Epidemics in Networks:
Modeling, Optimization and Security Games

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 13 september 2010 om
10:00 uur
door

Jasmina OMIĆ

magister Elektrotehnički fakultet Beograd, Srbija
geboren te Zagreb, Croatia.

2

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. P.F.A. Van Mieghem

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. P.F.A. Van Mieghem	Technische Universiteit Delft, promotor
Prof. dr. ir. R.E. Kooij	Technische Universiteit Delft
Prof. dr. A. Orda	Technion Israel Institute of Technology
Prof. dr. C. Scoglio	Kansas State University
Prof. dr. ir. S. Heemstra de Groot	Technische Universiteit Delft
Reader dr. A. Ganesh	University of Bristol
Prof. Dr. R.D. van der Mei	Vrije Universiteit Amsterdam

To my mother

Thesis Summary

Epidemics in Networks: Modeling, Optimization and Security Games

Epidemic theory has wide range of applications in computer networks, from spreading of malware to the information dissemination algorithms. Our society depends more strongly than ever on such computer networks. Many of these networks rely to a large extent on decentralization and self-organization. While decentralization removes obvious vulnerabilities related to single points of failure, it leads to a higher complexity of the system. A more complex type of vulnerability appears in such systems. For instance, computer viruses are imminent threats to all computer networks. We intend to study the interaction between malware spreading and strategies that are designed to cope with them.

The main goals of this thesis are:

1. to analyze influence of network topology on infection spread

2. to determine how topology can be used for network protection

3. to formulate and study optimization of malware protection problem with respect to topology

4. to investigate non-cooperative game of security

We used analytical tools from various fields to answer these questions. First of all, we have developed homogeneous and heterogeneous N-intertwined, susceptible - infected - susceptible (SIS) model for virus spread. This model is used to determine the influence of topology on the spreading process. For the N-intertwined model, we show that the largest eigenvalue of the adjacency matrix of the graph rigorously defines the epidemic threshold. The results of the model also predict the upper and lower bounds on epidemics as a function of nodal degree. The epidemic threshold is found to be a consequence of the mean field approximation. However, slow convergence to the steady-state justifies the application of the threshold concept. We used the exact 2^N-state Markov chain model to explore the phase transition phenomenon for two contrasting cases, namely the line graph and the complete graph.

The N-intertwined model assumes that the infection spreading over a link is a Poisson process. By introducing infection delay, we studied the influence of deviation from Poisson process assumption on epidemic threshold for the special case of a complete bi-partite graph. Due to the special structure of bi-partite graphs we were also able to derive approximate formula for the extinction probability in the first phase of the infection.

In the case of SIS epidemic models, the effects of infection depend on the protection of individual nodes. We studied optimization of protection scheme for different networks. We use the results from heterogeneous N-intertwined model to determine the global optimum at the threshold. Above the threshold, the problem is a sum of ratios fractional programming problem, which is **NP**-complete. Therefore, we only determine the upper bound on the optimum. Contrary to the common sense, reducing the probability of infection for higher degree nodes pushes the network out of the global optimum. For the case of complete bi-partite graphs, we derive optimal threshold if only 2 fixed protection rates are available.

Computer networks are generally distributed systems and protection cannot be globally optimized. The Internet is an extreme example: there is no global control center, and obtaining complete information on its global state is an illusion. To approach the issue of security over decentralized network, we derived a novel framework for network security under the presence of autonomous decision makers. The problem under the consideration is the N players non-cooperative game. We have established the existence of a Nash equilibrium point (NEP). The willingness of nodes to invest in protection depends on the price of protection. We showed that, when the price of protection is relatively high for all the nodes, the only equilibrium point is that of a completely unprotected network; while if this price is sufficiently low for a single node, it will always invest in protecting itself. We determine bounds on the Price of Anarchy (PoA), that describes how far the NEP is from the global optimum. We have also proposed two methods for steering the network equilibrium, namely by influencing the relative prices and by imposing an upper bound on infection probabilities.

A quarantine is another possible measure against the epidemic. A quarantine on a set of network nodes separates them from the rest of the network by removing links. The concept of threshold and the N-intertwined model provides a tool to analyze how quarantine improves the network protection. We studied several different networks from artificially generated to real-world examples using the modularity algorithm. The real-world networks tend to show a better epidemic threshold after clustering than artificially generated graphs. The real-world networks have typically two or three big clusters and several smaller ones, while Barabási-Albert (BA) and Erdös-Rényi (ER) graphs have several smaller clusters comparable in size. However, the number of removed links in a graph using modularity algorithm is unjustifiably high, suggesting that complete quarantine is not a viable solution for real-world networks.

Jasmina Omić

Samenvatting

Epidemieën in Netwerken: Modellering, Optimalisering en Veiligheids Spelen

De epidemietheorie heeft een brede waaier van toepassingen in computernetwerken, van het uitspreiden van malware tot aan algoritmen voor informatieverspreiding. Onze maatschappij hangt nu meer dan ooit af van computernetwerken. Veel van deze netwerken zijn voor een groot deel decentraal en de zelforganiserende. Terwijl decentralisatie enige duidelijke punten van kwetsbaarheid verwijdert, leidt het tot een hogere complexiteit van het systeem. Een complexer type van kwetsbaarheid verschijnt in dergelijke systemen. Bijvoorbeeld, zijn de computervirussen belangrijke bedreigingen voor alle computernetwerken. In dit proefschrift onderzoeken wij de interactie tussen het verspreiden van malware en strategien die dreigende ontwikkelingen dienen te behandelen. De belangrijkste doelstellingen van deze thesis zijn:

1. de invloed van de netwerktopologie op verspreide besmetting te analyseren

2. te bepalen hoe de topologie voor netwerkbescherming kan worden gebruikt

3. het malwarebescherming optimalisatieprobleem vaststellen en de invloed van de topologie op optimalisatie bepalen

4. een niet-coöperatief spel van veiligheid te onderzoeken

Wij gebruikten analytische hulpmiddelen van diverse gebieden om deze vragen te beantwoorden. Eerst en vooral, hebben wij homogeen en heterogeen N-verstrengeld, vatbaar - besmet - vatbaar (SIS) model voor virusverspreiding ontwikkeld. Dit model wordt gebruikt om de invloed van de topologie op het spreidingsproces te bepalen. Voor het N-verstrengeld model, tonen wij aan dat de grootste eigenwaarde van de nabijheidmatrix van de graaf de epidemische drempel bepaalt. De resultaten van het model voorspellen ook de hogere en lagere grenzen aan epidemien als functie van knoopgraad. De epidemiedrempel is een

uitkomst van gemiddelde-gebiedsbenadering te zijn. Echter, de langzame convergentie naar de evenwichtstoestand rechtvaardigt de toepassing van het drempelconcept. Wij gebruikten het exacte 2^N-staat Markov proces model om het fenomeen van de faseovergang voor twee verschillende gevallen te onderzoeken, namelijk de lijngraaf en de volledige graaf. Het N- verstrengeld model veronderstelt dat de besmetting die over een verbinding uitspreidt een Poisson proces is. Door besmettingsvertraging te introduceren, bestudeerden wij de invloed van afwijking op de Poisson proces aanname op de epidemiedrempel voor het speciale geval van de bi-partite graaf. Vanwege de speciale structuur van bi-partite graven konden wij ook benaderende formules voor de uitroeiingwaarschijnlijkheid in de eerste fase van de besmetting afleiden. In het geval van SIS epidemische modellen, hangen de gevolgen van besmetting van de bescherming van de individuele knopen af. Wij bestudeerden optimalisering van de beschermingsregeling voor verschillende netwerken. Wij gebruiken de resultaten van het heterogeen N- verstrengeld model om het globale optimum bij de drempel te bepalen. Boven de drempel, is het probleem een 'som van verhoudingen' fractie-programmeringprobleem, die **NP**- complete is. Daarom bepalen wij slechts het boven limiet op het optimum. Tegen verwachtingen in, drijft verlagen van de besmettingswaarschijnlijkheid voor hogere graad knooppunten het netwerk uit het globale optimum. In het geval van volledige bi-partite graven, leiden wij de optimale drempel af indien slechts 2 vaste beschermingsniveaus beschikbaar zijn.

Computernetwerken zijn in het algemeen gedistribueerde systemen en de bescherming kan niet globaal worden geoptimaliseerd. Het Internet is een extreem voorbeeld: er is geen globaal controlecentrum en het verkrijgen van volledige informatie over de globale toestand is een illusie. Om de kwestie van veiligheid voor gedecentraliseerde netwerken te benaderen, leiden wij een nieuw kader voor netwerkveiligheid onder de aanwezigheid van autonome besluitvormers af. Het probleem in overweging is het N-speler niet-coöperatieve spel. Wij hebben het bestaan van een Nash evenwichtspunt (NEP) aangetoond. De bereidheid van knopen om in bescherming te investeren hangt van de prijs van bescherming af. Wij toonden aan dat, wanneer de prijs van bescherming voor alle knopen vrij hoog is, het enige evenwichtspunt dat van een volledig onbeschermd netwerk is; terwijl als deze prijs voor n enkele knoop voldoende laag is, het altijd in bescherming zal investeren. Wij bepalen grenzen aan de Prijs van Anarchie (PoA), die beschrijft hoe ver het NEP van het globale optimum ligt. Wij hebben ook twee methodes voorgesteld om het netwerkevenwicht te sturen, namelijk door de relatieve prijzen te benvloeden en door een verbindend limiet aan besmettingswaarschijnlijkheid op te leggen.

Een quarantaine is een andere mogelijke maatregel tegen de epidemie. Een quarantaine op een verzameling netwerkknopen scheidt hen van de rest van het netwerk door verbindingen te verwijderen. Het concept van een drempel en het N- verstrengeld model verstrekken een hulpmiddel dat analyseert hoe de quarantaine de netwerkbescherming verbetert. Gebruikenmakend van het modulariteitsalgoritme bestudeerden wij verscheidene netwerken van kunstmatig ge-

produceerde tot real-world voorbeelden. De real-world netwerken neigen een betere epidemiedrempel te tonen na het groeperen dan kunstmatig geproduceerde graven. De real-world netwerken hebben typisch twee of drie grote clusters en verscheidene kleinere delen, terwijl Barabási-Albert (BA) and Erdös-Rényi (ER) graven verscheidene kleinere clusters hebben van vergelijkbare grootte. Het aantal verwijderde verbindingen in een graaf die het modulariteitsalgoritme gebruikt is echter onweerlegbaar hoog, voorstellend dat de volledige quarantaine geen haalbare oplossing voor real-world netwerken is.

Jasmina Omić

Contents

III   Influence of quarantine on epidemic spread   107

8   Virus spread in social networks   109

9   Conclusions   125

List of Abbreviations   139

Acknowledgments   141

Curriculum Vitae   143

Publications by the Author   145

NGInfra PhD Thesis Series   147

Chapter 1

Introduction

In the course of this thesis, we will explore dynamic process of spread on networks. The networks that we will consider are part of our every day life like the Internet, the World Wide Web, power grids, transportation system, social networks etc. These networks are constructed as the interaction networks, for example, social networks represent the interactions between friends. The networks also represent physical infrastructure used by a dynamic process, the Internet and power grids are two most prominent examples.

Dynamic process or dynamics of flows of some quantities over the network and its interaction with the network elements is important dimension of network complexity. A full characterization of a network is not possible without considering interplay between structural and dynamical aspects.

The dynamic process that is focus of our work is the spread of malware over computer networks. We are especially interested how epidemic interacts with protection strategies.

Although the modeling of diseases is an old discipline [1], the epidemic theory was first applied to computer diseases - computer malware by Kephart and White (KW) in 1993 [2]. The KW model was a homogeneous models. The malware was spreading with the same speed in every part of the network and nodes were cured from infection with the same frequency. The model also assumed that the underlaying network of interactions was a regular graph where every node has the same number of neighbors. This assumption was later shown to be inadequate for malware spreading. Pastor-Satorras and Vespignani [3] discussed discrepancy between the data of virus spread on the Internet and theoretical results of the KW model for homogeneous networks. They introduced a model that underlines the influence of degree heterogeneity in the networks. At that point, the influence of topology in the application of epidemic theory in computer networks became an important issue. Recently, malware started spreading over on-line social networks, emphasizing the importance of underlying structure. On-line social networks

have specific power-low network structure which makes them prone to epidemic spreading [3], [4], [5], [6].

Epidemic modeling was not used only for malware spreading. Engineers used epidemic paradigm to design distributed and scalable algorithms for ad-hoc and P2P networks [7]. The Erdös-Rényi graph or a hypercube were often used to represent such networks. The adequate functioning of the epidemic algorithm is responsible for the resilience of the network [8].

Propagation of faults and failures is yet another application for epidemic models in networks. If an element in a network fails it can lunch an avalanche of failures in the neighboring elements. One example is cascading BGP failures on a fully connected topology [9].

Since there are many different applications and the range of topologies appearing in applications varies significantly, an abstract epidemic model should be capable of capturing any epidemic process on any given finite graph.

Different epidemic processes on computer networks are not only heterogeneous in the sense of underlining topology. Usually, spreading is heterogeneous in the sense of protection ability and spreading power of each individual network element. The difference in bandwidth and computational power between different hosts on the Internet will determine the speed with which the malware is spreading. The protection is also highly heterogeneous. A protection strategy can be an antivirus software, with its signature quality and the speed of response to new virus strains. An important property of a protection strategy is the frequency with which the host is checked and secured. Several factors influence the choice of the protection strategy, most notably the significance and value of the protected information, the probability of infection, the overhead of employing the protection strategy and its (monetary) price.

