analysis and data-based model parameter inference in this study.

1.2.3 Large-scale complex epidemic simulations

Simulations are an indispensable part of epidemic spreading research. Numerical simulations are conducted to validate the results from theoretical modelling and analysis. The challenge in epidemic spreading simulations is that they can be very computationally intensive and time-consuming, particularly for large-scale epidemics. Worse still, a simulation with the same set of parameters often needs to be repeated thousands of times to obtain an averaged result, in order to mitigate the stochastic effects that arise in each run of a simulation.

When it comes to hybrid epidemics, multiple spreading mechanisms should be simulated. For example, the computer worm Conficker employs three distinctive spreading mechanisms: (1) local spreading among computers in a same local area network, (2) neighbourhood spreading among computers in adjacent area networks, and (3) global spreading among all computers on the Internet [26]. In addition, networks underlying the local spreading of hybrid epidemics can also be more complicated. Most epidemic studies simulate epidemics spreading on one, two, or several networks [3, 19, 37–40]. The worm, Conficker, however, spreads on more than $92k$ local area networks.

The simulation scenario of hybrid epidemics is exceedingly more complex than traditional single-mechanism-based epidemics. Thus, efficiently simulating how these particular epidemics spread is a challenging task.

1.2.4 Interdisciplinary research encompassing computer worms and HIV infection

This thesis conducts case studies on two real-world epidemics: the computer worm Conficker [21] spreading on the Internet and HIV infection within the human body [41]. In order to better understand the two real epidemics, it is essential to develop mathematical models that can describe and predict their spreading dynamics.

Computer worm Conficker is an epidemic on the Internet. By analysing the program of Conficker, the computer security company Symantec reveals that the worm uses a sophisticated spreading strategy to achieve two objectives: (1) to infect the maximum number of computers and (2) to simultaneously prevent itself from being detected on infected computers [26]. However, existing modelling research on Conficker [42, 43] has only considered one spreading mechanism, thus failing to represent the true spreading dynamics of the worm. The worm's sophisticated spreading strategy needs to be understood and properly abstracted in a new mathematical model.

The case of modelling the HIV infection within human body is even more complex. An HIV infection model not only has to describe the viral infection, but also needs to take into account the effects of the immune system and clinical treatments. There are complex interactions between the immune system and the virus. The immune system tries to kill the virus and the virus in return degrades the functionality of the immune system [44].

Additional difficulties lie in an incomplete understanding of HIV pathogenesis [25]. Despite three decades of intensive research on the disease, mathematical models still often stumble over questions related to the quantitative behaviours of the virus because they have yet to be clinically investigated. For example, there is still limited understanding of the HIV latent reservoir, which is the principle obstacle related to the eradication of the disease. As a result, HIV infection models often need to include some reasonable assumptions and speculation about the behaviour of the virus.

1.3 Contributions

The thesis has contributed to push forward our understanding of hybrid epidemics. In particular, it has showed and explained the existence of the critically hybrid epidemics, where limited spreading mechanisms complement each other and can be combined to

make a highly infectious epidemic. Two real world examples of such epidemics have been analysed in detail, and the results have important implications for the Internet security and human health. Here we provide a brief summary of our contributions in this thesis.

1.3.1 Model and explanation of critically hybrid epidemics

This thesis presents a novel theoretical hybrid epidemic model that incorporates both local spreading among directly connected individuals (nodes) and global spreading among all individuals. The model provides a unified approach for analysing hybrid epidemics in both a single population and a metapopulation consisting of many subpopulations.

Specifically the model incorporates a hybrid tradeoff that represents the strategy of allocating fixed spreading power between two spreading mechanisms. The model enables the interaction and the joint impact of the two spreading mechanisms on epidemic dynamics, ranging from a completely local spreading scenario to a completely global spreading scenario, to be examined systematically.

This hybrid epidemic model also provides the foundation to explain the critically hybrid epidemic phenomena. And the predictions of the model agree well with stochastic simulation results.

1.3.2 The crucial role of tradeoff between spreading mechanisms for hybrid spreading optimisation

Results from this thesis demonstrate that it is indeed possible to have a highly contagious epidemic by mixing simple, ineffective spreading mechanisms. The properties of such epidemics are critically determined by how the fixed spreading power is allocated to two different spreading mechanisms, and usually there is an optimal allocation (i.e. tradeoff between two mechanisms) that leads to a maximal outbreak size. We provide a method to help predict this optimal allocation (tradeoff) base on other known epidemic parameters.

The study is practically oriented in that it can be utilised for designing new and more efficient information dissemination systems whose recipients exist in a metapopulation. For example, a large distributed database, which includes numerous data nodes on the Internet, needs to deliver updates to all of its data nodes. Those data nodes in

local area networks collectively form a metapopulation for the local spreading, while being connected globally through the Internet.

1.3.3 Inference of epidemic parameters and dynamics of computer worm Conficker

Computer worm Conficker is an example of a *critically* hybrid epidemic. This study is the first to model and analyse all three of the worm's spreading mechanisms including local, neighbourhood, and global spreading. By applying the proposed Conficker spread model to the Internet measurement data [35, 45] collected by the Center for Applied Internet Data Analysis (CAIDA), our study inferred the worm's epidemic parameters, including the infection rates of its three different spreading mechanisms, recovery rate, spreading network structure, etc.

The inferred epidemic parameters demonstrate that the infection rate of Conficker's global probing is extremely low with less than one successful infection in every ten million random probes across the Internet. The infection rates of local and neighbourhood probing, however, are around a million times larger than that of global probing. The worm spends around 90% of its time on global probing, which allows it to explore the vast expanse of the Internet, although with a very low infection rate. These inferred results help explain why the worm was able to infect millions of computers and why it remains active on the Internet years after its detection [28].

1.3.4 State-of-the-art mathematical model for HIV infection within human body

This thesis presents a novel HIV infection model that can explain the complex progress of the HIV infection in all its phases and its variable timescale. Such a unified model is important not only to understand HIV infection dynamics, but also to estimate the long term effects of therapeutic strategies on HIV progression.

Notably, the model additionally includes explicit terms for the two modes of virus spread, parameterised from experimental observation. The model output shows excellent agreement to two sets of clinical data from a cohort of HIV infected patients and from the Short Pulse Anti-Retroviral Therapy at Seroconversion trial [46].

1.3.5 Novel insight for developing future HIV treatments

We show that the two modes of spread of HIV infection, cell-to-cell spread and cell-free spread, complement each other and both make important contributions to disease progression. This research is the first to theoretically demonstrate that cell-to-cell spread becomes increasingly effective as the infection progresses and thus may present a considerable treatment barrier for eradicating the disease.

Our predictions emphasise the potential benefits of early treatment, and suggest that drugs with the ability to effectively block cell-to-cell spread may provide significant therapeutic benefit in long term management or eradication of HIV infection.

Reviewers of our journal article relevant to this contribution have suggested that these predictions could provide a major advance in the understanding of HIV dynamics, if confirmed by future clinical trials.

1.3.6 Other contribution - LeoTask

This study involves intensive computation for theoretical predictions and numerical simulations. To facilitate such computation, we developed a Java framework for computation-intensive and time-consuming research tasks. The name of the framework is LeoTask and its source code is freely available at `http://github.com/mleoking/leotask`.

LeoTask has a unique combination of features that are expected to be useful to the broader research community. It automatically and simultaneously executes tasks on multiple CPU cores of a computing facility. It uses a configuration file to enable automatic exploration of parameter space and flexible aggregation of results, and therefore allows researchers to focus on programming the key logic of a computing task. It also supports reliable recovery from interruptions, dynamic and cloneable networks and integration with the plotting software Gnuplot. Appendix C provides more details about this framework.

1.4 Thesis outline

The rest of this thesis is organised as follows. Chapter 2 begins with a briefly literature review of research directions on epidemic spreading and provides an overview of the two most classic epidemic spreading models. Chapter 3 introduces hybrid epidemics

and critically hybrid epidemics, reviews research directly relevant to the study, and provides a brief introduction to two real epidemics: the computer worm Conficker spreading on the Internet and HIV infection spreading within human body.

In Chapter 4, a new mathematical model for hybrid epidemics is proposed and validated. The model incorporates both local and global spreading mechanisms and introduces a tradeoff between them. Chapter 5 uses the model to investigate the interaction and the joint impact of the two spreading mechanisms on epidemic dynamics, and demonstrates that it is indeed possible to have a highly contagious epidemic by mixing simple, ineffective spreading mechanisms.

Chapter 6 analyses the spreading behaviour of a critically hybrid epidemic - the computer worm Conficker. Epidemic parameters for the worm are initially inferred from measurement data, and then intensive simulations are used to highlight the complex interactions among Conficker's three spreading mechanisms. The results demonstrate that utilising multiple spreading mechanisms plays an important role in the worm's huge outbreak and lasting survival.

Chapter 7 presents a novel HIV infection model that can explain the complex progress of the infection in all its phases and its variable timescale. The model is used in Chapter 8 to investigate the HIV infection dynamics and estimate the long term effects of therapeutic strategies on HIV progression. We show that the two modes of spread of the HIV infection complement each other and both make important contributions to disease progression. Chapter 8 also provides novel insight into future HIV treatments.

Chapter 4 provides the analytical solutions for the final outbreak size and the threshold of hybrid epidemics. The case studies in Chapter 6 and Chapter 7 use numerical calculations to investigate the epidemic dynamics. They together provide a useful combination to the literature.

Chapter 9 discusses the limitations of the study and relevant areas of future research. Finally, Chapter 10 sums up and concludes the thesis.

Chapter 2

Epidemic spreading

Epidemic spreading phenomena are ubiquitous in the real world. And some of them are reviewed in [1–3]. Well-known examples include human disease spreading, computer virus proliferation, and information propagation. Despite the distinct nature and environments in which they spread, all these phenomena can be abstracted as packages of information (e.g. a virus) spreading through networks (e.g. computer networks) [1, 4].

2.1 Research directions in literature

The research directions of the studies on epidemic spreading are introduced in this section. In additional, a discussion of hybrid epidemic spreading and existing research that is directly relevant to this study are introduced in Chapter 3 in greater detail.

There is an extensive history of research into the modelling of epidemic spreading. Recent reviews about epidemic spreading studies can be found in [17, 47]. The multidisciplinary approach to studying the problem began with collaborative efforts between biologists and mathematicians. Their early studies assumed that epidemics spread in a fully-mixed population whereby an infected person can randomly spread a disease to any other person [30–33]. We refer to this kind of spreading as *global spreading*. Later on, epidemic spreading was studied more from the perspective of statistical physics. The epidemic is considered to spread in a network where a computer virus or a piece of information spreads through specifically defined connections between computers or people [19, 34]. We refer to this kind of spreading as *local spreading*.

Since the introduction of network based epidemic analysis [19,34], epidemics have been studied in various networks and populations, including structured households [48–51], metapopulations that consist of a number of weakly connected subpopulations [52–

58], clustered networks [59–67], dynamic and temporal networks [68–75], multiplex networks [39, 40, 76–81], inter-connected networks [37, 38, 82–93], etc.

Human behaviour is also increasingly considered in the modelling and analysis of epidemic spreading, including awareness of the disease [80, 94–100], mobility patterns [29, 101–103], and so on.

Another important research stream on epidemics has focussed on how to control and immunise against an epidemic. Various immunisation strategies have been proposed, including the simplest uniform immunisation [104] whereby individuals in a given population are randomly immunised [30], the more efficient targeted immunisation that selectively immunises those individuals most likely to influence epidemic spreading [105–117], the more practical acquaintance immunisation that considers a more realistic scenario whereby there is no global knowledge about the spreading network structure and only local connection information can be used to immunise individuals [109, 118–124], etc. There are also studies about intervention strategies to control the spread of epidemics [125, 126].

In addition, there is a great deal of interest in using the knowledge obtained from studying epidemic spreading mechanisms in systems in order to solve real world problems. Example applications that have taken advantage of knowledge from epidemic spreading studies include viral marketing that promotes advertisements by spreading them through social networks [5–7], epidemic routing that aims to find and maintain a path to send information from one device to another [8–10], epidemic-based replicated database which is a distributed system that stores multiple copies of data in a large number of machines [11–13], and epidemic-based information dissemination [14–16].

2.2 Classic epidemic spreading models

There are a number of epidemic spreading models [3, 17]. Here the two most widely studied classic epidemic models are introduced, including the Susceptible-Infected-Susceptible (SIS) model and the Susceptible-Infected-Recovered (SIR) model.

2.2.1 Susceptible-Infected-Susceptible model

For the Susceptible-Infected-Susceptible (SIS) model, there are two potential states for each individual: *susceptible* and *infected*. Fig. 2.1 shows the classic SIS epidemic

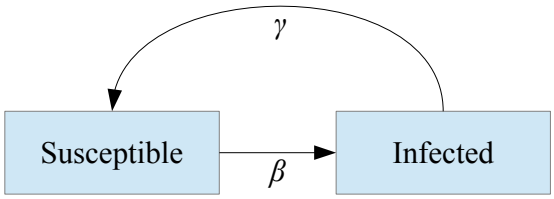

Figure 2.1: The classic Susceptible-Infected-Susceptible (SIS) epidemic model.

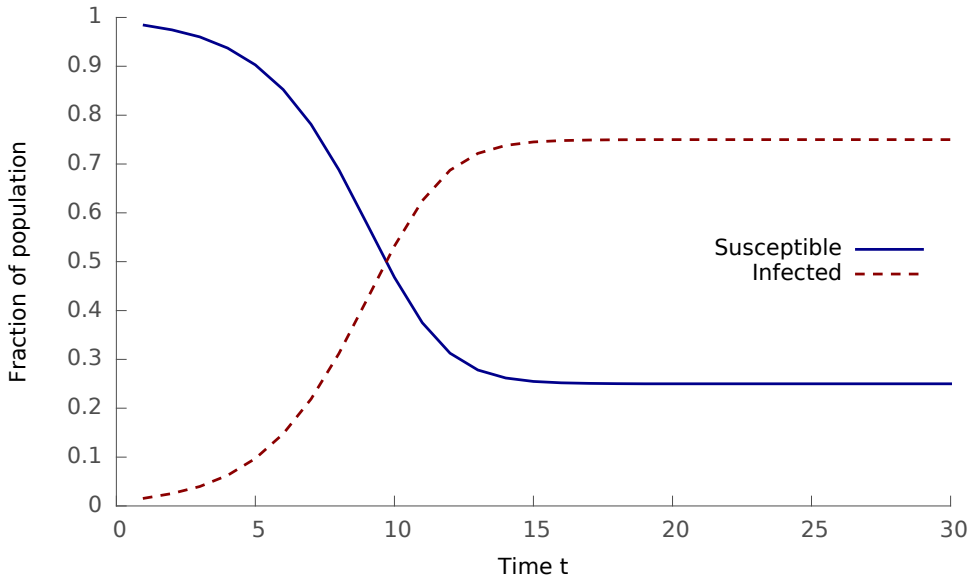

Figure 2.2: A typical time evolution of the SIS model. The parameter values are: infection rate $\beta = 0.8$ and recovery rate $\gamma = 0.2$. Initially the fraction of infected nodes are 1% and the other 99% nodes are susceptible. The results are numerically calculated by choosing an interval $dt = 1$.

model: for each unit time, (1) each *susceptible* individual becomes infected by a neighbouring infected individual with an infection rate of β; and (2) each *infected* individual becomes *susceptible* again with a recovery rate of γ. A more detailed glossary of terms used in epidemic modelling can be found in [47, 127].

Let s, and i respectively represent the fraction of *susceptible* and *infected* individuals. In a fully-mixed population where an infected individual can infect any other individual [3, 20], the differential equations for the SIS model are

$$\begin{aligned} \frac{ds}{dt} &= \gamma i - \beta si \\ \frac{di}{dt} &= \beta si - \gamma i \end{aligned} \tag{2.1}$$

with

$$s + i = 1 \tag{2.2}$$

And by setting $ds/dt = 0$ in Eq. (2.1), we can get the stable value of s as γ/β.

Fig. 2.2 illustrates a typical time evolution of the SIS model, showing the fraction of susceptible and infected individuals as a function of time.

2.2.2 Susceptible-Infected-Recovered model

Figure 2.3: The classic Susceptible-Infected-Recovered (SIR) epidemic model.

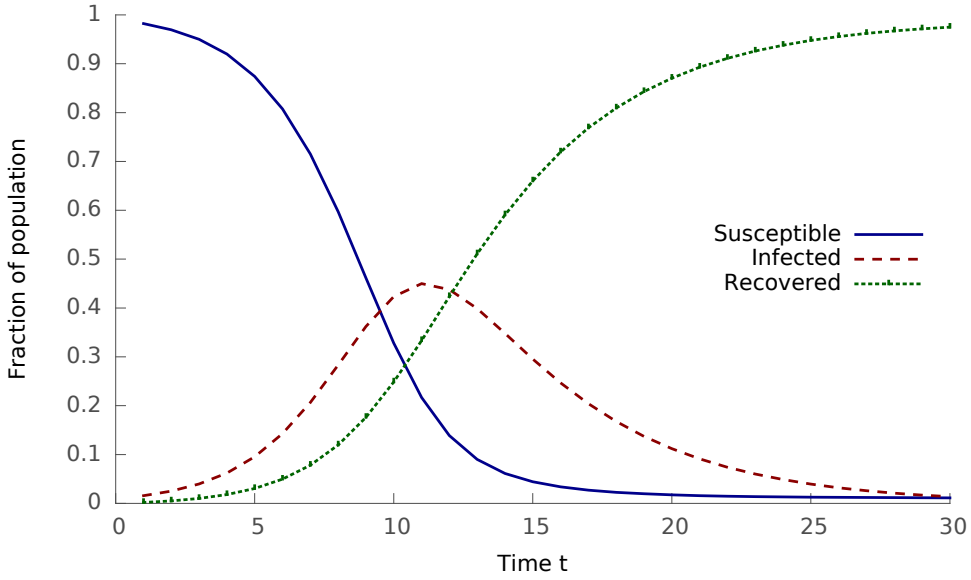

Figure 2.4: A typical time evolution of the SIR model. The parameter values are: infection rate $\beta = 0.8$ and recovery rate $\gamma = 0.2$. Initially the fraction of infected nodes are 1% and the other 99% nodes are susceptible. The results are numerically calculated by choosing an interval $dt = 1$.

Similarly, for the classic Susceptible-Infected-Recovered (SIR) epidemic model, as shown in Fig. 2.3, each individual is in one of three states: *susceptible*, *infected*, and *recovered*. At every time step, (1) each susceptible individual gets infected by a neighbouring infected individual with an infection rate of β; (2) each infected individual recovers at a rate γ; and (3) each recovered individual remains in the same state.

Let s, i, and r respectively represent the fraction of *susceptible*, *infected*, and *recovered* individuals. In a fully-mixed population where an infected individual can infect

any other individual [3, 20], the differential equations for the SIR model are

$$\begin{aligned} \frac{ds}{dt} &= -\beta si \\ \frac{di}{dt} &= \beta si - \gamma i \\ \frac{dr}{dt} &= \gamma i \end{aligned} \tag{2.3}$$

with

$$s + i + r = 1 \tag{2.4}$$

Fig. 2.4 illustrates a typical time evolution of the SIR model, showing the fraction of susceptible, infected, and recovered individuals as a function of time.

Our study focusses on the SIR epidemic spreading model.

2.3 Summary

Epidemic spreading phenomena exist in a wide range of domains [1–3]. Understanding and modelling the epidemic dynamics can have a significant practical impact on health care, technology and the economy. This section briefly introduced the research directions of epidemic spreading, together with two typical epidemic spreading models: the Susceptible-Infected-Susceptible (SIS) model and the Susceptible-Infected-Recovered (SIR) model.

Chapter 3

Hybrid epidemic spreading

Many epidemics are *hybrid* in the sense that they spread via two or more spreading mechanisms simultaneously. A hybrid epidemic can use both local spreading and global spreading mechanisms. The key difference between the two spreading mechanisms is that in *global spreading*, the population is fully mixed, i.e. each individual contacts any other individual with equal probability, whereas in *local spreading*, the epidemic must follow a limited set of connections among individuals, thus the structure, or topology, of the network plays a role in the spreading process.

There are many real examples. Computer worm Code-red spends 1/8th of its time probing any computers on the Internet at random (global spreading) and the rest of the time probing computers located in the local area networks of an infected computer (local spreading) [22]. HIV infection of cells inside human body is reported to spread by two parallel routes: cell-free infection through virus particles that are released by infected cells and by cell-to-cell infection through direct contact between infected and susceptible cells [128–130].

It is clear that hybrid epidemics are more complex than simple epidemics. Their behaviour is affected not only by the multiple spreading mechanisms that they use, but also by the population's structure on which they spread. Studying hybrid epidemics may provide crucial clues for developing a better understanding of many real epidemics.

3.1 Relevant works

Early relevant studies investigated epidemics spreading in populations whose individuals mix at both local and global levels ("two levels of mixing") [131]. These early

studies [131] did not incorporate the structure of the local spreading network, instead they assumed both local and global spreading are fully-mixed. Since network based epidemic analysis was introduced [3, 19], considerable effort has been made to understand how the structure of the local or global spreading network affects the dynamics of hybrid epidemics.

For example, Kiss et al. [23] studied epidemic spreading on a network with an additional process of "casual contacts", and provided a deterministic prediction on the final size of the epidemic outbreak. While Ball et al. [132] studied a similar case and presented a stochastic analysis on the probability distribution of the outbreak size.

Wang et al. [102] simulated hybrid mobile phone virus spreading, taking into account the mobility patterns of mobile phone users. A hybrid mobile phone virus employs both spreading through Multimedia Messages (MMS) and spreading through short distances communication between adjacent devices through Bluetooth technology (BT). This work also considered people's travelling pattern in the simulation. Cheng et al. [133] and Gao et al. [134] provided a relatively simple [133] and detailed [134] model for hybrid mobile phone viruses respectively.

Stone et al. [135, 136] studied epidemic spreading on dynamic small-world networks, which combines fixed short-range links within an individual's local neighbourhood with time-varying stochastic long-range links outside of that neighbourhood. Estrada et al. [137] studied the epidemic spreading with both "close" and "casual" contacts, where the "casual" contacts between two individuals are determined by their social proximity (e.g. the shortest path distance). These works focussed on analysing the role of the long-range links and social proximity, respectively, on the dynamics of epidemic spreading.

Studies on epidemic spreading in clustered random networks [59–67] and networks incorporating households [48–51, 138–140], are relevant to the study. In these studies, epidemic spreading is examined in networks with local structures (clusters/households). However, some of their assumptions are not valid in fragmented networks. For example, the household assumption is predicated on the idea that local structures are like households, i.e. fully connected, which is often not the case in real world scenarios.

Studies have also been conducted for epidemic spreading on metapopulations [52–

58]. The term metapopulation was originally defined in biological research as "a lot of populations" by Levins [141]. The term was then adopted to study epidemic spreading in a metapopulation that consists of a number of separated subpopulations that are only connected via the migration of individuals. This group of studies focus on analysing how the connection structure among subpopulations affects the epidemic spreading in a metapopulation. This thesis also studies hybrid epidemic spreading in metapopulations, but our definition of the metapopulation is different (See the Introduction of Chapter 4).

Recent works on two relatively new topics: multiplex networks [39, 40, 76–81] and inter-connected networks [37, 38, 82–93] are also relevant to this thesis. However, the focus of these two topics in previous research is different from the studies conducted herein. More specifically, most of the research into epidemics on multiplex networks and inter-connected networks investigate how dependence between networks affect spreading dynamics. While the focus of this research is on analysing how the interplay between multiple spreading mechanisms in a hybrid epidemic affects spreading. In addition, some of the assumptions that underpin the previous works, e.g. networks are tree-like, are also often violated in real spreading scenarios and so this was avoided in this current study.

Furthermore, most previous research has focussed on *non-critically* hybrid epidemics, where at least one of the two spreading mechanisms alone can cause an infection outbreak and therefore the combination of two mechanisms is not a necessary condition for an epidemic outbreak. In this case, given the same spreading power, a hybrid epidemic using two spreading mechanisms is often less contagious than an epidemic using only the more infectious of the two mechanisms. [23, 24].

However, many real examples of hybrid epidemics suggest the existence of *critically* hybrid epidemics where, using the same spreading power, a mixture of spreading mechanisms may be more infectious than using only one mechanism. We study the *critically* hybrid epidemics that previous studies have hardly investigated.

3.2 Critically hybrid epidemics

We define *critically* hybrid epidemics as hybrid epidemics where each of the spreading mechanisms alone is not able to cause any significant infection while conversely a

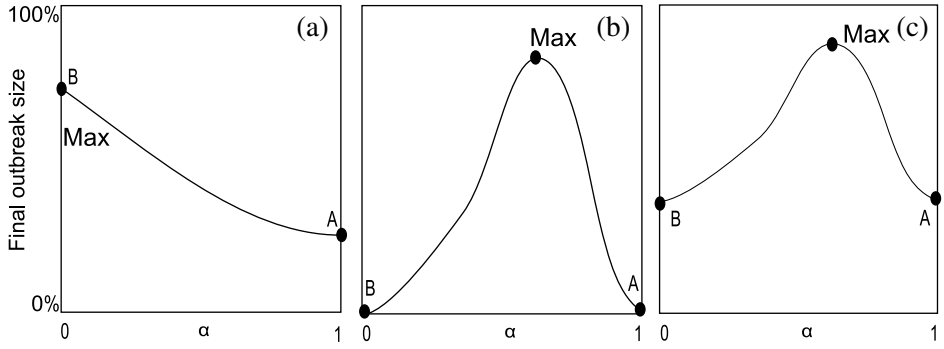

Figure 3.1: Hybrid epidemics, where two spreading mechanisms A and B are mixed at the ratio of α to $(1-\alpha)$, where $0 \leqslant \alpha \leqslant 1$. **(a)** Non-critically hybrid epidemic, where at least one of the two mechanisms can cause an outbreak on its own (i.e. when $\alpha = 1$ or $\alpha = 0$). **(b)** Critically hybrid epidemics, where each mechanism alone cannot cause any significant infection whereas a combination of them produces an epidemic outbreak. There exists an optimal α that produces the maximum outbreak. **(c)** Another type of hybrid epidemics, where each mechanism alone can cause a moderate outbreak while the combination of them leads to the largest outbreak. Our study focuses on the critically hybrid epidemics illustrated in **(b)**.

combination of the mechanisms can cause an enormous outbreak. In this case, the combination of different spreading mechanisms is a critical (and necessary) condition for a large-scale outbreak (see Fig. 3.1).

Critically hybrid epidemics are widely observed in nature and society. This thesis includes case studies for two critically hybrid epidemics: the computer worm Conficker spreading on the Internet and HIV infection within the human body.

Computer worms like Conficker [21] and Code-Red [22] employ both local spreading among computers in local area networks and global spreading to any random computers on the Internet. The local spreading has a high infection rate and is efficient at infecting computers in a local area network [142]. One reason for the high infection rate among computers in a local area network is that there is a tendency for similar software to be installed on these computers. Consequently, if the software has vulnerabilities that can be exploited by a worm, computers with the software inside the local area network are all vulnerable. However, worms with only local spreading cannot cause a major outbreak: it can only infect several local area networks. By comparison, global spreading is inefficient at infecting a computer, but each successful global infection spreads the worm to a random computer on the Internet. Global spreading creates colonies for a worm and enables it to spread all over the Internet. While worms that utilise either local or global spreading alone are not able to cause major infections,

worms employing both simultaneously can infect millions of computers and last for years [28]. Our case study on a recent computer worm Conficker in Chapter 6 supports the above insight.

The HIV virus spreads between T-cells within human body via two routes: (1) *cell-free* spreading, where susceptible cells are infected by the virus (released by other infected cells) in body fluid; and (2) *cell-to-cell* spreading, where infected cells infect and transfer the HIV to susceptible cells though direct contacts. The theoretical analysis herein suggests that if either of the two routes is blocked, the virus might be eliminated by the human immune system (see Chapter 8). Conversely, the combination of the two spreading routes makes HIV extremely contagious and difficult to cure.

3.3 Case studies

This thesis conducts case studies on two real world epidemics: (1) the spread of a recent computer worm Conficker on the Internet; and (2) the spread of the HIV infection within the human body. This chapter is a brief introduction to the two case studies, while more specific details are outlined for readers in Chapter 6, Chapter 7 and Chapter 8.

3.3.1 Computer worm Conficker spreading on the Internet

Computer worms are malicious applications that automatically propagate themselves through computer networks. The first recognised worm was "Morris" [143]. It was a benign worm created by a Cornell University student to measure the size of the Internet in 1988. Years later, a number of worms appeared, including Code-red [144, 145] in 2001, Witty [146] in 2004, and a recent one - Conficker [21, 147] in late 2008. Most of these later worms carry some destructive functions and consequently pose a serious threat to the security of the Internet via infecting millions of computers. Among all types of computer attacks, worms spreading on the Internet have led to the most rampant and extensive damage (See [148] for a recent survey of computer worms).

Conficker is one of the most contagious computer worms on record. It exploded on the Internet on 21 November 2008 and infected millions of computers in just a few days [21]. The worm's ability to spread to such a large number of computers in such a short period of time and the fact [28] that it is still alive years after being detected have

caused serious concern for protecting the security of the Internet.

Various studies [21, 143–147, 149–160] have been conducted on the spread of many computer worms, including Conficker. Most existing modelling studies assume that a worm only conducts global spreading, ignoring the localised spreading mechanisms. While many of the existing theoretical models can be useful in estimating the total outbreak size of a worm, they do not accurately reflect the hybrid spreading mechanisms of a worm.

3.3.2 HIV infection spreading within human body

HIV is a virus that attacks human immune system - the defence mechanism that protects the body against diseases. Without medical treatment, people infected by HIV can develop Acquired Immune Deficiency Syndrome (AIDS), which is a condition whereby the immune system is for all intents and purposes - failing. AIDS exposes a patient to opportunistic infections (caused by viruses, bacteria, parasites, fungi, etc.) and threatens his/her life [161].

Tremendous amounts of time, money and research has been spent over the past three decades to gain a better understanding of and improve the treatment of the virus. And while the state-of-the-art HIV treatments can manage and control the virus's infection, there is still no permanent cure for the disease [25]. This calls for further collaboration among researchers across different disciplines to help tackle the disease.

Mathematical models provide an important tool for understanding and predicting the course of natural HIV infection, complementing clinical studies.

Early studies conducted by Ho [162], Wei et al. [163], Perelson [164], and Nowak et al. [165] used simple mathematical models to fit clinical data to estimate the parameters associated with HIV infection. These parameters include virus production and clearance rates, life-span of infected T cells, etc. An important estimation from [162] and [164] is that HIV has a short life span but a high reproduction rate. This helps explain why HIV virus can evolve so quickly that the human immune system fails to even recognise the threat and respond appropriately.

Perelson et al. [41, 164] proposed a simple and classic model for the spread of the HIV infection within the human body, which is a modified version of the classic Susceptible-Infected-Recovered (SIR) epidemic spreading model that was introduced

in Section 2.2. The model studies how the densities of uninfected T cells (T), infected T cells (I) and free virus (V) change over time. The following differential equation system [41] describes this model:

$$\begin{aligned} \frac{dT}{dt} &= \lambda - d_T T - \beta V T \\ \frac{dI}{dt} &= \beta V T - \delta I \\ \frac{dV}{dt} &= pI - cV \end{aligned} \tag{3.1}$$

where free viruses are assumed to infect T cells at rate β following a mass action term βVT. Uninfected T cells are created at a constant rate λ and die at a rate d_T per cell. Infected T cells release free viruses at rate p per cell and die at rate δ per cell. δ is higher than d_T, which reflects the effect of HIV infection on shortening T cells' lifespan [41]. Free virus particles are cleared at rate c per virion (virus particle). This simple classic model can explain the dynamics of the acute infection phase of HIV [166] and the forming of a steady level of virus load [41].

These early studies illustrated the potential of using mathematical models to understand and explain the virus's infection. Results from these studies elicited wide interests from multiple disciplines and further research about the HIV infection, from both theoretical and empirical perspectives.