We will introduce a model which captures the influence of network topology and extend this model to full heterogeneous settings including heterogeneous spreading power and protection abilities of individual nodes.

Further more, systems that have to be protected are usually large and distributed with autonomous decision makers. A key-point is that the security of each host depends not only on the protection strategies it chooses to adopt but also on those chosen by other hosts in the network with example of the Internet malware. This class of problems has two aspects. First, it deals with epidemic processes, and as such calls for the employment of epidemic theory. Second, the distributed and autonomous nature of decision-making in major classes of networks (e.g., P2P, ad-hoc, and most notably the Internet) call for the employment of game theoretical approaches. Indeed, the trade-offs between the damage infection and the price and overhead of a protection strategy may be vastly different across users, hence placing certain nodes in an unfair position to protect much of the network by investing more than other nodes. In this thesis, we will discuss the global optimization problem of protection in a network and employ the game theory to determine the outcome of the non-cooperative security game.

Except individual protection of each host in a computer network, hosts can also be quarantined. A quarantine on a network is created by removing links that connect host or group of hosts to the rest of the network. This clustering allows limited intra community communication between nodes to continue, while possibly quarantining the rest of the network. Removal of links was studied in diseases modeling [10], [11], [12], however the network was not separated in disconnected clusters. In the last part of the thesis, we will explore how the change of network topology induced by quarantine can be used to improve network protection.

Although, epidemic theory has many nice results, there are still many open questions. In the following section, we present a set of questions which are the scope of this thesis.

1.1   Scope

This thesis concentrates on N-intertwined epidemic model of SIS type, on any network, in continuous time. Using N-intertwined model, we attempt to answer the following questions:

1. **What is the influence of the network of contacts on epidemic spreading?** The sharp epidemic threshold , which is a consequence of mean field theory, is rigorously shown to be equal to the largest eigenvalue (the spectral radius) of the adjacency matrix. Positive eigenvalues and their eigenvectors of adjacency matrix are more important than negative ones above the threshold. (ch. 4)

2. **What is the role of topology in heterogeneous protection settings?** Minimum threshold in the sense of the total protection is reached if the protection is applied proportional to the node degree. (ch. 5)

3. **What is the solution to the global optimum in protection against epidemic?** The global optimum in the case of protection on the threshold exists and the sum of all protection is equal to twice the number of links in the graph. Global optimum above the threshold is bounded from above by the protection proportional to the node degree. (ch. 6)

4. **What is the outcome of non-cooperative security game that rises as a consequence of malware spreading?** Multiple NEP (Nash Equilibrium Point) exist for the sum of protections above the minimum threshold. If there is not enough protection for minimum threshold to be reached, at least one NEP exists. PoA (Price of Anarchy) can be very large, but it is not unbounded. (ch. 7)

5. **What is the influence of quarantine on epidemic spread?** Network protection against epidemic can be improved using quarantine for any kind of graph. However, the number of removed links is large. (ch. 8)

1.2   Thesis outline

The thesis consist of 9 chapters organized in 3 parts. The first part presents and discusses the N-intertwined model. The second describes the global protection optimization problem and non-cooperative security protection game. The last part considers consequences of quarantine application on epidemic spread.

In **Chapter 2**, we present an introduction to epidemic modeling: the concept of threshold and interaction matrix and related work in the field of malware modeling.

1.2.1   Part I: N-intertwined model

Chapter 3 Presents the exact 2^N-state Markov chain model as a comparison model for N-intertwined model. In particular, we discuss the convergence to the absorbing steady state.

The major part of **Chapter 4** is devoted to our N-intertwined Markov model. We derive the model, assess the influence of the mean field approximation, derive precise relations and upper bounds for the steady-state. Further, we characterize the exponential die out and the role of the spectrum of adjacency matrix A. The accuracy of the Kephart and White model is evaluated and our model is compared with exact computations. We conclude with application of the model to complete bi-partite graph.

Chapter 5 extends the N-intertwined model for virusspread in any network with N nodes to a full heterogeneous setting. The steady-state infection probabilities are specified in terms of a generalized Laplacian, that possesses analogous properties as the classical Laplacian in graph theory. The critical threshold that separates global network infection from global network health is characterized via an N dimensional vector that makes the largest eigenvalue of a modified adjacency matrix equal to unity. We apply these results to two special case graphs, the complete bi-partite and the regular graph.

1.2.2   Part II: Optimization of protection and game theory

In **Chapter 6**, we consider two protection optimization problems for heterogeneous spread using N-intertwined model, namely the optimization at the threshold and above the threshold. We solve the optimization problem at the threshold. Above the threshold problem belongs to the class of sum of ratios fractional programming problems and global optimization is necessary. We determine the upper bond on global optimum and show that for strong epidemics increase in protection

of well connected nodes to the detriment of nodes with low degree is not a good strategy.

In **Chapter 7**, we introduce a novel framework for network security under the presence of autonomous decision makers with multiple (possibly infinite) protection strategies. The model encompasses general (arbitrary) topologies. Further, we establish the existence of a Nash equilibrium point and characterization of its properties. We discus the related global (i.e., social) optimization problem, and establishment of an upper bound on the price of anarchy. Finally, we propose schemes for a network manager to influence the game, resulting in a potentially major improvement in the level of network security.

1.2.3 Part III: Influence of quarantine on epidemic spread

In **Chapter 8**, we quantify the improvements of protection that the network quarantine is introducing, using the epidemic threshold concept and the N-intertwined Susceptible Infected Susceptible (SIS) epidemic model on a large set of networks, both real-world and artificial ones.

Chapter 2

Epidemic modeling

Initially a part of the biology, mathematical modeling of epidemic has evolved and became a fine tool that is applied in other scientific fields. The modeling of diseases is thee hundred years old, but application to the Internet virus and worms modeling, fault propagation and epidemic algorithms for information spread in distributed systems is just emerging. This is partly due to the fact that the Internet itself has just being born and partly because the spreading mechanisms and the nature of medium are different. As a young discipline, the epidemic modeling in computer networks can learn a lot from already flourished biological models.

Epidemic modeling has three main goals. The first one is to understand the mechanisms of spreading and how different parameters influence its course. The second goal is to be able to predict the course of epidemic in future, which includes among others, the final size of the epidemic and convergence time to the steady state. Finally, the third goal is to determine mechanisms to control and stop epidemic and study their influence on the process.

The epidemic model consists of a set of assumption about the nature of the population and the spreading mechanism. The assumptions regarding population usually belong to the following categories [1]:

- **General structure of the population.** The population can be homogeneous - such that every individual reacts to infection and spreads infection in the same manner. There can be several different homogeneous populations or stratas interacting or completely heterogeneous population.

- **Population dynamics.** The set of individuals can be closed or open. In the closed set, the number of individuals does not change over time, there are no new births, deaths, emigrations or immigrations.

- **Diseases statuses of an individual.** There are many possible states in epidemic models. An individual can be susceptible to infection, incubating,

infectious, carrier without symptoms, immune or removed.

The spreading mechanisms determines how exactly the infection is transmitted. For example, in diseases, the infection can be airborne or it can be transmitted by blood.

Based on the disease statuses of individuals and transition between statuses or states, the most studied models are SI (Susceptible - Infected), SIS (Susceptible - Infected - Susceptible) and SIR (Susceptible - Infected - Recovered). In the SI model, the individuals are susceptible to infection and than they become infected and stay in that state forever. In SIS model, the individuals can transit from infected state to susceptible state again. The spread of common cold can be modeled with this model. In the case of SIR model, after some time spent in infected state, an individual dies or becomes immune (recovered) and can not spread infection further.

When there is a well-described set of characteristics of the spreading process and population, second step is to choose specific mathematical modeling approach. Although the spreading is stochastic process in its nature, deterministic models are frequently used. It is used to approximate the mean of the random process which implies that the number of individuals has to be large enough. The stochastic models are used for the small populations and in the cases when the fluctuation of the infected population is of interest.

2.1   Epidemic modeling in biology

Biological epidemiology has produced significant number of deterministic and stochastic models as well as relevant conclusion about prevention and prediction of epidemics. One of the first epidemic models can be traced to Daniel Bernoulli in 1760, who worked on a simple deterministic mode of smallpox. Subsequently, more complex deterministic models developed. Stochastic models appeared later in 19th and 20th century.

The epidemic modeling of diseases in the population is based on the *law of mass action*, which states that in the homogeneously mixed population the rate of interaction is proportional to the product of the infected and susceptible. The law of mass action is a superposition of all contributions of individual components in the interaction, and if more than one process is involved effects are additive.

In diseases modeling, SIR model is used most frequently as the most realistic model for disease spread. SIR model exhibits threshold behavior, in other words if the epidemic parameters are belove some critical value, the virus in the network with N nodes dies out before large population is infected. If parameters are above the threshold, epidemic will reach most of the population. In the case of deterministic modeling, Kermack – McKendrick theorem [13] determine the threshold and the portion of infected nodes. For the stochastic SIR model, Whittle's theorems [14] determines the threshold behavior.

In the following sections, we will present most important results for deterministic and probabilistic models without going into details. We will discuss concepts of threshold and interaction matrix that appear in heterogeneous, multi-strata models.

2.1.1 Kermack-McKendrick threshold theorem

In [13], Kermack and McKendrick (KM) established deterministic epidemic model with fixed population of N individuals and three states (SIR). The results constitute a benchmark for a range of epidemic models. Their main result treats epidemic threshold that separates epidemic from a small infection. We will state the theorem formally.

Theorem 1. *For the deterministic SIR model with x denoting the fraction of susceptible, y the fraction of infectives and z the fraction of removals we can write equations:*

$$\frac{dx}{dt} = -\beta xy; \quad \frac{dy}{dt} = \beta xy - \gamma y; \quad \frac{dz}{dt} = \gamma y; \quad \frac{1}{x}\frac{dx}{dt} = -\frac{\beta}{\gamma}\frac{dz}{dt}$$

where β denotes the pairwise rate of infection and γ is the removal rate. For this system of equations it holds that:

1. *(Survival and Total size) When the infection stops spreading the fraction of susceptible that was never infected is x_∞ and the fraction of individuals ultimately removed is $z_\infty = x_0 + y_0 - x_\infty$ and z_∞ is a unique root of the equation:*

$$N - z_\infty = x_0 + y_0 - z_\infty = x_0 e^{-z_\infty \frac{\beta}{\gamma}}$$

   where x_0, y_0 are initial fraction of susceptible and infected nodes.

2. *(Threshold theorem) A major outbreak occurs if and only if $\frac{dy}{dt}|_{t=0} > 0$ which is equivalent to the $x_0 > \frac{\gamma}{\beta}$.*

3. *(Second threshold theorem) If x_0 exceeds $\frac{\gamma}{\beta}$ by a small value ϵ, then the final fraction of susceptible left in the population is approximately $x_\infty = \frac{\gamma}{\beta} - \epsilon$, and $z_\infty \approx 2\epsilon$.*

Whether major outbreak will occur depends on the initial condition i.e. the fraction of susceptibles at the start of the epidemic. Dependency of the spread on initial condition is specific feature of the SIR model, in SI and SIS models, the steady-state does not depend on initial conditions. This work is followed by a similar theorem for stratified population which involves interaction or transmission matrix [15].

Theorem 2. *For the deterministic SIR model in stratified population with m strata, x_i denotes the fraction of susceptible, y_i the fraction of infectives and z_i the fraction of removals in strata i, we can write equations:*

$$\frac{dx_i}{dt} = -x_i(\beta_{1i}y_1 + ... + \beta_{1i}y_1);$$

$$\frac{dy_i}{dt} = x_i(\beta_{1i}y_1 + ... + \beta_{1i}y_1) - \gamma_i y_i;$$

$$\frac{dz_i}{dt} = \gamma_i y_i;$$

where β_{ji} is the pairwise rate of infection between strata j and i and γ_i is the removal rate of strata i.

Provided that solution trajectory lies in region χ defined by

$$x_i, y_i, z_i \geq 0; \quad x_i + y_i + z_i = x_{i0} + y_{i0}; \quad (j = 1..m)$$

where x_{i0}, y_{i0} are initial fractions of susceptible and infected nodes in strata i. If the transmission matrix $\mathbf{B}$ with elements β_{ij} is primitive[1] and the removal rate vector has all elements positive, than it holds that:

1. *(Survival and Total size) When the infection stops spreading the fraction of susceptible that was never infected is $x_{i\infty}$ in each of the strata $i = 1..m$ and the fraction of individuals ultimately removed $z_{i\infty}$ constitutes a unique solution in χ.*

2. *(Threshold theorem) A major outbreak occurs if and only if the largest eigenvalue λ_{max} of the non-negative matrix $diag(\mathbf{x_0})\mathbf{B}'diag(\gamma^{-1})$ lies strictly outside of the unit circle.*

We have omitted second threshold theorem for the stratified population model because it is out of the scope of this discussion. The stratified model is the fist one to introduce interaction matrix, which plays important role in our model. The basic result that relates the largest eigenvalue of interaction matrix with the threshold is significant contribution to the epidemic theory. Using the concept of interaction matrix it is possible to explore influence of the interaction graph on the epidemic spreading.