Since then, numerous works have been proposed to model HIV infection [166–178]. Most studies, however, focus on a specific period of HIV infection, e.g. the acute infection phase or the chronic stable phase [179]. While existing research was able to provide some insight into the detailed infection mechanisms, the entire multiple phase course (See 7 for details) and the progression of HIV infection have still not been convincingly captured and explained in either modelling or empirical studies [179].

3.4 Summary

Epidemics have traditionally been studied as a single spreading process, either in a fully mixed population or on a network. Many epidemics, however, are hybrid, employing more than one spreading mechanism.

For existing works relevant to hybrid epidemics, most have studied *non-critically* hybrid epidemics, where at least one of the two spreading mechanisms alone can cause

an infection outbreak and therefore the mix of two mechanisms is not a necessary condition for an epidemic outbreak.

The focus of this thesis is about *critically* hybrid epidemics where each of the spreading mechanisms alone is not able to cause any significant infection whereas a combination of the mechanisms can cause an epidemic outbreak. In this case, the mix of different spreading mechanisms is a critical condition for an outbreak (see Fig. 3.1). The dynamics of *critically* hybrid epidemics have not been properly investigated so far.

Two real examples of *critically* hybrid epidemics are the spread of the computer worm Conficker across the Internet and the HIV infection within the human body. Their spreading dynamics have not been properly investigated in the existing literature on epidemics.

For the spread of Conficker, existing research has mostly focussed on its overall statistical properties, such as the size of outbreak in different countries. Modelling studies have analysed it as a single-mechanism epidemic. The epidemic parameters, including infection rates for different spreading mechanisms, were still unknown for this worm before our study.

Numerous modelling studies have been conducted on the spread of the HIV infection within the human body [41, 179, 180]. Most of them have focussed on a specific singular period of the HIV infection [179], and by default have assumed that HIV is a single-mechanism epidemic, overlooking the realistic possibility that the virus spreads through direct cell-to-cell contacts. The life course including multiple phases (See Chapter 7 for details) and the progression of HIV infection have yet to be captured and explained by either modelling or empirical studies in a satisfactory manner [179].

These two real cases will be studied in this thesis using a hybrid modelling approach. Our work will show that an appreciation of the epidemics' hybrid nature is crucial for a correct understanding of their spreading dynamics and properties.

Chapter 4

Modelling hybrid epidemics

In this chapter we develop a theoretical framework for studying hybrid epidemics that employ both *local spreading* and *global spreading*, both in a single population and in a metapopulation.

4.1 Introduction

Hybrid epidemics often take place in a *metapopulation*, consisting of a number of subpopulations. Each subpopulation is a collection of densely or strongly connected nodes, whereas nodes from different subpopulations are weakly connected. This is illustrated in Fig. 4.1, where the model considers two spreading mechanisms: 1) local spreading where an infected node can infect nodes in its subpopulation and 2) global spreading, where an infected node can infect all nodes in the metapopulation. Note that our definition of metapopulation is different from the classical metapopulation defined in ecology where subpopulations are connected via flows of agents [54, 57].

In the model each subpopulation for local spreading can be either fully-mixed or a network. Because a fully-mixed subpopulation can also be represented as a fully connected network, we describe each subpopulation as a network, for mathematical convenience. At each time step, an infected node has a fixed total spreading power that must be allocated between the two spreading mechanisms. Let the hybrid tradeoff, α, represent the proportion of spreading power spent in local spreading. The proportion of global spreading power is $1-\alpha$. A variable α enables us to investigate the interaction and the joint impact of the two spreading mechanisms on epidemic dynamics, ranging from a completely local spreading scenario (with $\alpha = 1$) to a completely global spreading scenario (with $\alpha = 0$).

We consider the hybrid epidemic spreading in terms of the Susceptible-Infected-Recovered (SIR) model [3, 47], where each node is in one of three states: *susceptible* (s), *infected* (i), and *recovered* (r). At each time step, each infected node spreads both locally and globally; it infects 1) each directly connected node in the same subpopulation with probability $b_1 = \alpha\beta_1$ and 2) each susceptible node in the metapopulation with probability $b_2 = (1-\alpha)\beta_2$. β_1 is the local infection probability when all spreading power is allocated to local spreading ($\alpha = 1$). And β_2 is the global infection probability when all spreading power is allocated to global spreading ($\alpha = 0$). Each infected node recovers at a probability γ, and then remains permanently in the recovered state. A node can infect other nodes and then recover in the same time step.

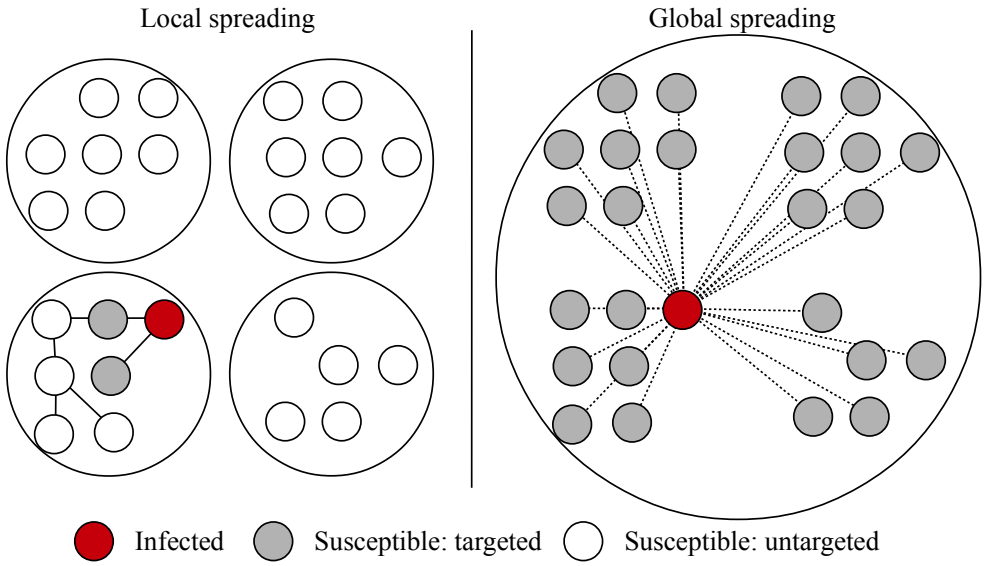

Figure 4.1: Hybrid epidemic spreading in a metapopulation. At each time step, an infected node has a fixed total spreading power that must be allocated between local spreading and global spreading. The proportion of spreading power spent in local spreading is α and that in global spreading is $1-\alpha$. Local spreading occurs between infected and susceptible nodes that are connected in individual subpopulations; global spreading happens between an infected node and any susceptible node in the metapopulation.

4.2 Hybrid spreading in a single-population

Before analysing the hybrid spreading in a metapopulation, we study a relatively simple case where the epidemic process takes place in a *single population*. That is, there is only one population, where local spreading is via direct connections on a network structure and global spreading can reach any node in the population.

Here we extend the edge-based compartmental modelling system [181] for the analysis. The system in [181] was proposed to analyse single-mechanism based epidemics for the continuous time case. Here the system is extended to analyse 1) hybrid epidemics, and 2) for the discrete time case. We calculate the probability that a random test node u is in each state: susceptible $s(t)$, infected $i(t)$, and recovered $r(t)$.

We denote $p(k)$ as the probability that a node has degree (i.e. number of neighbours) k. The generating function [182] of degree distribution $p(k)$ is defined as $g_0(x) = \sum_{k=0}^{\infty} p(k)x^k$. Let $p_n(k)$ represent the probability that a random neighbour of u has k neighbours. We assume the network is *uncorrelated*: the degree of each node is independent from degrees of other nodes [3]. In an uncorrelated network $p_n(k) = p(k)k/\langle k\rangle$, where $\langle k\rangle$ is the average degree of the network and $\langle k\rangle = g_0'(1)$ [3].

Let $\theta(t)$ be the probability that a random neighbour v has not infected u through local spreading. Let $\vartheta(t)$ be the probability that a random node w has not infected u through global spreading. Suppose u has k neighbours, the probability that it is susceptible is $s_k(t) = \vartheta(t)^{n-1}\theta(t)^k$ where n is the total number of nodes in the population. Then by averaging $s_k(t)$ over all degrees, we have,

$$s(t) = \vartheta(t)^{n-1} \sum_{k=0}^{\infty} p(k)\theta(t)^k = \vartheta(t)^{n-1} g_0(\theta) \tag{4.1}$$

and

$$r(t+1) = r(t) + \gamma i(t) \text{ and } i(t) = 1 - s(t) - r(t)$$

As shown in Fig. 4.2, the probability θ can be broken into three parts: v is susceptible at t, ϕ_s; v is infected at t but has not infected u through local spreading, ϕ_i; v is recovered at t and has not infected u through local spreading, ϕ_r. Neighbour v cannot be infected by u and itself, then $\phi_s = \vartheta^{n-2}\sum_k p_n(k)\theta^{k-1} = \vartheta^{n-2} g_0'(\theta)/g_0'(1)$. In a time step, neighbour v 1) infects u with probability $b_1\phi_i$ through local spreading and 2) recovers without infecting u through local spreading at probability $\gamma(1-b_1)\phi_i$, i.e. after every time step: $(1-\theta)$ increases by $b_1\phi_i$ and ϕ_r increases by $\gamma(1-b_1)\phi_i$. The increase of ϕ_r here, $\gamma(1-b_1)\phi_i$, is different from that $(r\phi_i)$ in the original system in [181]. Because the original system was designed for the continuous time case, and in the discrete time case in this study, neighbour v can infect u and recover at the same time step. Given that ϕ_r and $1-\theta$ are both approximately 0 in the beginning ($t = 0$),

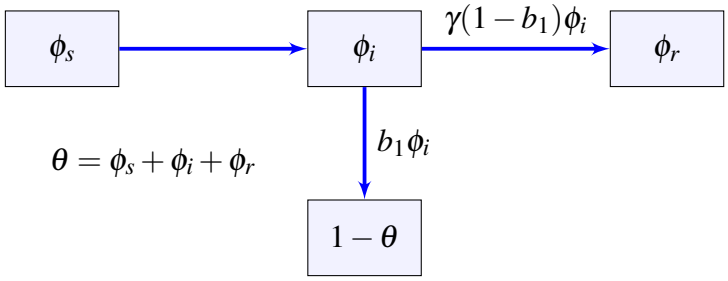

Figure 4.2: Relations between ϕ probabilities. For a random test node u, θ is the probability that u has not been infected by a random neighbour v through local spreading. θ can be broken into three parts: ϕ_s, ϕ_i, and ϕ_r, which respectively represent the probability that v has not infected u when v is susceptible, infected, and recovered. In a time step, neighbour v 1) infects u with probability $b_1\phi_i$ through local spreading and 2) recovers without infecting u through local spreading at probability $\gamma(1-b_1)\phi_i$, i.e. after every time step: $(1-\theta)$ increases by $b_1\phi_i$ and ϕ_r increases by $\gamma(1-b_1)\phi_i$.

we have $\phi_r = \gamma(1-b_1)(1-\theta)/b_1$. Then

$$\phi_i = \theta - \phi_s - \phi_r = \theta - \vartheta^{n-2}\frac{g_0'(\theta)}{g_0'(1)} - \frac{\gamma(1-b_1)}{b_1}(1-\theta) \tag{4.2}$$

For global spreading, the probability ϑ can also be broken into three parts: w is susceptible at t, φ_s; w is infected at t but has not infected u through global spreading, φ_i; w is recovered at t but has not infected u through global spreading, φ_r. Using a similar derivation process, we have $\varphi_s = \vartheta^{n-2}\sum_k p(k)\theta^k = \vartheta^{n-2}g_0(\theta)$ and $\varphi_r = (1-\vartheta)\gamma(1-b_2)/b_2$, and

$$\varphi_i = \vartheta - \varphi_s - \varphi_r = \vartheta - \vartheta^{n-2}g_0(\theta) - \frac{\gamma(1-b_2)}{b_2}(1-\vartheta) \tag{4.3}$$

When the epidemic stops spreading, $\phi_i = 0$ and $\varphi_i = 0$. By setting $\phi_i = 0$ in Eq. (4.2) we get

$$\vartheta^{n-2} = \frac{g_0'(1)}{g_0'(\theta)}(\theta + \frac{\gamma(1-b_1)}{b_1}\theta - \frac{\gamma(1-b_1)}{b_1}) \tag{4.4}$$

Substituting Eq. (4.4) and $\varphi_i = 0$ into Eq. (4.3), we have

$$\vartheta = w(\theta) = \frac{g_0'(1)(\theta + \theta\gamma(1-b_1)/b_1 - \gamma(1-b_1)/b_1)g_0(\theta)/g_0'(\theta)}{1+\gamma(1-b_2)/b_2} + \frac{\gamma(1-b_2)/b_2}{1+\gamma(1-b_2)/b_2} \tag{4.5}$$

By setting $\phi_i = 0$ and substituting Eq. (4.5) in Eq. (4.2) and rearranging Eq. (4.2) we

have

$$\theta = [\frac{w(\theta)^{n-2} g_0'(\theta)/g_0'(1) + \gamma(1-b_1)/b_1}{1+\gamma(1-b_1)/b_1} \equiv f(\theta)] \tag{4.6}$$

Then θ_∞ - stationary value of θ is a fixed point of $f(\theta)$. A fixed point of a function is a value that is mapped to itself by the function. Strogatz's book [183] provides more detailed methods to analyse nonlinear dynamics.

4.2.1 Analysis of the threshold condition

The iterative scheme about $f(\theta)$ in Eq. (4.6) has a known fixed point of $\theta = 1$ which represents no epidemic outbreak. We test the stability of this fixed point. By substituting Eq. (4.2) and Eq. (4.5) into $d\theta/dt = -b_1\phi_i$, setting $\theta = 1+\varepsilon$ and take the leading order (Taylor Series), we have $d\varepsilon/dt = \varepsilon h$ and

$$h = \frac{b_1 A + b_2 B + b_1 b_2 C - \gamma^2}{b_2 + \gamma - b_2\gamma} \tag{4.7}$$

where $A = \gamma^2 + g_0''(1)\gamma/g_0'(1) - \gamma$, $B = \gamma^2 + n\gamma - 3\gamma$, and $C = -\gamma^2 - g_0''(1)\gamma/g_0'(1) - n\gamma + 4\gamma - ng_0''(1)/g_0'(1) + 3g_0''(1)/g_0'(1) + ng_0'(1) - 2g_0'(1) + n - 3$. Then $\varepsilon = Ce^{ht}$ where C is a constant. When h is negative, $|\varepsilon|$ gradually decreases and approaches 0 as t increases; while when h is positive, $|\varepsilon|$ gradually increases and approaches $+\infty$ with the increase of t. That is the fixed point $\theta = 1$ turns from stable to unstable when h changes from negative to positive. The threshold condition for an epidemic outbreak is then $h > 0$:

$$h(\beta_1, \beta_2, \gamma, \alpha, p(k)) > 0 \tag{4.8}$$

This epidemic threshold represents a condition which, when *not* satisfied, results in an epidemic that vanishes at an exponentially fast pace [20, 184]. There are two special cases.

- For completely local spreading ($\alpha = 1, b_1 = \beta_1, b_2 = 0$), the threshold reduces to $\beta_1 g_0''(1)/[g_0'(1)(\beta_1 + \gamma - \gamma\beta_1)] > 1$. Here $g_0'(1) = \langle k\rangle$ and $g_0''(1) = \langle k^2\rangle - \langle k\rangle$ where $\langle k\rangle$ is the average degree of the network and $\langle k^2\rangle$ is the average degree square of the network [3]. In the end of this section, we show that this threshold agrees with previous threshold results [34] for single-mechanism epidemics spreading on networks for the discrete time case. For infinite scale-free networks, we have $(\langle k^2\rangle - \langle k\rangle)/\langle k\rangle \to \infty$ such that the threshold 'vanishes' (i.e. $\infty > 1$ is

always satisfied), in agreement with previous observation [3, 19].

- For completely global spreading ($\alpha = 0, b_1 = 0, b_2 = \beta_2$), the threshold reduces to $\beta_2(n-3+\gamma)/\gamma > 1$, and when n is large it is approximate to $\beta_2 n/\gamma > 1$. $\beta_2 n/\gamma$ is the basic reproduction number, R_0, for single-mechanism epidemics spreading in a fully mixed population [20]. R_0 is the average number of nodes that an infected node can infect before it recovers. Thus the threshold is equivalent to $R_0 > 1$, in agreement with previous work [20].

Here we use Newman's method [34] to obtain the threshold condition for the local spreading to verify our results. Firstly we need to calculate the "transmissibility" T which is the average probability that an epidemic is transmitted between two connected nodes, of which one is infected and the other is susceptible. According to [34], for the discrete time case T can be calculated as

$$T = 1 - \int_0^\infty d\beta_1 \sum_{\tau=0}^{\infty} p(\beta_1)p(\tau)(1-\beta_1)^\tau \tag{4.9}$$

where τ is the time steps that an infected node remains infected, $p(\tau)$ and $p(\beta_1)$ respectively are the probability distribution of τ and β_1. For the model in this chapter, β_1 is a constant and $p(\tau) = (1-\gamma)^{\tau-1}\gamma$, in which $(1-\gamma)^{\tau-1}$ is the probability that an infected node has not recovered until $\tau - 1$ steps after infection, and γ is the probability that the node recovers at the τth step after infection. Also for the model in this chapter, each infected node at least remains infected for 1 time step. So that T for our model can be obtained as

$$T = 1 - \sum_{\tau=1}^{\infty} (1-\gamma)^{\tau-1}\gamma(1-\beta_1)^\tau = \frac{\beta_1}{\beta_1 + \gamma - \gamma\beta_1} \tag{4.10}$$

According to [34] the epidemic threshold for completely local spreading is $Tg_0''(1)/g_0'(1) > 1$ i.e. $\beta_1 g_0''(1)/[g_0'(1)(\beta_1 + \gamma - \gamma\beta_1)] > 1$. This is the same as the epidemic threshold for completely local spreading obtained in this section.

Note that treating each edge as having this value of T independently will lead to the correct epidemic threshold and final size calculation, but there are further discussions on its correctness in calculating the infection probabilities [185–188].

4.2.2 Analysis of the final outbreak size

The final outbreak size, r_∞, is the fraction of nodes that are recovered when all epidemic activities cease, i.e. when all nodes are either recovered or susceptible. When $t \to \infty$, the probability that a node is infected $i(t) \to 0$. Thus $r_\infty = 1 - s_\infty = 1 - \vartheta_\infty^{n-1} g_0(\theta_\infty)$ and

$$r_\infty = 1 - w(\theta_\infty)^{n-1} g_0(\theta_\infty) \tag{4.11}$$

where the value of θ_∞ can be numerically calculated by conducting the fixed-point iteration of Eq. (4.6). Eq. (4.11) can be viewed as a function of the hybrid epidemic parameters and the network degree distribution. To be noted here, for completely global spreading ($\alpha = 0, b_1 = 0, b_2 = \beta_2$), θ_∞ cannot be calculated from Eq. (4.6) (because $b_1 = 0$). In this case, $\theta_\infty = 1$, $g_0(\theta_\infty) = 1$, and $r_\infty = 1 - \vartheta_\infty^{n-1}$ where ϑ_∞ can be obtained by setting $\varphi_i = 0$, $g_0(\theta) = 1$ and solving the Eq. (4.3) in the range $0 < \vartheta < 1$.

4.2.3 Numerical evaluation

Numerical simulations were performed to verify the above theoretical predictions for hybrid epidemics in a single population. Two typical topologies for local spreading in the single-population are considered: (1) A fully connected network that represents a fully mixed population; and (2) a random network with Poisson degree distribution, which are generated by the Erdős-Rényi (ER) model [189] with an average degree of five. Each of the two networks has 1000 nodes.

At the beginning, five randomly selected nodes are infected and all others are susceptible to infection. At each time step, each infected node infects 1) each directly connected susceptible node with probability $\alpha\beta_1$ and 2) each susceptible node in the population with probability $(1-\alpha)\beta_2$. Each infected node recovers with the probability γ. There are two steps to simulate when an action (infection or recovery) happens with a probability p: 1) *generate* a random number x in the range of $0 \leq x < 1$; 2) *if* $x < p$ *then* the action happens *otherwise* the action does not happen.

We run simulations for different values of $\alpha \in [0, 1]$. The global infection probability is set to $\beta_2 = 10^{-4}$ and the recovery probability $\gamma = 1$ (i.e. an infected node can only spread the epidemic in one time step). For epidemics on the fully connected network, the local infection probability is $\beta_1 = 6 \times 10^{-3}$. And for epidemics on the random network the probability is $\beta_1 = 0.8$. We choose typical values for these param-

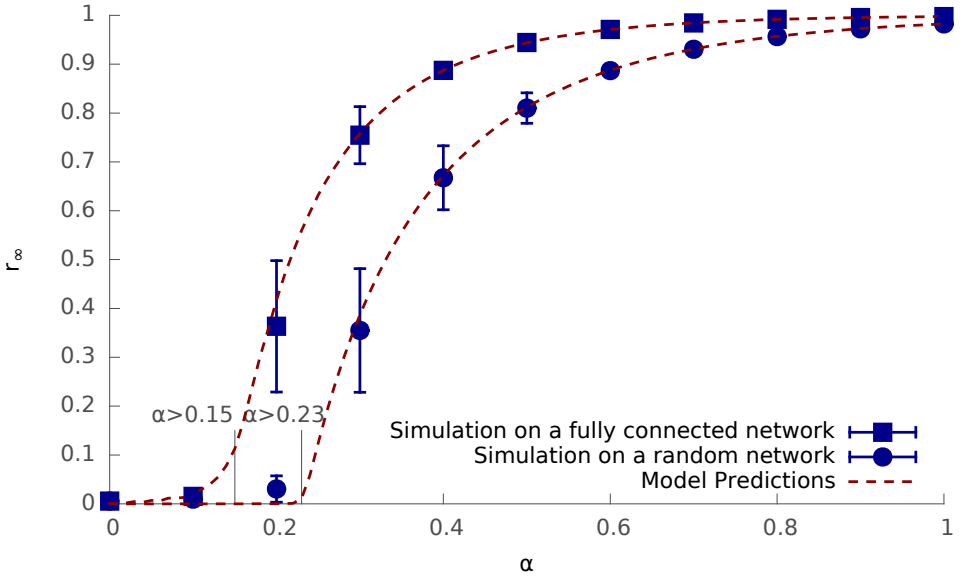

Figure 4.3: Final outbreak size r_∞, as a function of the hybrid tradeoff α, for hybrid epidemics in a single-population in which nodes form a fully connected network (i.e. fully mixed) or a random network with an average degree of five. The population has 1000 nodes. The global infection probability $\beta_2 = 10^{-4}$ and recovery probability $\gamma = 1$ are the same for epidemics on these two types of networks. The local infection probability β_1 is 6×10^{-3} for epidemics on the fully connected network; and it is 0.8 for epidemics on the random network. Initially five random nodes are infected. Simulation results are shown as points and theoretical predictions of Eq. (4.11) are dashed curves. The simulation results are averaged over 1,000 runs and the bars around the points show the associated standard deviation. The epidemic thresholds predicted by Eq. (4.8) are marked as vertical bars on the horizontal axis.

eters, with an additional consideration to make the results for the two networks clearly separated in the figure. Fig. 4.3 shows that the size of the predicted final outbreak using Eq. (4.11) is closely approximated by simulation results. It is also evident that the hybrid epidemic is characterised by a phase change, where the threshold is well predicted by Eq. (4.8).

4.3 Hybrid spreading in a metapopulation

Next the above theoretical results for a single-population are extended upon to analyse hybrid spreading in a metapopulation that consists of a number of subpopulations. Local infection happens only between nodes in the same subpopulation whereas global infection occurs both within and between subpopulations.

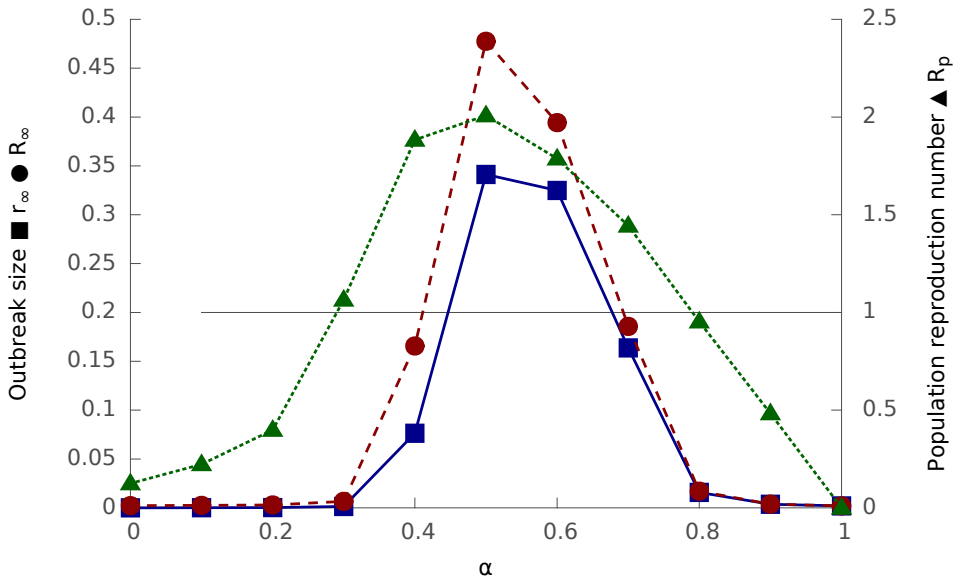

Figure 4.4: Simulation results of hybrid epidemics in a metapopulation. The three results are: (**i**) the final outbreak size as the fraction of recovered nodes r_∞ (squares); (**ii**) the final outbreak size as the fraction of recovered subpopulations R_∞ (circles); (**iii**) the population reproduction number, R_p (triangles, right y-axis). Each variable is plotted as a function of the hybrid tradeoff α. The metapopulation contains 500 subpopulations and each subpopulation is a random network with 100 nodes and an average degree of five. The local infection probability $\beta_1 = 0.8$, the global infection probability $\beta_2 = 10^{-6}$ and the recovery probability $\gamma = 1$. Initially three random nodes in a subpopulation are infected. Simulation results are shown as points and each result is averaged over 1,000 runs.

4.3.1 Analysis of spreading at population level

We define a subpopulation as susceptible if it contains only susceptible nodes. A subpopulation is infected if it has at least one infected node. A subpopulation is recovered if it has at least one recovered node and all other nodes are susceptible. Only global spreading enables infection between subpopulations, whereas spreading within a subpopulation can occur via both local and global spreading.

The final outbreak size at the population level R_∞, is defined as the proportion of subpopulations that are recovered when the epidemic stops spreading. We define that a subpopulation A directly infects another subpopulation B if an infected node in A infects a susceptible node in B. We define the population reproduction number, R_p, as the average number of other subpopulations that an infected subpopulation directly infects before it recovers. Note that our definition of R_p is similar to R_* in [54] but the definition of a metapopulation in [54] is different. A metapopulation includes many subpopulations. In order for an epidemic to spread in a metapopulation, an infected

subpopulation should infect at least one other subpopulation before it recovers, i.e. the threshold condition of the hybrid epidemic at the population-level is $R_p > 1$.

Fig. 4.4 shows simulation results of the final outbreak sizes r_∞ and R_∞ and the population reproduction number R_p (right y axis) as a function of the hybrid tradeoff α. Epidemic parameter values are included in the caption of Fig. 4.4. All three results show a bell shape curve regarding α. The results clearly suggest that the epidemic will not cause any significant infection if it uses only local spreading ($\alpha = 1$) or only global spreading ($\alpha = 0$). The maximal outbreak at the node level $r^*_\infty = 0.34$ is obtained around the optimal hybrid tradeoff $\alpha^* = 0.5$. That is, if 50% of the infection events occur via local spreading (and the rest via global spreading), the epidemic will ultimately infect 34% of all nodes in the metapopulation. At the population level, the total percentage of recovered subpopulations R_∞ follows a very similar trend to r_∞, and the maximum epidemic size in terms of subpopulations occurs at the same optimal α^*. The population reproduction number R_p follows a similar trend to the final outbreak sizes R_∞ and r_∞. The threshold $R_p > 1$ defines the range of α for which the final outbreak sizes are significantly larger than zero.

It is important to appreciate that although the maximal R^*_p is uniquely defined by the optimal α^*, other R_p can be obtained by *two* different α values, on either side of the optimal α^*, potentially representing different epidemic dynamics.

4.3.2 Analysis of the population reproduction number R_p

The population reproduction number R_p is a fundamental characteristic of hybrid epidemics in a metapopulation. We consider a metapopulation with $N+1$ subpopulations, which are denoted as p_i where $i = 0, 1, 2...N$. Each subpopulation has n nodes connected to a similar structured local spreading network. p_0 is the subpopulation where the epidemic starts from.

We assume the infection inside the initially infected subpopulation p_0 is caused entirely by infected nodes inside p_0. That is, we neglect the effects of global spreading of other N subpopulations on p_0. This is an acceptable assumption when the metapopulation has a larger number of subpopulations. Under these conditions, hybrid spreading within p_0 is the same as spreading in a single-population, which has been analysed in previous sections. To predict R_p, we first analyse the expected number of nodes outside

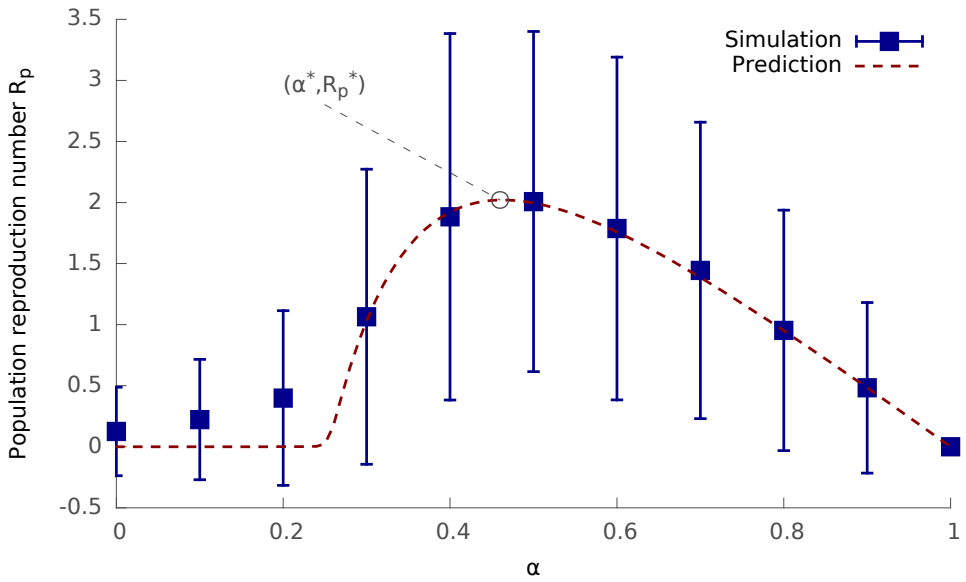

Figure 4.5: Population reproduction number R_p as a function of the hybrid tradeoff α. Epidemics spread in a metapopulation that contains 500 subpopulations and each subpopulation is a random network with 100 nodes and an average degree of five. The local infection probability $\beta_1 = 0.8$, the global infection probability $\beta_2 = 10^{-6}$ and the recovery probability $\gamma = 1$. Initially three random nodes in a subpopulation are infected. Theoretical prediction from Eq. (4.13) is shown as a dashed curve. Simulation results are shown as points (average over 1,000 runs) and bars (one standard deviation).

p_0 that will be infected by p_0. We then estimate the number of other subpopulations that these infected nodes should belong to. Let $s_N(t)$ represent the probability that a random test node in other subpopulations are susceptible at time t. Using the same parameters defined in the analysis on hybrid epidemics in a single population, we have $s_N(t) = \vartheta(t)^n$ where n is the number of nodes in p_0.