---

[1]A nonnegative square matrix $\mathbf{B}$ is said to be a primitive matrix if there exists k such that for all i, j, the (i, j) entry of $\mathbf{B}^k$ is positive. A sufficient condition for a matrix to be a primitive matrix is for the matrix to be a nonnegative, irreducible matrix with a positive element on the main diagonal.

2.1.2   Whittle's threshold theorem

While deterministic models are very good at characterizing the epidemic in large population, they are not satisfactory for smaller populations like that of the household size. A general epidemic process of the SIR type, modeled by the simple stochastic (Markovian) epidemic model obeys

$$\Pr[(X,Y)(t+\Delta t) = (i-1, j+1)|(X,Y)(t) = (i,j)] = \beta ij\Delta t + o(\Delta t)$$
$$\Pr[(X,Y)(t+\Delta t) = (i, j-1)|(X,Y)(t) = (i,j)] = \gamma j\Delta t + o(\Delta t)$$
$$\Pr[(X,Y)(t+\Delta t) = (i, j)|(X,Y)(t) = (i,j)] = 1 - (\beta i + \gamma)j\Delta t - o(\Delta t)$$

Using forward Kolmogorov equations [16], the state probabilities can be determined. Even with a small population, threshold exists and it is formulated in the following theorem.

Theorem 3. *For this process with initial number of susceptible N and infectives I and relative removal rate $\frac{\gamma}{\beta}$ for any ϵ in $(0,1)$, let $p(\epsilon)$ denote the probability that at most $[N\epsilon]$ of the susceptible are ultimately infected i.e. that the intensity of epidemic does not exceed ϵ then*

1. *If $\frac{\gamma}{\beta} < N(1-\epsilon)$, then*

$$\left(\frac{\gamma}{\beta}\frac{1}{N}\right)^I \leqslant p(\epsilon) \leqslant \left(\frac{\gamma}{\beta}\frac{1}{N(1-\epsilon)}\right)^I$$

2. *If $N(1-\epsilon) \leq \frac{\gamma}{\beta} < N$, then*

$$\left(\frac{\gamma}{\beta}\frac{1}{N}\right)^I \leqslant p(\epsilon) \leqslant 1$$

3. *If $\frac{\gamma}{\beta} > N$ then $p(\epsilon) = 1$ and the infection can not reach more than ϵN individuals.*

Whittle *et al.* used non-negative integer-valued Markov process in continuous time [17]. Similar result exist for stratified population [18].

The threshold theorem distinguishes epidemics with **sub-** and **super-critical** conditions. Sub-critical case behaves roughly like birth and death process (BD). Duration and the final size are the same as in the BD process. Super-critical is more complex – it is possible that only a minor epidemic occurs with probability $(\frac{\gamma}{N\beta})^I$ which is similar to the super-critical BD conditioned on extinction. The early behavior of the epidemic is similar to the super-critical BD process conditioned on non-extinction. This is till the number of susceptible approaches

borderline value $\frac{\gamma}{\beta}$. At this stage, process follows approximately the solution of the deterministic epidemic model $E[X(t)] \approx x(t)$. If $\frac{\gamma}{\beta} < N$ and $\frac{\gamma}{\beta}$ and N are sufficiently large; the number of not infected is random variable with normal distribution. If $N > 3\frac{\gamma}{\beta}$ than number of never-infected is Poisson distributed.

Models in biology are adjusted for specific needs of diseases spread. Similarly, application of epidemic models in computer networks will reflect particularities of the field. In the following section, we will explore several important epidemic models for computer networks.

2.2   Epidemic Modeling and the Internet

In computer networks, epidemic modeling is applied in three areas. First of all, it is used for modeling computer virus and worm propagation [2]. Second, it is employed in epidemic algorithms and information dissemination in distributed networks [7], [8]. Finally, propagation of faults and failures is modeled using epidemic theory. One of the examples is the work of Coffman *et al.* [9], who models cascading BGP failures on a fully connected topology using birth and death process.

Epidemic algorithms for information dissemination are also referred to as gossip dissemination. Epidemic algorithms are simple and easy to deploy and mathematical tools predict the system behavior. Usually, either the information is spread forever, modeled by SI or each node spreads the information for some time and than it stops, following the SIR model [7]. Unreliable networks which use gossip algorithms can be modeled with SIS model.

The computer viruses are defined as small programs that can reproduce and copy themselves on other systems or on other files. The worm, also called network virus, does not need user intervention to spread, but instead it uses protocol and software bugs. In the beginning, before the computer networks became common, the viruses were propagating by means of removable media like floppy discs. The first PC virus was the Brain developed by two Pakistan brothers in 1986 [19]. The program was originally designed to protect the software copyright. In the '90s, macro viruses became common [20]. The macro viruses are programs written in script languages of the text editors. After infecting a file, every time the file is opened the virus script was executed. One of the first viruses that used Internet to spread were e-mail viruses. Traveling over the network as an e-mail attachment, they were written in different script languages or they used software bags to execute them selves on the computer. One of the most well known is the LOVE BUG first discovered in 2000 [20]. This virus raised concerns worldwide about a problem that was emerging. The newest species use cross-site scripting – a code injected into web page that execute on the client side [21], like Java scripts.

Today viruses and worms use different methods for spreading and different security vulnerabilities. Some of them use several propagation vectors at the same time.

We discuss several modeling approaches that appear in the literature.

2.2.1 Kephart and White model (KW)

One of the first milestones in the modeling of viruses on the Internet was set by the paper of Kephart and White [2]. The model belongs to the homogeneous models of the SIS type. KW have introduced a fixed, directed network as a part of the model for the first time. Kephart and White [2] considered a connected, regular graph[2] of N nodes. Using deterministic and probabilistic analyzes they have developed a number of useful analytical techniques.

The number of infected nodes in the population at time t is denoted by $I(t)$. If the population is sufficiently large, we can convert $I(t)$ to $y(t) \equiv I(t)/N$, a continuous quantity representing the fraction of infected nodes. Hence, the implicit assumption is that the number of states is sufficiently large such that the asymptotic regime for an infinite number of states is reached. Similarly to the models in previous section, the rate at which the fraction of infected nodes changes, is determined by two processes: (a) infected nodes are being cured and (b) susceptible nodes are infected. For process (a), the cure rate of a fraction y of infected nodes is δy. The rate at which the fraction y grows in process (b) is proportional to the fraction of susceptible nodes, i.e. $1 - y$. For every susceptible node, the rate of infection is the product of the infection rate β per link, the number of infected neighbors (i.e. the degree k) of the node, which is ky. Combining all contributions yields the time evolution of $y(t)$ in the Kephart and White model, described by the differential equation

$$\frac{dy(t)}{dt} = \beta k y (1 - y) - \delta y \tag{2.1}$$

with solution

$$y(t) = \frac{y_0 y_\infty}{y_0 + (y_\infty - y_0) e^{-(\beta k - \delta)t}} \tag{2.2}$$

where y_0 is the initial fraction of infected nodes whereas the steady-state fraction is $y_\infty = \lim_{t \to \infty} y(t)$ obeying $\frac{dy_\infty}{dt} = 0$.

The Kephart and White differential equation (2.1) is the basis of a large class of mean field models that, apart from some variations, possess the same type of solution, specified by a "steady-state" epidemic threshold,

$$\tau_{c,KW} = \frac{1}{k} \tag{2.3}$$

---

[2]Kephart and White have modeled an Erdös-Rényi random graph $G_p(N)$ with average degree $p(N-1)$, which tends, for large N, to a regular graph. Hence, to first order in N, the properties of virus spread in Erdös-Rényi random graphs and regular graphs are the same.

where $\tau_{c,KW}$ is critical effective spreading rate. Since each node has (on average) the same degree, the Kephart and White model is also termed a "homogeneous" model. Many variations on and extensions of the Kephart and White model have been proposed (see e.g. [22]). The logistic model of population growth, that was first introduced by Verhulst in 1838 as mentioned by Daley and Gani [1, p. 20], is, in fact, the same as the simple Kephart and White model. Moreover, the simplest stochastic analogon [1, p. 56-63] – a pure birth process with transition rate $\lambda_{n,n+1} = \beta n\,(N - n)$– is mathematically identical to the shortest path problem [17, Chapter 16] in the complete graph with i.i.d. exponential link weights. This observation and relation to the complete graph shows that these earlier models do not take the confining way of actual virus transport into account.

The deterministic analyzes provide basic overview of the epidemic features, however it does not provide information about fluctuation of the number of infected individuals which can result in the extinction of infection before it reaches the steady state. KW proposed probabilistic model of time dependent probability distribution of the infected nodes [2].

Using probabilistic model, KW estimate the survival probability distribution as well as the variance in the steady state.

2.2.2 Pastor-Satorras and Vespignani model

Vespignani and Pastor-Satorras (VP) [3] have noticed discrepancy between the data of the virus spread on the Internet and theoretical results from the SIS model for a homogeneous network. They have noticed the long lasting low prevalence even with the aggressive antivirus campaigns. This would suggest that relative spreading rate is very close to the critical value for a great number of the virus strains and for the all cleaning rates which is impossible. The SIS model for homogeneous networks would suggest that for fast cleaning rate the threshold would be reached and the virus should extinct exponentially fast.

They have proposed in [3] that spreading media is responsible for this paradox. They have concluded that viruses are spreading in the networks were the degree fluctuations are unbounded.

For the uncorrelated scale free networks the equation of the spreading dynamics can be written as

$$\frac{dy_k(t)}{dt} = -\delta y_k(t) + \beta k(1 - y_k(t))\Theta_k[\{y_{k'}(t)\}]$$

where $y_k(t)$ is relative density of infected vertices with degree k and $\Theta_k[\{y_{k'}(t)\}]$ is the probability that the node with k vertices is connected to the infected node. They assumed that Θ_k depends only on the degree k and the density of infected vertices $\{y_{k'}(t)\}$. The threshold can be derived as

$$\tau_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

which for $N \longrightarrow \infty$ and $\langle k^2 \rangle \longrightarrow \infty$ the threshold $\tau_c \longrightarrow 0$.

The actual social and virtual networks on the Internet are not infinite and they are very far from the thermodynamic limit. In that sense VP have discussed the effect of the maximum degree k_c on the threshold τ_c. For the scale-free exponent $2 < \gamma < 3$ threshold scales with

$$\tau_c(k_c) \simeq \left(\frac{k_c}{m} \right)^{\gamma - 3}$$

where m is the minimum degree of any vertex. They have also discussed the threshold for the correlated networks which does not have an exact solution.

The VP showed that topology is very important for the virus spread, however several questions stay unanswered. The cleaning rate is different for different nodes, some users are not well informed or have illegal software and can not download updates, having $\delta \longrightarrow 0$. The same effect of the long prevalence was also observed in the cases where the worm does not use social or virtual networks to spread, but the complete graph topology. An example of virus that spreads on completely connected graph is IP scanning worms.

2.2.3 Modeling scanning worms

Contemporary worms, also called network viruses are self-replicating programs which scan automatically IP address space for new victims. Population of hosts compose a full meshed, homogeneous network, where every user can contact every other. The paper of Zou et al. [23] gives a survey of the existing scanning types based on the logistic model of population growth. The considered model does not take into account curing or removing of nodes, thus using the SI model. We will present the most important scanning strategies.

1. *Uniform scanning* – The worm is scanning IP addresses uniformly. Any host can infect any other with the rate $\beta = \frac{\eta}{\Omega}$, where η is an average number of scans in time unit, Ω is number of IP addresses that worm is able to scan. The basic equation is

$$\frac{dI(t)}{dt} = \frac{\eta}{\Omega} I(t)(N - I(t))$$

2. *Uniform flash worm* – An example of idealistic worm which knows who vulnerable hosts are and scans the list uniformly. The worm is the fastest, but also very hard to deploy. Since it is fast spreading worm average propagation delay ϵ must be taken into account. The propagation equation is

$$\frac{dI(t)}{dt} = \frac{\eta}{\Omega} I(t - \epsilon)(N - I(t))$$

where $I(t - \epsilon) = 0, \forall t < \epsilon$.

3. *Destructive worm* – Most of the worms do not destruct the infected host, but some of them do. The destructive worm propagation model is derived based on the Witty worm that writes data at a random point of a hard disc after every $20,000$ scans till the infected computer crashes. The propagation equation is

$$\begin{aligned}
\frac{dI(t)}{dt} &= \frac{\eta}{\Omega} I(t)(N - I(t) - D(t)) - \lambda I(t) \\
\frac{dD(t)}{dt} &= \lambda I(t)
\end{aligned}$$

Scanning worms are one of the most prosperous sorts of malware. They spread quickly and automatically. However, they are also easiest to detect and stop, leaving Internet for stealth sorts of malware. New sorts use social engineering and spread over social networks. With the introduction of new Web applications for exchange of information and data, the number of security incidents increased. Scansafe Web STAT [21] reported that between May 2007 and May 2008 the risk of exposure to the exploits and compromised Web sites increased 407%.

The problem arises from the use of Web 2.0 which introduces client content into the Web pages. Applications like AJAX [3] and Adobe Flash provide excellent platform for malware development. Machines are infected simply by visiting a Web page, no user interaction is necessary [24]. Several worms have appeared that use social networking Web sites to spread. Many of them use AJAX scripts like Samy [24], Yamanner [25] and Mikeyy [26]. Other involve user interaction in order to download worm payload on the local machine as Koobface [27].

2.2.4   Topological aspects of epidemics

As far as the IP-scanning worms are concerned, individuals interact homogeneously without special fixed relations that can be modeled by a relation graph. One of the most important novelties in diseases modeling is the introduction of a network of contacts and traveling patterns. The homogeneous mixing assumption was shown not to be adequate [28]. Great attention is given to epidemics on different networks from random graphs [29], small world graphs [30], to scale free networks [3]. Generalization of the epidemic modeling to any network structure was recently proposed by Ganesh *et al.* in [31], Wang *et al.* [32] and Newman [6].

The physical layer of the telecommunication systems are often fixed networks. Higher communication layers can also be modeled by relation graphs, like P2P networks or DNS servers structure for the error propagation. Therefore, models

---

[3]asynchronous JavaScript and XML

that aim to analyze spreading in computer networks or on application layer networks have to take into account its topology. We will present several influential models that take into account topology of computer networks.

2.2.5 Sufficient conditions for fast recovery and lasting infection

Ganesh *et al.* [31] derived sufficient conditions for fast recovery and lasting infection of an epidemic on an arbitrary network for SIS and SIR models. For the effective spreading rate $\tau < \frac{1}{\lambda_{max}}$, the lifetime of the epidemic is of the order $O(\log(N))$, where N is the number of nodes in the network and λ_{max} is the largest eigenvalue of adjacency matrix of the underling graph.

Further, Ganesh *et al.* bound the expected survival time with the worst case scenario of an epidemic. If the relative infection rate τ is larger than the generalized isoperimetric constant[4] $\eta(G,m)$ of a graph G, the mean epidemic lifetime [31] is of the order $\Omega(e^{N^a})$, for some positive constant a.

The sufficient condition for lasting infection for any initial condition is

$$
\begin{aligned}
\tau &> \frac{1}{\eta(G,m)} \\
m &= \Theta(N^a) \text{ and } m \leqslant \lfloor N/2 \rfloor \\
E[T] &= \Omega(e^{N^a})
\end{aligned}
\qquad (2.4)
$$

where T is the time till absorbing state is reached and the maximal size of infected cluster m is tightly bounded by N^a, (from below ($m = \Omega(N^a)$) and from above ($m = O(N^a)$)).

The sufficient conditions (2.4) hold for any initial condition, any set of initially infected nodes. The proof consists of lower bounding the actual process with another stochastically dominating process - the standard coupling argument.

The result for the SIR epidemic model, Ganesh *et al.* explore topological aspects in [33].

2.2.6 The Model of Wang *et al.*

The major merit of the model of Wang *et al.* [32] is the incorporation of an arbitrary network characterized by the adjacency matrix A. This is generalizes the homogeneous Kephart and White model, where the only network characteristic

---

[4]Let S denote the cluster of maximum m nodes and $e(S, S^c)$ the number of links between cluster S and the rest of the network and $|S|$ denotes the number of nodes in set S. Generalized isoperimetric constant is the minimum of the ratios between the number of links between clusters and the size of the cluster for different maximal sizes of the infected cluster m,

$$
\eta(G,m) = \inf_{S \subset \{1,..,n\}, |S| \leq m} \frac{e(S,S^c)}{|S|}, \qquad 0 < m \leqslant \lfloor N/2 \rfloor
$$

was the (average) degree. The discrete-time model of Wang *et al.* belongs to the class of mean field models. Their major and intriguing result is that the epidemic threshold is specified by

$$\tau_{c,\text{WCWF}} = \frac{1}{\lambda_{\max}(A)}$$

Unfortunately, this result is proved in an approximate manner which questions to what extent this remarkable result holds in general.

For a finite power law graphs studied by Vespignani and Pastor-Satorras epidemic threshold is more precisely indicated by the largest eigen value then by the term $\frac{\langle k \rangle}{\langle k^2 \rangle}$ as shown in [32].

We will use Wang *et al.* notation in order to distinguish discrete model from our N-intertwined model. Without going into detailed derivations, we present here the evolution equation. Probability of node i to be infected at the step t is denoted by $p_{i,t}$, probability that at step t infected node i attempts to infect neighbor j is denoted by β_w and probability of curing during the same step with δ_w.

$$p_{i,t} = 1 - \prod (1 - \beta_w p_{j,t-1})(1 - p_{i,t-1} - \frac{1}{2}\delta_w p_{i,t-1}) - \frac{1}{2}\delta_w p_{i,t-1}$$

which leads to the steady state solution denoted by $p_{i,\infty} = p_i$

$$p_i = \frac{1 - \prod (1 - \beta_w p_{j,t-1})}{1 - (1 - \frac{1}{2}\delta_w)\prod (1 - \beta_w p_{j,t-1}) + \frac{1}{2}\delta_w} \tag{2.5}$$

We will discuss in details differences between Wang *et al.* and N-intertwined model in the sections 4.5.3.

2.2.7   Interactive Markov Chains

One of the most comprehensive frameworks that uses interaction matrix to describe relations between individuals, is the *Interactive Markov Chains (IMC)*. Garetto et al. [34] developed a stochastic model of virus propagation based on IMC [35].

IMC is specified for the directed graph $G(N, L)$. Each node on the graph has a state that evolves over time according to an internal discrete-time Markov chain. The transition probabilities depend on the internal state of a node as well as on the states of neighboring nodes. The system can be completely described by the *global Markov chain* with m^N states where m is the number of states of a node. However, because of the exponential growth of the number of states Garetto et al. use the *influence model* which is a special case of IMC developed by C. Asavathiratham [36]. With the influence model it is possible to obtain the marginal state probabilities of each node with at most mN states.

The model is discrete and in order to emphasize this fact we will use different notation. Let $\boldsymbol{\pi}_j[k]$ be a state probability row vector of a node j at a given time step k. The evolution of each node has a form

$$\boldsymbol{\pi}_j[k+1] = \sum_{i=1}^{N} w_{ij}\boldsymbol{\pi}_i[k]\boldsymbol{P}_{ij} \tag{2.6}$$

Where w_{ij} is weight of a link connecting node i and j taking values from the interval $[0,1]$, and $\sum_{i=1}^{N} w_{ij} = 1$. $\boldsymbol{P}_{ij}$ determines how specific states of node j influence specific states of node i and the number of rows equal to the number of states in i and the number of columns equal to the number of states in j. For a homogeneous influence model number of states is equal for all nodes and is marked by m. In this case transition matrix is $m \times m$ matrix. The element $(\boldsymbol{P}_{ij})_{gh}$ (h column and g row of the matrix $\boldsymbol{P}_{ij}$) defines how the state g of the neighbor i can influence the state h of the node j. The overall influence of a neighboring node is represented with the weight w_{ij}. In this way network topology is extracted. Garetto et al. obtain marginal state probabilities of all sites at any given time step from the equation

$$H = \begin{bmatrix} w_{11}\boldsymbol{P}_{11} & \cdots & w_{1N}\boldsymbol{P}_{1N} \\ \vdots & & \vdots \\ w_{N1}\boldsymbol{P}_{N1} & \cdots & w_{NN}\boldsymbol{P}_{NN} \end{bmatrix}$$
$$\Pi[k] = \Pi[0]H^k$$

where $\Pi[0]$ is initial node probabilities.

This model is in many aspects similar to the model of Wang et al. The model of Garetto et al. separates the topology influence determined by matrix W, and state influence determined by matrix P, and analyze behavior for different topologies and relates eigenvalue structure to the spreading process.

2.2.8 Pair-approximation models

Pair-approximation models take into account states of node pairs together for all node pairs in the network. With only two states per node - infected or susceptible number of states for one pair becomes 4, while the number of pairs that have to be considered for a network of N nodes is $\binom{N}{2}$. The pair-approximation model is used for computer virus modeling by Nikoloski *et al.* [37]. Although model is more precise, additional complexity is not justified.

2.2.9 Percolation on a graph

Percolation theory is devised as a model of a fluid flow and other similar processes in disordered media. Using infinite lattice topology, porous rocks are modeled by adding or removing link/node at random. Two different type of percolation exist, namely the site percolation and the bond percolation. In the model with bond

percolation, each link in lattice exist with the probability p. For such a lattice, a percolation threshold exist, defined in the following way. If p is 0, all nodes in lattice are isolated. If $p = 1$, the system is connected from one side to the other, since there are paths that go completely across the system, linking one sub-unit to the next along the spanning cluster. If probability p changes from 1 to 0, at one point there is no longer an unbroken path from one side of the system to the other.

Epidemic can be modeled using percolation theory on any graph G with site percolation. Garetto et al. [34] used site percolation on the small-world graph to set boundaries of the final size of infected population. The shortcomings of percolation theory is assumption of infinite set and impossibility to apply any epidemic model except that of a SI type or special case of SIR with nodes immune from the start.

In this section, we gave and overview of important results in epidemic modeling in biology and computer networks. We concentrated mostly on the results that determine threshold and how the underlining graph of contacts influences spread.

Part I

Epidemic spreading on networks and N-intertwined model

Chapter 3

The exact 2^N state Markov chain

Epidemic spreading with two states – Susceptible and Infected can be completely described by the global Markov chain with 2^N states. We will first study the global Markov chain for SIS model.

We consider the virus spread in an undirected graph $G(N, L)$ characterized by a symmetric adjacency matrix A. We assume that the arrival of an infection on a link and the curing process of an infected node are independent Poisson processes with rate β and with rate δ, respectively. As soon as a node i receives an infection at time t, it is considered to be infected and infectious and in state $X_i(t) = 1$. Similarly, an infected node i is cured with rate δ, and in the healthy state $X_i(t) = 0$ at time t. At each time t a node is in one of these two states.

The state $Y(t)$ of the network at time t is defined by all possible combinations of states in which the N nodes can be at time t,

$$Y(t) = \begin{bmatrix} Y_0(t) & Y_1(t) & \dots & Y_{2^N-1}(t) \end{bmatrix}^T$$

and

$$Y_i(t) = \begin{cases} 1, & i = \sum_{k=1}^{N} X_k(t)\, 2^{k-1} \\ 0, & i \neq \sum_{k=1}^{N} X_k(t)\, 2^{k-1} \end{cases}$$

Hence, the state space of the Markov chain is organized with $x_k \in \{0, 1\}$ as

State number i	$x_N x_{N-1}...x_2 x_1$
0	00...000
1	00...001
2	00...010
3	00..011
...	...
$2^N - 1$	11...11

The number of the states with j infected nodes is $\binom{N}{j}$. Figure 3.1 shows an example of the Markov state diagram in a graph with $N = 4$ nodes.
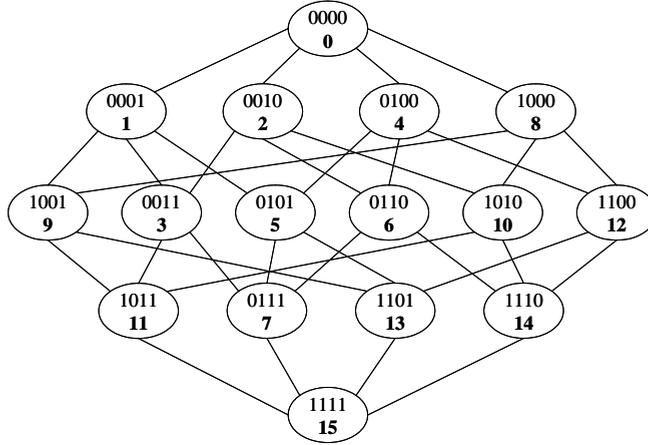
Figure 3.1: *The state diagram in a graph with $N = 4$ nodes and the binary numbering of the states.*

The defined virus infection process is a continuous-time Markov chain with 2^N states specified by the infinitesimal generator Q with elements

$$q_{ij} = \begin{cases} \delta & \text{if } i = j + 2^{m-1}; m = 1, 2...N; x_m = 1 \\ \beta \sum_{k=1}^{N} a_{mk} x_k & \text{if } i = j - 2^{m-1}; m = 1, 2...N; x_m = 0 \\ -\sum_{k=1; k \neq j}^{N} q_{kj} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

and $i = \sum_{k=1}^{N} x_k 2^{k-1}$. The time dependence of the probability state vector $s(t)$, with components

$$\begin{aligned} s_i(t) &= \Pr[Y(t) = i] \\ &= \Pr[X_1(t) = x_1, X_2(t) = x_2, ..., X_n(t) = x_n] \end{aligned}$$

and normalization $\sum_{i=0}^{2^N - 1} s_i(t) = 1$, obeys [17, p. 182] the differential equation

$$\frac{ds^T(t)}{dt} = s^T(t)Q$$

whose solution is

$$s^T(t) = s^T(0)e^{Qt}$$

The definition of $s_i(t)$ as a joint probability distribution shows that, if we sum over all the states of all nodes except for the node j, we obtain the probability that a node j is either healthy $x_j = 0$ or infected $x_j = 1$,

$$\Pr[X_j(t) = x_j] = \sum_{i=0; i \neq j}^{2^N - 1} s_i(t)$$

where, in the index $i = \sum_{k=1}^{N} x_k 2^{k-1}$, every x_k with $k \neq j$ takes both values from the set $\{0, 1\}$, while for $k = j$, $x_k = x_j$ is either 0 (healthy) or 1 (infected). Defining $v_j(t) = \Pr[X_j(t) = 1]$, then the relation between the vectors $s(t)$ and $v(t)$ is

$$v^T(t) = s^T(t)M$$

where the $2^N \times N$ matrix M contains the states in binary notation, but bit-reversed:

$$M = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

The binary representation of the network states determines the structure of the Q matrix. The upper triangular part of Q, denoted by Q_A, depends on the adjacency matrix elements a_{ij}, while the lower triangular part Q_δ does not. The diagonal elements of any Q matrix are the negative sum of the row elements, such that $Q_{\text{diag}} = \text{diag}(q_{00}, q_{11}, \ldots, q_{2^N - 1, 2^N - 1})$ with $q_{jj} = -\sum_{k=1; k \neq j}^{N} q_{kj}$ as in (3.1). It is thus instructive to write Q as a sum of three matrices $Q = Q_\delta + Q_A + Q_{\text{diag}}$. The structure of the matrix Q_δ is shown in the Fig. 3.2, where the block matrix $B(j) = \delta I_{2^j \times 2^j}$ and the nondefined elements are zeros. This nested structure is the consequence of the binary representation.