When p_0 recovers at time T, the *fraction* of nodes in other subpopulations that have been infected by (infected nodes in) p_0 (via global spreading) is $x_N = 1 - s_N(T) = 1 - \vartheta(T)^n = 1 - w(\theta_T)^n$ where we have used Eq. (4.5). Then the *number* of such infected nodes is

$$X_N = x_N nN = (1 - w(\theta_T)^n)nN \tag{4.12}$$

where nN is the total number of nodes in other N subpopulations and θ_T can be numerically calculated as θ_∞ by fixed-point iteration of Eq. (4.6). As the nodes are infected randomly via the global spreading, the probability that an infected node does not belong to a particular subpopulation i is $1 - 1/N$; and the probability that none of these infected nodes belongs to the subpopulation i is $(1 - 1/N)^{X_N}$. So the probability that

at least one infected node belongs to the subpopulation i is $1-(1-1/N)^{X_N}$. Thus the population reproduction number R_p, which is the number of other subpopulations that these infected nodes should belong to, is:

$$R_p = N(1-(1-1/N)^{X_N}) \tag{4.13}$$

4.3.3 Numerical evaluation

Fig. 4.5 compares the predicted R_p against simulation results as a function of the hybrid tradeoff α. R_p is characterised by a bell-shaped curve. It peaks at the optimal hybrid tradeoff α^* where the population reproduction number achieves its maximal value R_p^*. This optimal point is of particular interest as it represents the optimal trade-off between the two spreading mechanisms, where the hybrid epidemic is most infectious and therefore has the most extensive outbreak.

4.4 Discussion

We assume the hybrid epidemic employs local and global spreading, where local spreading can only potentially infect nodes inside a subpopulation, and global spreading can potentially infect all nodes in all subpopulations. Our analysis for the population reproduction number does not rely on the tree-like spreading assumption but instead we assume the global infection probability is very small. These assumptions are applicable to spreading of most computer worms, e.g. Conficker [21] and CodeRed [146] spreading on the Internet, and the spreading of some diseases, e.g. HIV infection spreading within the human body. This assumption, nevertheless, can be inappropriate for the spreading of many other diseases, where there is no global spreading mechanism that enables an individual to infect all other individuals, e.g. spreading of flu in a country.

The total spreading power of a node at each time step is assumed to be fixed. The results of this analysis regarding how the tradeoff between two spreading mechanisms affects the outbreak of a hybrid epidemic only makes sense under this assumption. And this assumption is valid for the majority of computer worms.

The model proposed in this chapter does not aim to capture every detail of a particular real world epidemic. Instead, the model is designed to be simple, so that it can

be used to theoretically explore the general properties of many hybrid epidemics. By simplifying the spreading scenario, we can focus on analysing the key and common properties of hybrid epidemics.

It is also important to understand the particular properties of different epidemics. In Chapter 6 and Chapter 7, we propose more specific and detailed epidemic models for two real hybrid epidemics. These case studies provide examples for adopting and extending the basic hybrid epidemic model in this chapter to analyse more complicated epidemics.

Specifically, in Chapter 6, we conduct a case study on a real computer worm - Conficker. Because the computer worm employs three, rather than two spreading mechanisms, the model proposed in this chapter needs to be extended to include an additional spreading mechanism. In addition, Chapter 6 also infers the epidemic parameter values from real epidemic spreading data.

4.5 Summary

In this chapter, a new mathematical model for hybrid epidemic was presented and validated. The model incorporates two typical spreading mechanisms including local spreading among directly connected individuals (nodes) and global spreading among all individuals. Our model can analyse epidemics spreading in a single population or a metapopulation consisting of numerous weakly connected subpopulations. We analyse the outbreak of hybrid epidemics on both the individual level and the subpopulation level. In addition, the epidemic threshold condition for hybrid epidemics was also derived in this chapter. Predictions from the mathematical model are in accordance with the stochastic simulation results, both in single populations and in metapopulations.

Results from both numerical simulations and theoretical analysis in this chapter indicated that, in a metapopulation, a mix of local and global spreading mechanisms can significantly increase the outbreak of a hybrid epidemic than using either spreading mechanism alone. And in addition, there exists an optimal tradeoff between the two spreading mechanisms that enables a hybird epidemic to cause the largest outbreak in a metapopulation. The relation between the largest outbreak and the optimal epidemic parameters, in particular the tradeoff, is of practical significance and value because they may provide new clues for controlling hybrid epidemics. The next chapter will

investigate this in detail, studying how to optimise the spreading of a hybrid epidemic.

Chapter 5

Optimisation of hybrid epidemics

Previous work on hybrid epidemics has focussed on what we refer to as *non-critically* hybrid epidemics, where at least one of the spreading mechanisms alone is able to cause an epidemic outbreak, and a mixture of mechanisms brings no additional advantage.

Results in the previous chapter suggest there could exist an optimal hybrid tradeoff for hybrid epidemics in a metapopulation. This indicates the existence of a *critically* hybrid epidemic, where each individual spreading mechanism alone is unable to cause any significant spreading whereas the combination of such mechanisms leads to a huge epidemic outbreak. This is relevant to the design of a complex system [190, 191]. This chapter takes advantage of the model developed in the previous chapter to analyse how the hybrid tradeoff between the two spreading mechanisms affects the outbreak of a hybrid epidemic in greater detail.

5.1 Introduction

Hybrid spreading, the propagation of infectious agents using two or more alternative mechanisms, is a common feature of many real world epidemics. Widespread epidemics (e.g. computer worms) typically spread efficiently via local spreading through connections within a subpopulation, but also use global spreading to probe more distal targets typically with much lower effectiveness. In many cases, the amount of resources (e.g. time, energy or money) that an infectious agent can devote to all modes of propagation is limited. This chapter focusses on the tradeoff, α, between local and global spreading, and the effect of this tradeoff on the outbreak of an epidemic.

This chapter investigates the relationships between α, the relative weight given to each spreading mechanism, and the other epidemic properties. These properties include

epidemic infectivity, subpopulation structure, epidemic threshold, and the population reproduction number.

We investigate the maximum epidemic outbreak in the context of varying infectivity and recovery probabilities. For a given set of epidemic variables, we calculate the theoretical prediction of R_p as a function of α using Eq. (4.13), and then obtain the optimal α^* and the maximal R_p^*. For ease of analysis, we fix the global infection probability β_2 at a small value of 10^{-6} and then focus on the local infection probability β_1 and the recovery probability γ. The metapopulation used in the calculation is as in Fig. 4.4.

5.2 Optimal hybrid tradeoff

5.2.1 Numerical results for optimal hybrid tradeoff

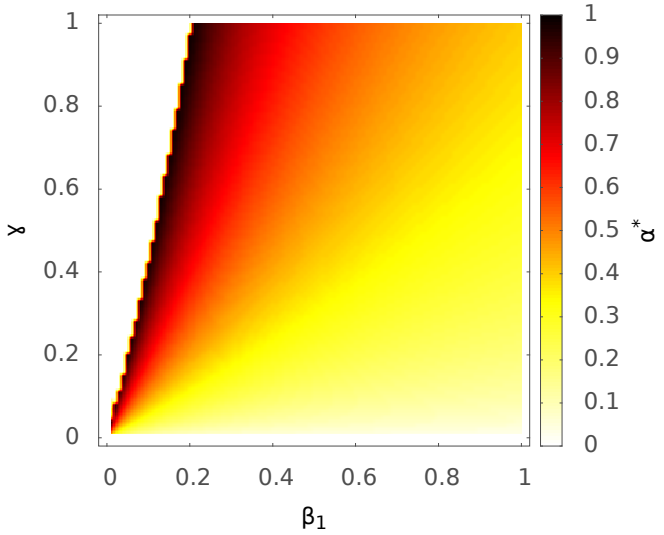

Figure 5.1: Estimated optimal hybrid tradeoff α^* for hybrid epidemics as a function of local infection probability β_1 and recovery probability γ. Epidemics spread in a metapopulation that contains 500 subpopulations and each subpopulation is a random network with 100 nodes and an average degree of five. The global infection probability $\beta_2 = 10^{-6}$.

Fig. 5.1 shows the optimal hybrid tradeoff α^* as a function of β_1 and γ. For a given γ, a larger β_1 results in a smaller α^*. Intuitively this can be understood as when the efficiency of local spread increases, less power needs to be devoted to this spreading mechanism, and more can be allocated to global spreading. On the other hand, for a given β_1, a larger γ results in an increase in α^*. When the recovery probability is higher, nodes remain infectious for shorter times. In this case, in order to achieve the

maximum epidemic outbreak, more local infection is favoured, since this will allow an infected subpopulation to remain infected for longer, and hence increase the probability of infecting other subpopulations before it recovers.

5.2.2 Method to predict optimal hybrid tradeoff

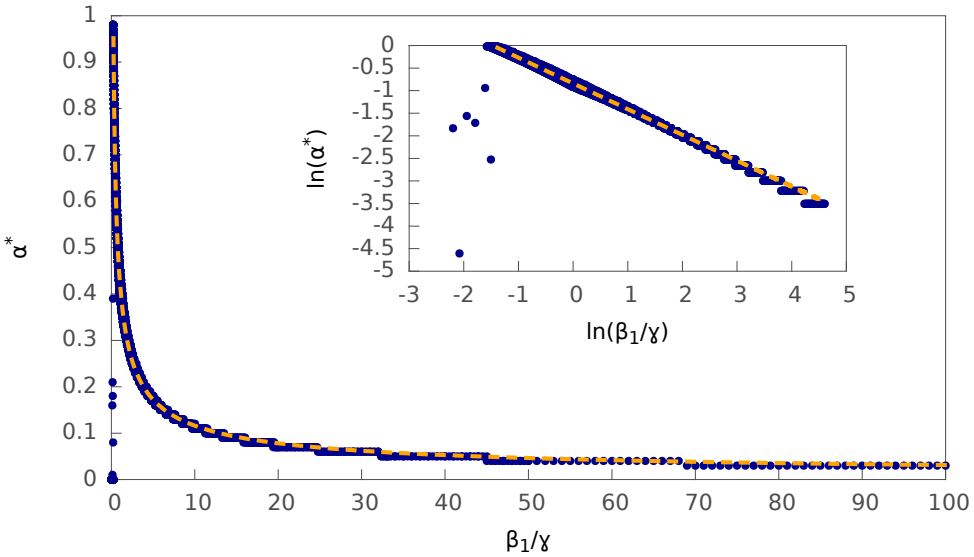

Figure 5.2: Estimated optimal hybrid tradeoff α^* for hybrid epidemics as a function of β_1/γ (local infection probability divided by recovery probability), where the inset is on log-log scale fitted as $ln(\alpha^*) = -0.84 - 0.57 \cdot ln(\beta_1/\gamma)$. Epidemics spread in a metapopulation that contains 500 subpopulations and each subpopulation is a random network with 100 nodes and an average degree of five. The global infection probability $\beta_2 = 10^{-6}$.

A plot of α^* versus β_1/γ is shown in Fig. 5.2. The fitting on the log-log scale in the inset indicates the two quantities have a power-law relationship, i.e. α^* is determined by β_1/γ. This means the optimal hybrid tradeoff α^* can be predicted when β_1/γ is known.

The finding that α^* is determined by β_1/γ is in line with the results in Fig. 5.1. Because Fig. 5.1 shows that each straight line starting from ($\beta_1 = 0, \gamma = 0$) has the same β_1/γ value and the same α^* value. We also observe that in Fig. 5.2 the power-law relationship does not hold when β_1/γ is close to 0; these outlier points however only represent extreme cases where there is no epidemic outbreak (β_1 is close to 0).

5.3 Maximal population reproduction number

5.3.1 As a function of local infection probability and recovery probability

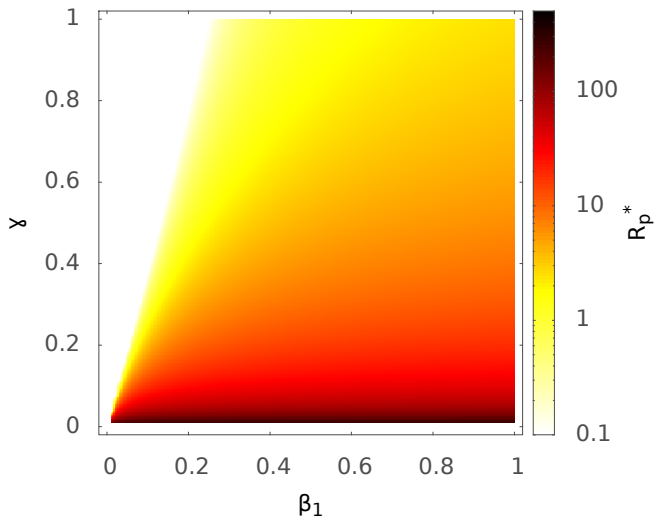

Figure 5.3: Estimated maximal population reproduction number R_p^* for hybrid epidemics as a function of local infection probability β_1 and recovery probability γ. Epidemics spread in a metapopulation that contains 500 subpopulations and each subpopulation is a random network with 100 nodes and an average degree of five. The global infection probability $\beta_2 = 10^{-6}$.

Fig. 5.3 shows the maximal R_p^* as a function of β_1 and γ, where the R_p^* is obtained when the corresponding value of α^* in Fig. 5.1 is used. R_p^* is very sensitive to the recovery probability γ. As γ approaches zero, the value of R_p^* increases dramatically (note that R_p^* uses a log-scale colour-map) regardless of value of β_1. This is consistent with the notion that a low recovery probability will favour any type of epidemic spreading. For a fixed γ, R_p^* increases with β_1. An increased infection probability of local spreading increases the reproductive number, if other parameters are kept constant, but the effect is much smaller than that of changing the recovery probability, because global spreading maintains the reproductive number when local spreading falls to lower values.

5.3.2 As a function of local infection probability and hybrid trade-off

Fig. 5.4 plots R_p as a function of β_1 and α on a log-log scale while fixing $\gamma = 0.1$. For given values of β_1, the corresponding optimal α^* are shown as points. The points

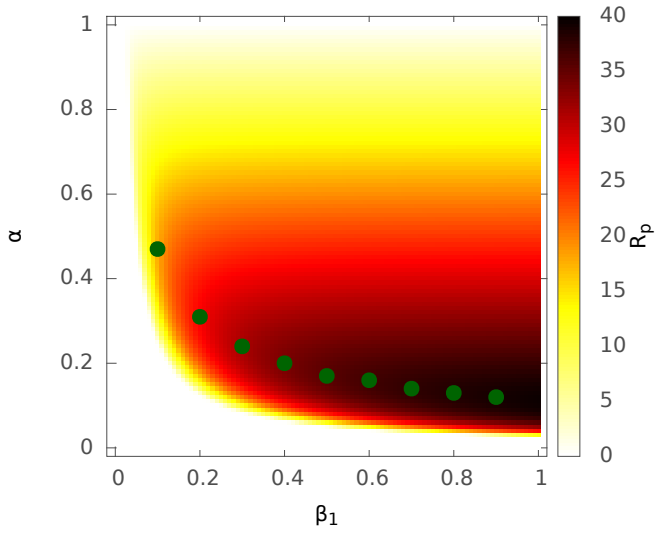

Figure 5.4: Estimated maximal population reproduction number R_p^* for hybrid epidemics as a function of hybrid tradeoff α and local infection probability β_1 with recovery probability $\gamma = 0.1$. The points are the corresponding optimal α^* for given β_1. Epidemics spread in a metapopulation that contains 500 subpopulations and each subpopulation is a random network with 100 nodes and an average degree of five. The global infection probability $\beta_2 = 10^{-6}$.

always fall in the area of the maximal R_p^* for the given β_1. Each point represents a local optimum. The global optimum, the largest possible value of R_p, is in the bottom-right corner, where the local infection probability is high but the epidemic expends the most power on global spreading. Infection across subpopulations can only be achieved by global spreading. Since global spreading has a low infection probability, the epidemic should spend most of its time (or resources) on global spreading. Considerably less time is spent on local spreading but its infection probability is high anyway.

Fig. 5.1 shows a clear phase shift between areas where an epidemic occurs (the coloured area) and areas where it does not (the white area towards the top-left corner). Accordingly, the corresponding R_p^* in Fig. 5.3 in the area where no epidemic occurs is very small. The boundary between the epidemic and non-epidemic phase is defined by the line $\beta_1/(\beta_1+\gamma-\gamma\beta_1) \approx 0.2$. This is the threshold for completely local spreading in a single-population: $\beta_1/(\beta_1+\gamma-\gamma\beta_1) > g_0'(1)/g_0''(1)$ and $g_0'(1)/g_0''(1) \approx 0.2$ for the network topology used. Since the global infection probability β_2 is fixed at a small value, no major spreading will occur either within or between subpopulations below this threshold.

5.4 Discussion

A better understanding of hybrid spreading mechanisms also has practical implications in the sense that it has opened up the opportunity for strategies to protect against the spread of epidemics. It is clear from both theoretical analysis and numerical simulations that epidemics can spread with extremely low global infection probabilities (far below individual recovery probabilities), provided there is efficient local infection. Such conditions are common for cyber epidemics (as computers within infected local networks tend to be more vulnerable to infection [142]). Protection strategies that target local networks collectively (for example intensive local vaccination around individual disease incidents, as was used in the final stages of smallpox eradication [192]) may be a key element for future strategies to control mixed spreading epidemics.

The above results are of practical relevance when the total amount of time or capacity that is allocated to spreading is limited by some resource constraint. For example, the total probing frequency of Internet worms is often capped at a low rate to prevent them from being detected by anti-virus software. Furthermore, other epidemic parameters, such as local or global infection probabilities can be difficult to change because they are derived from inherent properties of the infectious agent. For example it would be difficult to increase the global infection probability of an Internet worm. The tradeoff between different types of spreading therefore becomes a key parameter in terms of design strategy, which can be manipulated to maximise outbreak size.

5.5 Summary

This chapter demonstrated that it is indeed possible to have a highly contagious epidemic by combining ineffective spreading mechanisms. The properties of such epidemics are critically determined by the tradeoff at which the different spreading mechanisms are mixed, and usually there is an optimal tradeoff that leads to a maximal outbreak size. In addition, we provided the method to predict the optimal tradeoff when other epidemic parameters (including the infection probabilities, recovery probability, and the metapopulation structure) are known.

Results in this chapter provide new strategies for maximising beneficial epidemics (e.g. disseminating information) and estimating the worst outcome of damaging hybrid epidemics (e.g. computer worms). The results can be utilised in practise to design

more efficient information dissemination systems through integrating multiple spreading mechanisms and optimising the tradeoff between them.

Chapter 6

Analysis of the computer worm Conficker

Conficker is a computer worm that exploded on the Internet in 2008. It is unique in combining three different spreading strategies: local probing, neighbourhood probing, and global probing. This chapter proposes a mathematical model that combines these three modes of spreading to capture the worm's spreading behaviour. The parameters of the model are inferred directly from network data obtained during the first day of the Conifcker epidemic. The model is then used to explore how the trade-off between spreading modes determines the worm's effectiveness. Our results show that the Conficker epidemic is an example of a critically hybrid epidemic, in which the different modes of spreading in isolation do not lead to successful epidemics. Such hybrid spreading mechanisms may be used beneficially to provide the most effective strategies for promulgating information across a large population. When used maliciously, however, they can present a dangerous challenge to current internet security protocols.

6.1 Introduction

Each computer on the Internet is associated with an Internet Protocol (IP) address. Conficker views the Internet as a meta-population, where computers are located in subpopulations, i.e. Local Area Networks (LAN), each consisting of computers whose IP addresses share the same prefix.

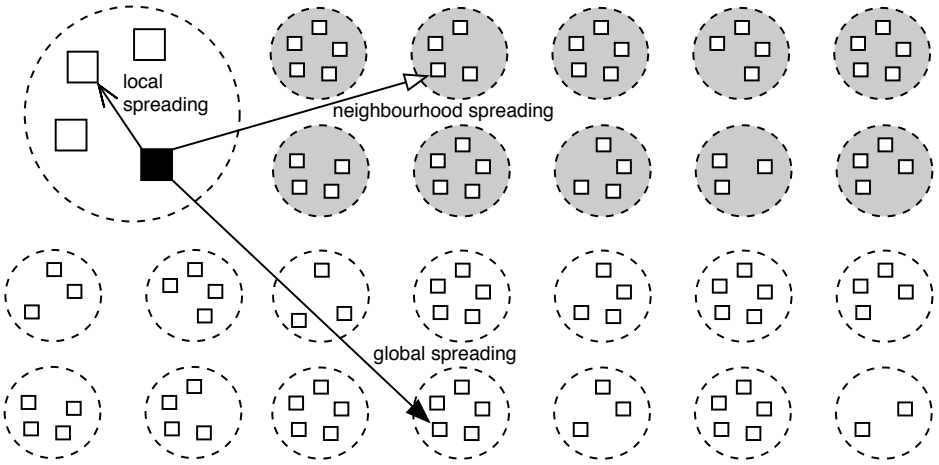

Figure 6.1: Conficker's three probing strategies: (1) global spreading, where it probes any computer on the Internet at random; (2) local spreading, where it probes computers in the same local network; (3) neighbourhood spreading, where it probes computers in ten neighbouring local networks.

6.1.1 Conficker spreading mechanisms

There are several versions of the Conficker worm on the Internet. Here we only study the first version of the worm whose initial outbreak was captured by the Internet measurement data [35, 45] collected by the Center for Applied Internet Data Analysis (CAIDA) in 2008.

According to the computer security company Symantec [26], Conficker uses three spreading mechanisms (see Fig. 6.1):

- *Global* spreading, where the worm probes computers with random IP addresses on the Internet;
- *Local* spreading, where the worm on an infected computer probes computers in the same Local Area Network (LAN) with the same IP address prefix;
- *Neighbourhood* spreading, where it probes computers in ten neighbouring LANs (with smaller consecutive IP address prefixes).

6.1.2 Related works

Porras et al. [193] conducted a detailed analysis on the inner working flow of a *single* Conficker worm by reverse engineering the captured Conficker worm sample. This study also monitored the network activities of a host (computer) infected by the Conficker worm. In addition, Porras et al. provided some potential evidence about the

origin of the worm.

Both Li et al. [42] and Yao et al. [43] conducted *population-level* modelling and analysis of the Conficker worm spreading. Li et al. [42] used a small self collected dataset and Yao et al. [43] used the Conficker dataset from the Center for Applied Internet Data Analysis (CAIDA) network telescope database [35, 45]. Both studies only consider global spreading of the worm, ignoring its localised spreading behaviours. Weaver [194] also conducted population-level modelling and analysis for the third Conficker worm variant - Conficker C that was detected on 20 February 2009.

Shin et al. [21] analysed the distribution of IP addresses infected by the Conficker worm based on a self collected dataset including 24,912,492 unique victim IP addresses captured from 1 January 2010 to 8 January 2010. In their work, the distribution of these victim IP addresses were calculated according to a number of measures including the countries they were located in, the Internet Service Providers they were using, and so on. Irwin [195] conducted a similar study based on a different dataset of the Conficker worm. Its dataset was collected between August 2005 and September 2009, using a monitoring network located within South Africa. Irwin [195] also provided statistics about several properties of received attack packets, including their Time to Live (TTL), size, and operation system. These two works [21, 195] do not model the spreading dynamics of the Conficker worm.

Most of the previous modelling studies on the Conficker worm spread assume the worm only conducts global spreading, ignoring the two localised spreading mechanisms: local and neighbourhood spreading. While existing theoretical models might be useful in estimating the total outbreak size of Conficker, they do not accurately replicate its spreading mechanisms. Consequently, they are limited in their ability to provide further insight and estimation for spreading dynamics of the worm should future variants of the worm alter the combination of different spreading mechanisms. In addition, the parameters of Conficker's hybrid spreading and how they affect the epidemic dynamics of the worm can help explain why the worm is so contagious but have not been the focus of previous studies and so they have garnered little attention.

6.2 Our model of Conficker spreading

Here we use a node to represent an IP address at which a computer (or computers) connects to the Internet. For convenience, a sub-population (or a LAN) is referred to as a subnet. A node must be in one of three states: *susceptible*, *infected*, and *recovered* [3]. Initially only a small number of nodes are infected and all others are susceptible. At each time step, an infected node attempts to spread the worm to susceptible nodes using one of the three probing strategies:

- Global spreading with probability α_g, where the worm probes nodes on the Internet at random with the global infection rate $\beta_g \in [0,1]$.

- Local spreading with probability α_l, where it probes nodes in the local subnet with the local infection rate $\beta_l \in [0,1]$;

- Neighbourhood spreading with the probability α_n, where it probes nodes in ten neighbouring subnets with the neighbourhood infection rate $\beta_n \in [0,1]$;

The mixing probabilities satisfy $\alpha_g + \alpha_l + \alpha_n = 1$.

An infected node is recovered with recovery rate $\gamma \in [0,1]$. A recovered node remains recovered and cannot be infected again. Note that for mathematical analysis, the mixing probabilities could be incorporated into the infection rates. However, this study has treated them as separate parameters, considering that an infection rate reflects inherent properties of a computer worm in the context of a specific target population, whereas mixing probabilities are settings that can be easily modified in the worm's code. The mixing probabilities are used as controlling parameters in the study while the other parameters are kept the same.

Only nodes that can potentially be infected by Conficker are relevant to the study. They are referred to as the *relevant* nodes. A subnet is relevant if it contains at least one relevant node. Irrelevant nodes include unused IP addresses and those computers that do not have the vulnerabilities that the worm can exploit. Note that although the irrelevant nodes and subnets do not participate in the spreading of Conficker, they will be probed by the worm as the worm does not have a priori knowledge regarding vulnerable nodes.

Let n represent the total number of relevant nodes and N the number of relevant subnets. The average number of relevant nodes in a subnet is $n_N = n/N$. Let N^+ represent the average number of relevant subnets in ten neighbouring subnets.

At time t, the total number of susceptible, infected, and recovered nodes at time t are $S(t)$, $I(t)$, and $R(t)$, respectively. Then the average number of infected nodes in a subnet is $I_N(t) = I(t)/N$, and the average number of infected nodes in ten neighbouring subnets is $I^+(t) = I_N(t)N^+$. Hence on average a susceptible node can be infected via (1) global probing by $I(t)$ infected nodes in the Internet; (2) local probing by $I_N(t)$ infected nodes in the local subnet; (3) neighbourhood probing by $I^+(t)$ infected nodes in the neighbouring subnets.

The average probabilities that a susceptible node is *not* infected by the global, local and neighbourhood probing are respectively:

$$\begin{aligned} P_g(t) &= (1-\alpha_g\beta_g)^{I(t)} \\ P_l(t) &= (1-\alpha_l\beta_l)^{I_N(t)} \\ P_n(t) &= (1-\alpha_n\beta_n)^{I^+(t)}. \end{aligned} \tag{6.1}$$

The average probability of *not* being infected by any probing is $P(t) = P_g(t)P_l(t)P_n(t)$. Thus the discrete evolution of Conficker spreading can be described as:

$$\begin{aligned} S(t+1) &= S(t) - S(t)[1-P(t)] \\ I(t+1) &= I(t) + S(t)[1-P(t)] - \gamma I(t) \\ R(t+1) &= R(t) + \gamma I(t) \end{aligned} \tag{6.2}$$

where $S(t)[1-P(t)]$ is the number of new infections at step t.

Compared with the hybrid epidemic model in Chapter 4, the epidemic model here incorporates an additional spreading mechanism - neighbourhood spreading. We here do not aim to obtain the analytical solutions for the final outbreak size and the epidemic threshold of this model. The theoretical predictions in this chapter are all numerically and iteratively calculated.

6.3 Inference of model parameters from measurement data

We infer the parameter values of our Conficker model from the Internet measurement data [35, 45] collected by the Center for Applied Internet Data Analysis (CAIDA) in 2008. This is the only publicly available dataset that has captured the *initial* outbreak process of the worm. The CAIDA Network Telescope project [35,45] monitors Internet traffic sent to a large set of *unusable* IP addresses, which account for around $1/256$ of all addresses. No legitimate traffic should be sent to these monitored addresses because they are not allocated for normal usage [196]. Thus the traffic data captured by this project provides important insight into various abnormal behaviours on the Internet.

When Conficker spreads on the Internet, its global spreading mechanism sends out probing packets to randomly generated IP addresses, some of which are unused IP addresses and therefore are monitored by the Network Telescope project. Conficker's probing packets are characterised by the Transmission Control Protocol (TCP) with destination port number 445. This feature can be used to distinguish Conficker packets from other packets in the Network Telescope data.

For each record of Conficker's probing packet, we are interested in two things: (1) the time when the packet is monitored by the Network Telescope project, and (2) the packet's source IP address, which gives the location of a Conficker-infected node. The destination address is ignored, as it is a randomly-generated, unused IP address.

We make use of the Network Telescope project's dataset collected on 21 November 2008, the day of the Conficker outbreak on the Internet. Before the outbreak day, the nearest two days with datasets available are 12 and 19 November 2008. The datasets on these two days are used to filter out background "noise" that was already occurring before the outbreak. That is, in the outbreak dataset, packets that were sent from any source address that had already sent packets to any of the unusable addresses in the two earlier datasets are discarded in the analysis. The prefix of /24 (i.e. IP address mask of 255.255.255.0) is used to distinguish different subnets [21]. The analysis utilises a 10-minute window.

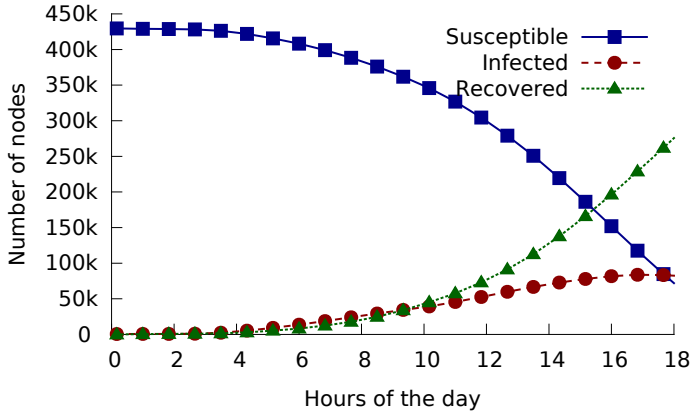

Figure 6.2: Numbers of susceptible nodes $S(t)$, infected nodes $I(t)$ and recovered nodes $R(t)$ as a function of time t, as inferred from CAIDA's dataset on 21 November 2008, the day of Conficker's outbreak.

6.3.1 Node status at a given time

The status of each node at time t from the CAIDA data is first identified. On the day of the Conficker outbreak, all relevant nodes are initially susceptible. A node is assumed to be infected by the worm when it has first been observed to have a Conficker probing packet coming from it; and the node is recovered when its last probing packet is observed before the end of the day. Fig. 6.2 shows the number of susceptible, infected and recovered nodes as observed in a 10-minute window.

6.3.2 New infections by each spreading mechanism

Let $dI_l(t)$, $dI_n(t)$ and $dI_g(t)$ represent the numbers of newly infected nodes through local, neighbourhood and global spreading, respectively, at time step t. Our analysis on the data suggests that 84% of new infections occur within already infected subnets or their neighbourhood subnets, i.e. only 16% of new infections appear outside the reach of local and neighbourhood probing. This agrees with the understanding that local and neighbourhood probing are significantly more effective than global probing [21]. And 73% of those new infections within the reach of local and neighbourhood probing (i.e. $73\% \times 84\%$ of all new infections) occur in already infected subnets. This finding suggests that local probing is more effective than neighbourhood probing. Based on the above analysis we can then approximately identify the probing mechanism that is responsible for a newly infected node by analysing the state of other relevant nodes at the time when the new infection occurs.

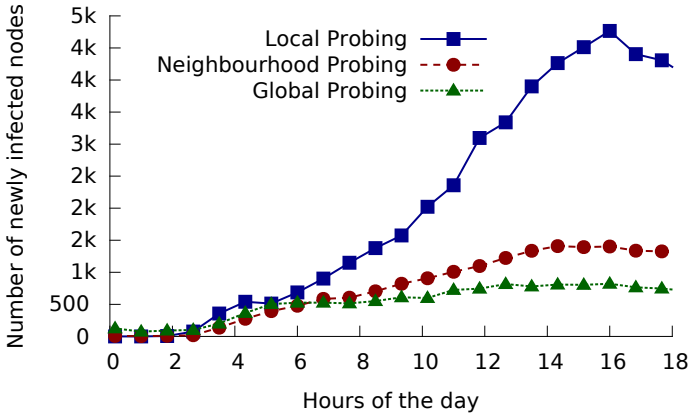

Figure 6.3: Numbers of nodes newly infected by Conficker via each of the three spreading mechanisms in 10-minute windows on the day of Conficker's outbreak, as inferred from CAIDA's dataset on 21 November 2008.

- IF there is an infected node already in the same subnet, the new infection is caused by that infected node via local spreading.
- ELSE IF there is an infected node in the 10 neighbouring subnets, then the new infection is via neighbourhood spreading.
- OTHERWISE, the newly infected node is infected via global spreading.

Fig. 6.3 shows the results, plotting the number of new infections caused by each spreading mechanism as a function of time.

6.3.3 Epidemic parameters of Conficker

Based on the above results, the average probabilities that a susceptible node at time t is *not* infected by the local, neighbourhood and global probing can be calculated as, respectively:

$$P_l(t) = 1 - \frac{dI_l(t)}{S(t)}, P_n(t) = 1 - \frac{dI_n(t)}{S(t)}, P_g(t) = 1 - \frac{dI_g(t)}{S(t)}. \tag{6.3}$$

$I_N(t)$ and $I^+(t)$ can also be calculated from the data according to their definitions. We define the effective infection rate as $b_l = \alpha_l\beta_l$, $b_n = \alpha_n\beta_n$ and $b_g = \alpha_g\beta_g$. Then we can use Eq. (6.1) and Eq. (6.3) to calculate b_l b_n and b_g.

Let λ denote the average total number of probes an infected node conducts during each time step. The average number of local, neighbourhood and global probes in a time step are $\alpha_l\lambda$, $\alpha_n\lambda$, and $\alpha_g\lambda$ respectively. The number of nodes (relevant and

irrelevant) probed by the local, neighbourhood and global probing are 256, 10×256 and 2^{30} (it is not 2^{32} due to a bug in the worm's random number generation algorithm [26]). We can express the effective infection rates as:

$$b_l = \alpha_l \lambda / 256,\ b_n = \alpha_n \lambda / 2560,\ b_g = \alpha_g \lambda / 2^{30}. \tag{6.4}$$

By solving Eq. (6.4) together with $\alpha_g + \alpha_l + \alpha_n = 1$, λ, α_l, α_n and α_g are obtained. Then the infection rates of three spreading mechanisms can be calculated as $\beta_l = b_l/\alpha_l$, $\beta_n = b_n/\alpha_n$, and $\beta_g = b_g/\alpha_g$. And the recovery rate can be obtained as $\gamma = dR(t)/I(t)$, where $dR(t) = R(t+1) - R(t)$.

6.3.4 Inference results and evaluation

The inferred values of the Conficker model parameters are shown in Table 6.1, including the mixing probability α and the infection rate β for the three spreading mechanisms, the recovery rate γ, the recovery time $\tau = 1/\gamma$ which is the average time it takes for an infected node to recover, and the probing frequency λ. The parameter values are averaged over time windows between 04:00 and 16:00 when the spreading behaviour is stable. Computers are online and offline on a daily basis following a diurnal pattern [156]. This factor only has a marginal impact on our results.

Table 6.1: Conficker parameters inferred from data.

Global spreading	$\alpha_g = 89.1\%$	$\beta_g = 7.7 \times 10^{-8}$
Local spreading	$\alpha_l = 5.3\%$	$\beta_l = 0.32$
Neighbourhood spreading	$\alpha_n = 5.6\%$	$\beta_n = 0.032$
Recovery rate	$\gamma = 0.064$	
Recovery time	$\tau = 156$ mins	
Probing frequency	$\lambda = 82.5 per 10 mins$	

All parameters are measured in a 10-minute window.

We observe in the data that the worm has infected in total n=430,135 nodes, which are located in N=92,267 subnets. On average, each subnet has n_N=4.7 relevant nodes, and N^+=4.3 of ten neighbouring subnets are relevant.

With these parameter values, our Conficker model (see Eq. (6.2)) can be used to theoretically predict the worm's outbreak process. As measured from the data, the number of nodes in the three statuses are $S = 423,899$, $I = 3,945$, and $R = 2,291$ at 04:00. The prediction begins at 04:00 and uses these numbers as the initial condition.

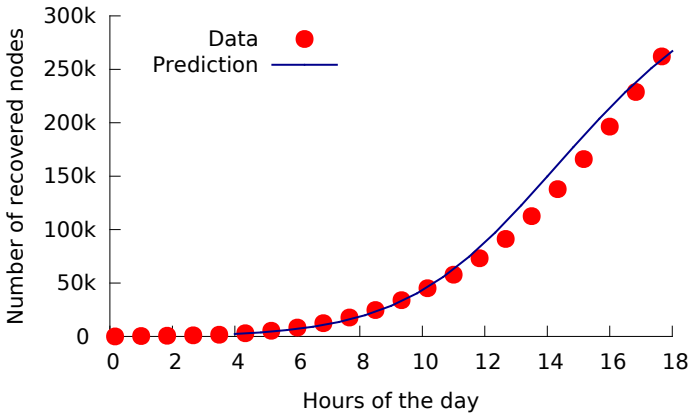

Figure 6.4: Outbreak of computer worm Conficker. Points are measured from Network Telescope's dataset collected on the outbreak day. The curve is a theoretical prediction from our Conficker model using the inferred parameters.

As shown in Fig. 6.4, our model predictions closely match the measurement data.

The inferred parameters are consistent with expectations. For example, local spreading has a high infection rate because if a computer is already infected, then other computers in the same subnet are likely to have a similar computer system and thus are also likely to be vulnerable to the worm. By comparison, global spreading has an extremely low infection rate. On average, more than 10 million global probings will produce only a single new infection. On average an infected node retains its status for approximately 2.5 hours (156 mins) before it recovers (e.g. switched off or updated with new anti-virus database). The worm only sends out 8 probing packets per minute. Such a deliberately low probing rate helps the worm to evade a computer's or network's security systems.

6.4 Dynamics of hybrid spreading

6.4.1 Mixing two spreading mechanisms

Simulations are run using our Conficker model with the parameter values derived from the previous analysis. The simulation network has 100k subnets. Each subnet contains five relevant nodes and has 4 relevant adjacent subnets. This topology setting resembles the spreading network observed in the Conficker data. Initially two random nodes are infected. The only controlling (variable) parameters are the mixing probabilities of the spreading mechanisms. Simulation results on mixture of two spreading mechanisms are shown in Fig. 6.5.

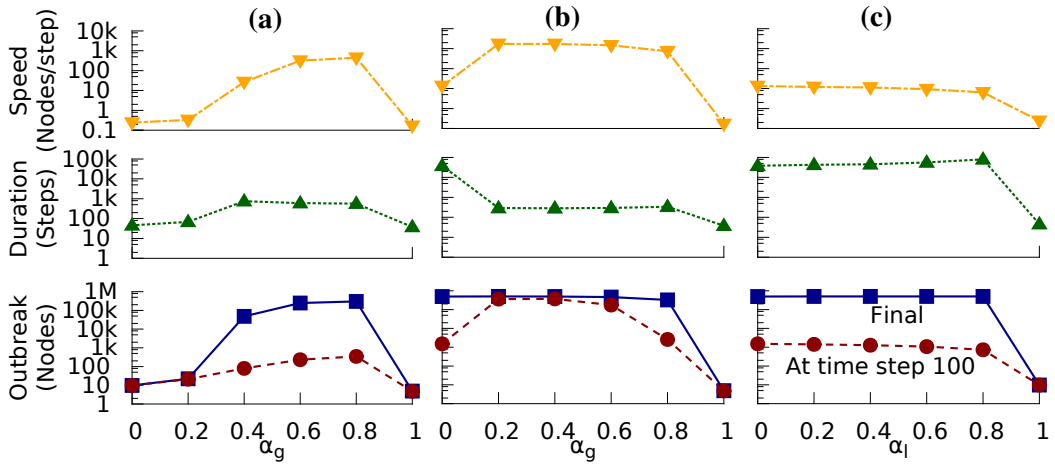

Figure 6.5: Simulation results for the mix of Conficker's two spreading mechanisms with different mixing probabilities. **(a)** Mix of global (α_g) and local ($1-\alpha_g$) mechanisms; **(b)** Mix of global (α_g) and neighbourhood (1-α_g) mechanisms; **(c)** Mix of local (α_l) and neighbourhood (1-α_l) mechanisms. In each case, the outbreak size, the total duration of the spreading, and the speed of spreading are measured. The outbreak results include both the final outbreak size (square) and the outbreak size at time step 100 (filled circle). Each data point is averaged over 100 simulation runs. Note the y axes are all logarithmic.

Fig. 6.5a shows that as explained above, global spreading or local spreading alone cannot cause an outbreak, whereas a mixture at a ratio of 0.8 to 0.2 produces a large and rapid outbreak. Fig. 6.5b shows that the neighbourhood spreading alone ($\alpha_g = 0$) can cause a large, but very slow outbreak, whereas the mix of neighbourhood spreading with just a small amount of global spreading can dramatically accelerate the spreading process. Fig. 6.5c shows that adding local spreading to neighbourhood spreading slows down the spreading process considerably. When they are mixed at the ratio of 0.8 to 0.2, the spreading reaches the same final outbreak size but the whole process lasts for the longest time.

6.4.2 Mixing three spreading mechanisms

Simulation results on mixing three spreading mechanisms are shown in Fig. 6.6. Fig. 6.6a shows it is not difficult to achieve a large final outbreak size when the three mechanisms are all present and local spreading or global spreading are not dominant. Fig. 6.6b shows spreading will last for a longer time if there is less global probing. Fig. 6.6c shows that the most contagious variation of the worm is a mix of global, local and neighbourhood spreading at the probabilities of 0.4, 0.2 and 0.4 (see circle on the plot), which causes the largest final outbreak with the highest spreading speed.

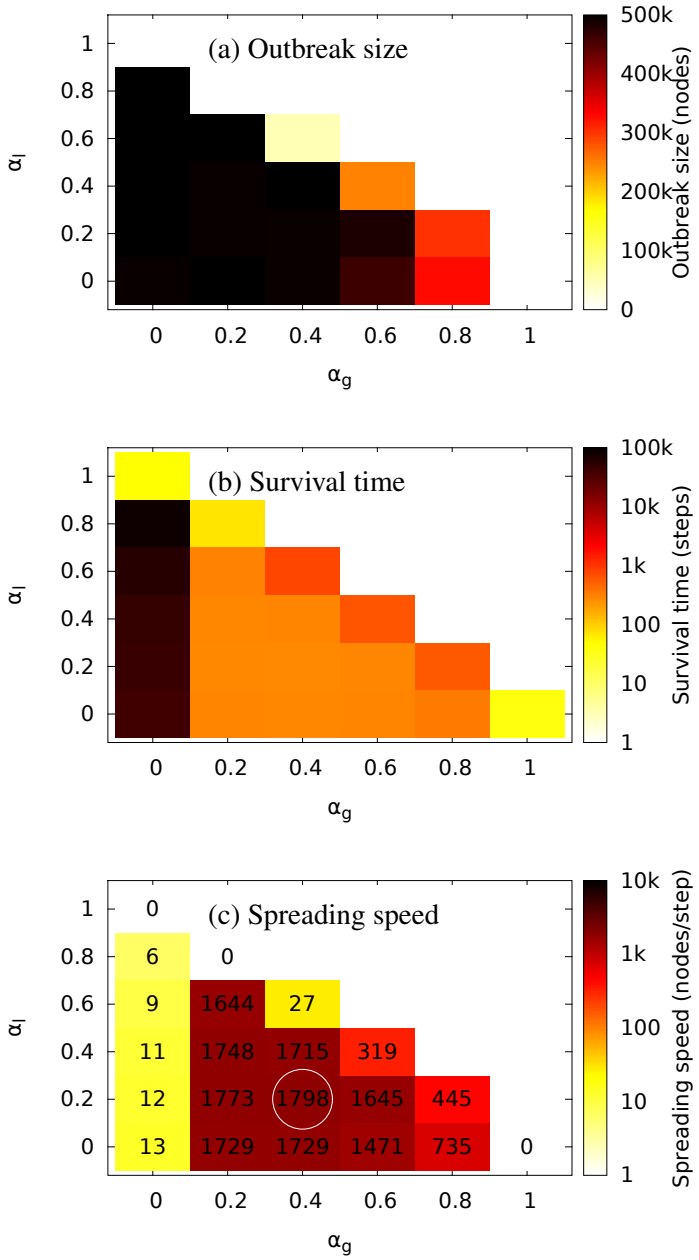

Figure 6.6: Simulation results when three of Conficker's spreading mechanisms are mixed at different probabilities. Spreading properties shown include the final outbreak size, the survival time and the spreading speed (see colour maps) as functions of the mixing probabilities of global spreading α_g (x axis) and local spreading α_l (y axis), where the mixing probability of neighbourhood spreading is $\alpha_n = 1 - \alpha_g - \alpha_l$.

6.5 Discussion

In this chapter, the epidemic spreading parameters of the Conficker worm are inferred from observed data collected during the first few hours of the epidemic. Simulations of worm spreading, based on these parameters, allow us to reach some important conclusions about the worm's use of hybrid spreading mechanisms.

6.5.1 Advantage of hybrid spreading mechanism

Conficker's global probing is extremely ineffective. The infection rate of global probing is many orders of magnitude smaller than the recovery rate. This means, if Conficker used only the global probing, it would not have caused any significant infection on the Internet at all.

Local probing has a remarkably high infection rate, $\beta_l = 0.32$, which means when an infected node conducts only local spreading, a susceptible node in the same subnet has an 1/3 chance of being infected in a step (10-mins). However, local probing is confined within a subnet. If the worm used only local probing, it would not have infected any other subnets apart from those containing initially infected nodes.

Neighbourhood probing is constrained to a neighbourhood of ten subnets. It has a high infection rate because computers in adjacent IP address blocks often belong to the same organisation with similar computer systems and therefore similar vulnerabilities that can be exploited by the worm. Since different nodes' neighbourhoods can partially overlap with each other, in theory it is possible for the worm to reach any node in the whole meta-population by using only neighbourhood probing. Such process, however, would be extraordinarily slow as demonstrated in Fig. 6.5b.

In summary, if Conficker used only a single spreading mechanism, it would have vanished on the Internet without causing any significant impact.

Thus the enormous outbreak of the worm lies in its ability to do two things. Firstly, it needs to devote great effort to explore every corner of the Internet to find a new vulnerable computer. Every new victim will open a new colony full of similarly vulnerable computers. Secondly, it needs to make the most out of each new colony.

This is exactly what Conficker does. It allocates most of its time on global probing with a mixing probability of $\alpha = 89\%$. This to a certain extent compensates for the ineffectiveness of global probing. Although the worm allocates small amounts of time

on local and neighbouring probing, their high infection rates allow them to exploit all possible victims in subnets with efficiency. And all newly infected nodes join the collective effort to flood the Internet with more global random probes.

In short, the Conficker worm is an example of a critically hybrid epidemic. It has the ability to cause an enormous outbreak not because it has an advanced ability to exploit weaknesses of a computer, but because it has a remarkable capacity for discovering all potentially vulnerable computers on the Internet, i.e. it is not the infectivity, but the hybrid spreading that makes Conficker one of the most infectious worms on record.

6.5.2 Challenges of critically hybrid epidemics

The analysis of critically hybrid epidemics such as Conficker has important general implications. Firstly, it demonstrates that it is possible to design a high impact epidemic based on mechanisms, each of relatively low efficiency. Indeed results in Fig. 6.6c suggest that Conficker could have had a larger outbreak with higher speed if it had used a different set of mixing probabilities, which requires only a few lines of Conficker's program code to be changed. Hybrid mechanisms may therefore be ideal for rapid efficient penetration of a network, for example in the context of an advertising campaign or in order to promulgate important public health or security information. An interesting example might be the use of media campaigns (global spreading) where the reader or viewer is specifically requested to pass on a message via Twitter or Facebook to their "local" group contacts.

Conversely, malicious hybrid epidemics can be extremely difficult to defend against, and many existing defence strategies may not be effective. For example, immunising a selected portion of a local population in order to isolate and hence protect the vulnerable nodes will not be effective, because vulnerable nodes can still be found by the worm through random global spreading.

Another possible preventative measure is to reduce the average time it takes for an infected node to recover, for example to speed up the release of anti-virus software updates or increase the frequency of security scanning on computers. Our theoretical predictions (using Eq. (6.2)) in Fig. 6.7 show that the final outbreak size (in terms of total recovered nodes) does not change significantly when the recovery time is reduced

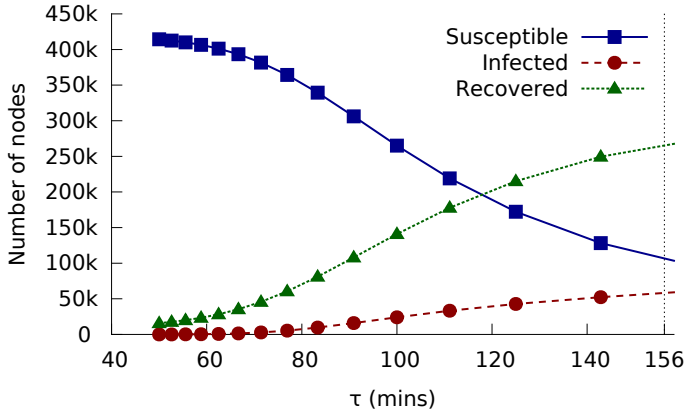

Figure 6.7: Predicted numbers of susceptible, infected and recovered nodes at 16:00 on the outbreak day as a function of the recovery time τ, which is the average time for an infected node to recover. Conficker's recovery time is 156 minutes.

from 156 minutes to 140 or 120 minutes. In practice, even achieving such reductions would represent a remarkably difficult technical challenge. It is clear from the discussion above that epidemics can spread with extremely low global infection rates (far below individual recovery rates), provided there is efficient local infection. The extremely efficient spreading achieved once a given subnet or a set of subnets has been penetrated is therefore obviously a key determinant of the worm's outbreak [21]. Thus, defence strategies that focus on security co-operation between nodes with a local network neighbourhood (a "neighbourhood watch" strategy [21]) may be the key to the prevention of similar outbreaks in the future.

In conclusion, the study uses data collected during the first day of the Conficker epidemic to parametrise a hybrid model to capture the worm's spreading behaviour. The study highlights the importance of mixing different modes of spreading in order to achieve large, rapid and sustained epidemics, and suggests that the trade-off between the different modes of spreading will be critical in determining the epidemic outcome.

6.6 Summary

This chapter introduced a mathematical model to describe the spreading behaviour of a critically hybrid epidemic - computer worm Conficker. The study was based on measurement data provided by the Center for Applied Internet Data Analysis (CAIDA)'s Network Telescope project [35, 45], which monitors Internet traffic anomalies. Using the proposed Conficker spreading model, we inferred the worm's epidemic parameters

from data.

The theoretical predictions in this chapter closely reproduced the outbreak process of Conficker. We demonstrated that the worm might be able to spread faster, reach a larger outbreak size or survive for longer time by just revising the ratios at which the worm allocated its time on each of the spreading mechanisms (while keeping everything else the same), which can be easily achieved by changing a few lines of code.

This study has both theoretical and practical significance for understanding and tackling hybrid epidemics: Firstly, it is amongst the first studies to explore a real-life critically hybrid epidemic, where the epidemic's parameter values are inferred from measurement data. Secondly, it highlighted the complex interactions among Conficker's three spreading mechanisms, and demonstrated how the worm can be more contagious if it combines its three spreading mechanisms in an optimal way.

Chapter 7

Modelling HIV infection within the human body

Human Immunodeficiency Virus (HIV) can propagate from infected to susceptible cells via two mechanisms: cell-free infection following fluid-phase diffusion of virions released by infected cells and by highly-efficient direct cell-to-cell transmission at cell contacts. The contribution of this hybrid spreading mechanism, which is also a characteristic of many computer worm outbreaks, to HIV progression in vivo remains unknown. In this chapter, we introduce a mathematical model of HIV dynamics that explicitly incorporates hybrid spreading. The model is validated against clinical data from a cohort of untreated patients.

7.1 Introduction

HIV is a virus that attacks the human immune system - the defence that protects the body against disease. Without medical treatment, people infected by HIV can develop Acquired Immune Deficiency Syndrome (AIDS), which is a condition whereby the immune system fails. AIDS exposes a patient to opportunistic infections (caused by viruses, bacteria, parasites, fungi, etc.) and threatens his/her life [161].

Tremendous effort has been spent in understanding and treating the virus over the past three decades. State-of-the-art HIV treatments can manage and control the virus's infection, but there is still no permanent cure for the disease [25]. This calls for further collaboration among the research communities and commitment from civil societies to tackle the disease.

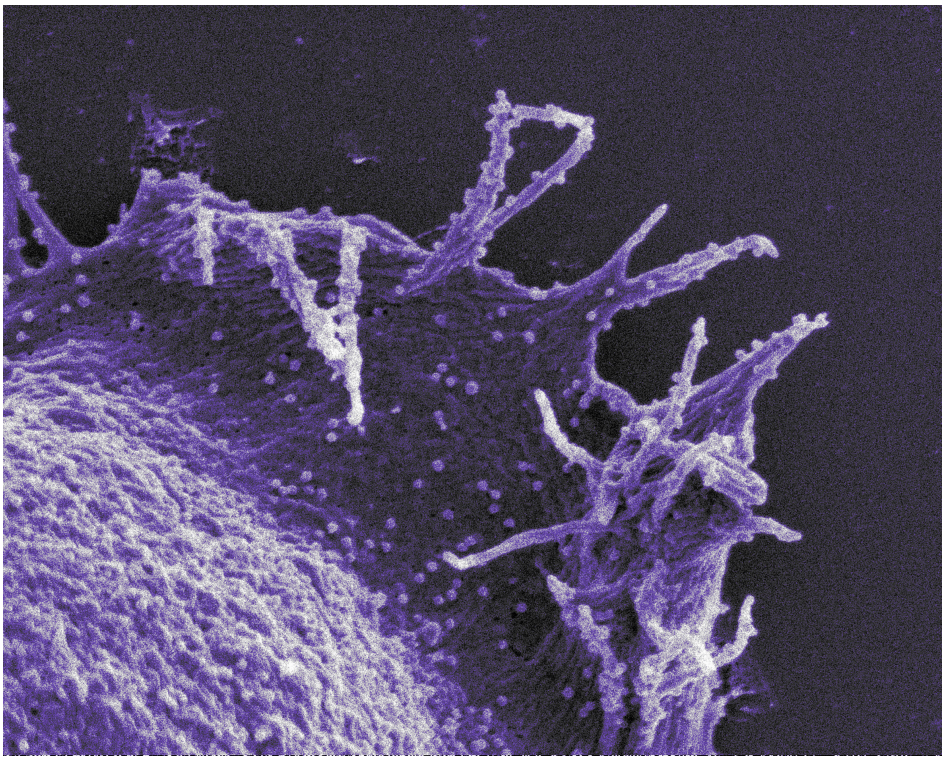

Figure 7.1: Scanning electron micrograph of HIV viruses budding from an infected $CD4^+$ T cell. Viruses are the dots in the figure. Such viruses travel freely in blood and body fluid. When they encounter a healthy (susceptible) T cell, they may infect the cell via the cell-free spreading. Image provided by Dr. Clare Jolly.

7.1.1 HIV spreading mechanisms

HIV primarily infects $CD4^+$ T cells inside the human body. $CD4^+$ T cells are helper cells that play an important role in the functioning of the immune system. To simplify the expression, in this thesis, unless otherwise stated, we use "T cells" to refer to $CD4^+$ T cells. The virus is now known to spread between T cells via two parallel routes: cell-free spread and cell-to-cell spread.

According to the classical model of HIV spread, as shown in Fig. 7.1, the virus particles bud from an infected T cell, enter the blood/extracellular fluid and then can infect another T cell following a chance encounter (termed cell-free spread). Because diffusion of virus particles is much faster than cell migration, and there is extensive flow of blood and fluid, this mode of spreading can be characterised by a well mixed epidemic spreading model. In this scenario, the probability of infection for a particular cell will be proportional to the concentration of extracellular infectious virus.

HIV infection can also transmit through direct cell-to-cell contacts between T cells. Two pathways of cell-to-cell transmission have been reported. Firstly, as il-

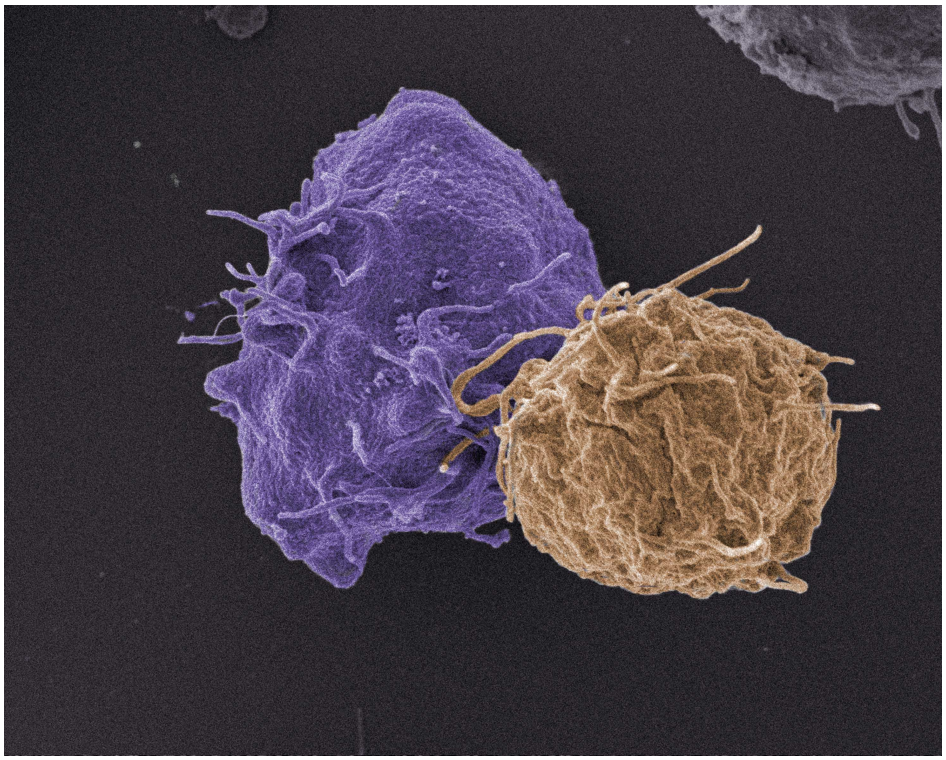

Figure 7.2: Scanning electron micrograph of a virological synapse between $CD4^+$ T cells. When a healthy T cell (light-yellow-coloured) forms a synapse with an infected cell (purple-coloured), HIV can be transmitted directly to the healthy cell via the cell-to-cell spreading. Image provided by Dr. Clare Jolly.

lustrated in Fig. 7.2, an infected T cell can transmit the virus directly to a target T cell via a virological synapse [128–130]. Secondly, an antigen presenting cell (APC) can also transmit HIV to T cells by a process that either involves productive infection or capture and transfer of virions *in trans* [128]. Whichever pathway is used, infection by cell-to-cell transfer is reported to be much more efficient than cell-free virus spread [197–199]. A number of factors contribute to this increased efficiency, including polarised virus budding towards the site of cell-to-cell contact, close apposition of cells which minimises fluid-phase diffusion of virions, and clustering of HIV entry receptors on the target cell to the contact zone [128, 129]. Cell-to-cell spread is thought to be particularly important in lymphoid tissues where $CD4^+$ T lymphocytes are densely packed and likely to frequently interact. Indeed, intravital imaging studies have supported the concept of the HIV virological synapse *in vivo* [200, 201].

The cell-free spreading resembles the global spreading analysed in previous chapters and the cell-to-cell spreading resembles the local spreading. There is no neighbourhood spreading mechanism in this HIV infection model.

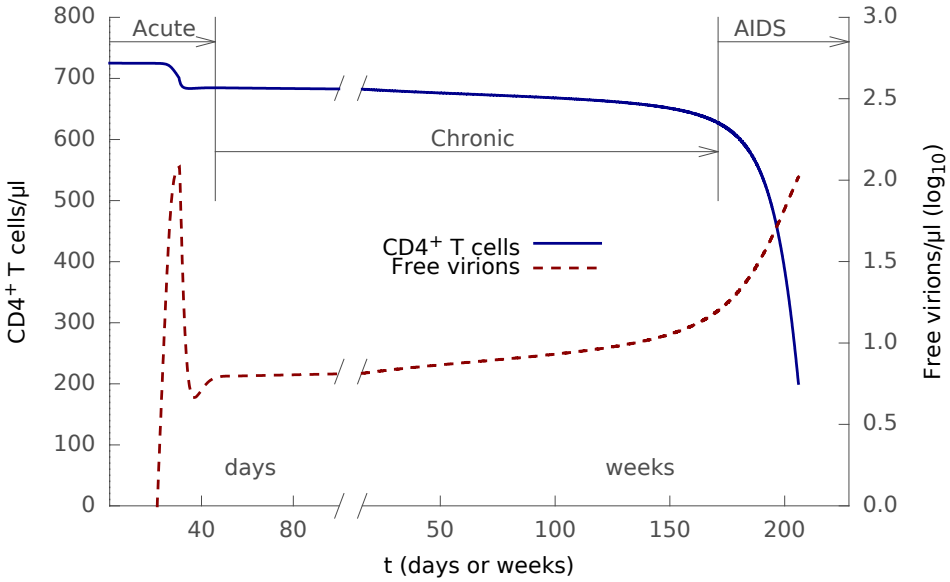

Figure 7.3: A typical HIV infection course including three phases. $CD4^+$ T cell density (left y axis) and virus load (right y axis) are two important clinical measurements that are regularly monitored for HIV infected patients. These two measurements show different dynamics during different phases of HIV infection. This typical HIV infection course is for illustration only. It is drawn based on the general pattern of HIV infection courses observed in clinical measurements.

7.1.2 Three phases of HIV infection

Despite some variants among different patients, the HIV infection course is found to follow a general pattern [179]. Fig. 7.3 shows a typical HIV infection course, including three phases: 1) acute infection characterised by a rapid peak in virus load (3-6 weeks post-infection) followed by a rapid fall in virus levels, 2) a stable chronic phase of variable length characterised by low level virus load and slowly declining $CD4^+$ T cell numbers, and 3) the final stage, Acquired Immune Deficiency Syndrome (AIDS), characterised by multiple opportunistic infections and a rapid fall in $CD4^+$ T cell count.

The cellular and viral changes that drive each phase of this complex infection have been the subject of intense debate, in which mathematical models have played an important role in delineating HIV pathogenesis and informing antiretroviral therapies [41, 162, 164].

7.1.3 Related works

The relatively recent finding that the HIV infection can also be transmitted via direct cell-to-cell contacts between T cells [128, 129] has yet to be properly examined and incorporated into modelling studies [41]. Most HIV infection models still assume the

virus is only transmitted via cell-free spreading through free viruses in the blood and body fluid. The models that take cell-to-cell infection into consideration are briefly introduced in this section.

Sigal et al. [202] used a simple mathematical model and clinical experiments to establish the connection between multiple infections in cell-to-cell transmission and the resistance of HIV to medication. Portillo et al. [203] compared the difference between cell-free and cell-to-cell infections in terms of the genetic diversity of transmitted viruses. Through both empirical experiments and theoretical modelling, the researchers concluded that cell-to-cell spread can transmit multiple copies of HIV viruses through an infection contact. This, as the researchers argued, is one of the mechanisms that helps HIV maintain genetic diversity. Wodarz et al. [204] used a simplified model, which assumed cells are located on a regular grid and infected cells can only pass the infection to its nearest neighbours, to demonstrate that multiple infections (coinfection) is important for HIV transmission at low virus load levels.