The matrix Q_A is shown in Fig. 3.3. The block matrices $C(j)$ in Q_A are diagonal matrices of size $2^j \times 2^j$ with diagonal elements depending on the adjacency matrix A. The first row of the matrix Q is zero, and as a consequence the largest block is $C(N-1)$. The elements of Q_A depend on the indices i, j where $i = \sum_{k=1}^{N} x_k 2^{k-1}$ as $Q_A(i, j) = \beta \sum_{k=1}^{N} a_{mk} x_k$ where $i = j - 2^{m-1}; m = 1, 2 \ldots N; x_m = 0$. The exact 2^N- state Markov chain has an absorbing state because the first row in Q is a zero row and the absorbing state is the zero state in which all nodes are healthy. The steady-state is just this absorbing state, with steady-state vector $s_\infty = \pi = (1, 0, \ldots, 0)$. The probability state vector requires the insights in the eigenstructure of Q because [38]

$$s(t) = s(0)e^{Qt} = \pi + \sum_{k=1}^{2^N - 1} e^{\lambda_k t} \sum_{m=0}^{n_k - 1} r_{k,m} \frac{t^m}{m!}$$

where n_k denotes the multiplicity of the eigenvalue λ_k (with $\text{Re}\,\lambda_k < 0$) and the vector $r_{k,m}$ is related to the left- and right eigenvector belonging to λ_k and the
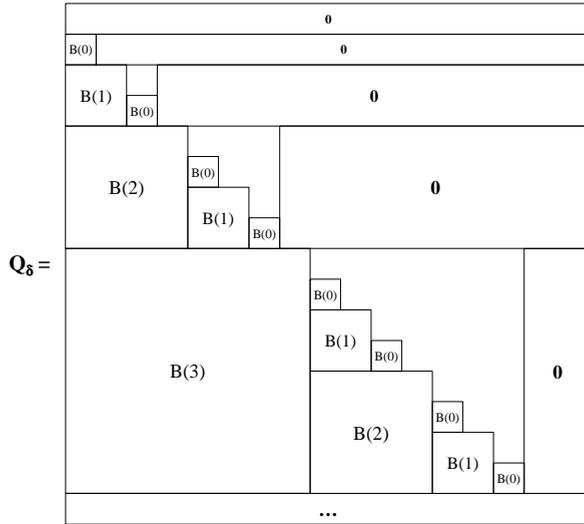
Figure 3.2: *The lower triangular part Q_δ of the infinitesimal generator Q.*

initial conditions. Since $v_j(t) = \left(s^T(t)M\right)_j = \sum_{k=0}^{2^N-1} s_k M_{kj}$ is a sum of certain rows of $s(t)$, we may write

$$v_j(t) = \sum_{k=1}^{2^N-1} e^{\lambda_k t} \sum_{m=0}^{n_k-1} \left(\sum_{i \in M_j} (r_{k,m})_i \right) \frac{t^m}{m!}$$

where M_j denotes the j-th column in the matrix M. Let μ_j be the largest eigenvalue λ_k of the set where $(r_{k,m})_i \neq 0$, then $v_j(t)$ is dominated (for not too small t) by

$$v_j(t) \sim e^{\mu_j t} \sum_{m=0}^{n_{\mu_j}-1} \gamma_m \frac{t^m}{m!} \tag{3.2}$$

which shows that a "bell-shape" distribution of $v_j(t)$ can only occur if that largest eigenvalue $\mu_j < 0$ has a multiplicity larger than 1.

3.1 Spectrum of Q

For all infinitesimal generators, it holds that $\det Q = 0$, and, hence, the largest eigenvalue is $\lambda = 0$.

Theorem 4. *For $\beta = 0$, the eigenvalues of the matrix Q, defined by (3.1), are $\lambda(Q_{\beta=0}) = -k\delta$ with multiplicity $\binom{N}{k}$, where $0 \le k \le N$.*

Figure 3.3: *The upper triangular part Q_A of Q.*

Proof. For $\beta = 0$, the infinitesimal generator $Q = Q_\delta + Q_{\text{diag}} + Q_A$ reduces to the lower-triangular matrix $Q_\delta + Q_{\text{diag}}$, whose eigenvalues are identical to the diagonal elements of Q_{diag}, which are multiples of δ. In fact, the structur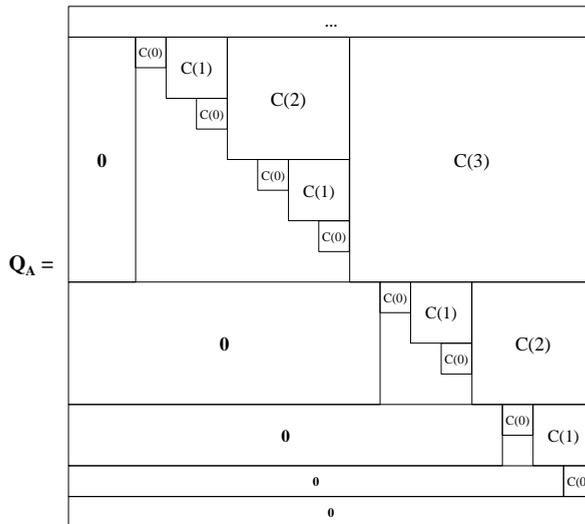e of Q_δ shows that each block row j has a row sum equal to $k\delta$ for $1 \leq k \leq N$ whose value appears $\binom{j}{k-1}$ times. Hence, $Q_{\beta=0}$ has an eigenvalue at $\lambda = -k\delta$ with multiplicity $\sum_{j=0}^{N-1} \binom{j}{k-1} = \binom{N}{k}$. These contain all the non-zero eigenvalues of $Q_{\beta=0}$ because $\sum_{k=1}^{N} \binom{N}{k} = 2^N - 1$. $\qquad\square$

For small values of τ, Q tends thus to a discrete, binomial spectrum. Fig. 3.4 illustrates that, also for larger τ, the spectrum of Q for the complete graph K_N is still discrete[1], containing many eigenvalues with high multiplicity.

Proposition 5. *For constant δ and increasing β (and $\tau = \frac{\beta}{\delta}$), the eigenvalues of Q shift, on average, to more negative values than those of $Q_{\beta=0}$.*

Proof. We apply Gershgorin's Theorem[2] to $Q = Q_\delta + Q_{\text{diag}} + Q_A$, where $Q_A = \beta T_A$ and T_A only contains (non-zero) integer elements related to the adjacency matrix A as observed from (3.1). Hence, $q_{ii} < 0$ decreases with β which implies

---

[1] Random matrices of this size exhibit an almost continuous spectrum.

[2] Every eigenvalue of a matrix B lies in at least one of the circular discs with centers b_{jj} and radii $R_j = \sum_{k=1; k \neq j} |b_{jk}|$. For any infinitesimal generator Q, Gershgorin's Theorem shows that $|\lambda_i - q_{ii}| \leq |q_{ii}|$ and that the maximum possible interval for real eigenvalues of Q is $[0, 2\max_i |q_{ii}|]$.
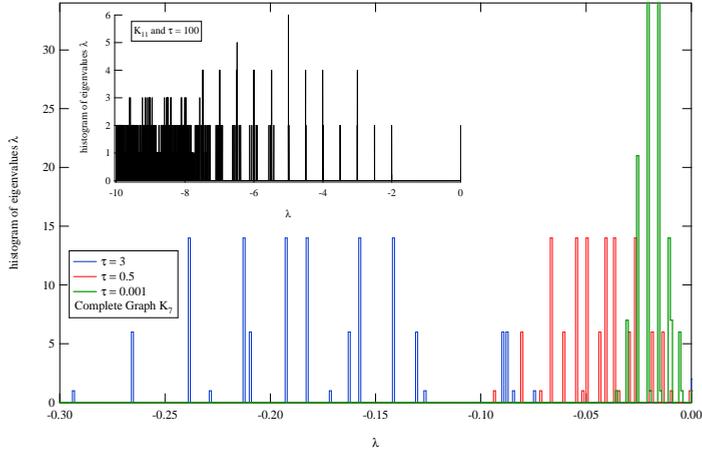
Figure 3.4: *(in color) The histogram eigenvalues λ of Q of the in the complete graph K_N for three values of τ gives the number of times an eigenvalue λ occurs. The insert shows the spectrum of K_{11} for an extremely high $\tau = 100$.*

that both the center position and the possible range of each eigenvalue $\lambda_i(Q)$ increases with β.   □

Corollary 6. *The eigenvalues of Q for the complete graph K_N and line graph spread over the largest, respectively smallest possible range among all connected graphs. The maximum possible range of the real part of eigenvalues of Q for any connected graph is $\left(-\frac{(\beta N + \delta)^2}{2\beta}, 0\right]$*

Proof. From $Q_A = \beta T_A$, defined in the proof of Theorem 5, it follows that the maximum possible sum of row elements occurs for K_N (all $a_{ij} = 1$ except for $a_{ii} = 0$) and the minimum one for line graph (only one 1-element on each row in the adjacency matrix A). Gershgorin's Theorem then provides the first statement. Since the maximum eigenvalue range thus occurs for a complete graph, we consider in the Q-matrix for K_N the i-th row with k one-bits in the binary representation. The row elements, except from the diagonal element, represents the transitions from and to a state with $N - k$ healthy and k infected nodes. The row sum of these positive elements equals $\beta k(N-k) + k\delta$, and, hence, $q_{ii} = -\beta k(N-k) - k\delta$. Optimizing with respect to k proves the corollary.   □

As shown in the Section 3.1.3, also for the line graph, the maximum of the diagonal elements q_{ii} can be computed.

Yet, there are open questions regarding the spectrum of Q. (a) Although Q is not symmetric, computations reveal that all eigenvalues of Q are real (and negative). (b) Perturbation theory of Q for small β (or τ) expresses the eigenvalues

in terms of those of $Q_{\beta=0}$ and of the corresponding right- and left-eigenvectors of $Q_{\beta=0}$. However, the multiplicity of the eigenvalues of $Q_{\beta=0}$ further complicates the perturbation analysis. (c) The recursive block-structure (due to the binary representation) of Q needs to be exploited.

In the sequel of this section, we confine to explicit computation of the Q matrix for two extreme types of graphs, the complete graph which has the smallest average hopcount (or the fastest virus penetration), and the line graph that possesses the largest possible average hopcount.

3.1.1 The complete graph K_N

Fig. 3.5 shows the four largest eigenvalues of Q for the complete graph K_N for $N = 5, 8$ and 10. The second largest eigenvalue seems the only eigenvalue that increases – contrary to the expectations of Gershgorin's Theorem – roughly exponentially in τ and with rate increasing for increasing size N. This second largest eigenvalue determines the speed of convergence towards the steady-state. Fig. 3.5 also shows that, initially for small τ, the third and fourth eigenvalue are the same and bifurcate (see dots) into distinct values roughly around $\tau_c = \frac{1}{\lambda_{\max}(A)} = \frac{1}{N-1}$. Hence, (3.2) indicates that below τ_c, the dominant eigenvalue is simple causing exponential decay, while above τ_c, it has multiplicity larger than 1 creating a bell-shape.