Komarova et al. conducted a series of modelling studies relevant to cell-to-cell infection [130, 205–207]. In [205], they studied a hypothetical scenario where an infected T cell transmits a fixed number of viruses during its lifetime and it can decide the number of viruses to be transmit at each cell-to-cell infection contact. In the above scenario, the authors' model suggested that transmitting only one virus particle per infection contact maximises the HIV infection. Another two studies [130, 206] analysed a similar topic, but utilised a set of assumptions that led to a different set of conclusions. These studies concluded that, depending on specific assumptions, the optimised strategy could be either transmitting only one or an intermediate number of viruses per cell-to-cell infection. Research in [207] used a simplified HIV infection model to fit clinical records and concluded that cell-free and cell-to-cell infections nearly equally contribute to the spread of HIV among T cells.

The rich literature related to the mathematical modelling of HIV infection within the human body has been reviewed several times recently [41, 179, 180, 208]. Recent studies incorporate sophisticated models of immune selection [178, 209], as well as the formation of a latent reservoir of quiescent infected cells [173, 210]. However, the interplay of cell-to-cell spread and increased $CD4^+$ T cell activation, that is likely to have profound influences on the progression of the disease has been hitherto little

studied.

While existing works were able to provide some insight on the detailed infection mechanisms, the entire course including the three phases and progression of the HIV infection have still not been satisfyingly described and explained by either modelling or empirical studies [179].

7.2 Our hybrid HIV infection model

Here we develop a new mathematical model that incorporates the basic principles of previous host-centric models including a virus-dependent immune response [209], viral latency and a progressive increase in cell activation [211, 212]. Notably, the model additionally includes explicit terms for the two modes of virus spread, parametrised from experimental observation.

An additional feature that is introduced into the model is the role of increased T cell activation as the HIV infection progresses. It is well established that robust HIV replication occurs in activated T cells [211,213], while resting (quiescent) cells are not productively infected by HIV [214], due in part to high levels of the enzyme SAMHD1 (SAM domain and HD domain-containing protein 1) that acts as a viral restriction factor by metabolising nucleotides required for HIV reverse transcription [215, 216]. The number of activated T cells may therefore limit the number of cells susceptible to productive HIV infection during the asymptomatic chronic phases of the disease, when the proportion of activated cells is small [213]. Conversely, there is increasing evidence that HIV progression is accompanied by widespread immune activation, and development of a chronic inflammatory process [211, 212].

The model is depicted in Fig. 7.4. We consider four distinct $CD4^+$ T cell states: activated, uninfected susceptible (S) cells; activated and productively infected (I) cells; quiescent, uninfected (Q) cells; and quiescent latently (L) infected cells. The total $CD4^+$ cell density (N) is given by the sum of these four terms. The Quiescent state is introduced to take into account the clinical finding that quiescent (i.e. resting) T cells are not productively infected by HIV [214]. Similarly the Latent state is introduced to incorporate the fact that infected T cells can become latent and establish a latent reservoir for the virus [170]. These two states are also incorporated in some earlier modelling studies [173, 179].

Table 7.1: Variables and parameters of our model

	Variables	Initial value	Unit	
Q	Density of quiescent CD4$^+$ T cells [a]	$Q_0 = 708$	$cells/\mu l$	
S	Density of susceptible CD4$^+$ T cells [a]	$S_0 = 17$	$cells/\mu l$	
I	Density of infected CD4$^+$ T cells	$I_0 = 0$	$cells/\mu l$	
L	Density of latent CD4$^+$ T cells	$L_0 = 0$	$cells/\mu l$	
N	Density of all CD4$^+$ T cells	$N_0 = 725$	$cells/\mu l$	
V	Density of free HIV virus	$V_0 = 10^{-6}$	$virions/\mu l$	
	Parameters	Default value	Unit	Reference
γ	Death rate of quiescent and latent cells	0.001	$/day$	[217]
γ_S	Death rate of susceptible cells	0.0625	$/day$	[218]
γ_I	Death rate of infected cells	0.5	$/day$	[217]
b	Production rate of new quiescent cells [b]	0.17	$cells/\mu l/day$	[219]
r_S	Resting rate of susceptible cells	0.5	$/day$	
r_I	Resting rate of infected cells	0.0001	$/day$	
a	Cell activation coefficient	0.01	$/day$	
p	Cell proliferation coefficient	1	$/day$	[217]
N_M	Cell threshold density [c]	800	$cells/\mu l$	
g	Virus release rate from an infected cell	100	$virions/cell/day$	[217]
γ_V	Virus death rate	3	$/day$	[217, 219]
c	Cell contact rate [d]	330	$/day$	[197, 220, 221]
θ	Cell synapse rate [e]	0.56		[129, 197, 220, 222]
β_1	Cell-to-cell infection rate [f]	0.19		[198, 203, 221, 223]
β_2	Cell-free infection rate	0.00135	$\mu l/virions/day$	[217]
D	Immune response delay [g]	30	$days$	[224]
κ	Killing coefficient (when $t \geqslant D$) [h]	1.5	$/day$	

[a] Q_0 and S_0 are chosen to maintain a stable density of CD4$^+$ T cells ($N_0 = 725\,cell/\mu l$) in the absence of an infection; [b] Production from sources including thymus; [c] Beyond this threshold CD4$^+$ T cell proliferation stops; [d] The average number of effective contacts between two CD4$^+$ T cells. Cells contact each other frequently, but only a proportion $(20-25\%)$ of these random contacts are sufficiently stable and "effective" for a synapse to be formed at rate θ; [e] The average probability of forming a virological synapse when two cells have an effective contact; [f] The average probability that a susceptible cell is infected when it forms a synapse with an infected cell; [g] Between the initial infection and the onset of cellular immune response, including the period before symptoms begin; [h] $\kappa = 0$ when $t < D$.

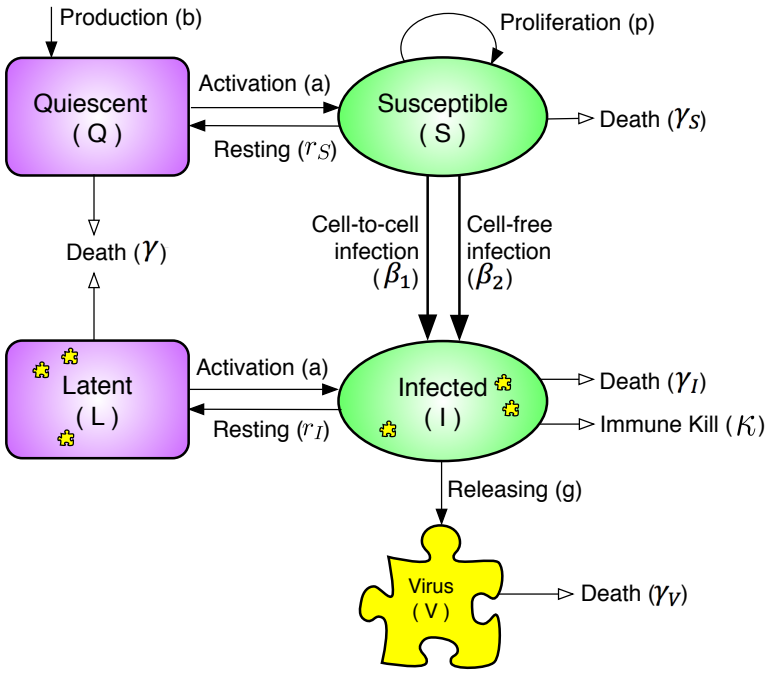

Figure 7.4: Diagrammatic representation of our hybrid HIV infection model.

7.2.1 Ordinary differential equation system

The model can be described by an Ordinary Differential Equation (ODE) system (Eq. (7.1)) and is illustrated in Fig. 7.4.

$$\begin{aligned}
\frac{dQ}{dt} &= -\gamma Q + r_S S - a\frac{N_M}{N}Q + b \\
\frac{dS}{dt} &= -\gamma_S S - r_S S + a\frac{N_M}{N}Q - cI\frac{S}{N}\theta\beta_1 - SV\beta_2 + p\frac{N_M - N}{N_M}S \\
\frac{dI}{dt} &= -\gamma_I I - r_I I + a\frac{N_M}{N}L + cI\frac{S}{N}\theta\beta_1 + SV\beta_2 - \kappa\frac{I}{I+0.1}\frac{N}{N_M}I \\
\frac{dL}{dt} &= -\gamma L + r_I I - a\frac{N_M}{N}L \\
\frac{dV}{dt} &= -\gamma_V V + gI
\end{aligned} \tag{7.1}$$

The density variables (Q, S, I, L, V) and parameters are defined in Table 7.1. The densities are measured as numbers of cells or virions in a μl of blood/extracellular fluid. We set a density variable to zero when it drops to below $10^{-12}/\mu l$, accounting for the fact that when the density of cells or virions drops to such low level, there is a high probability that it would die out (density becomes zero). The default value of parameters, shown in Table 7.1, are taken from the literature or estimated from clinical and experimentally observed data. The killing coefficient κ is equal to its value in Table

1 when $t \geq D$; otherwise $\kappa = 0$ when $t < D$.

The production rate of new quiescent T cells from sources, such as the thymus, within the human body is represented by b. Quiescent T cells are activated and become susceptible at a variable activation rate $a(N_M/N)$, where a is the activation coefficient and N_M is the density of T cells at which proliferation stops. The activation rate $a(N_M/N)$ increases as the total T cell density (N) falls (caused by HIV progression). The detailed mechanism of how the activation rate increases with the progression of HIV is still unclear, and may include increased rates of co-infections, danger signals from dying T cells or homeostatic regulatory loops. This term, $a(N_M/N)$, here is an approximation, which encompasses the combined effects of all these different mechanisms. Quiescent T cells die at a rate of γ.

Susceptible T cells turns into quiescent T cells at a rate r. They proliferate at a variable rate $p(1 - N/N_M)$, where p is the proliferation coefficient, N is the total T cell density, and N_M is the T cell density at which proliferation stops. This variable proliferation rate is a reasonable approximation [217] of the real T cell proliferation process, based on evidence [162] that T cell proliferation rate is density-dependent and would slow as the T cell density becomes high. Susceptible T cells die at a rate γ_S.

T cells and HIV viruses move freely and make contacts with each other randomly. We define β_2 as the infection rate of a susceptible cell being infected by a free virus. Note that β_2 incorporates two probabilities: (1) the probability of a T cell contacting a HIV virus in a given time; and (2) the probability that an HIV infection takes place during the contact between them. Thus we expect the total number of new cell-free infections caused by free HIV viruses to be $\beta_2 SV$.

T cells move randomly and contact each other at a relatively constant rate. Some of such contacts are "effective" such that they have a chance to form a more stable contact, called "synapse". When a synapse is formed between two cells, if one cell is HIV infected and the other is susceptible, then there is a probability that the infected cell can infect the susceptible cell. We define c as the average number of effective contacts a T cell makes with *all* other T cells in a unit period of time. Biological studies suggest that this number remains stable in a living person where the total number of T cells (N) must stay above a certain level [220, 221]. Then cS/N is the average number of effective contacts that a T cell makes with *susceptible* T cells. We define θ as the

average probability that an effective contact becomes a synapse; and β_1 as the average infection rate that an infected T cell infects a susceptible cell while they form a synapse. Thus we expect each infected T cell to infect $c(S/N)\theta\beta_1$ susceptible cells in a unit period via cell-to-cell infection; hence the total number of new cell-to-cell infections caused by all infected T cells are $cI(S/N)\theta\beta_1$.

In reality, cell-to-cell transfer occurs locally involving only the infected cell and its immediate neighbours. The model abstracts this process by averaging infection over all cells. In practice, local effects will only distort this average when target cells in the vicinity of an infected cell become limiting. This limit seems unlikely to be reached except very late in infection, given that infected cells continue to migrate, albeit at a slower rate [200], and uninfected target cells continue to migrate into the vicinity of an infected cell. More complex spatial models will be required, however, to understand the detailed anatomical distribution of HIV infected cells over time. The amount of cell-to-cell transfer in the model depends on the number of infected T cells I and the proportion of susceptible cells S/N at any given time. Since N, the total density of $CD4^+$ cells, is not held constant, but in fact declines over time, cell-to-cell spread becomes increasingly effective as HIV progresses.

The model does not explicitly distinguish antigen-presenting cell to T cell (APC/T) from T cell to T cell (T/T) transmission. APC/T transmission may potentially be most important very early in establishing infection [225], a process that is not examined in detail in this model. These two types of transmission are in fact both likely to occur most frequently and efficiently in the microenvironment of an APC/T cluster, where APC/T interactions lead to T cell activation, and hence favour also T/T interaction. Furthermore, there are very few quantitative estimates of the parameters of APC/T interaction *in vivo*. The incorporation of an additional cell type is therefore unlikely to have a major effect on the model behaviour, but would add significantly to model complexity and uncertainty.

The cause of T cell death in HIV infection continues to be controversial, and probably includes several effects including lysis of infected cells by effector cells such as CD8 T cells and NK (Natural Killer) cells, apoptosis/pyroptosis and bystander death [212]. Non-immunological death of infected T cells is represented by a death rate of γ_I. And the term $\kappa\frac{I}{I+0.1}\frac{N}{N_M}$ is used to model the death of infected T cells by the

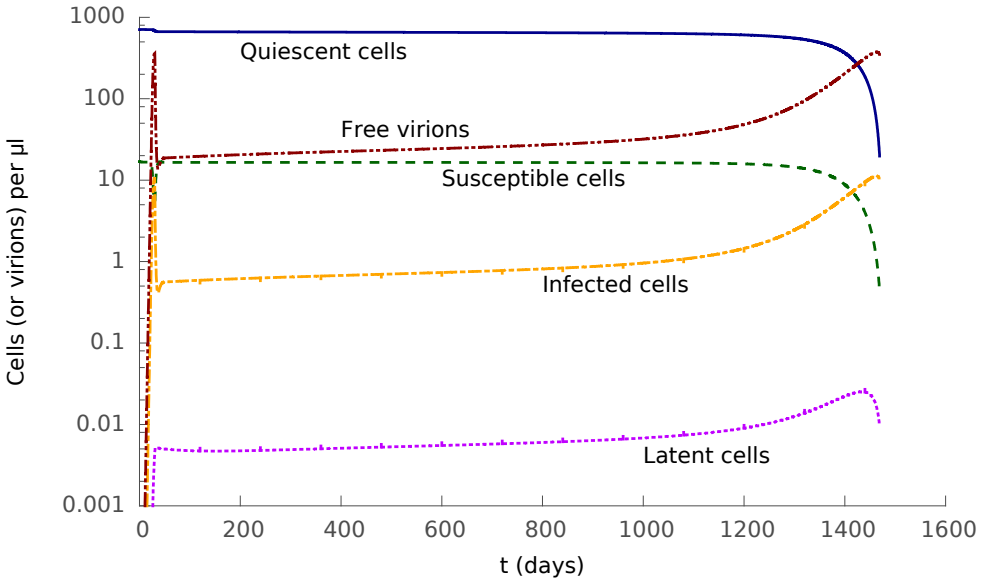

Figure 7.5: Density of quiescent, susceptible, latent and infected CD4$^+$ T cells, and the density of free virions as a function of time (in days).

cellular immune response. κ is initially 0 and changes to a higher value in Table 7.1 when the cellular immune response kicks in D (default value: 30) days after the initial infection. The term $\frac{I}{I+0.1}$ captures the relationship between the strength of the immune response and the density of infected CD4$^+$ T cells [209]. The term N/N_M captures immune exhaustion caused by HIV infection. It falls from around 1 before infection towards 0 as HIV progresses (because N_M is a constant and N, the total density of CD4$^+$ T cells, gradually declines with the progression of HIV).

Infected cells return to a quiescent state, and become latent, at rate r_I. Latent cells die and are activated (i.e. become infected cells) at the same rates as quiescent cells. Infected T cells release free viruses at rate g. Free viruses die at a rate of γ_V.

The abortive infection of quiescent cells is not considered in this simplified model, similarly to most previous modelling studies [41, 164, 206, 207]. HIV immune escape mutants [226] are not directly modelled in this study but their effects on degrading cellular immune response are reflected in the immune exhaustion in our model.

7.2.2 Reproduction of the full course of HIV infection

The numerical solutions of each of the variables are shown in Fig. 7.5. Fig. 7.6 shows the combined CD4$^+$ T cell counts (N) and virus load (V), which are measured routinely in the clinic to monitor HIV infection. Notably, the qualitative behaviour of the model

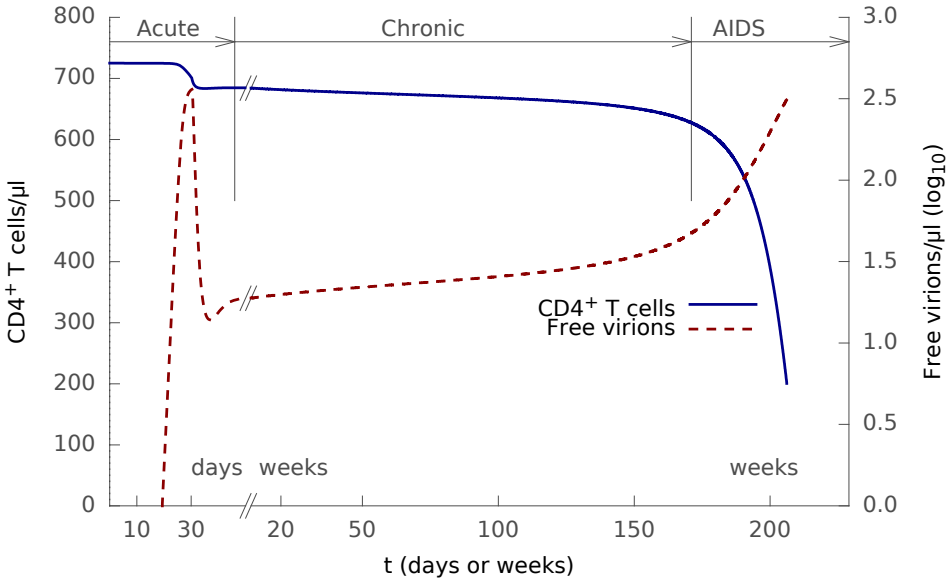

Figure 7.6: Numerical solutions of the model, plotting N, the density of all $CD4^+$ T cells (y axis on the left) and V, the density of free virions (y axis on the right in log scale) as a function of time (in days or weeks), respectively. Parameter values are in Table 7.1.

accurately reflects the three main phases of disease that are observed clinically. The model reproduces an acute infection phase, where the virus replicates rapidly (reflecting the absence of any pre-existing adaptive immunity), peaks and then returns to a low level by approximately five weeks. This metastable level of virus represents the clinical "set-point". The virus then remains stable for a prolonged period (note interruption and change of scale in x axis), during which time T cells decline very slowly. Finally, T cell numbers start to drop faster, and viral loads rise. The model calculation is stopped when $CD4^+$ level reaches $200 cells/\mu l$.

7.3 London clinical data

7.3.1 Recruitment of HIV infected patients

The London clinical data is collected by P. Pellegrino and I. Williams in the Mortimer Market Centre for Sexual Health and HIV Research (London, UK), where individuals acutely-infected with HIV were recruited. Subjects were mostly male Caucasians who presented with symptoms of acute retroviral illness. Patient viral loads and $CD4^+$ T cell counts were measured longitudinally at serial time-points following infection using standard clinical tests. All subjects were offered anti-retroviral treatment at diagnosis.

The subjects selected for inclusion in the data all chose not to receive anti-retroviral therapy in acute or early infection, and remained untreated until disease progression, evidenced by a substantial decline in their circulating $CD4^+$ T cell count occurred.

7.3.2 Criteria of patient selection

There were 39 patients in total in the data from Mortimer Market Centre for Sexual Health and HIV Research (London, UK). For this study, we focussed the analysis only on patients with more than ten data points for both HIV load and $CD4^+$ measurements (29 out of the 39 patients). We also excluded from the analysis a further 12 patients who showed no overall decrease in $CD4^+$ count, or no increase in viral load at later timepoints. The focus of the model described above is to capture the "typical" characteristics of HIV infection, which include that $CD4^+$ count falls in general, and viral load increases in general as infection progresses. It is widely accepted that in some patients (for example the so-called "elite controllers") viral load remains low or undetectable and $CD4^+$ count remains unchanged for long periods. The mechanisms responsible for these phenomena are still incompletely understood. The current model does not attempt to incorporate any such mechanism, and this group of patients was therefore not included in the study. Further elaboration of the current model to include additional features of viral control will be informative in helping to understand such patients. The identifiers for the remaining 17 patients are: MM1, MM4, MM8, MM9, MM12, MM13, MM23, MM24, MM27, MM33, MM39, MM40, MM42, MM43, MM45, MM57, MM60.

7.4 Model evaluation

7.4.1 Calibrating model parameters for each patient

Most of the parameters of our model, and especially those determining HIV infectivity via cell-to-cell or cell-free spreading, were obtained from experimental observations (Table 7.1). It was therefore important to test that the model with this parametrization accurately fit real clinical datasets. We therefore evaluated our model against a set of longitudinal T cell count and virus load measurements obtained from a cohort of HIV-infected individuals who were recruited following clinical presentation with symptomatic acute HIV infection and followed over time with serial measurement of

Table 7.2: Model parameters calibrated from a cohort of treatment-naive HIV patients.

Patient	$Q_0\,(cells/\mu l)$	$S_0\,(cells/\mu l)$	$N_M\,(cells/\mu l)$	$\kappa\,(day^{-1})$	$D(days)$
MM1	879	21	996	1.753604	30
MM4	684	16	773	1.474027	33
MM8	392	9	439	1.076508	49
MM9	440	10	494	1.349088	61
MM12	542	13	611	1.982885	72
MM13	537	13	605	1.250311	31
MM23	419	10	470	1.150800	51
MM24	547	13	617	1.494605	54
MM27	606	14	684	1.353939	30
MM33	781	18	884	1.623077	33
MM39	553	13	623	1.276949	32
MM40	435	10	488	1.137114	46
MM42	641	15	724	1.801382	57
MM43	582	14	656	2.511129	80
MM45	542	13	611	1.273904	33
MM57	406	9	455	1.207627	58
MM60	587	14	662	1.325221	33

Results are shown for all 17 patients. In the data, time is recorded relative to the first appearance of symptoms of HIV infection. As the actual initial infection date is unknown, we assumed a constant "eclipse" phase of 20 days between initial infection and first appearance of symptoms.

plasma viral levels and circulating T cell counts. The subjects selected for inclusion in this study all chose not to receive antiretroviral treatment in acute or early infection, and remained untreated until progression towards AIDS, evidenced by a substantial decline in their circulating $CD4^+$ T cell count.

7.4.2 Model prediction vs. clinical data

We use our model to theoretically reproduce the HIV infection courses in the data. The values for HIV infectivity were fixed, as derived from the literature or our own observations (Table 7.1). Values for five parameters (Q_0, S_0, N_M, κ and D) describing the characteristics of the immune response were chosen for each patient to minimise the error of the predicted quasi-stable level of T cell counts (N_s) and viral load (V_s), and the time of progression to AIDS (t_A). Of the five parameters, four of them (Q_0, S_0, N_M and D) can be determined in a relative straightforward manner from the patient data; Q_0, S_0, and N_M are determined by averaging the density of $CD4^+$ T cells within 100 days after infection for each patient; D can be determined as the time when the peak of the virus density appears in the data. We need to infer the κ value. We use a search method with a target to minimise the error in T cell density (N) between measurement data on

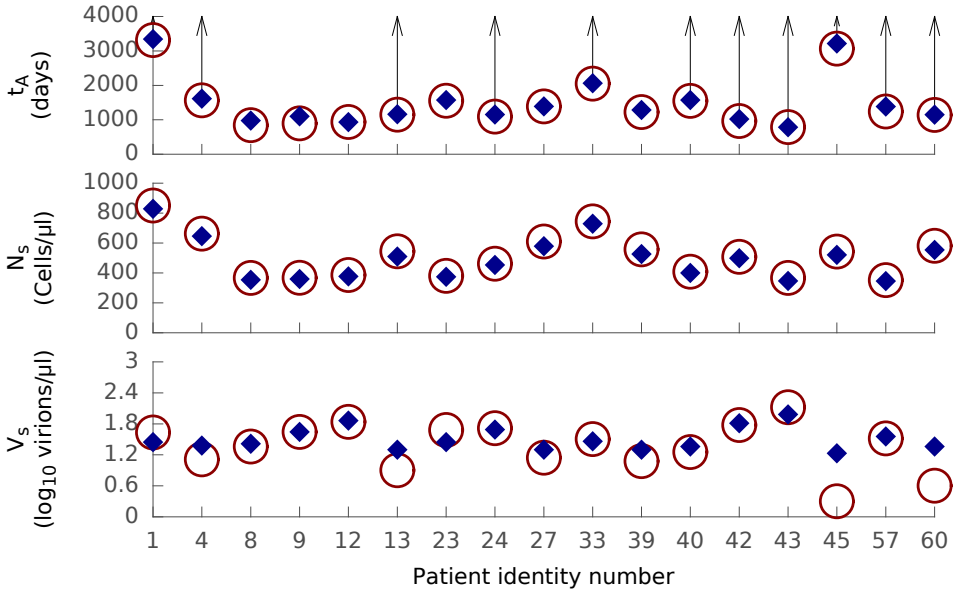

Figure 7.7: Clinical data (circle and arrow) for all patients under study comparing against model prediction (diamond) for the time to AIDS (t_A), the quasi-steady density of CD4$^+$ T cells (N_s) and the quasi-steady density of free virions (V_s). An arrow represents that t_A is greater than a particular value (represented by the connected circle) for a patient as his / her CD4$^+$ count did not reached AIDS level ($200\,cells/\mu l$) in the data. Full prediction results are shown in Table 7.3.

the last day and the predicted value. Each step of our method improves the accuracy of prediction. We stop when the κ value is refined to the precision level shown in the Table 7.2. All other values remained fixed at the default values in Table 7.1.

The predicted progression results are compared against the actual measurements in Fig. 7.7, Fig. 7.8, and Table 7.3. The predicted V_s and t_A for each patient were negatively correlated (correlation coefficient = -0.46), in agreement with the well-established relationship between these two clinical values. Remarkably, the model can fit all patients by modifying the five immune-relevant parameters over a narrow range. Furthermore, the parameter values that gave the best results for the patients (see Table 7.2) are all very close to those in Table 7.1, which were derived independently from experimental measurements.

For the five parameters Q_0, S_0, N_M, κ and D, when other parameters are fixed: 1) Q_0, S_0, and N_M determine the initial density of CD4$^+$ T cells; 2) D determines the time of the transition from the acute phase (phase 1) to the chronic phase (phase 2); 3) κ determines the time of the transition from the chronic phase (phase 2) to the AIDS phase (phase 3). κ controls the long term development of the infection towards AIDS.

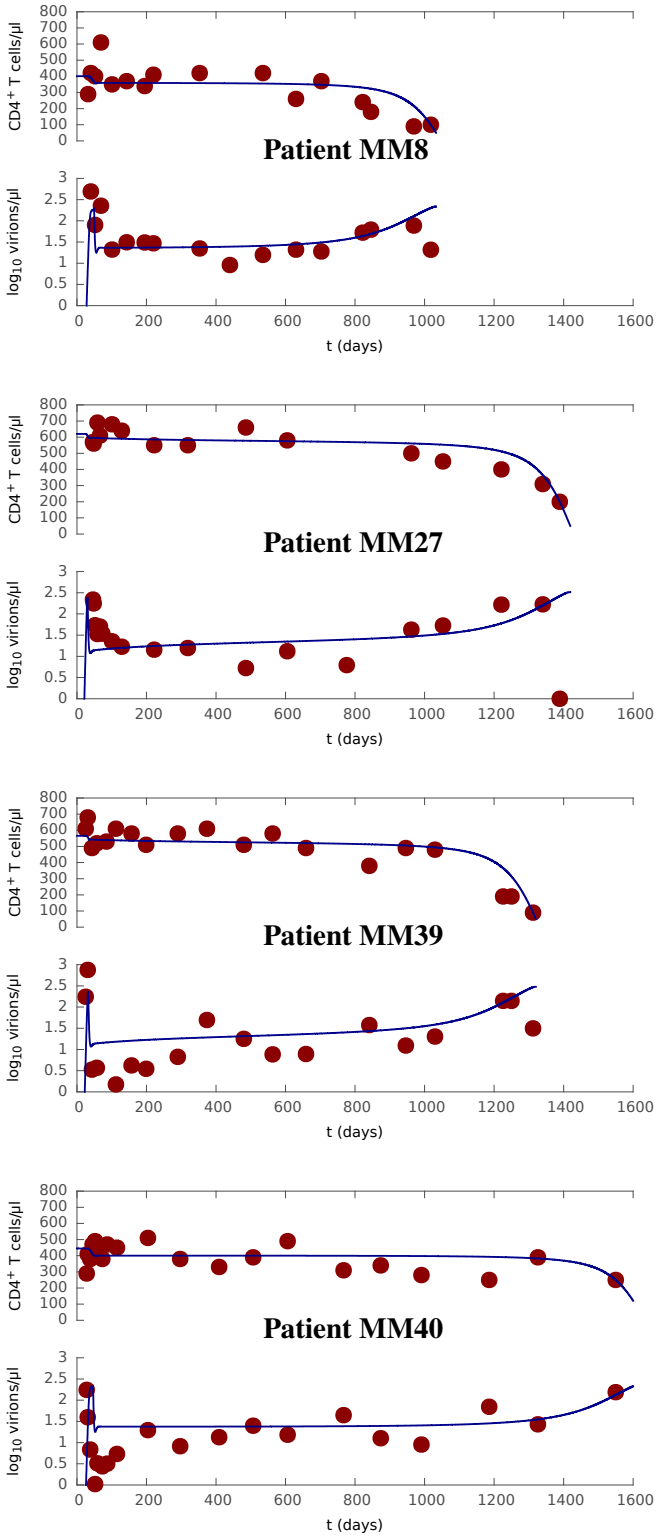

Figure 7.8: Prediction (curve) of HIV progression course (N and $log_{10}V$) for four typical patients, where clinical data are shown as dots. The model predictions reproduce the HIV infection courses in the clinical data. There are several outlier data points for the virus level. Further investigations are needed for these points.

Table 7.3: Model prediction of a cohort of treatment-naive HIV patients.

Patient	$N_s\,(cells/\mu l)$		$log_{10}V_s\,(virions/\mu l)$		$t_A\,(days)$	
	Data	Model	Data	Model	Data	Model
MM1	851	829	1.63	1.45	>3317	3342
MM4	663	646	1.11	1.38	>1567	1612
MM8	367	354	1.36	1.41	846	978
MM9	368	359	1.64	1.64	895	1101
MM12	386	377	1.84	1.86	935	931
MM13	545	510	0.90	1.30	>1151	1162
MM23	381	374	1.68	1.45	1554	1574
MM24	462	453	1.72	1.69	>1093	1152
MM27	610	579	1.15	1.30	1389	1389
MM33	745	729	1.51	1.46	>2051	2061
MM39	558	526	1.08	1.30	1226	1286
MM40	408	400	1.26	1.36	>1550	1572
MM42	508	499	1.78	1.81	>966	1022
MM43	367	347	2.12	1.98	>787	786
MM45	543	521	0.30	1.23	>3068	3217
MM57	352	346	1.52	1.56	>1240	1387
MM60	582	554	0.60	1.36	>1140	1149

N_S is the quasi-steady $CD4^+$ T cell density and V_S is the quasi-steady density of free virions, which are average densities between the 100th and 800th days after initial infection. t_A is the time to AIDS, which is defined as the time between initial infection and when the density of $CD4^+$ T cells falls to 200 $cells/\mu l$.

7.5 Discussion

The aim of this study was to develop a model that reduces the enormous biological complexity of the course of HIV infection to a few well-defined equations but nevertheless retains the main dynamic features of the disease. In particular, the model was required to predict the evolution of a long lived metastable state of low level viral infection, which ultimately breaks down to uncontrolled viral growth and a precipitous fall in $CD4^+$ T cells, two hallmarks of AIDS. The model incorporates both an immune response, which is believed to be a major factor limiting viral expansion by killing of infected cells, and the formation of long lived latently infected cells, which are believed to play an important role in limiting the long-term effects of antiviral therapy. The key distinguishing features of the model are that it explicitly incorporates cell-to-cell spread of virus as well as classical spread via cell-free virus. The motivation of building such a model was to investigate the role of these two modes of spread in determining the outcome of infection, and thus complement the limited *in vivo* experimental and clinical data available addressing this question.

The model parameters were set initially according to published experimental data obtained from a range of previous studies. This parameterization resulted in a model that met the criteria outlined above, and reproduced all three main phases of HIV infection. The robustness of the model was tested by examining its ability to predict a set of longitudinal clinical data (virus load and $CD4^+$ T cell count) obtained from a cohort of patients who were followed over extended periods in the absence of any antiviral therapy. As expected, the progression of disease in these patients was highly variable, both in terms of viral set point and disease progression. Despite its relative simplicity, the model predicted this independent dataset well, requiring only small, and biologically realistic changes to a few parameters.

7.6 Summary

This chapter presented a novel HIV infection model that can explain the complex progress of the infection in all its phases and its variable timescale. Such a unified model is important not only to understand the HIV infection dynamics, but also to estimate the long term effects of therapeutic strategies on HIV progression.

Notably, the model additionally includes explicit terms for the two modes of virus spread, parametrised from experimental observation. The model predictions are consistent with a set of longitudinal data (viral load and $CD4^+$ T cell count) from a cohort of untreated HIV infected patients.

Chapter 8

Analysis of HIV dynamics and treatments

Mathematical models provide an important tool for understanding and predicting the natural course of HIV infection that complements clinical studies. The most appropriate therapeutic regimens for HIV therapy continue to be the subject of disagreement and debate. In addition, many new therapeutic modalities aimed at achieving viral eradication, such as the HDAC inhibitors [227], or therapeutic vaccination [228] are being proposed. However, testing new treatment regimens is a costly and time consuming task, and the logistic challenges and expense of running clinical trials to evaluate and compare treatments remain a major bottleneck to translational advances in HIV therapy. Mathematical models have proved of value in the past, but have suffered from omitting important biological processes, thus compromising their ability to accurately recapitulate clinical observations.

8.1 Introduction

Hybrid spreading is in fact a feature of other viral infections [229], but is also shared in other "epidemic" scenarios such as the spread of computer worms [22, 230], or of mobile phone viruses [102]. The mathematical analysis of hybrid spreading has received significant previous attention [29, 57, 102, 231]. However, the importance of hybrid spread to HIV dissemination and disease progression, has not been explored from a mathematical point of view.

The HIV infection model presented in the Chapter 7 explicitly incorporates cell-to-cell transmission, and changing rates of cellular activation, two processes that are

known to be key features of HIV infection. With increased sophistication, our HIV infection model allows us to theoretical investigate the dynamics of HIV infection progression including the roles of cell-to-cell spread and cellular activation. This chapter will also investigate the long term effects of various treatment strategies on HIV infection using our model.

8.2 Dynamics of HIV infection

Having confirmed that the model gives realistic estimations and predictions of real clinical data in Chapter 7, we investigated the behaviour of the model in greater detail.

8.2.1 Role of cell-to-cell spread

The role of the two spreading routes was further examined by systematic variation of the cell-to-cell infection rate, β_1, and the cell-free infection rate, β_2. The predicted outcome of infections are shown in Fig. 8.1. When either route is abolished, infection is blocked completely; T cell level returns to normal and the virus is cleared after the cellular immune response kicks in. If cell-to-cell spread is removed from the model ($\beta_1 = 0$) even a doubling in cell-free infection rate does not result in infection progression. In contrast, a doubling of the cell-to-cell infection rate increases the set-point of viral load, and greatly speeds up the progression of infection even in the absence of cell-free infectivity. Thus the model suggests cell-to-cell spread may be an important force in allowing the virus to establish an infection in lymphoid tissue [225].

Fig. 8.2 depicts the number of $CD4^+$ T cells newly infected via either cell-to-cell spread or cell-free spread as the infection progresses. The model predicts that cell-to-cell transfer becomes increasingly dominant as the total number of $CD4^+$ T cells falls, the proportion of susceptible cells rises (Fig. 8.2 inset left y axis) and the strength of immune response falls (because of immune exhaustion, see Fig. 8.2 inset right y axis).

8.2.2 Effect of cellular activation

In the context of the model, the transition from phase 1 (acute) to phase 2 (stable chronic) is driven by a balance between several processes, including viral spreading through two parallel modes, and the cellular immune response, i.e. killing of infected cells as the cytotoxic $CD8^+$ T cell response becomes active. Paradoxically, in the stable chronic phase, the activation of T cells, which is the hallmark of adaptive immunity and

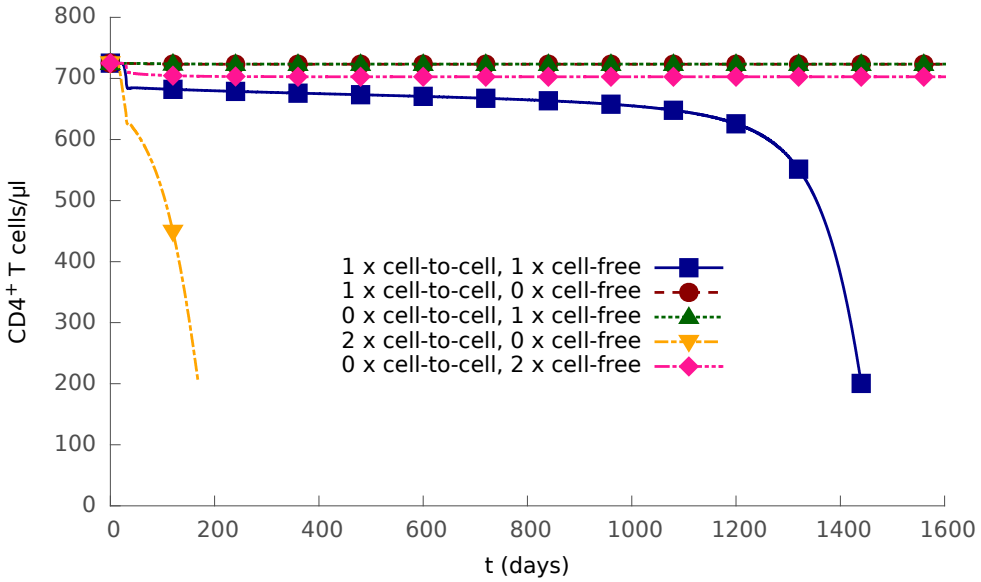

Figure 8.1: Two modes of HIV infection. The density of CD4$^+$ T cells as a function of time for different values of cell-to-cell infection rate β_1 and cell-free infection rate β_2: (1) both use their default value, (2) β_1 uses its default value and $\beta_2 = 0$, (3) $\beta_1 = 0$ and β_2 uses its default value, (4) β_1 is twice its default value and $\beta_2 = 0$, (5) $\beta_1 = 0$ and β_2 is twice its default value.

is aimed at protecting the host, in fact contributes to the persistence of HIV. The role of CD4$^+$ T cell activation is explored in Fig. 8.3. In this model, the rate of T cell activation $a(N_M/N)$ increases as the number (N) of T cells falls. In the absence of this feedback (i.e. when the activation rate is fixed), the HIV infection would not progress to AIDS after the onset of the cellular immune response. In contrast, if the activation rate is doubled, then infection progresses significantly faster to AIDS. These results confirm and extend the findings of DeBoer and Perelson [217], which suggested an increasing rate of cellular activation was important in the establishment of chronic infection and progression to AIDS. The results are also consistent with evidence that non-pathogenic SIV infection in the natural host species results in viral replication in the absence of chronic immune activation and no AIDS [232].

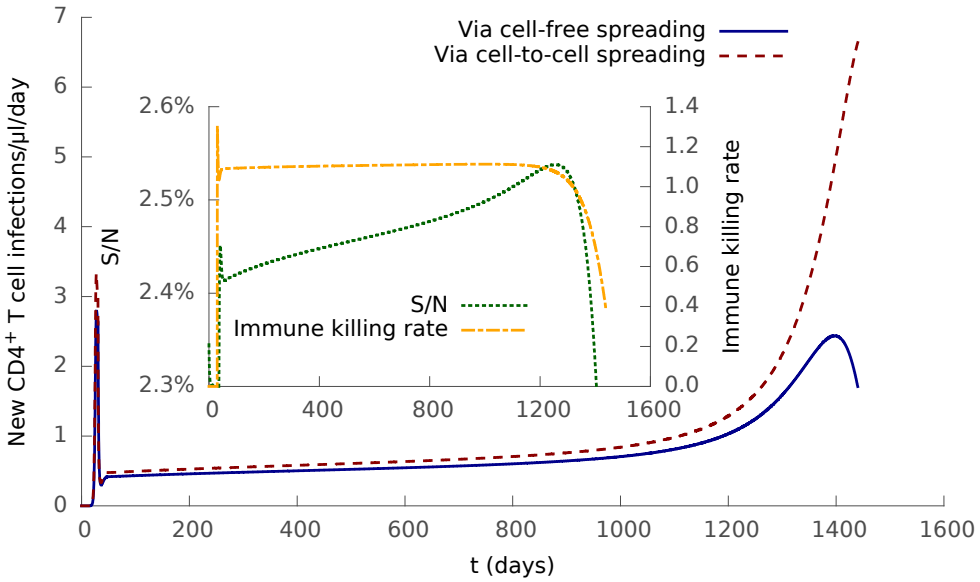

Figure 8.2: Numbers of newly infected cells in a day via cell-to-cell spreading and cell-free spreading, respectively. The inset shows the ratio of susceptible cells to all cells (S/N, left y axis) and the strength of immune response ($\kappa \frac{I}{I+0.1} \frac{N}{N_M}$, right y axis) as a function of time, respectively.

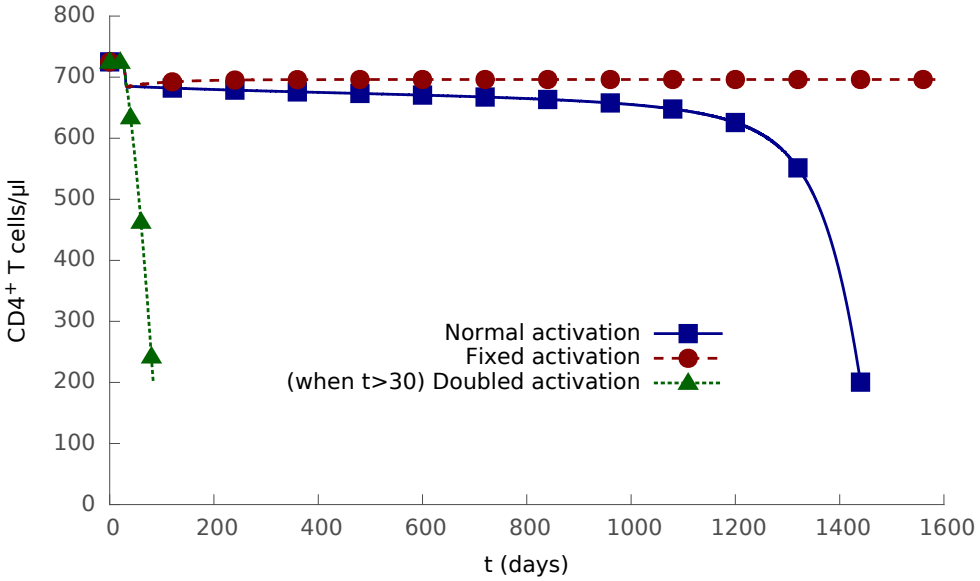

Figure 8.3: Progression of the HIV infection for different cell activation rates, including (1) normal activation (aN_M/N in Eq. (7.1)), (2) fixed activation (aN_M/N_0, where N_0 is the initial density of CD4$^+$ T cells), and (3) doubled activation ($2 \times aN_M/N$ when $t > D$).

8.3 Model prediction based on treatment trial data

The proposed model can help evaluate the long term effects of different treatments on HIV progression. Once the start time, duration, and effectiveness against two modes of

HIV spread are known for a treatment, its effects on HIV progression can be evaluated by the model.

8.3.1 The SPARTAC data

To validate the model's ability to evaluate HIV treatments, we use it to theoretically reproduce the results of the Short Pulse Anti-Retroviral Therapy at Seroconversion (SPARTAC) trial [46]. The SPARTAC trial aims to evaluate how the short-course antiretroviral therapy (ART) delays HIV progression. The patients (366 in total) who participated in the trial were randomly assigned to three groups: standard care, 12-week ART treatment, and 48-week ART treatment. For these three groups of patients, the primary end point t_p (defined as when CD4$^+$ count $\leq 350 cells/\mu l$ or the start of long-term ART) on average reached 157 weeks (standard care), 184 weeks (12-week ART), and 222 weeks (48-week ART) after randomisation. Randomisation is the time when the trial starts.

8.3.2 Calibration of model parameters

We first estimate time points of the SPARTAC trial (randomisation, start of the treatments, and primary end points t_p) in terms of days after the initial infection. According to [46], the median interval between seroconversion and randomisation (start of trial) was 12 weeks. The exact average time of seroconversion for patients in the SPARTAC trial is not directly available from [46]. We assume it is two weeks after initial infection, as seroconversion normally happens within a few weeks after HIV-1 infection. Then the time of randomisation (start of trial) can be estimated as $7(2+12) = 98 days$ after infection. The treatment is estimated to start three days after randomisation [46], i.e. $98+3 = 101 days$ after infection. And the primary end points for patients in standard care, 12-week ART, and 48-week ART groups are respectively: $98+7\times 157 = 1197 days$, $98+7\times 184 = 1386 days$, and $98+7\times 222 = 1652 days$ after infection.

We then use the average CD4$^+$ count and virus load of all patients at randomisation, and the primary end point of the patients in standard care group to fit the five immune-relevant model parameters (Q_0, S_0, N_M, κ and D). These fitted parameters represent an average patient in the trial.

8.3.3 Theoretical evaluation of the trial results

We then evaluated the effects of the 12 and 48 week therapies using the model. The therapy was assumed to be 100% effective against cell-free transmission, but we evaluated the model for both 100% and 50% efficiency against cell-to-cell transmission, since cell-to-cell transmission has been reported as being more resistant to some forms of therapy [202, 233]. Both modalities reproduced the observed effects of therapy well, and the model results were robust to changes in the efficiency (Table 8.1). The model therefore not only fits known datasets (standard care group) but also accurately predicts the effects of new treatment regimes on two independent patient groups (12-week and 48-week).

Table 8.1: Model prediction of the SPARTAC trial

Patient groups:		Standard-care	12-week-treatment	48-week-treatment
N_r ($cells/\mu l$)	Data:	559		
	Model:	560		
V_r ($virons/\mu l$)	Data:	17 ($log_{10}V_r = 1.2$)		
	Model:	12 ($log_{10}V_r = 1.1$)		
t_p (days)	Data:	1,197	1,386	1,652
	Model:	1,197	**1,392** | 1,394	**1,654** | 1,656

N_r and V_r are the density of CD4$^+$ T cells and free virions at the *randomisation* (when patients groups are randomly divided and the trial starts). t_p is the time between the initial infection and the *primary end point* (when density of CD4$^+$ T cells decreases to 350 $cells/\mu l$ or the start of long-term ART [46]). The t_p estimated from the model is the time when CD4$^+$ T cells decreases to 350 $cells/\mu l$. Section 8.3.2 describes how t_p is inferred from the data [46] in terms of days after infection. The model parameters calibrated from the trial data (by fitting the average N_r, V_r for all patients and t_p for the standard care group) are $Q_0 = 566\ cells/\mu l$, $S_0 = 13\ cells/\mu l$, $N_M = 638\ cells/\mu l$, $\kappa = 1.292165\ /day$ and $D = 28\ days$. The trial treatment is assumed to be 100% effective against cell-free HIV spread. We estimate the t_p for 12-week, and 48-week treatment groups if the treatment is 50% (value in bold font) or 100% (value in normal font) effective against cell-to-cell spread of HIV.

8.4 Analysis of HIV treatment strategies

This section investigates how a treatment's starting time and efficiency affect its effect on HIV infection progression.

8.4.1 Early treatment intervention

We further explored the sensitivity of the model to perturbation as a function of treatment starting time (Fig. 8.4). The "treatment" lasts for 30 days, during which both cell-

free and cell-to-cell infection are completely blocked. Once "treatment" is finished, two modes of HIV infection resume. Early treatment in this model (3 days after infection, i.e. post-exposure prophylaxis) leads to no decline in CD4$^+$ T cell density, and no chronic infection phase. The same treatment applied when T cell density reaches the levels (500 CD4$^+$ T $cells/\mu l$ and 350 CD4$^+$ T $cells/\mu l$) at which the World Health Organization recommends ART initiation [234] is followed by a rapid virus rebound after the treatment stops, and the disease progresses according to its normal course. Thus, as HIV progresses it becomes increasingly difficult to control infection in this model.

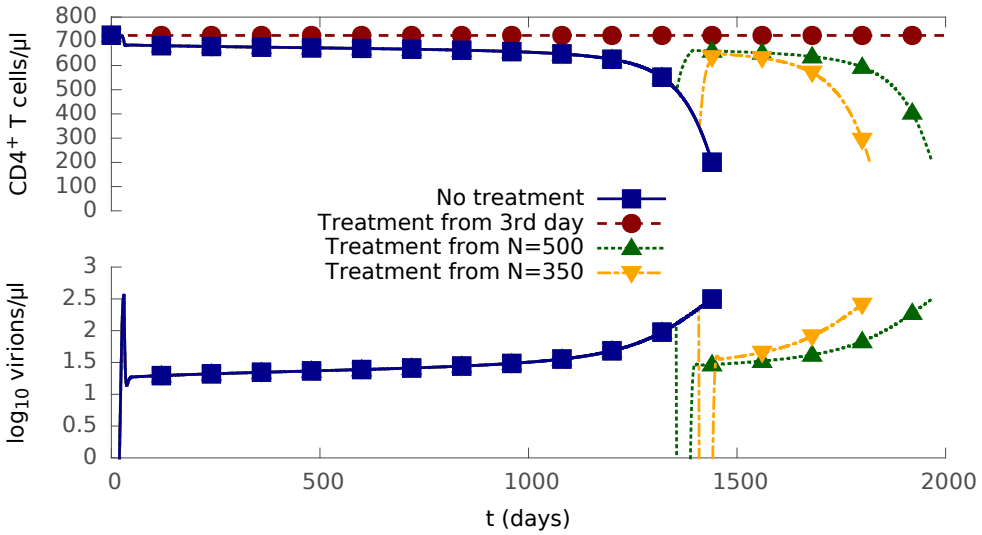

Figure 8.4: Impact of treatment starting time on HIV progression. HIV progression for a 30-day "perfect" treatment starting at three different times after the initial infection: (1) on the 3rd day when the density of all CD4$^+$ T cells is $N = 725\,cells/\mu l$, (2) when $N = 500\,cells/\mu l$; (3) when $N = 350\,cells/\mu l$. The 'prefect' treatment here means both cell-to-cell infection and the cell-free infection are completely blocked (i.e. $\beta_1 = 0$, $\beta_2 = 0$ and the virus release rate $g = 0$) for 30 days.

8.4.2 Treatments on both cell-to-cell spread and cell activation

Finally, we looked at the interactions between treatment starting time, activation rate and efficiency of therapy against cell-to-cell spread (Fig. 8.5). In general, increased efficiency of therapy and earlier treatment both prolonged time to progression to AIDS. However, the effects of altered activation depend in a complex way on the context of the intervention. Blocking activation early is beneficial, since it will reduce the number of susceptible cells HIV can infect; while blocking activation late, when the latent HIV reservoir has been established, will prevent latent HIV from being activated and erad-

icated. Increasing cellular activation, which has been proposed as a means of flushing out the latent reservoir [227], can be effective in increasing time to AIDS when given in the context of efficient anti-viral therapy, but can shorten the time if concomitant anti-viral therapy is incomplete. This is because increasing cellular activation increases both the number of susceptible cells (activated from quiescent cells) and the number of infected cells (activated from latent cells) at the same time. Thus if it is used together with an effective anti-viral therapy, the latent HIV reservoir will be flushed out and killed by the antiviral drugs. But if the concomitant anti-viral therapy is not efficient enough to clear the increased number of infected cells, the spread of HIV will speed up.

8.5 Discussion

A fundamental prediction of the analysis is that, given the known experimentally derived parameters of cell-to-cell and cell-free spreading, the two modes of spread complement each other and both make important contributions to disease progression. In addition, it is clear that there is a close relationship between the proportion of activated T cells, cell-to-cell spread and disease progression. Specifically, cell-to-cell spreading is strong when the percentage of activated, and therefore susceptible cells is high in the population, since an infected cell is then more likely to encounter an effective target (a susceptible cell) to infect. When the percentage of susceptible cells is low, infected cells will mostly encounter quiescent/resting cells that will provide ineffective targets. These conditions may occur both early and late in HIV infection. The site of infection itself (for example the vaginal mucosa) may contain a large proportion of activated T cells some of which may be interacting with APC, particularly if there is a concomitant sexually transmitted infection. The well-documented association between HIV infection and other mucosal infections may therefore reflect the key importance of a high concentration of APC and activated T cells in early transmission of the virus [235]. There is also convincing evidence that gut associated lymphoid tissue is a major site of viral replication early on in the disease [236,237]. This tissue is characterised by an unusually large proportion of T cells capable of supporting HIV replication, presumed to result from chronic exposure to the gut microbiome. Since cell-to-cell spread is much more efficient, and under these conditions the number of activated target cells are not

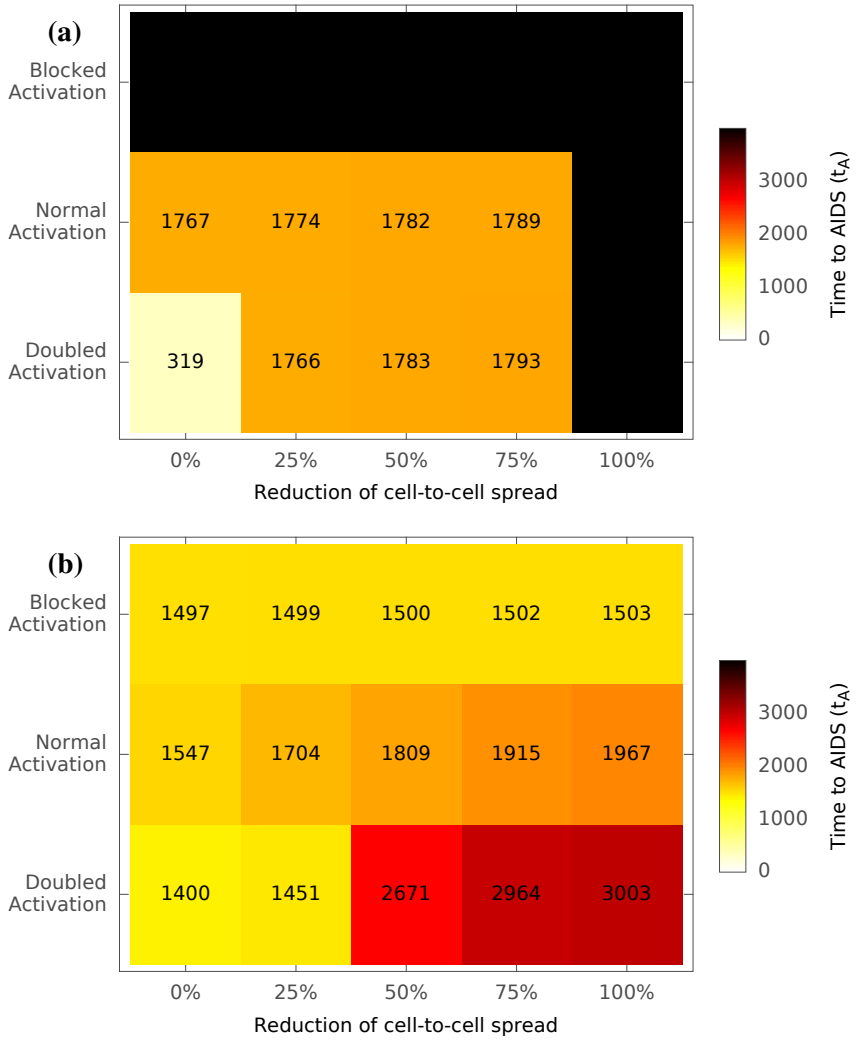

Figure 8.5: Prediction of the time to AIDS, t_A, for different treatment schemes. We consider 30-day treatments starting **(a)** on the 3rd day and **(b)** when $N = 500\,cells/\mu l$, respectively. All treatments block cell-free infection completely ($\beta_2 = 0$ and $g = 0$). The treatments also affect cell-to-cell infection (x axis) and/or manipulate the $CD4^+$ T cell activation process (y axis, see Fig. 8.3). The black squares represent treatments that reduced the density of infected cells, latent cells and free virions all below 10^{-12} per μl.

limiting, our model predicts that gut lymphoid tissue would provide an ideal microenvironment for rapid propagation of HIV, at least until the majority of target cells are infected or die. Further development of the basic model, to allow heterogeneity associated with different anatomical compartments would allow this prediction to be tested directly.

The model also predicts a dominant role for cell-to-cell spread in the late phase of HIV. As in the scenario of an infection early on in the gut lymphoid tissue, our model predicts that the late stage of the disease will be associated with a large number of infected cells, combined with a large proportion of activated target cells. A high number of infected cells is a simple corollary of the very high levels of free virus late in infection in the absence of treatment. There is also substantial experimental evidence for increased immune activation in the late phases of HIV [211], and indeed this has been proposed as an important contribution to pathogenesis. Thus cell-to-cell spread is likely to become the dominant mode of transmission in the late stages of HIV. This may be important in light of recent data showing that the different components of current HAART display variable efficacy against cell-to-cell spread [202, 233].

Interestingly, the hybrid spreading mechanism employed by HIV is reminiscent of those of some computer worms such as Red Code II and Cornficker [22, 230], which allocate their resources between probing for susceptible target computers in local area networks and globally across the Internet. Like HIV, local interactions have a high chance of success but access only a limited number of targets while global spread targets a much larger number of targets with a much lower probability of success. Modelling studies have shown that this hybrid spreading is required to explain the large outbreak of such worms on the Internet [231]. It is tempting to speculate that hybrid spreading may contribute to the pathogenicity and dynamics of infection of other viruses that employ parrallel spreading mechanisms, for example the Hepatitis C virus [229].

The current model incorporates some simplifying assumptions. Nevertheless, the model does provide some insight into the effectiveness of therapy at different stages of disease. Specifically, the model strongly supports the hypothesis that interfering with viral infection early in HIV progression is likely to have a major impact on the subsequent progress of the disease. Interfering with viral transmission is predicted

by our model to be much more effective early in HIV infection. The clinical decision about when to start therapy remains a matter of debate. Current WHO guidelines suggest commencing treatment at $CD4^+$ T cell density of $> 350 cells/\mu l$ and $< 500 cells/\mu l$ [234]. However, studies exploring much earlier commencements of treatment, have claimed increased efficacy [46, 238]. For example, the recent "Short Pulse Anti-Retroviral Therapy at Seroconversion" (SPARTAC) trial, demonstrated a long term clinical benefit of a limited period of ART soon after seroconversion [46]. Our model accurately predicted the results of the SPARTAC trial providing further support for the model's generality and robustness.

Anti-viral therapies that specifically target cell-to-cell spread are not currently available, but are clearly important therapeutic goals [233]. A number of previous studies have proposed combining antiviral therapies with therapies that either limit CD4+ T cell activation (thus reducing the number of target susceptible cells) or increased T cell activation, thus flushing out residual latent cells. These approaches have not provided clear cut clinical benefits [239,240]. Indeed our model suggests that the outcome of such manipulation of cellular activation will be critically dependent on the time at which it is administered, and the efficiency of concomitant antiviral therapy. Targeted suppression of $CD4^+$ T cell activation, in combination with antiviral therapies may nevertheless offer a useful approach, if used early on in infection.

8.6 Summary

This chapter showed that the two modes of spread of HIV infection, cell-to-cell spread and cell-free spread, complement each other and both make important contributions to disease progression. This work is the first to reasonably predict that cell-to-cell spread becomes increasingly effective as infection progresses and thus may present a considerable treatment barrier for eradicating the disease.

The analysis in this chapter provided specific predictions that emerge from the close links between $CD4^+$ T cell activation and cell-to-cell spread, and their combined contribution to both early and late phases of HIV infection. These predictions emphasise the potential benefits of early or prophylactic treatment with antiretroviral therapies, and suggest that drugs with the ability to effectively block cell-to-cell spread may provide significant therapeutic benefit in long term management or eradication of HIV

infection.

Chapter 9

Discussion

This chapter discusses the limitations of our study and suggests future research directions.

9.1 Limitations

It is important to note that the research presented in this thesis has its limitations. This is mainly due to two reasons.

Firstly, like many other modelling studies, we need to simplify the real systems and introduce assumptions to a model in order to reduce the model's complexity and conduct mathematical analysis. For example the hybrid epidemic model proposed in Chapter 4 assumes the local spreading network is uncorrelated. This assumption significantly simplifies theoretical derivation but it prevents the model from accurately describing hybrid epidemics with local spreading on correlated networks (e.g. clustered networks [65, 67]). More sophisticated mathematical models for epidemic spreading in correlated networks [17, 241] can be adopted to overcome this limitation.

Secondly, there are many issues about the real phenomena that we still do not understand entirely. This is particularly true for HIV. For example, clinical studies indicate that the cell activation rate increases with the progression of HIV infection [211, 212], but the quantitative mechanism of such relation is still unclear. In the HIV infection model in Chapter 7, we assume that the cell activation rate is linearly related to the progress of HIV progression. Further clinical investigations on the detailed mechanisms of HIV infection can help clarify the relation and improve the model.

9.2 Future directions

Many important future studies can be built upon our models and results.

9.2.1 Spreading speed and survival time of hybrid epidemics

Most existing hybrid epidemic spreading studies largely overlook the duration of an epidemic. They focus on modelling the final outbreak size and how it is affected by other epidemic parameters. Survival time, nevertheless, is a practical and important factor in helping to better understand and control many real epidemics. Some computer worms increase their survival by intentionally reducing their spreading speed to avoid detection [26]. Survival time can be as important, if not more important than outbreak size for computer worms because many of them seek to establish a lasting network that can be controlled and exploited by the worm designer thereafter. The computer worm Conficker, for example, is still active on the Internet, 6 years after its outbreak [28]. The HIV infection (spread) within human body also persists for years before it eventually takes a patient's life. Consequently, these empirical realities throw light on the practical importance of systematically modelling and in turn helping understand the epidemic survival time in the future.

This is especially true for hybrid epidemics where the tradeoff among spreading mechanisms can be manipulated. Many interesting and useful aspects about how and why the tradeoff affects a hybrid epidemic's survival are yet to be investigated. For example in Chapter 4 and Chapter 6, we found that, while causing the same final outbreak size, a hybrid epidemic can either spread quickly and survive for a short temporal period, or spread slowly and survive for a considerably longer temporal period. As shown in Fig. 6.5 and Fig. 4.4, two different tradeoff values can lead to a same outbreak size of a hybrid epidemic, but with different survival time and spreading speed. This is an interesting finding that should be further investigated. The simulation results in Chapter 6 about the dynamics of the computer worm Conficker's spreading have provided some relevant insight. We aim to further model, explain, and utilise this finding in the future.

9.2.2 Additional case studies

We have conducted case study on two distinct real hybrid epidemics: the computer worm Conficker spreading on the Internet and HIV infection within the human body.

However, these only represent only two of the many real world hybrid epidemics, many of which require new and specific hybrid epidemic models to incorporate their unique characteristics.

For example the Hepatitis C Virus (HCV), similar to the Human Immunodeficiency Virus (HIV), also employs both cell-free and cell-to-cell spreading mechanisms [242]. But they have different infection targets. While HIV infects $CD4^+$ T cells of the human immune system, HCV infects the liver's tissue cells. Compared with HIV, HCV also has different interactions with the immune system. So that, even though HCV shares similar spreading mechanisms with HIV, HIV's epidemic model cannot be directly applied to HCV. A new and more specific hybrid epidemic model is necessary to take into account HCV's unique characteristics.