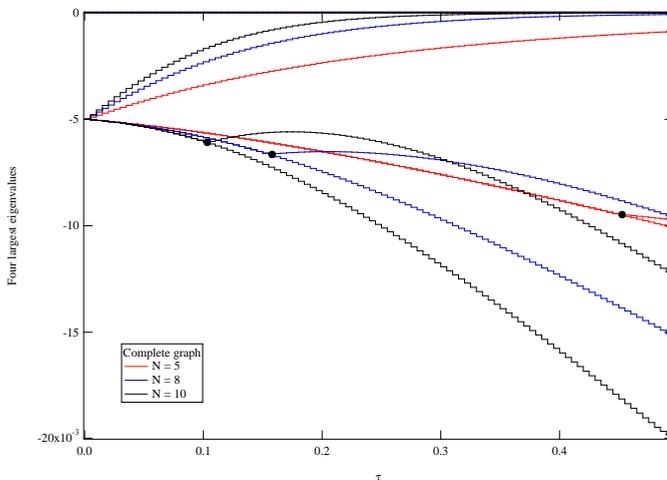**Figure 3.5:** *The four largest eigenvalues of the infinitesimal generator Q for the complete graph with size $N = 5, 8$ and 10 as a function of τ with $\delta = 5 \ 10^{-3}$. The second largest eigenvalues are increasing with τ as $\lambda_2(5) \approx -\delta e^{-3.5\tau}, \lambda_2(8) \approx -\delta e^{-8.8\tau}$ and $\lambda_2(10) \approx -\delta e^{-14.7\tau}$.*

In Fig. 3.6, the eigenvalues of Q for all computable complete graphs (up to $N = 13$) have been numerically calculated. The second largest eigenvalue seems well fitted (for $\tau \geq 0.05$) by

$$\lambda_2 = -\delta \, e^{-b(\tau)L} \tag{3.3}$$

where $L = \binom{N}{2}$ denotes the number of links in the complete graph K_N. The dependence on τ is approximately given by $b(\tau) \approx 0.17\tau(1 + 2\tau)$. Assuming that the scaling law (3.3) of λ_2 holds for any N, the convergence time T of the virus spread in K_N towards the steady-state (the zero state), defined by $r_2 e^{-|\lambda_2|T} = 10^{-\epsilon}$ is found as $T = O\left(e^{b(\tau)L}\right) = O\left(e^{\frac{b(\tau)}{2}N^2}\right)$. In other words, for large size N and $\tau > 0$, the convergence time T is so large that convergence towards the zero state is in reality never reached, which explains the appearance of the so-called "metastable state".

Ganesh *et al.* [31] show that, for $\tau < \tau_c$ – a regime that is not covered by (3.3) –, the mean epidemic lifetime $E[T]$ scales as $O(\log N)$ while, for $\tau > \tau^* > \tau_c$ where τ^* is the generalized isoperimetric constant, $E[T] = O\left(e^{N^a}\right)$, for some constant a. If we may extrapolate (3.3) to large N, it shows that the constant $a = 2$ for K_N.
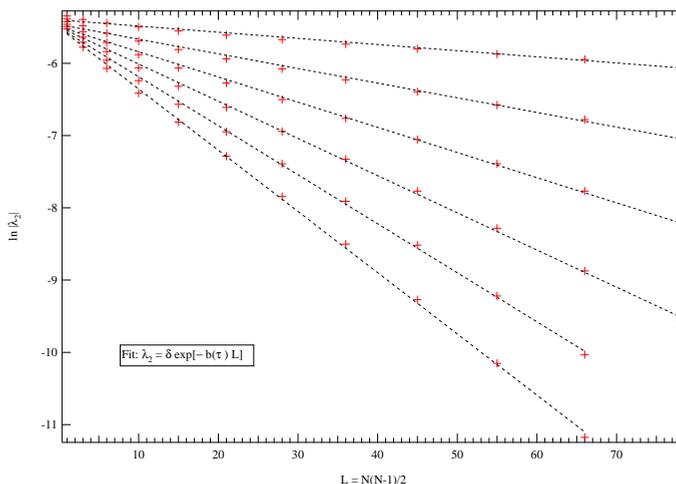

Figure 3.6: *The logarithm of $-\lambda_2$ versus the number links in K_N for $\tau = 0.05, 0.1, 0.15, \ldots, 0.3$ and $\delta = 5 \; 10^{-3}$.*

3.1.2   The line graph

Fig. 3.7 plots the second largest eigenvalue λ_2 of Q for the line graph. The largest eigenvalue of the adjacency matrix A of the line graph, where each row

has precisely one non-zero element in the upper triangular part of A, is $\lambda_{\max}(A) = 2\cos\left(\frac{\pi}{N+1}\right) < 2$. Fig. 3.7 (axis on the right) also shows the epidemic threshold of the line graph $\tau_c = \frac{1}{\lambda_{\max}(A)} > \frac{1}{2}$ versus N.
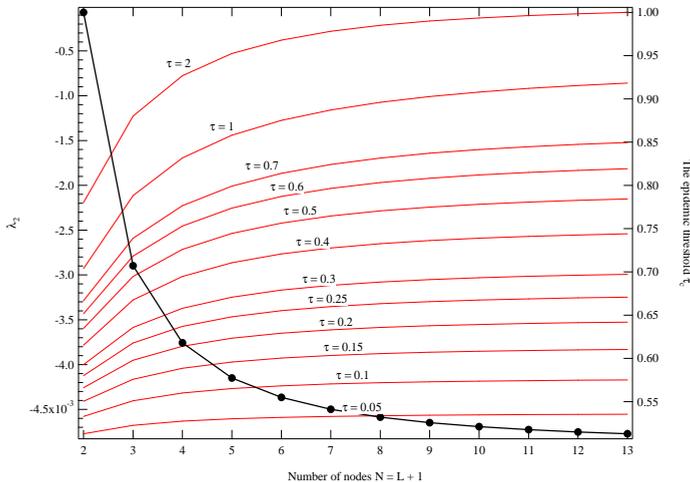
Figure 3.7: *The second largest eigenvalue λ_2 of Q in the line graph versus the number of nodes N for various τ and $\delta = 5 \cdot 10^{-3}$. The epidemic threshold τ_c is shown in dotted line on the right hand side axis*

As observed from Fig. 3.7, the curves λ_2 increase very slowly with N. Via curve fitting in the range $N \in [8, 13]$, we found that
$$\lambda_2(\tau, N) \approx -\delta e^{-\tau(1.184 + 0.0413N)},$$
which shows the exponential dependence on τ (accurate) and the less accurate dependence on N. If extrapolation to large N is allowed, the convergence time T of the virus spread in the line graph towards the steady-state (the zero state) is $T = O\left(\frac{1}{\lambda_2}\right) = O\left(e^{\tau(1.184 + 0.0413N)}\right)$, which is considerably smaller than in K_N, the other extreme case.

3.1.3 The row sum of Q for the line topology

We compute the upper bound of the sum of the rows in Q for the line topology. First, let us consider two cases with the same number of infected nodes on the same line graph as shown in Fig. 3.8.

Case a) has two nodes that can be infected by two neighbors and one that can be infected by only one neighbor. In the case b) only one node can be infected by one neighbor. Thus, in the case a) all healthy nodes can be infected by two neighbors in contrast to case b) where one node can be infected by only
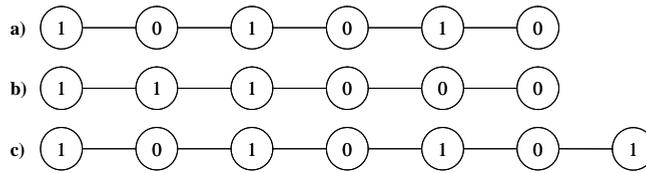
Figure 3.8: *a) and b): Line graph with $N = 6$ and 3 infected nodes. The '1 ' refers to an infectected and a '0 ' to healthy node. c) Line graph with $N = 7$ (odd number of nodes) and 4 infected nodes.*

one neighbor. Since, from the viewpoint of curing, both cases are equal, we will consider only the cases analogous to a), where nodes are alternately infected. There is also a difference between the line graphs with odd and even number of nodes N, as observed from case c). We can now write the sum of the non-diagonal elements of such a i-th row in Q as a function of the number of infected nodes k. We have for odd N,

$$\max |q_{ii}| = (2\beta(k-1) + \beta + \delta k), \qquad k < \frac{N+1}{2}$$

$$\max |q_{ii}| = (2\beta(N-k) + \delta k), \qquad k \geqslant \frac{N+1}{2}$$

and when N is even,

$$\max |q_{ii}| = (2\beta(k-1) + \beta + \delta k), \qquad k \leqslant \frac{N}{2}$$

$$\max |q_{ii}| = (2\beta(N-k) + \delta k), \qquad k > \frac{N}{2}$$

Chapter 4

N-intertwined model

In this chapter, we focus on simple continuous-time model which belongs to the class of susceptible-infected-susceptible (SIS) models. Our major motivation is to understand the influence of graph characteristics on epidemic spreading.

By separately observing each node, we will model the virus spread in a bi-directional network specified by a symmetric adjacency matrix A. Every node i at time t in the network has two states: infected with probability $\Pr[X_i(t) = 1]$ and healthy with probability $\Pr[X_i(t) = 0]$. At each moment t, a node can only be in one of two states, thus $\Pr[X_i(t) = 1] + \Pr[X_i(t) = 0] = 1$. If we apply Markov theory straight away, the infinitesimal generator $Q_i(t)$ of this two-state continuous Markov chain is,

$$Q_i(t) = \left[\begin{array}{cc} -q_{1;i} & q_{1;i} \\ q_{2;i} & -q_{2;i} \end{array} \right] \tag{4.1}$$

with $q_{2;i} = \delta$ and

$$q_{1;i} = \beta \sum_{j=1}^{N} a_{ij} 1_{\{X_j(t)=1\}}$$

where the indicator function $1_x = 1$ if the event x is true else it is zero. The coupling of node i to the rest of the network is described by an infection rate $q_{1;i}$ that is a random variable, which essentially makes the process doubly stochastic. This observation is crucial. For, using the definition of the infinitesimal generator [17, p. 181],

$$\Pr[X_i(t + \Delta t) = 1 | X_i(t) = 0] = q_{1;i}\Delta t + o(\Delta t)$$

the continuity and differentiability shows that this process is not Markovian any-more. The random nature of $q_{1;i}$ is removed by an additional conditioning to all possible combinations of rates, which is equivalent to conditioning to all possible combinations of the states $X_j(t) = 1$ (and their complements $X_j(t) = 0$) of the

neighbors of node i. Hence, the number of basic states dramatically increases. Eventually, after conditioning each node in such a way, we end up with a 2^{N-} state Markov chain, defined in Chapter 3.

Instead of conditioning, we replace the actual, random infection rate by an effective or average infection rate, which is basically a mean field approximation,

$$E[q_{1;i}] = E\left[\beta \sum_{j=1}^{N} a_{ij} 1_{\{X_j(t)=1\}}\right] \tag{4.2}$$

In general, we may take the expectation over the rate β, the network topology via the matrix A and the states $X_j(t)$. Since we assume that both the infection rate β and the network are constant and given, we only average over the states. Using $E[1_x] = \Pr[x]$ (see e.g. [17]), we replace $q_{1;i}$ by

$$E[q_{1;i}] = \beta \sum_{j=1}^{N} a_{ij} \Pr[X_j(t) = 1]$$

which results in an effective infinitesimal generator,

$$\overline{Q_i(t)} = \begin{bmatrix} -E[q_{1;i}] & E[q_{1;i}] \\ \delta & -\delta \end{bmatrix}$$

The effective $\overline{Q_i(t)}$ allows us to proceed with Markov theory. Denoting $v_i(t) = \Pr[X_i(t) = 1]$ and recalling that $\Pr[X_i(t) = 0] = 1 - v_i(t)$, the Markov differential equation [17, (10.11) on p. 182] for state $X_i(t) = 1$ turns out to be non-linear

$$\frac{dv_i(t)}{dt} = \beta \sum_{j=1}^{N} a_{ij} v_j(t) - v_i(t) \left(\beta \sum_{j=1}^{N} a_{ij} v_j(t) + \delta\right) \tag{4.3}$$

Each node obeys a differential equation as (4.3),

$$\begin{cases} \frac{dv_1(t)}{dt} = \beta \sum_{j=1}^{N} a_{1j} v_j(t) - v_1(t) \left(\beta \sum_{j=1}^{N} a_{1j} v_j(t) + \delta\right) \\ \frac{dv_2(t)}{dt} = \beta \sum_{j=1}^{N} a_{2j} v_j(t) - v_2(t) \left(\beta \sum_{j=1}^{N} a_{2j} v_j(t) + \delta\right) \\ \qquad\qquad\qquad\qquad \vdots \\ \frac{dv_N(t)}{dt} = \beta \sum_{j=1}^{N} a_{Nj} v_j(t) - v_N(t) \left(\beta \sum_{j=1}^{N} a_{Nj} v_j(t) + \delta\right) \end{cases}$$

Written in matrix form, with $V(t) = \begin{bmatrix} v_1(t) & v_2(t) & \cdots & v_N(t) \end{bmatrix}^T$, we arrive at

$$\frac{dV(t)}{dt} = \beta A V(t) - \text{diag}(v_i(t))(\beta A V(t) + \delta u) \tag{4.4}$$

where u is the all-one vector and $\text{diag}(v_i(t))$ is the diagonal matrix with elements $v_1(t), v_2(t), \ldots, v_N(t)$.

We rewrite (4.4) with $V(t) = \text{diag}(v_i(t)) u$ as

$$\frac{dV(t)}{dt} = \beta A V(t) - \delta \text{diag}(v_i(t)) u - \text{diag}(v_i(t)) \beta A V(t)$$

$$= (\beta A - \delta I) V(t) - \beta \text{diag}(v_i(t)) A V(t)$$

or

$$\frac{dV(t)}{dt} = (\beta \text{diag}(1 - v_i(t)) A - \delta I) V(t) \tag{4.5}$$

An extension of the N-intertwined model where the curing and infection rates are node specific is examined in more details in Chapter 5.