Other hybrid epidemic cases include the spreading of mobile phone viruses [102], Code-Red [22] and many other computer worms [148].

9.2.3 Research on future HIV treatments

Our HIV model provides a strategy to theoretically evaluate the long term effect of treatment strategies on HIV infection progression. This is typified by our reproduction and evaluation of the results of the Short Pulse Anti-Retroviral Therapy at Seroconversion (SPARTAC) trial [46] in Chapter 8. Following our analysis on the SPARTAC trial, we expect researchers to use our model to evaluate more future treatment strategies.

A very important value of our HIV model is that it can help pre-select the most promising treatment strategies among a potential long list of candidates before conducting a treatment trial. Trials are often expensive and time-consuming. Being able to filter out unpromising treatment candidates can increase the success rate of treatment trials, and save time and resources.

9.2.4 Controlling strategies for hybrid epidemics

It is of practical importance to be able to control the spreading of epidemics. On the one hand, positive epidemics, such as the dissemination of useful information, should be promoted as they contribute to society. While, on the other hand, negative epidemics, such as the spreading of computer worms and viruses on the Internet, should be mitigated because they have damaging consequences to the society or economy.

Our results in Chapter 5 demonstrated that it is indeed possible to have a highly

contagious epidemic on a metapopulation by mixing local and global spreading mechanisms. And the tradeoff between local and global spreading plays a critical role in determining a hybrid epidemic's outbreak size. In the future, more efficient hybrid spreading systems can be built for positive epidemics, by using our model to optimise the tradeoff between different spreading mechanisms.

In Chapter 6, we show for the computer worm Conficker, an example of a negative hybrid epidemic, mixing multiple spreading mechanisms plays an important role in its enormous outbreak and long lasting survival. Most of the worm's infections were caused by localised spreading mechanisms (See Fig. 6.3). A promising direction to mitigate such epidemics can be community-based protection strategies [21].

9.2.5 Immunisation against hybrid epidemics

As mentioned in Chapter 2, immunisation strategies protect a network against harmful epidemics and they are an important branch of epidemic spreading research. Most existing immunisation strategies try to fragment an epidemic spreading network into many separated components [111]. The level of fragmentation of a network is often the criterion to compare different immunisation strategies. These strategies largely work for epidemics employing only a local spreading mechanism where infection is transmitted solely through connected individuals on a network. Their effectiveness against hybrid epidemics, however, is highly questionable. This is because, hybrid epidemics also employ global spreading that is not constrained by a spreading network. Although normally with a very low infection rate, global spreading enables an infected individual to infect any other individuals. That is, even if the local spreading network can be fragmented into separate components by existing immunisation strategies, global spreading can persist and consequently still spread the epidemic between these components.

Therefore, there is a practical need to re-evaluate existing immunisation strategies against hybrid epidemics. Furthermore, more efficient immunisation strategies specifically targeting hybrid epidemics should be investigated. This thesis does not directly provide these immunisation strategies, which is an important direction for future research.

Chapter 10

Conclusions

Hybrid epidemics are ubiquitous in nature and society but we still lack understanding of its important dynamics. Especially the *critically* hybrid epidemics, where a mix of different mechanisms is a necessary condition for an epidemic outbreak, have rarely been explored and explained. The motivation of this research was to improve our understanding of hybrid epidemics.

Firstly a general framework to model hybrid epidemics was introduced in Chapter 4. Results based on theoretical analysis and numerical simulations in Chapter 5 showed that properties of hybrid epidemics are critically determined by a hybrid tradeoff. The hybrid tradeoff defines the proportion of resource allocated to two spreading mechanisms: local spreading among individuals in a community and global spreading among all individuals in all communities. For an epidemic spreading in a metapopulation with a small global infection rate, our results indicate that the optimal hybrid tradeoff that leads to the largest outbreak is determined by β_1/γ (local infection probability/recovery probability).

To investigate hybrid epidemics in real world scenarios, case studies were conducted on two distinct epidemics: the computer worm Conficker spreading on the Internet and the Human Immunodeficiency Virus (HIV) infection within the human body.

Chapter 6 analysed the spreading of the computer worm Conficker in detail, inferring the worm's epidemic parameters from Internet measurement data. Our simulations and analysis results revealed that individually each of Conficker's spreading mechanisms is not sufficient for spreading, however, the worm successfully produced one of the most serious outbreaks on the Internet by combining multiple simple and ineffective spreading mechanisms.

By collaborating with immunologists, we not only were able to model the spread of HIV infection within the human body, but also provided novel insight into HIV pathogenesis and treatment. Chapter 7 proposed a mathematical epidemic model that incorporates HIV's two infection mechanisms: cell-free spread and cell-to-cell spread. The model can accurately recapitulate the entire HIV infection course as observed in clinical trials. The model was then used in Chapter 8 to investigate HIV infection dynamics and estimate the long term effect of therapeutic strategies on HIV infection progression. Our results suggested that the two spreading modes of HIV infection complement each other and both make important contributions to disease progression. In addition, results in this thesis emphasise the potential benefits of early HIV treatment, and suggest that drugs with the ability to effectively block cell-to-cell spread may provide significant therapeutic benefit in the long term management or eradication of the HIV infection.

In summary, this PhD research has: (1) advanced our understanding of hybrid epidemics; (2) provided mathematical frameworks for future analysis; and (3) demonstrated, with two successful case studies, that such research can have a significant impact on important issues such as cyberspace security and human health.

This study laid the foundation for analysing hybrid epidemics, especially critically hybrid epidemics. Future studies can build upon our models and results to further advance our understanding of hybrid epidemics in general as well as with specific real world cases.

Appendix A

Author's Publications

- Changwang Zhang, Shi Zhou, Elisabetta Groppelli, Pierre Pellegrino, Ian Williams, Persephone Borrow, Benjamin M. Chain, Clare Jolly. "Hybrid spreading mechanisms and T cell activation shape the dynamics of HIV-1 infection". *PLoS Computational Biology*, 11(4):e1004179, 2015.

  Author Contributions: Conceived and designed the experiments: **CZ** SZ CJ BMC. Performed the experiments: **CZ** EG PP IW PB. Analyzed the data: **CZ** SZ BMC CJ. Contributed reagents/materials/analysis tools: IW PB. Wrote the paper: **CZ** SZ PB BMC CJ.

- Changwang Zhang, Shi Zhou, Joel C. Miller, Ingemar J. Cox, and Benjamin M. Chain. "Optimizing Hybrid Spreading in Metapopulations". *Scientific Reports*, 5:9924, 2015.

  Author Contributions: **CZ**, SZ, IJC, and BMC designed the study. **CZ** and JCM conducted the mathematical modelling and derivation. **CZ** performed the computational analysis and simulations. SZ and BMC wrote the manuscript with contributions from **CZ** and JCM and IJC.

- Changwang Zhang, Shi Zhou, and Benjamin M. Chain. "Hybrid Epidemics - A Case Study on Computer Worm Conficker". *PLoS ONE*, 2015. (Accepted)

  Author Contributions: Conceived and designed the experiments: **CZ** SZ BMC. Performed the experiments: **CZ**. Analyzed the data: **CZ** SZ BMC. Wrote the paper: **CZ** SZ BMC.

- Changwang Zhang, Shi Zhou, Jie Li and Benjamin M. Chain. "LeoTask: a productive and reliable framework for computational research", *Bioinformatics*, 2015. (Under review)

- Changwang Zhang, Zhiping Cai, Weifeng Chen, Xiapu Luo, and Jianping Yin,"Flow level detection and filtering of low-rate DDoS". *Computer Networks*, 56(15):3417-3431, 2012.

- Changwang Zhang, Shi Zhou, and Benjamin M. Chain. "Immunising the Internet". *Second Annual CLMS Symposium*, London, UK, 2012.

Appendix B

Glossaries

B.1 About computer worms

- **Conficker**: Conficker is a computer worm which erupted on the Internet in 2008.
- **IP addresses**: Each computer on the Internet is associated with an Internet Protocol (IP) address. An IP address represents a computer.
- **LAN**: Local Area Network (LAN) consisting of computers whose IP addresses share the same prefix.
- **Probing**: A computer worm probes IP addresses. If an IP address associated with a computer susceptible to the worm is probed, the computer will be infected.
- **Subnet**: A sub network of the Internet. A subnet consists of computers among which a worm can spread from one infected to all others through local probing.

B.2 About Human Immunodeficiency Virus (HIV)

- **HIV**: Human Immunodeficiency Virus (HIV) is a virus that attacks human immune system - the defence to protect body against diseases.
- **AIDS**: Without medical treatment, people infected by HIV can develop Acquired Immune Deficiency Syndrome (AIDS), which is a condition representing failure of the immune system.
- **Cell-free infection**: Virus particles bud from an infected T cell, enter the blood/extracellular fluid and then can infect another T cell.

- **Cell-to-cell infection**: HIV infection transmits through direct cell-to-cell contacts between T cells.
- **Virion**: virus particle
- **CD4$^+$ T cells**: A type of cells that facilitate the function of the immune system. They are the primary target for HIV infection.
- **CD4$^+$ count**: The density of CD4$^+$ T cells in a blood sample.
- **Viral load**: The density of HIV virions in a blood sample.
- **Cellular immune response**: A response, in which cells of the immune system kill infected cells, including HIV-infected cells.
- **HIV treatment naive**: A patient is HIV treatment naive if he/she has never taken HIV medicine.

Appendix C

LeoTask - a fast, flexible and reliable framework for computational research

This study involves intensive computation for theoretical predictions and numerical simulations. To facilitate such computation, we developed a Java framework for computation-intensive and time-consuming research tasks.

C.1 Introduction

Research tasks, especially in the field of bioinformatics, are increasingly computationally intensive [243]. Computational research (e.g. simulations and data analysis) typically explores results over a large parameter space and often needs to repeat a task a sufficiently large number of times to get an average result. Many complex computational tasks can last for days, or even weeks. As a result, they are prone to artificial (e.g. a colleague sharing the same computing facility stops your program) or natural (e.g. power outage) interruptions.

To accelerate computational research, it is imperative to conduct tasks in parallel, fully utilising the processing power of computing facilities [243]. Presently, computing facilities typically have multiple cores in their Central Processing Unit (CPU), and each of the cores can individually conduct a processing task [244, 245]. For example, the latest desktop computer can have 4 to 8 cores, while a computing sever can have more than 16 cores.

While there are built-in mechanisms for running simultaneous tasks in all major programming languages, the level of complexity often makes these built-in mechanisms difficult to use and time consuming to program with. For example, the built-in parallel

running mechanism in Java requires choosing for each function of a program whether it can be executed by multiple tasks simultaneously or by only one task at a time. Such choices affect not only the speed of a program but also its accuracy, i.e. a poor choice could result in a slow program and furthermore a program that provides inaccurate results.

Reliability is a critical but often overlooked feature of many computational programs and frameworks. A reliable program should be able to recover and continue running despite interruptions. Considerable effort is often needed to make a program reliable, especially if the program runs in parallel.

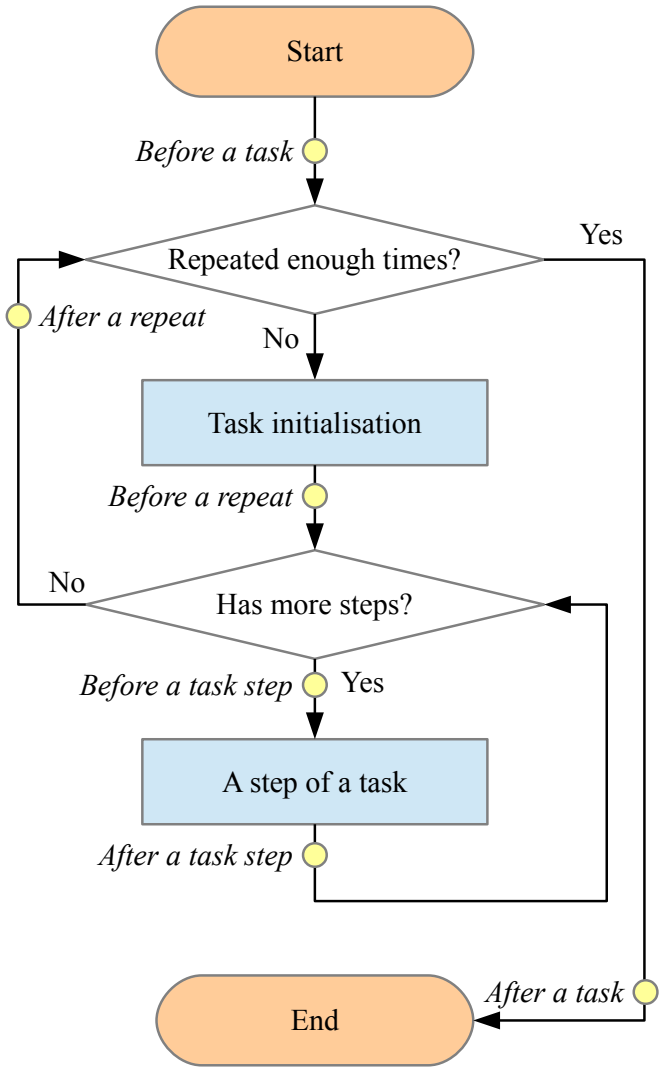

Figure C.1: Default flow and time points for a task. There are 6 default built-in time points for result collection: 1) before the start of a task, 2) before starting a repeated run of a task, 3) before starting a step in a run, 4) after finishing a step in a run, 5) after finishing a repeated run of a task, and 6) after finishing a task.

C.2 Features

Here we present a framework, LeoTask, to reliably facilitate conducting research tasks in parallel. It has the following combination of features that will be attractive to members of the computational research community:

- *Automatic & parallel parameter space exploration*

  LeoTask uses an intuitive configuration file to specify the value ranges of task parameters. The framework will figure out all possible combinations of values for all parameters and then run tasks with different parameter value combinations in parallel, i.e. LeoTask automatically explores the parameter space. The framework maps and runs all the tasks to all available processing cores of a computing facility.

- *Flexible & configuration-based result aggregation*

  The configuration file also sets when and how task results are aggregated. As shown in Fig. C.1, LeoTask has a default task flow with 6 default time points for specifying when the results are collected. Applications using the framework can also use different task flows and define additional time points. The framework supports aggregating results conditioned on a set of parameters. For example, for a task with parameter x_1, x_2 and result y. Given the value range of x_1 and x_2, the framework can aggregate y conditioned on the value of x_1, x_2, value pair (x_1, x_2), or any mathematical function of x_1 and x_2.

- *Programming model focussing only on the key logic*

  The framework separates the key logic of a task from other "bookkeeping" parts of the program, which include parameter setting, result aggregation, result plotting, etc. Users only need to program the key logic and use the configuration file to set other "bookkeeping" aspects of a task. This facilitates sharing a program among a community where many end users only need to rerun a program with different parameter values and different ways to aggregate results. The framework frees those end users from reading the program and they only need to modify a more intuitive configuration file.

- *Reliable & automatic interruption recovery*

LeoTask periodically writes the current state and results of the tasks to checkpoint files and can then recover and continue running from a checkpoint file when necessary. To recover from an interruption and still guarantee the completeness and correctness of results, a program has to track the tasks running in parallel. The framework handles these technique details automatically.

Considering that Java programs can run *directly* on all major computing facilities (with different operating systems), this feature provides not only reliability but also flexibility for running a program. The following scenario is possible for a program using LeoTask: a user first started the program on a laptop; after several hours the user determined that it was running too slow, thus copied the program and a check point file to a server A and continued running the program there; after another several hours, the running speed decreased on server A because other colleagues began to use server A and made it busy, so the user again copied the program and a check point file from sever A to another unused server B and finished the remaining part of the program there.

- *Dynamic & cloneable networks structures*

  Network analysis is widely used in research, particularly among bioinformatic studies [181,246]. For example, there are transcription networks [247], signalling networks [248], epidemic spreading networks [3], etc. LeoTask provides data structures to represent a node, a link, a network, and a network set. A network consists of nodes and links. A network set includes multiple networks and they can overlap (share nodes) with each other. All structures can be dynamic [249], supporting adding, deleting, and updating elements. The framework also allows merging multiple networks.

  LeoTask can create copies of a large network system including the states of each node, link, network, and network set. This enables users to investigate many common scenarios that were previously time consuming to program. For example, a user can simulate epidemic spreading for the first 10 minutes and make five copies of the current network system (preserving the states of networks, nodes, and links) to test five different intervention strategies on these five copies *in parallel* so that the initial conditions are random but at the same time *exactly* the

same for these different strategies tested.

- *Integration with Gnuplot*

  Gnuplot is a piece of widely-used open-source plotting software. LeoTask can output aggregated results directly as Gnuplot scripts which can be processed by Gnuplot to produce publication-quality figures. In addition, LeoTask includes a unique hybrid programming engine that compiles Gnuplot scripts and enables users to refer to values of Java objects or call Java functions within Gnuplot scripts.

C.3 Availability

The source code for LeoTask is freely available under FreeBSD License at `https://github.com/mleoking/leotask`

C.4 Discussion

The LeoTask framework has been used in epidemic spreading simulations [231], large dataset analysis [230], and HIV model parameter fitting from clinical records.

We believe LeoTask's combination of features makes it useful for the broader research community. We also welcome suggestions or direct contributions to improve LeoTask and make it more widely accessible.

Bibliography

[1] A. Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32, Jan 2012.

[2] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275, Oct 2008.

[3] M. Newman. *Networks: An Introduction*. Oxford University Press, May 2010.

[4] M. Boguá, R. Pastor-Satorras, and A. Vespignani. Epidemic Spreading in Complex Networks with Degree Correlations. *Statistical Mechanics of Complex Networks*, 625:127–147. Springer Berlin Heidelberg, 2003.

[5] M. R. Subramani and B. Rajagopalan. Knowledge-sharing and Influence in Online Social Networks via Viral Marketing. *Commun. ACM*, 46(12):300–307, Dec 2003.

[6] J. Leskovec, L. A. Adamic, and B. A. Huberman. The Dynamics of Viral Marketing. *ACM Trans. Web*, 1(1), May 2007.

[7] W. Chen, C. Wang, and Y. Wang. Scalable Influence Maximization for Prevalent Viral Marketing in Large-scale Social Networks. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1029–1038. KDD'10. ACM, 2010.

[8] X. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing. *Computer Networks*, 51(10):2867–2891, Jul 2007.

[9] Y. Lin, B. Li, and B. Liang. Stochastic analysis of network coding in epidemic routing. *IEEE Journal on Selected Areas in Communications*, 26(5):794–808, Jun 2008.

[10] Z. Ren, W. Liu, X. Zhou, J. Fang, and Q. Chen. Summary-Vector-Based Effective and Fast Immunization for Epidemic-Based Routing in Opportunistic Networks. *IEEE Communications Letters*, 18(7):1183–1186, Jul 2014.

[11] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic Algorithms for Replicated Database Maintenance. *Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing*, pages 1–12. PODC '87. ACM, 1987.

[12] P. Eugster, R. Guerraoui, A.-M. Kermarrec, and L. Massoulie. Epidemic information dissemination in distributed systems. *Computer*, 37(5):60–67, May 2004.

[13] A.-M. Kermarrec and M. v. Steen. Gossiping in Distributed Systems. *SIGOPS Oper. Syst. Rev.*, 41(5):2–7, Oct 2007.

[14] M. Vojnović, V. Gupta, T. Karagiannis, and C. Gkantsidis. Sampling Strategies for Epidemic-Style Information Dissemination. *IEEE/ACM Transactions on Networking*, 18(4):1013–1025, Aug 2010.

[15] P.-Y. Chen, S.-M. Cheng, and K.-C. Chen. Optimal Control of Epidemic Information Dissemination Over Networks. *IEEE Transactions on Cybernetics*, 44(12):2316–2328, Dec 2014.

[16] F. Sahneh, F. Chowdhury, G. Brase, and C. Scoglio. Individual-based Information Dissemination in Multilayer Epidemic Modeling. *Mathematical Modelling of Natural Phenomena*, 9(02):136–152, Jan 2014.

[17] K. Rock, S. Brand, J. Moir, and M. J. Keeling. Dynamics of infectious diseases. *Reports on Progress in Physics*, 77(2):026602, Feb 2014.

[18] M. J. Keeling and K. T. D. Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, Sep 2005.

[19] R. Pastor-Satorras and A. Vespignani. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, 86(14):3200, Apr 2001.

[20] R. M. Anderson. Discussion: The Kermack-McKendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, 53(1-2):1–32, Mar 1991.

[21] S. Shin, G. Gu, N. Reddy, and C. Lee. A Large-Scale Empirical Study of Conficker. *IEEE Transactions on Information Forensics and Security*, 7(2):676–690, Apr 2012.

[22] D. Moore, C. Shannon, and K. Claffy. Code-Red: a case study on the spread and victims of an internet worm. *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 273–284. IMW. ACM, 2002.

[23] I. Z. Kiss, D. M. Green, and R. R. Kao. The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Mathematical Biosciences*, 203(1):124–136, Sep 2006.

[24] Y. Wang and Z. Jin. Global analysis of multiple routes of disease transmission on heterogeneous networks. *Physica A: Statistical Mechanics and its Applications*, 392(18):3869–3880, Sep 2013.

[25] F. Barré-Sinoussi, A. L. Ross, and J.-F. Delfraissy. Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology*, 11(12):877–883, Dec 2013.

[26] E. Chien. Downadup: Attempts at Smart Network Scanning. Jul 2010. `http://www.symantec.com/connect/blogs/downadup-attempts-smart-network-scanning`. Accessed Dec 2014.

[27] The Rendon Group. Conficker Working Group: Lessons Learned. Jan 2011. `http://www.confickerworkinggroup.org/wiki/uploads/Conficker_Working_Group_Lessons_Learned_17_June_2010_final.pdf`. Accessed Dec 2014.

[28] ESET Virusradar. Win32/Conficker Charts. Jun 2014. `http://www.virusradar.com/en/Win32_Conficker/chart/week`. Accessed Dec 2014.

[29] D. Balcan and A. Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature physics*, 7(7):581–586, Jul 2011.

[30] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Aug 1992.

[31] N. T. J. Bailey. *The mathematical theory of epidemics*. Hafner, 1957.

[32] W. O. Kermack and A. G. McKendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A*, 115(772):700–721, Aug 1927.

[33] H. W. Hethcote. The Mathematics of Infectious Diseases. *SIAM Review*, 42(4):599, 2000.

[34] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, Jul 2002.

[35] Center for Applied Internet Data Analysis. UCSD Network Telescope – Three Days Of Conficker Dataset. 2008. `http://www.caida.org/data/passive/telescope-3days-conficker_dataset.xml`. Accessed Dec 2014.

[36] Center for Applied Internet Data Analysis. The UCSD Network Telescope. 2012. `http://www.caida.org/projects/network_telescope/`. Accessed Dec 2014.

[37] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin. Networks formed from interdependent networks. *Nature Physics*, 8(1):40–48, 2012.

[38] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Robustness of a Network of Networks. *Physical Review Letters*, 107(19):195701, Nov 2011.

[39] J. Gómez-Gardeñes, I. Reinares, A. Arenas, and L. M. Floría. Evolution of Cooperation in Multiplex Networks. *Scientific Reports*, 2, Aug 2012.

[40] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas. Diffusion Dynamics on Multiplex Networks. *Physical Review Letters*, 110(2):028701, Jan 2013.

[41] A. S. Perelson and R. M. Ribeiro. Modeling the within-host dynamics of HIV infection. *BMC Biology*, 11(1):96, Sep 2013.

[42] R. Li, L. Gan, and Y. Jia. Propagation Model for Botnet Based on Conficker Monitoring. *International Symposium on Information Science and Engineering (ISISE)*, pages 185–190. Dec 2009.

[43] Y. Yao, W.-l. Xiang, H. Guo, G. Yu, and F.-X. Gao. Diurnal Forced Models for Worm Propagation Based on Conficker Dataset. *International Conference on Multimedia Information Networking and Security (MINES)*, pages 431–435. Nov 2011.

[44] C. Fraser, K. Lythgoe, G. E. Leventhal, G. Shirreff, T. D. Hollingsworth, S. Alizon, and S. Bonhoeffer. Virulence and Pathogenesis of HIV-1 Infection: An Evolutionary Perspective. *Science*, 343(6177):1243727, Mar 2014.

[45] Center for Applied Internet Data Analysis. UCSD Network Telescope – Two Days in November 2008 Dataset. 2008. `http://www.caida.org/data/passive/telescope-2days-2008_dataset.xml`. Accessed Dec 2014.

[46] S. Fidler, K. Porter, F. Ewings, J. Frater, G. Ramjee, D. Cooper, H. Rees, M. Fisher, M. Schechter, and e. al. Short-course antiretroviral therapy in primary HIV infection. *The New England journal of medicine*, 368(3):207–217, Jan 2013.

[47] T. House. Modelling epidemics on networks. *Contemporary Physics*, 53(3):213, 2012.

[48] T. House and M. J. Keeling. Deterministic epidemic models with explicit household structure. *Mathematical Biosciences*, 213(1):29–39, May 2008.

[49] F. Ball, D. Sirl, and P. Trapman. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Mathematical Biosciences*, 224(2):53–73, Apr 2010.

[50] F. Ball. An SIR epidemic model on a population with random network and household structure, and several types of individuals. *Advances in Applied Probability*, 44(1):63–86, Mar 2012.

[51] J. Ma, P. v. d. Driessche, and F. H. Willeboordse. Effective degree household network disease model. *Journal of Mathematical Biology*, 66(1-2):75–94, Jan 2013.

[52] A. Vazquez. Epidemic outbreaks on structured populations. *Journal of Theoretical Biology*, 245(1):125–129, Mar 2007.

[53] D. J. Watts, R. Muhamad, D. C. Medina, and P. S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32):11157–11162, 2005.

[54] V. Colizza and A. Vespignani. Invasion Threshold in Heterogeneous Metapopulation Networks. *Physical Review Letters*, 99(14):148701, Oct 2007.

[55] A. S. Mata, S. C. Ferreira, and R. Pastor-Satorras. Effects of local population structure in a reaction-diffusion model of a contact process on metapopulation networks. *Physical Review E*, 88(4):042820, Oct 2013.

[56] Y. Min, X. Jin, Y. Ge, and J. Chang. The Role of Community Mixing Styles in Shaping Epidemic Behaviors in Weighted Networks. *PLoS ONE*, 8(2):e57100, Feb 2013.

[57] M. J. Keeling, L. Danon, M. C. Vernon, and T. A. House. Individual identity and movement networks for disease metapopulations. *Proceedings of the National Academy of Sciences*, 107(19):8866–8870, May 2010.

[58] A. Apolloni, C. Poletto, J. J. Ramasco, P. Jensen, and V. Colizza. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling*, 11(1):3, Jan 2014.

[59] J. P. Gleeson. Bond percolation on a class of clustered random networks. *Physical Review E*, 80(3):036107, Sep 2009.

[60] B. Bollobás, S. Janson, and O. Riordan. Sparse random graphs with clustering. *Random Structures & Algorithms*, 38(3):269–323, 2011.

[61] T. Britton, M. Deijfen, A. N. Lagerås, and M. Lindholm. Epidemics on random graphs with tunable clustering. *Journal of Applied Probability*, 45(3):743–756, 2008.

[62] J. P. Gleeson and S. Melnik. Analytical results for bond percolation and k-core sizes on clustered networks. *Physical Review E*, 80(4):046121, Oct 2009.

[63] J. C. Miller. Spread of infectious disease through clustered populations. *Journal of The Royal Society Interface*, 6(41):1121–1134, Mar 2009.

[64] J. C. Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901, Aug 2009.

[65] M. E. J. Newman. Random Graphs with Clustering. *Physical Review Letters*, 103(5):058701, Jul 2009.

[66] M. J. Tildesley, T. A. House, M. C. Bruhn, R. J. Curry, M. O'Neil, J. L. E. Allpress, G. Smith, and M. J. Keeling. Impact of spatial clustering on disease transmission and optimal control. *Proceedings of the National Academy of Sciences*, 107(3):1041–1046, Jan 2010.

[67] E. Volz, J. Miller, A. Galvani, and L. Meyers. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Computational Biology*, 7(6):e1002042, 2011.

[68] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, Dec 2009.

[69] S. Bansal, J. Read, B. Pourbohloul, and L. A. Meyers. The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics*, 4(5):478–489, Sep 2010.

[70] J. Fernández-Gracia, V. M. Eguíluz, and M. San Miguel. Update rules and interevent time distributions: Slow ordering versus no ordering in the voter model. *Physical Review E*, 84(1):015103, Jul 2011.

[71] P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, Oct 2012.

[72] R. Lambiotte, L. Tabourier, and J.-C. Delvenne. Burstiness and spreading on temporal networks. *The European Physical Journal B*, 86(7):1–4, Jul 2013.

[73] N. Masuda and P. Holme. Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Reports*, 5, Mar 2013.

[74] B. Min, K.-I. Goh, and A. Vazquez. Spreading dynamics following bursty human activity patterns. *Physical Review E*, 83(3):036102, Mar 2011.

[75] Y. Zhu, D. Li, W. Guo, and F. Zhang. Effect of Heterogeneity of Vertex Activation on Epidemic Spreading in Temporal Networks. *Mathematical Problems in Engineering*, 2014:e409510, Apr 2014.

[76] O. Yağan and V. Gligor. Analysis of complex contagions in random multiplex networks. *Physical Review E*, 86(3):036103, Sep 2012.

[77] K.-M. Lee, J. Y. Kim, W.-k. Cho, K.-I. Goh, and I.-M. Kim. Correlated multiplexity and connectivity of multiplex random networks. *New Journal of Physics*, 14(3):033027, Mar 2012.

[78] C. D. Brummitt, K.-M. Lee, and K.-I. Goh. Multiplexity-facilitated cascades in networks. *Physical Review E*, 85(4):045102, Apr 2012.

[79] M. Szell, R. Lambiotte, and S. Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, Aug 2010.

[80] C. Granell, S. Gómez, and A. Arenas. Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks. *Physical Review Letters*, 111(12):128701, Sep 2013.

[81] K.-M. Lee, B. Min, and K.-I. Goh. Towards real-world complexity: an introduction to multiplex networks. *The European Physical Journal B*, 88(2):120, Feb 2015.

[82] C. D. Brummitt, R. M. D'Souza, and E. A. Leicht. Suppressing cascades of load in interdependent networks. *Proceedings of the National Academy of Sciences*, 109(12):E680–E689, Mar 2012.

[83] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, Apr 2010.

[84] J. Chen, H. Zhang, Z.-H. Guan, and T. Li. Epidemic spreading on networks with overlapping community structure. *Physica A: Statistical Mechanics and its Applications*, 391(4):1848–1854, Feb 2012.

[85] M. Dickison, S. Havlin, and H. E. Stanley. Epidemics on interconnected networks. *Physical Review E*, 85(6):066109, Jun 2012.

[86] G. Dong, J. Gao, L. Tian, R. Du, and Y. He. Percolation of partially interdependent networks under targeted attack. *Physical Review E*, 85(1):016112, Jan 2012.

[87] E. A. Leicht and R. M. D'Souza. Percolation on interacting networks. *arXiv:0907.0894*, Jul 2009.

[88] W. Li, A. Bashan, S. V. Buldyrev, H. E. Stanley, and S. Havlin. Cascading Failures in Interdependent Lattice Networks: The Critical Role of the Length of Dependency Links. *Physical Review Letters*, 108(22):228702, May 2012.

[89] R. Parshani, S. V. Buldyrev, and S. Havlin. Interdependent Networks: Reducing the Coupling Strength Leads to a Change from a First to Second Order Percolation Transition. *Physical Review Letters*, 105(4):048701, Jul 2010.

[90] R. Parshani, S. V. Buldyrev, and S. Havlin. Critical effect of dependency groups on the function of networks. *Proceedings of the National Academy of Sciences*, 108(3):1007–1010, Jan 2011.

[91] A. Saumell-Mendiola, M. A. Serrano, and M. Boguñá. Epidemic spreading on interconnected networks. *Physical Review E*, 86(2):026106, Aug 2012.

[92] S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski. Percolation theory on interdependent networks based on epidemic spreading. *EPL (Europhysics Letters)*, 97(1):16006, Jan 2012.

[93] A. Vespignani. Complex networks: The fragility of interdependency. *Nature*, 464(7291):984–985, Apr 2010.

[94] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16):6872–6877, Apr 2009.

[95] S. Funk, M. Salathé, and V. A. A. Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of The Royal Society Interface*, 7(50):1247–1256, Sep 2010.

[96] S. Funk, E. Gilad, and V. a. A. Jansen. Endemic disease, awareness, and local behavioural response. *Journal of Theoretical Biology*, 264(2):501509, May 2010.

[97] Z. Ruan, M. Tang, and Z. Liu. Epidemic spreading with information-driven vaccination. *Physical Review E*, 86(3):036117, Sep 2012.

[98] F. D. Sahneh, F. N. Chowdhury, and C. M. Scoglio. On the existence of a threshold for preventive behavioral responses to suppress epidemic spreading. *Scientific Reports*, 2, Sep 2012.

[99] Q. Wu, X. Fu, M. Small, and X.-J. Xu. The impact of awareness on epidemic spreading in networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(1):013101, Mar 2012.

[100] L. Yan-Ling, J. Guo-Ping, and S. Yu-Rong. Epidemic spreading on a scale-free network with awareness. *Chinese Physics B*, 21(10):100207, Oct 2012.

[101] D. Balcan, V. Colizza, B. Gonalves, H. Hu, J. J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, Dec 2009.

[102] P. Wang, M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding the Spreading Patterns of Mobile Phone Viruses. *Science*, 324(5930):1071–1076, May 2009.

[103] S. Meloni, N. Perra, A. Arenas, S. Gómez, Y. Moreno, and A. Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific Reports*, 1, Aug 2011.

[104] R. Pastor-Satorras and A. Vespignani. Epidemics and immunization in scale-free networks. *Handbook of Graphs and Networks*, pages 111–130. Wiley-VCH Verlag GmbH & Co. KGaA, 2002.

[105] D. H. Zanette and M. Kuperman. Effects of immunization in small-world epidemics. *Physica A: Statistical Mechanics and its Applications*, 309(3):445–452, 2002.

[106] Z. Dezső and A.-L. Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65(5):055103, May 2002.

[107] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Physical Review E*, 65(3):036104, Feb 2002.

[108] Z. Liu, Y.-C. Lai, and N. Ye. Propagation and immunization of infection on general networks with both homogeneous and heterogeneous components. *Physical Review E*, 67(3):031911, Mar 2003.

[109] N. Madar, T. Kalisky, R. Cohen, D. b. Avraham, and S. Havlin. Immunization and epidemic dynamics in complex networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2):269–276, Mar 2004.

[110] W.-J. Bai, T. Zhou, and B.-H. Wang. Immunization of susceptibleinfected model on scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 384(2):656–662, 2007.

[111] Y. Chen, G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley. Finding a Better Immunization Strategy. *Physical Review Letters*, 101(5):058701, Jul 2008.

[112] Y. Wang, G. Xiao, J. Hu, T. H. Cheng, and L. Wang. Imperfect targeted immunization in scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 388(12):2535–2546, 2009.

[113] K. T. D. Eames, J. M. Read, and W. J. Edmunds. Epidemic prediction and control in weighted networks. *Epidemics*, 1(1):70–76, Mar 2009.

[114] C. M. Schneider, T. Mihaljev, S. Havlin, and H. J. Herrmann. Suppressing epidemics with a limited amount of immunization units. *Physical Review E*, 84(6):061911, Dec 2011.

[115] B. Mirzasoleiman, M. Babaei, and M. Jalili. Immunizing complex networks with limited budget. *EPL (Europhysics Letters)*, 98(3):38004, May 2012.

[116] C. M. Schneider, T. Mihaljev, and H. J. Herrmann. Inverse targeting - An effective immunization strategy. *EPL (Europhysics Letters)*, 98(4):46002, May 2012.

[117] G. Tanaka, C. Urabe, and K. Aihara. Random and Targeted Interventions for Epidemic Control in Metapopulation Models. *Scientific Reports*, 4, Jul 2014.

[118] R. Cohen, S. Havlin, and D. b. Avraham. Efficient Immunization Strategies for Computer Networks and Populations. *Physical Review Letters*, 91(24):247901, Dec 2003.

[119] P. Holme. Efficient local strategies for vaccination and network attack. *EPL (Europhysics Letters)*, 68(6):908, Dec 2004.

[120] J. Gómez-Gardeñes, P. Echenique, and Y. Moreno. Immunization of real complex communication networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 49(2):259–264, Feb 2006.

[121] L. K. Gallos, F. Liljeros, P. Argyrakis, A. Bunde, and S. Havlin. Improving immunization strategies. *Physical Review E*, 75(4):045104, Apr 2007.

[122] M. Salath and J. H. Jones. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput Biol*, 6(4):e1000736, Apr 2010.

[123] F. Nian and X. Wang. Efficient immunization strategies on complex networks. *Journal of Theoretical Biology*, 264(1):77–83, May 2010.

[124] P. Yuan and S. Tang. Community-based immunization in opportunistic social networks. *Physica A: Statistical Mechanics and its Applications*, 420:85–97, 2015.

[125] N. M. Ferguson, C. A. Donnelly, and R. M. Anderson. The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science*, 292(5519):1155–1160, May 2001.

[126] E. M. Cottam, J. Wadsworth, A. E. Shaw, R. J. Rowlands, L. Goatley, S. Maan, N. S. Maan, P. P. C. Mertens, K. Ebert, Y. Li, and e. al. Transmission Pathways of Foot-and-Mouth Disease Virus in the United Kingdom in 2007. *PLoS Pathog*, 4(4):e1000050, Apr 2008.

[127] S. Mishra, D. N. Fisman, and M.-C. Boily. The ABC of terms used in mathematical models of infectious diseases. *Journal of Epidemiology and Community Health*, Oct 2010.

[128] Q. Sattentau. Avoiding the void: cell-to-cell spread of human viruses. *Nature Reviews Microbiology*, 6(11):815–826, Nov 2008.

[129] C. Jolly, K. Kashefi, M. Hollinshead, and Q. J. Sattentau. HIV-1 cell to cell transfer across an Env-induced, actin-dependent synapse. *The Journal of experimental medicine*, 199(2):283–293, Jan 2004.

[130] N. L. Komarova, D. N. Levy, and D. Wodarz. Synaptic transmission and the susceptibility of HIV infection to anti-viral drugs. *Scientific Reports*, 3, Jul 2013.

[131] F. Ball, D. Mollison, and G. Scalia-Tomba. Epidemics with two levels of mixing. *The Annals of Applied Probability*, 7(1):46–89, Feb 1997.

[132] F. Ball and P. Neal. Network epidemic models with two levels of mixing. *Mathematical Biosciences*, 212(1):69, Mar 2008.

[133] S.-M. Cheng, W. C. Ao, P.-Y. Chen, and K.-C. Chen. On Modeling Malware Propagation in Generalized Social Networks. *IEEE Communications Letters*, 15(1):25 –27, Jan 2011.

[134] C. Gao and J. Liu. Modeling and Restraining Mobile Virus Propagation. *IEEE Transactions on Mobile Computing*, PP(99):1, 2012.

[135] T. E. Stone and S. R. McKay. Critical behavior of disease spread on dynamic small-world networks. *EPL (Europhysics Letters)*, 95(3):38003, Aug 2011.

[136] T. E. Stone, M. M. Jones, and S. R. McKay. Comparative effects of avoidance and vaccination in disease spread on a dynamic small-world network. *Physica A: Statistical Mechanics and its Applications*, 389(23):5515, Dec 2010.

[137] E. Estrada, F. Kalala-Mutombo, and A. Valverde-Colmeiro. Epidemic spreading in networks with nonrandom long-range interactions. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 84(3), 2011.

[138] F. Ball. Stochastic and deterministic models for SIS epidemics among a population partitioned into households. *Mathematical Biosciences*, 156(1-2):41–67, Mar 1999.

[139] F. Ball and N. G. Becker. Control of transmission with two types of infection. *Mathematical Biosciences*, 200(2):170–187, Apr 2006.

[140] F. Ball and O. Lyne. Epidemics among a population of households. *mathematical approaches for emerging and re-emerging infectious diseases part II: models, methods and theory*, pages 115–142, 2002.

[141] R. Levins. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America*, 15(3):237240, Sep 1969.

[142] C. C. Zou, D. Towsley, and W. Gong. On the performance of Internet worm scanning strategies. *Performance Evaluation*, 63(7):700–723, Jul 2006.

[143] E. H. Spafford. The internet worm program: an analysis. *SIGCOMM Comput. Commun. Rev.*, 19(1):17–57, Jan 1989.

[144] C. C. Zou, W. Gong, and D. Towsley. Code red worm propagation modeling and analysis. *Proceedings of the 9th ACM conference on Computer and communications security*, pages 138–147. CCS'02. ACM, 2002.

[145] S. Staniford, V. Paxson, and N. Weaver. How to Own the Internet in Your Spare Time. *Proceedings of the 11th USENIX Security Symposium*, pages 149–167. USENIX Association, 2002.

[146] C. Shannon and D. Moore. The spread of the Witty worm. *IEEE Security & Privacy*, 2(4):46– 50, Aug 2004.

[147] J. Giles. Conficker: the inside story. *New Scientist*, 202(2712):36–39, Jun 2009.

[148] P. Li, M. Salour, and X. Su. A survey of internet worm detection and containment. *IEEE Communications Surveys Tutorials*, 10(1):20–35, 2008.

[149] Z. Chen, L. Gao, and K. Kwiat. Modeling the spread of active worms. *IEEE International Conference on Computer Communications*, 3:1890 – 1900 vol.3. Apr 2003.

[150] Z. Chen and C. Ji. Spatial-temporal modeling of malware propagation in networks. *IEEE Transactions on Neural Networks*, 16(5):1291 –1303, Sep 2005.

[151] Z. Chen and C. Ji. Importance-scanning worm using vulnerable-host distribution. *IEEE Global Telecommunications Conference (GLOBECOM)*, 3:6 pp. Dec 2005.

[152] Z. Chen and C. Ji. Optimal worm-scanning method using vulnerable-host distributions. *International Journal of Security and Networks*, 2(1):71–80, Jan 2007.

[153] Z. Chen and C. Ji. Measuring Network-Aware Worm Spreading Ability. *IEEE International Conference on Computer Communications (INFOCOM)*, pages 116 –124. May 2007.

[154] Z. Chen, C. Ji, and P. Barford. Spatial-Temporal Characteristics of Internet Malicious Sources. *IEEE INFOCOM 2008. The 27th Conference on Computer Communications*, pages 2306 –2314. Apr 2008.

[155] A. Dainotti, A. Pescape, and G. Ventre. Worm Traffic Analysis and Characterization. *IEEE International Conference on Communications, 2007. ICC '07*, pages 1435 –1442. Jun 2007.

[156] D. Dagon, C. Zou, and W. Lee. Modeling botnet propagation using time zones. *Annual Network & Distributed System Security Symposium (NDSS)*. 2006.

[157] P. Manna, S. Chen, and S. Ranka. Inside the Permutation-Scanning Worms: Propagation Modeling and Analysis. *IEEE/ACM Transactions on Networking*, 18(3):858 –870, Jun 2010.

[158] K. Perumalla and S. Sundaragopalan. High-fidelity modeling of computer network worms. *Annual Computer Security Applications Conference (ACSAC)*, pages 126 – 135. Dec 2004.

[159] M. A. Rajab, F. Monrose, and A. Terzis. On the effectiveness of distributed worm monitoring. *USENIX Security Symposium*, pages 15–15. SSYM'05. USENIX Association, 2005.

[160] N. Weaver, I. Hamadeh, G. Kesidis, and V. Paxson. Preliminary results using scale-down to explore worm dynamics. *ACM workshop on Rapid malcode*, pages 65–72. WORM'04. ACM, 2004.

[161] J. H. L. Playfair and B. M. Chain. *Immunology at a Glance*. Wiley-Blackwell, 9th edition edition edition, Feb 2009.

[162] D. D. Ho, A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, 373(6510):123–126, Jan 1995.

[163] X. Wei, S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, and e. al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373(6510):117–122, Jan 1995.

[164] A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science*, 271(5255):1582–1586, Mar 1996.

[165] M. A. Nowak and C. R. M. Bangham. Population Dynamics of Immune Responses to Persistent Viruses. *Science*, 272(5258):74–79, Apr 1996.

[166] M. A. Stafford, L. Corey, Y. Cao, E. S. Daar, D. D. Ho, and A. S. Perelson. Modeling plasma virus concentration during primary HIV infection. *Journal of theoretical biology*, 203(3):285–301, Apr 2000.

[167] A. V. Herz, S. Bonhoeffer, R. M. Anderson, R. M. May, and M. A. Nowak. Viral dynamics in vivo: limitations on estimates of intracellular delay and virus decay. *Proceedings of the National Academy of Sciences*, 93(14):7247–7251, Jul 1996.

[168] A. S. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Hurley, K. Saksela, M. Markowitz, and D. D. Ho. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*, 387(6629):188191, May 1997.

[169] S. Bonhoeffer, R. M. May, G. M. Shaw, and M. A. Nowak. Virus dynamics and drug therapy. *Proceedings of the National Academy of Sciences*, 94(13):6971–6976, Jun 1997.

[170] D. Finzi, J. Blankson, J. D. Siliciano, J. B. Margolick, K. Chadwick, T. Pierson, K. Smith, J. Lisziewicz, F. Lori, C. Flexner, and e. al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature Medicine*, 5(5):512–517, May 1999.

[171] R. M. Ribeiro and S. Bonhoeffer. Production of resistant HIV mutants during antiretroviral therapy. *Proceedings of the National Academy of Sciences*, 97(14):7681–7686, Jul 2000.

[172] A. S. Perelson. Modelling viral and immune system dynamics. *Nature Reviews Immunology*, 2(1):28–36, Jan 2002.

[173] L. Rong and A. S. Perelson. Modeling Latently Infected Cell Activation: Viral and Latent Reservoir Persistence, and Viral Blips in HIV-infected Patients on Potent Therapy. *PLoS Comput Biol*, 5(10):e1000533, Oct 2009.

[174] N. K. Vaidya, L. Rong, V. C. Marconi, D. R. Kuritzkes, S. G. Deeks, and A. S. Perelson. Treatment-Mediated Alterations in HIV Fitness Preserve CD4+ T Cell Counts but Have Minimal Effects on Viral Load. *PLoS Comput Biol*, 6(11):e1001012, Nov 2010.

[175] E. E. Giorgi, B. Funkhouser, G. Athreya, A. S. Perelson, B. T. Korber, and T. Bhattacharya. Estimating time since infection in early homogeneous HIV-1 samples using a poisson model. *BMC Bioinformatics*, 11(1):532, Oct 2010.

[176] C. L. Althaus and R. J. D. Boer. Intracellular transactivation of HIV can account for the decelerating decay of virus load during drug therapy. *Molecular Systems Biology*, 6(1):348, Jan 2010.

[177] J. M. Conway and D. Coombs. A Stochastic Model of Latently Infected Cell Reactivation and Viral Blip Generation in Treated HIV Patients. *PLoS Comput Biol*, 7(4):e1002033, Apr 2011.

[178] M. Elemans, A. Florins, L. Willems, and B. Asquith. Rates of CTL Killing in Persistent Viral Infection In Vivo. *PLoS Comput Biol*, 10(4):e1003534, Apr 2014.

[179] S. Alizon and C. Magnus. Modelling the Course of an HIV Infection: Insights from Ecology and Evolution. *Viruses*, 4(12):1984–2013, Oct 2012.

[180] Y. Xiao, H. Miao, S. Tang, and H. Wu. Modeling antiretroviral drug responses for HIV-1 infected patients using differential equation models. *Advanced Drug Delivery Reviews*, 65(7):940–953, Jun 2013.

[181] J. C. Miller, A. C. Slim, and E. M. Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society Interface*, 9(70):890–906, May 2012.

[182] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, Jul 2001.

[183] S. H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, Aug 2014.

[184] C. Castellano and R. Pastor-Satorras. Thresholds for Epidemic Spreading in Networks. *Physical Review Letters*, 105(21):218701, Nov 2010.

[185] J. C. Miller. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E*, 76(1):010101, Jul 2007.

[186] M. Hastings. Systematic Series Expansions for Processes on Networks. *Physical Review Letters*, 96(14):148701, Apr 2006.

[187] E. Kenah and J. M. Robins. Second look at the spread of epidemics on networks. *Physical Review E*, 76(3):036113, 2007.

[188] J. C. Miller. Bounding the Size and Probability of Epidemics on Networks. 45(2):498–512, Jun 2008.

[189] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math.*, (6):290, 1959.

[190] D. Steward. The design structure system: A method for managing the design of complex systems. *IEEE Transactions on Engineering Management*, EM-28(3):71–74, Aug 1981.

[191] D. Helbing, S. Bishop, R. Conte, P. Lukowicz, and J. B. McCarthy. FuturICT: Participatory computing to understand and manage our complex world in a more sustainable and resilient way. *The European Physical Journal Special Topics*, 214(1):11–39, Dec 2012.

[192] C. C. De Quadros, L. Morris, E. A. Da Costa, N. Arnt, and C. H. Tigre. Epidemiology of variola minor in Brazil based on a study of 33 outbreaks. *Bulletin of the World Health Organization*, 46(2):165–171, 1972.

[193] P. Porras, H. Saïdi, and V. Yegneswaran. A foray into Conficker's logic and rendezvous points. *USENIX conference on Large-scale exploits and emergent threats: botnets, spyware, worms, and more*, pages 7–7. LEET'09. USENIX Association, 2009.

[194] R. Weaver. A probabilistic population study of the Conficker-C botnet. *Proceedings of the 11th international conference on Passive and active measurement*, pages 181–190. PAM'10. Springer-Verlag, 2010.

[195] B. Irwin. A network telescope perspective of the Conficker outbreak. *Information Security for South Africa (ISSA), 2012*, pages 1–8. 2012.

[196] E. Aben. Conficker/Conflicker/Downadup as seen from the UCSD Network Telescope. Feb 2009. `http://www.caida.org/research/security/ms08-067/conficker.xml`. Accessed Dec 2014.

[197] N. Martin, S. Welsch, C. Jolly, J. A. G. Briggs, D. Vaux, and Q. J. Sattentau. Virological synapse-mediated spread of human immunodeficiency virus type 1 between T cells is sensitive to entry inhibition. *Journal of virology*, 84(7):3516–3527, Apr 2010.

[198] M. Sourisseau, N. Sol-Foulon, F. Porrot, F. Blanchet, and O. Schwartz. Inefficient human immunodeficiency virus replication in mobile lymphocytes. *Journal of virology*, 81(2):1000–1012, Jan 2007.

[199] C. J. A. Duncan, R. A. Russell, and Q. J. Sattentau. High multiplicity HIV-1 cell-to-cell transmission from macrophages to CD4+ T cells limits antiretroviral efficacy. *AIDS*, 27(14):2201–2206, Sep 2013.

[200] T. T. Murooka, M. Deruaz, F. Marangoni, V. D. Vrbanac, E. Seung, U. H. v. Andrian, A. M. Tager, A. D. Luster, and T. R. Mempel. HIV-infected T cells are migratory vehicles for viral dissemination. *Nature*, 490(7419):283–287, Oct 2012.

[201] X. Sewald, D. G. Gonzalez, A. M. Haberman, and W. Mothes. In vivo imaging of virological synapses. *Nature Communications*, 3:1320, Dec 2012.

[202] A. Sigal, J. T. Kim, A. B. Balazs, E. Dekel, A. Mayo, R. Milo, and D. Baltimore. Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature*, 477(7362):95–98, Sep 2011.

[203] A. Del Portillo, J. Tripodi, V. Najfeld, D. Wodarz, D. N. Levy, and B. K. Chen. Multiploid Inheritance of HIV-1 during Cell-to-Cell Infection. *Journal of virology*, 85(14):7169–7176, Jul 2011.

[204] D. Wodarz and D. N. Levy. Effect of different modes of viral spread on the dynamics of multiply infected cells in human immunodeficiency virus infection. *Journal of The Royal Society Interface*, 8(55):289300, Feb 2011.

[205] N. L. Komarova, D. N. Levy, and D. Wodarz. Effect of Synaptic Transmission on Viral Fitness in HIV Infection. *PLoS ONE*, 7(11):e48361, Nov 2012.

[206] N. L. Komarova and D. Wodarz. Virus dynamics in the presence of synaptic transmission. *Mathematical Biosciences*, 242(2):161–171, Apr 2013.

[207] N. L. Komarova, D. Anghelina, I. Voznesensky, T. Benjamin, D. N. Levy, and D. Wodarz. Relative contribution of free-virus and synaptic transmission to the spread of HIV-1 through target cell populations. *Biology Letters*, 9(1), Feb 2013.

[208] M. Prague, D. Commenges, and R. Thiébaut. Dynamical models of biomarkers and clinical progression for personalized medicine: The HIV context. *Advanced Drug Delivery Reviews*, 65(7):954–965, Jun 2013.

[209] H. W. M. v. Deutekom, G. Wijnker, and R. J. d. Boer. The Rate of Immune Escape Vanishes When Multiple Immune Responses Control an HIV Infection. *The Journal of Immunology*, 191(6):3277–3286, Sep 2013.

[210] C. A. Spina, J. Anderson, N. M. Archin, A. Bosque, J. Chan, M. Famiglietti, W. C. Greene, A. Kashuba, S. R. Lewin, D. M. Margolis, and e. al. An In-Depth Comparison of Latent HIV-1 Reactivation in Multiple Cell Model Systems and Resting CD4+ T Cells from Aviremic Patients. *PLoS Pathog*, 9(12):e1003834, Dec 2013.

[211] F. Miedema, M. D. Hazenberg, K. Tesselaar, D. v. Baarle, R. J. d. Boer, and J. A. M. Borghans. Immune activation and collateral damage in AIDS pathogenesis. *Frontiers in HIV and AIDS*, 4:298, 2013.

[212] G. Doitsh, N. L. K. Galloway, X. Geng, Z. Yang, K. M. Monroe, O. Zepeda, P. W. Hunt, H. Hatano, S. Sowinski, I. M. n. Arias, and e. al. Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature*, 505(7484):509–514, Jan 2014.

[213] F. A. Koning, S. A. Otto, M. D. Hazenberg, L. Dekker, M. Prins, F. Miedema, and H. Schuitemaker. Low-Level CD4+ T Cell Activation Is Associated with Low Susceptibility to HIV-1 Infection. *The Journal of Immunology*, 175(9):6117–6122, Nov 2005.

[214] Z. Zhang, T. Schuler, M. Zupancic, S. Wietgrefe, K. A. Staskus, K. A. Reimann, T. A. Reinhart, M. Rogan, W. Cavert, C. J. Miller, and e. al. Sexual transmission and propagation of SIV and HIV in resting and activated CD4+ T cells. *Science*, 286(5443):1353–1357, Nov 1999.

[215] H.-M. Baldauf, X. Pan, E. Erikson, S. Schmidt, W. Daddacha, M. Burggraf, K. Schenkova, I. Ambiel, G. Wabnitz, T. Gramberg, and e. al. SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nature Medicine*, 18(11):1682–1687, Nov 2012.

[216] D. C. Goldstone, V. Ennis-Adeniran, J. J. Hedden, H. C. T. Groom, G. I. Rice, E. Christodoulou, P. A. Walker, G. Kelly, L. F. Haire, M. W. Yap, and e. al. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature*, 480(7377):379–382, Dec 2011.

[217] R. De Boer and A. Perelson. Target Cell Limited and Immune Control Models of HIV Infection: A Comparison. *Journal of Theoretical Biology*, 190(3):201–214, Feb 1998.

[218] H. Mohri, A. S. Perelson, K. Tung, R. M. Ribeiro, B. Ramratnam, M. Markowitz, R. Kost, A. Hurley, L. Weinberger, D. Cesar, and e. al. Increased turnover of T

lymphocytes in HIV-1 infection and its reduction by antiretroviral therapy. *The Journal of experimental medicine*, 194(9):1277–1287, Nov 2001.

[219] M. A. Stafford, L. Corey, Y. Cao, E. S. Daar, D. D. Ho, and A. S. Perelson. Modeling plasma virus concentration during primary HIV infection. *Journal of Theoretical Biology*, 203(3):285–301, Apr 2000.

[220] C. Jolly, I. Mitar, and Q. J. Sattentau. Requirement for an intact T-cell actin and tubulin cytoskeleton for efficient assembly and spread of human immunodeficiency virus type 1. *Journal of virology*, 81(11):5547–5560, Jun 2007.

[221] C. Jolly, N. J. Booth, and S. J. D. Neil. Cell-cell spread of human immunodeficiency virus type 1 overcomes tetherin/BST-2-mediated restriction in T cells. *Journal of virology*, 84(23):12185–12199, Dec 2010.

[222] W. Hübner, G. P. McNerney, P. Chen, B. M. Dale, R. E. Gordon, F. Y. S. Chuang, X.-D. Li, D. M. Asmuth, T. Huser, and B. K. Chen. Quantitative 3D Video Microscopy of HIV Transfer Across T Cell Virological Synapses. *Science*, 323(5922):1743–1747, Mar 2009.

[223] P. Chen, W. Hübner, M. A. Spinelli, and B. K. Chen. Predominant Mode of Human Immunodeficiency Virus Transfer between T Cells Is Mediated by Sustained Env-Dependent Neutralization-Resistant Virological Synapses. *Journal of virology*, 81(22):12582–12595, Nov 2007.

[224] A. J. McMichael, P. Borrow, G. D. Tomaras, N. Goonetilleke, and B. F. Haynes. The immune response during acute HIV-1 infection: clues for vaccine development. *Nature Reviews Immunology*, 10(1):11–23, Jan 2010.

[225] N. F. Parrish, F. Gao, H. Li, E. E. Giorgi, H. J. Barbian, E. H. Parrish, L. Zajic, S. S. Iyer, J. M. Decker, A. Kumar, and e. al. Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences of the United States of America*, 110(17):6626–6633, Apr 2013.

[226] H. R. Fryer, J. Frater, A. Duda, M. G. Roberts, R. E. Phillips, A. R. McLean, and T. S. T. Investigators. Modelling the Evolution and Spread of HIV Immune Escape Mutants. *PLoS pathogens*, 6(11):e1001196, Nov 2010.

[227] F. Wightman, P. Ellenberg, M. Churchill, and S. R. Lewin. HDAC inhibitors in HIV. *Immunology and cell biology*, 90(1):47–54, Jan 2012.

[228] W. Lu, L. C. Arraes, W. T. Ferreira, and J.-M. Andrieu. Therapeutic dendritic-cell vaccine for chronic HIV-1 infection. *Nature Medicine*, 10(12):1359–1365, Dec 2004.

[229] M. B. Zeisel, D. J. Felmlee, and T. F. Baumert. Hepatitis C Virus Entry. *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*, pages 87–112. Current Topics in Microbiology and Immunology. Springer Berlin Heidelberg, Jan 2013.

[230] C. Zhang, S. Zhou, and B. M. Chain. Hybrid spreading of the Internet worm Conficker. *arXiv:1406.6046 [cs.CR]*, Jun 2014.

[231] C. Zhang, S. Zhou, I. J. Cox, and B. M. Chain. Optimizing Hybrid Spreading in Metapopulations. arXiv:1409.7291 [physics.soc-ph], Sep 2014.

[232] G. Silvestri, M. Paiardini, I. Pandrea, M. M. Lederman, and D. L. Sodora. Understanding the benign nature of SIV infection in natural hosts. *The Journal of Clinical Investigation*, 117(11):3148–3154, Nov 2007.

[233] B. K. Titanji, M. Aasa-Chapman, D. Pillay, and C. Jolly. Protease inhibitors effectively block cell-to-cell spread of HIV-1 between T cells. *Retrovirology*, 10(1):161, Dec 2013.

[234] WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection. Jun 2013. `http://www.who.int/hiv/pub/guidelines/arv2013/download/en/index.html`.

[235] H. Ward and M. Rönn. Contribution of sexually transmitted infections to the sexual transmission of HIV. *Current opinion in HIV and AIDS*, 5(4):305–310, Jul 2010.

[236] J. M. Brenchley, T. W. Schacker, L. E. Ruff, D. A. Price, J. H. Taylor, G. J. Beilman, P. L. Nguyen, A. Khoruts, M. Larson, A. T. Haase, and e. al. CD4+ T Cell Depletion during all Stages of HIV Disease Occurs Predominantly in the

Gastrointestinal Tract. *The Journal of experimental medicine*, 200(6):749–759, Sep 2004.

[237] R. S. Veazey, M. DeMaria, L. V. Chalifoux, D. E. Shvetz, D. R. Pauley, H. L. Knight, M. Rosenzweig, R. P. Johnson, R. C. Desrosiers, and A. A. Lackner. Gastrointestinal Tract as a Major Site of CD4+ T Cell Depletion and Viral Replication in SIV Infection. *Science*, 280(5362):427–431, Apr 1998.

[238] A. Sáez-Cirión, C. Bacchus, L. Hocqueloux, V. Avettand-Fenoel, I. Girault, C. Lecuroux, V. Potard, P. Versmisse, A. Melard, T. Prazuck, and e. al. Post-Treatment HIV-1 Controllers with a Long-Term Virological Remission after the Interruption of Early Initiated Antiretroviral Therapy ANRS VISCONTI Study. *PLoS pathogens*, 9(3):e1003211, Mar 2013.

[239] G. P. Rizzardi, A. Harari, B. Capiluppi, G. Tambussi, K. Ellefsen, D. Ciuffreda, P. Champagne, P.-A. Bart, J.-P. Chave, A. Lazzarin, and e. al. Treatment of primary HIV-1 infection with cyclosporin A coupled with highly active antiretroviral therapy. *The Journal of Clinical Investigation*, 109(5):681–688, Mar 2002.

[240] M. M. Lederman, L. Smeaton, K. Y. Smith, B. Rodriguez, M. Pu, H. Wang, A. Sevin, P. Tebas, S. F. Sieg, K. Medvik, and e. al. Cyclosporin A provides no sustained immunologic benefit to persons with chronic HIV-1 infection starting suppressive antiretroviral therapy: results of a randomized, controlled trial of the AIDS Clinical Trials Group A5138. *The Journal of infectious diseases*, 194(12):1677–1685, Dec 2006.

[241] M. Boguñá and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66(4):047104, Oct 2002.

[242] F. Xiao, I. Fofana, L. Heydmann, H. Barth, E. Soulier, F. Habersetzer, M. Doffol, J. Bukh, A. H. Patel, M. B. Zeisel, and e. al. Hepatitis C Virus Cell-Cell Transmission and Resistance to Direct-Acting Antiviral Agents. *PLoS Pathog*, 10(5):e1004128, May 2014.

[243] A. Y. Zomaya. *Parallel Computing for Bioinformatics and Computational Biology: Models, Enabling Technologies, and Case Studies*. Wiley-Blackwell, May 2006.

[244] D. Geer. Chip makers turn to multicore processors. *Computer*, 38(5):11–13, May 2005.

[245] G. Blake, R. Dreslinski, and T. Mudge. A survey of multicore processors. *IEEE Signal Processing Magazine*, 26(6):26–37, Nov 2009.

[246] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175308, Feb 2006.

[247] S. Roy, M. Werner-Washburne, and T. Lane. A system for generating transcription regulatory networks with combinatorial control of transcription. *Bioinformatics*, 24(10):1318–1320, May 2008.

[248] J. A. Papin, T. Hunter, B. O. Palsson, and S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nature Reviews Molecular Cell Biology*, 6(2):99–111, Feb 2005.

[249] G. Zschaler and T. Gross. Largenet2: an object-oriented programming library for simulating large adaptive networks. *Bioinformatics*, 29(2):277–278, Jan 2013.