4.1 The steady-state

Assuming that the steady-state exists, we can calculate the steady-state probabilities of infection for each node. The steady-state, denoted by $v_{j\infty}$, implies that $\left.\frac{dv_j(t)}{dt}\right|_{t\to\infty} = 0$, and thus we obtain from (4.3) for each node j,

$$\beta \sum_{j=1}^{N} a_{ij} v_{j\infty} - v_{i\infty} \left(\beta \sum_{j=1}^{N} a_{ij} v_{j\infty} + \delta \right) = 0$$

Since all the diagonal elements of the adjacency matrix A are zero, $a_{jj} = 0$, we find

$$v_{i\infty} = \frac{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty}}{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty} + \delta} = 1 - \frac{1}{1 + \tau \sum_{j=1}^{N} a_{ij} v_{j\infty}} \tag{4.6}$$

This nodal steady-state is the ratio of the (average) infection rate induced by the node's direct neighbors $\sum_{j=1}^{N} a_{ij} v_{j\infty}$ over the total (average) rate of both the competing infection and curing process. Since $a_{jj} = 0$, (4.6) is equal to the steady-state probability in a two-state, continuous Markov chain (see e.g. [17, p. 196]), which exemplifies the local (or nodal) character of our N-intertwined Markov model. We observe the trivial solution $v_{i\infty} = 0$ for all i, which means that eventually, all nodes will be healthy. On the other hand, if $\delta = 0$, then all $v_{i\infty} = 1$, or slightly more precise, (4.6) shows that $v_{i\infty} = 1 - O(\tau^{-1})$ for large τ. Of course, if there is no curing at all ($\delta = 0$), all nodes will eventually be infected almost surely.

Lemma 7. *In a connected graph, either $v_{i\infty} = 0$ for all i nodes, or none of the components $v_{i\infty}$ is zero.*

Proof. If $v_{i\infty} = 0$ for one node i in a connected graph, then it follows from (4.6) that $\sum_{j=1}^{N} a_{ij} v_{j\infty} = 0$ which is only possible provided $v_{j\infty} = 0$ for all neighbors j of node i. Applying this argument repeatedly to the neighbors of neighbors in a connected graph proves the lemma. $\square$

Apart from the exact steady-state $v_{i\infty} = 0$ for all i, the non-linearity gives rise to a second solution, coined as the "metastable state". That second, non-zero solution can be interpreted as the fraction of time that a node is infected while the system is in the "metastable state", i.e. there is a long-lived epidemic.

Theorem 8. *For any effective spreading rate $\tau = \frac{\beta}{\delta} \geq 0$, the non-zero steady-state infection probability of any node i in the N-intertwined model can be expressed as a continued fraction*

$$v_{i\infty} = 1 - \cfrac{1}{1 + \tau d_i - \tau \sum_{j=1}^{N} \cfrac{a_{ij}}{1 + \tau d_j - \tau \sum_{k=1}^{N} \cfrac{a_{jk}}{1 + \tau d_k - \tau \sum_{q=1}^{N} \cfrac{a_{kq}}{1 + \tau d_q - \ddots}}}} \tag{4.7}$$

where $d_i = \sum_{j=1}^{N} a_{ij}$ is the degree of node i. Consequently, the exact steady-state infection probability of any node i is bounded by

$$0 \leq v_{i\infty} \leq 1 - \frac{1}{1 + \tau d_i} \tag{4.8}$$

Proof. We rewrite (4.6) as

$$
\begin{aligned}
v_{i\infty} &= 1 - \frac{1}{1 + \tau \sum_{j=1}^{N} a_{ij} v_{j\infty}} \\
&= 1 - \frac{1}{1 + \tau d_i - \tau \sum_{j=1}^{N} a_{ij}\left(1 - v_{j\infty}\right)} \\
&\leq 1 - \frac{1}{1 + \tau d_i}
\end{aligned}
$$

since $\tau \sum_{j=1}^{N} a_{ij}\left(1 - v_{j\infty}\right) \geq 0$ because $v_{j\infty} \in [0,1]$ for all j. This proves (4.8).

We proceed further by introducing $1 - v_{j\infty} = \frac{1}{1 + \tau \sum_{k=1}^{N} a_{jk} v_{k\infty}}$, such that

$$
\begin{aligned}
v_{i\infty} &= 1 - \frac{1}{1 + \tau d_i - \tau \sum_{j=1}^{N} \frac{a_{ij}}{1 + \tau \sum_{k=1}^{N} a_{jk} v_{k\infty}}} \\
&= 1 - \frac{1}{1 + \tau d_i - \tau \sum_{j=1}^{N} \frac{a_{ij}}{1 + \tau d_j - \tau \sum_{k=1}^{N} a_{jk}(1 - v_{k\infty})}} \\
&\leq 1 - \frac{1}{1 + \tau d_i - \tau \sum_{j=1}^{N} \frac{a_{ij}}{1 + \tau d_j}}
\end{aligned}
$$

This bound improves on (4.8). The third iteration gives

$$v_{i\infty} = 1 - \cfrac{1}{1 + \tau d_i - \tau \sum_{j=1}^{N} \cfrac{a_{ij}}{1 + \tau d_j - \tau \sum_{k=1}^{N} \cfrac{a_{jk}}{1 + \tau d_k - \tau \sum_{q=1}^{N} a_{kq}(1 - v_{q\infty})}}}$$

Ignoring $\sum_{q=1}^{N} a_{kq}(1 - v_{q\infty}) \geq 0$ yields a new upper bound that sharpens the previous upper bound of the second iteration. Each iteration provides a tighter upper bound by putting $\sum_{q=1}^{N} a_{kq}(1 - v_{q\infty}) = 0$ in the deepest fraction. Continuing the process leads to an infinite continued fraction expansion (4.7) for $v_{i\infty}$. $\qquad\square$

The continued fraction stopped at iteration k includes the effect of virus spread up to the $(k-1)$-hop neighbors of node i. As illustrated in Fig. 4.1 (and typical for other graphs that we have simulated), a few iterations in (4.7) already give an accurate approximation. The accuracy seems worst around $\tau = \tau_c$.
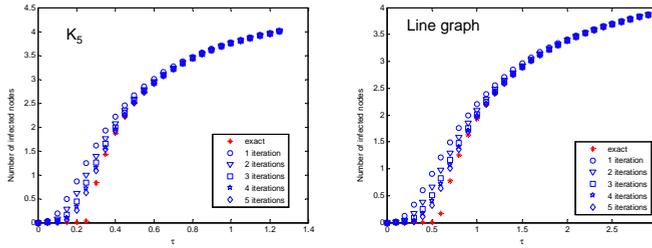
Figure 4.1: *Difference between the exact result and the k-iterations ($1 \leq k \leq 5$) of (4.7) for the complete graph and line graph (both with $N = 5$ nodes) versus the effective infection rate τ.*

Additional insight can be gained from (4.4), which in steady-state reduces to

$$AV_\infty - \text{diag}(v_{i\infty})\left(AV_\infty + \frac{1}{\tau}u\right) = 0$$

Define the vector $w = AV_\infty + \frac{1}{\tau}u$, then

$$w - \frac{1}{\tau}u = \text{diag}(v_{i\infty})w$$

or

$$(I - \text{diag}(v_{i\infty}))w = \frac{1}{\tau}u$$

Ignoring the absence of curing ($\delta = 0$ or $\tau \to \infty$), the bound (4.8) shows that $v_{i\infty}$ cannot be one such that the matrix $(I - \text{diag}(v_{i\infty})) = \text{diag}(1 - v_{i\infty})$ is invertible. Hence,

$$w = \text{diag}\left(\frac{1}{1 - v_{i\infty}}\right)\frac{1}{\tau}u$$

$$= \frac{1}{\tau}\left[\begin{array}{cccc} \frac{1}{1-v_{1\infty}} & \frac{1}{1-v_{2\infty}} & \cdots & \frac{1}{1-v_{N\infty}} \end{array}\right]^T$$

and we end up with the equation

$$\frac{1}{\tau} \begin{bmatrix} \frac{v_{1\infty}}{1-v_{1\infty}} & \frac{v_{2\infty}}{1-v_{2\infty}} & \cdots & \frac{v_{N\infty}}{1-v_{N\infty}} \end{bmatrix}^T = AV_\infty$$

Further, we expand each element as $\frac{v_{i\infty}}{1-v_{i\infty}} = \sum_{k=1}^{\infty} v_{i\infty}^k$, where the geometric series always converges since $v_{i\infty} < 1$.

With the notation $V_\infty^k = \begin{bmatrix} v_{1\infty}^k & v_{2\infty}^k & \cdots & v_{N\infty}^k \end{bmatrix}^T$, we arrive at the steady-state equation

$$\frac{1}{\tau} V_\infty + \frac{1}{\tau} \sum_{k=2}^{\infty} V_\infty^k = AV_\infty \tag{4.9}$$

Lemma 9. *There exists a value $\tau_c = \frac{1}{\lambda_{\max}(A)} > 0$ and for $\tau < \tau_c$, there is only the trivial steady-state solution $V_\infty = 0$. Beside the $V_\infty = 0$ solution, there is a second, non-zero solution for all $\tau > \tau_c$. For $\tau = \tau_c + \varepsilon$ where $\varepsilon > 0$ is an arbitrary small constant, $V_\infty = \varepsilon x$ where x is the eigenvector belonging to the largest eigenvalue of the adjacency matrix A.*

Proof. Theorem 8 shows that the only solution at $\tau = 0$ is the trivial solution $V_\infty = 0$. Let $V_\infty = \varepsilon x$, where $\varepsilon > 0$ is an arbitrary small constant and each component $x_i \geq 0$. Introduced in (4.9) gives, after division by ε,

$$Ax = \frac{1}{\tau} x + \frac{\varepsilon}{\tau} x^2 + O\left(\varepsilon^2\right)$$

For sufficiently small $\varepsilon > 0$, the steady-state equations reduce to the eigenvalue equation

$$Ax = \frac{1}{\tau} x \tag{4.10}$$

which shows that x is an eigenvector of A belonging to the eigenvalue $\frac{1}{\tau}$. Since A is a non-negative matrix, the Perron-Frobenius Theorem [17, p. 451] states that A has a positive largest eigenvalue $\lambda_{\max}(A)$ with a corresponding eigenvector whose elements are all positive and there is only one eigenvector of A with non-negative components. Hence, if $\frac{1}{\tau} = \lambda_{\max}(A) > 0$, then x (and any scaled vector $V_\infty = \varepsilon x$) is the eigenvector of A belonging to $\lambda_{\max}(A)$. If $\tau < \frac{1}{\lambda_{\max}(A)} = \tau_c$, then $\frac{1}{\tau}$ cannot be an eigenvalue of A and the only possible solution is $x = 0$, leading to the trivial solution $V_\infty = 0$. For $\tau > \tau_c$, Theorem 8 provides the non-zero solution of (4.6). $\qquad \square$

Canright *et al.* [39] proposed the eigenvector centrality (EVC) measure of a spreading power of a node

$$e_i = \frac{1}{\lambda_{\max}} \sum_{j=\text{neighbor}(i)} e_j$$

where e_k is the spreading power of a node k. Written in our notation as $v_{i\infty} = \frac{1}{\lambda_{max}} \sum_{j=1}^{N} a_{ij} v_{j\infty}$, the EVC is recognized as the component representation of the eigenvalue equation (4.10) for $\tau = \tau_c$. The steady-state infection probability is the long-run fraction of time during which the node is infected. The higher the probability $v_{i\infty}$, the faster the node i is prone to infection and the more important its role is in further spreading. This Markov steady-state interpretation may explain the term centrality analogously as the betweenness centrality of a node.

In passing by, we note that, by combining Theorem 8 and Lemma 9, a continued fraction expansion of the (scaled) largest eigenvector in any graph is found from (4.7) for $\tau = \tau_c = \frac{1}{\lambda_{max}(A)}$.

Lemma 10. *For any effective spreading rate $\tau = \frac{\beta}{\delta} \geq 0$, the components $v_{i\infty}$ of the steady-state infection probability vector obey*

$$\sum_{i=1}^{N} \left(\frac{1}{1 - v_{i\infty}} - \tau d_i \right) v_{i\infty} = 0 \qquad (4.11)$$

Proof. By summing all rows in (4.9), which is equivalent to multiplication of both sides in (4.9) by the all-one vector u^T yields

$$\frac{1}{\tau} \sum_{i=1}^{N} v_{i\infty} + \frac{1}{\tau} \sum_{k=2}^{\infty} \sum_{i=1}^{N} v_{i\infty}^k = u^T A V_\infty$$

$$= D^T V_\infty = \sum_{i=1}^{N} d_i v_{i\infty}$$

where $D = \begin{bmatrix} d_1 & d_2 & \cdots & d_N \end{bmatrix}^T$ is the degree vector. After rewriting the k-sum, we arrive at (4.11). □

Equation (4.11) is obeyed for the trivial solution $v_{i\infty} = 0$ and, if $v_{i\infty} = 1 - \frac{1}{\tau d_i}$. In the case of regular graphs (where $d_i = d$ for all $1 \leq i \leq N$), both $v_{i\infty} = 0$ and $v_{i\infty} = 1 - \frac{1}{\tau d}$ are exact solutions of (4.6). This shows that, in certain cases, the continued fraction (4.7) can be simplified.

The fraction $y_\infty(\tau) = \frac{1}{N} \sum_{i=1}^{N} v_{i\infty}(\tau)$ of infected nodes in the network, based on the estimate $v_{i\infty} \approx 1 - \frac{1}{\tau d_i}$, is

$$y_\infty(\tau) \approx 1 - \frac{1}{\tau N} \sum_{i=1}^{N} \frac{1}{d_i} \qquad (4.12)$$

Numerical computations in Fig. 4.2 assess the quality of the approximation (4.12).

Lemma 11. *For all i, $v_{i\infty} = 1 - \frac{1}{\tau d_i}$ cannot be a solution of (4.6) for $\tau \leq \frac{1}{d_{(2)}}$ where $d_{(2)} > d_{min}$ is the second smallest degree in the graph G.*
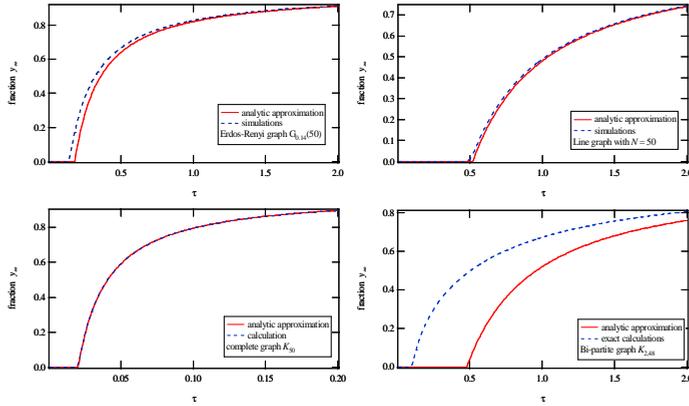
Figure 4.2: *Comparison of (4.12) and exact computations or precise simulations for different type of graphs with $N = 50$ nodes*

Proof. Indeed, $1 - v_{i\infty} = \frac{1}{\tau d_i} \leq \frac{d_{(2)}}{d_i}$ leads for $d_i = d_{\min} < d_{(2)}$ to $v_{i\infty} < 0$, which is impossible. $\qquad\qquad\square$

The strict inequality $d_{(2)} > d_{\min}$ is important. Lemma 11 explains that larger variations in the degree lead to worse results of (4.12) in Fig. 4.2.

Lemma 12. *In a connected graph G with minimum degree $d_{\min}$ and for $\tau \geq \frac{1}{d_{\min}}$, a lower bound of $v_{i\infty}$ for any node i equals*

$$1 - \frac{1}{1 + \frac{d_i}{d_{\min}}(\tau d_{\min} - 1)} \leqslant v_{i\infty} \tag{4.13}$$

Proof. Lemma 7 and Lemma 9 show that, for $\tau > \tau_c$, there exists a non-zero minimum $v_{\min} = \min_{1 \leq i \leq N} \{v_{i\infty}\} > 0$ of steady-state infection probabilities, which obeys (4.6), assuming that this minimum $v_{\min}$ occurs at node i,

$$v_{\min} = 1 - \frac{1}{1 + \tau \sum_{j=1}^{N} a_{ij} v_{j\infty}} \geq 1 - \frac{1}{1 + \tau \sum_{j=1}^{N} a_{ij} v_{\min}}$$

$$= 1 - \frac{1}{1 + \tau d_i v_{\min}} \geq 1 - \frac{1}{1 + \tau d_{\min} v_{\min}}$$

From the last inequality, it can be shown that

$$v_{\min} \geq 1 - \frac{1}{\tau d_{\min}} \tag{4.14}$$

which is only larger than zero provided $\tau > \frac{1}{d_{\min}} \geq \tau_c$. Introducing the bound (4.14), we also have for each node

$$v_i \geq v_{\min} \geq 1 - \frac{1}{1 + \tau d_i v_{\min}} \geqslant 1 - \frac{1}{1 + \frac{d_i}{d_{\min}}(\tau d_{\min} - 1)}$$

which is (4.13). $\qquad\qquad\square$

For $d_{\min} = 1$ the lowest possible lower bound for node i is

$$v_{i\infty} \geqslant 1 - \frac{1}{1 + (\tau - 1)d_i}$$

Finally, by combining the upper bound (4.8) and the lower bound (4.13) for $\tau \geq \frac{1}{d_{\min}}$, we find that $v_{i\infty}$ belongs to the interval

$$1 - \frac{1}{1 + \frac{d_i}{d_{\min}}(\tau d_{\min} - 1)} \leqslant v_{i\infty} \leqslant 1 - \frac{1}{1 + \tau d_i}$$

This shows clearly that for $\tau \to \infty$ variations between all values of v_i for all i will tend to 0.

4.2   Model approximations

At first glance, the averaging process – replacing $q_{1;i}$ in (4.2) by its mean $E[q_{1;i}]$ – seems quite accurate, because a sum S_N of independent indicators (Bernoulli random variables) is close – exactly if all Bernoulli random variables have the same distribution – to a binomial random variable, whose standard deviation $\sigma_{S_N} = \sqrt{\text{Var}[S_N]} = O\left(\sqrt{N}\right)$ is small compared to the mean $E[S_N] = O(N)$. The latter implies that the random variable S_N is closely approximated by its mean[1] for large N.

---

[1]More precisely, the central limit theorem for a sum $S_N = \sum_{j=1}^{N} R_j$ of *independent* random variables $R_1, \ldots, R_N$, each with finite variance $\text{Var}[R_j]$ (and small compared to $\text{Var}[S_N]$) states that, for large N,

$$\Pr\left[\frac{S_N - E[S_N]}{\sqrt{\sum_{j=1}^{N} \text{Var}[R_j]}} \leq x\right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

Applied to *independent* indicators with $\text{Var}\left[1_{\{X_j(t)=1\}}\right] = \Pr[X_j(t) = 1](1 - \Pr[X_j(t)]) \leq E\left[1_{\{X_j(t)=1\}}\right]$ shows that, for $x \geq 0$ and large N,

$$\Pr\left[|S_N - E[S_N]| \geq x\sqrt{E[S_N]}\right] \leq \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{t^2}{2}} dt \overset{\text{large } x}{\approx} \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}$$

where the last step follows after (successive) partial integration and retaining the $O(x^{-1})$ term in the series for large x. Hence, for *independent indicators*, large deviations from the mean are very unlikely.

However, $\frac{q_{1;i}}{\beta} = \sum_{j=\text{neighbor}(i)} 1_{\{X_j(t)=1\}} \in \{0, 1, \ldots, d_i\}$ is a sum of *dependent* indicators. In addition, if N is large, $q_{1;i}$ does not always increase with N. Indeed, $q_{1;i} \leq \beta d_{\max}(A)$ and the maximum degree $d_{\max}(A)$ in a graph can be independent of N, for example, in the line graph where $d_{\max} = 2$ for any N.

We will first elaborate on the dependence. Let us consider the time-dependent random variable $S_i(t) = 1_{\{X_i(t)=1\}}$, which is 1 if node i is infected, else it is zero. If the node i is infected ($X_i(t) = 1$) , $S_i(t)$ can change from 1 to 0 with curing rate δ. If the node i is healthy ($X_i(t) = 0$), $S_i(t)$ can change from 0 to 1 with rate $\beta \sum_{j=1}^{N} a_{ij} 1_{\{X_j(t)=1\}}$. The change of S_i in a sufficiently small time interval Δt is

$$\frac{S_i(t + \Delta t) - S_i(t)}{\Delta t} = (1 - S_i(t))\beta \sum_{j=1}^{N} a_{ij} 1_{\{X_j(t)=1\}} - \delta S_i(t)$$

After taking the expectation of both sides, we obtain
(with $E[S_i(t)] = \Pr[X_i(t) = 1] = v_i(t)$)

$$\frac{v_i(t + \Delta t) - v_i(t)}{\Delta t} = \beta \sum_{j=1}^{N} a_{ij} v_j(t) - \delta v_i(t)$$

$$- E\left[1_{\{X_i(t)=1\}}\beta \sum_{j=1}^{N} a_{ij} 1_{\{X_j(t)=1\}}\right]$$

Since $a_{ii} = 0$, only the case where $j \neq i$ appears in the remaining expectation, which is

$$E\left[1_{\{X_i(t)=1\}} 1_{\{X_j(t)=1\}}\right] = E\left[1_{\{X_i(t)=1\} \cap \{X_j(t)=1\}}\right]$$
$$= \Pr[X_i(t) = 1, X_j(t) = 1]$$
$$= c_{ij}(t) \Pr[X_i(t) = 1]$$

where the conditional probability $c_{ij}(t) = \Pr[X_j(t) = 1 | X_i(t) = 1]$. Hence, when $\Delta t \to 0$, we arrive at

$$\frac{dv_i(t)}{dt} = \beta \sum_{j=1}^{N} a_{ij} v_j(t) - v_i(t)\left(\beta \sum_{j=1}^{N} a_{ij} c_{ij}(t) + \delta\right)$$

Assuming that the graph is connected,

$$\Pr[X_j(t) = 1 | X_i(t) = 1] \geq \Pr[X_j(t) = 1]$$

because a given infection at node i cannot negatively influence the probability of infection at node j. When comparing with (4.3), we observe that the mean field approximation implicitly makes the assumption of independence that $\Pr[X_j(t) = 1, X_k(t) = 1] = \Pr[X_j(t) = 1] \Pr[X_k(t) = 1]$. Hence, the positive correlation is not incorporated appropriately. As a consequence, the rate of change in

$\frac{dv_i(t)}{dt}$ is always overestimated. *The N-intertwined Markov chain thus upperbounds the exact probability $v_i(t)$ of infection.*

Next, we will address the effect on the size N by computing the variance of $q_{1;i}$, $\text{Var}[q_{1;i}] = E[q_{1;i}^2] - (E[q_{1;i}])^2$. First, we have

$$E[q_{1;i}^2] = E\left[\beta\sum_{j=1}^{N}a_{ij}1_{\{X_j(t)=1\}}\beta\sum_{k=1}^{N}a_{ik}1_{\{X_k(t)=1\}}\right]$$

$$= \beta^2\sum_{j=1}^{N}\sum_{k=1}^{N}a_{ik}a_{ij}E\left[1_{\{X_j(t)=1\}}1_{\{X_k(t)=1\}}\right]$$

$$= \beta^2\sum_{j=1}^{N}\sum_{k=1}^{N}a_{ik}a_{ij}\Pr[X_j(t)=1, X_k(t)=1]$$

or, in terms of the conditional probabilities,

$$E[q_{1;i}^2] = \beta^2\sum_{j=1}^{N}\sum_{k=1}^{N}a_{ik}a_{ij}\Pr[X_j(t)=1 \,|\, X_k(t)=1]\Pr[X_k(t)=1]$$

Since $\Pr[X_j(t)=1 \,|\, X_k(t)=1] \leq 1$, an upperbound of $E[q_{1;i}^2]$ is

$$E[q_{1;i}^2] \leq \beta^2 d_i\sum_{k=1}^{N}a_{ik}\Pr[X_k(t)=1] = \max[q_{1;i}]E[q_{1;i}] \qquad (4.15)$$

The variance of $q_{1;i}$ is

$$\text{Var}[q_{1;i}] = \beta^2\sum_{j=1}^{N}a_{ij}\Pr[X_j(t)=1](1-\Pr[X_j(t)=1])$$

$$+ 2\beta^2\sum_{j=1}^{N}\sum_{k=j+1}^{N}a_{ik}a_{ij}(c_{kj}(t)-v_j(t))v_k(t) \qquad (4.16)$$

Since $c_{kj}(t) \geq v_j(t)$ as argued above, the second double sum consists of non-negative terms such that the variance $\text{Var}[q_{1;i}]$ is larger than in the case of independent random variables (where the double sum disappears). This fact is not in favor of the mean field approximation since larger variations around the mean $E[q_{1;i}]$ can occur which makes the mean a less good approximation for the random variable $q_{1;i}$. In particular, (4.16) shows that standard deviation $\sqrt{\text{Var}[q_{1;i}]} = O(N)$, whereas the standard deviation scales as $O(\sqrt{N})$ in case of independence! Especially in graphs with bounded maximum degree (such as the line graph), $\sqrt{\text{Var}[q_{1;i}]}$ may not decrease sufficiently fast in N compared to $E[q_{1;i}]$. Thus, we expect deviations between the N-intertwined and the exact model in those graphs to be largest.

For small τ (and t large enough to ignore the initial conditions),
$\Pr\left[X_k(t) = 1\right] \leq \varepsilon$ and (4.16) shows that the double sum is $O\left(\varepsilon^2\right)$. Hence, for small τ, the situation is close to the independence case, in which mean field theory performs generally well. An upperbound for $\mathrm{Var}[q_{1;i}]$ follows from (4.15) such that the coefficient of variation

$$\frac{\sqrt{\mathrm{Var}\ [q_{1;i}]}}{E\left[q_{1;i}\right]} \leq \sqrt{\frac{\max\left[q_{1;i}\right]}{E\left[q_{1;i}\right]} - 1}$$

This shows, that for large τ where $E\left[q_{1;i}\right] \to \max\left[q_{1;i}\right]$, the coefficient of variation is small, again in favor of the mean field approximation. *Hence, we expect that the deviations between the N-intertwined and the exact model are largest for intermediate values of* τ. As shown in Section 4.5.4, in some τ-region around τ_c, large deviations are indeed found.

The two observations, dependence and absence of a limiting process towards the mean as N increases, complicate a more precise assessment of the averaging process at this point. Since the mean field approximation is the only approximation made, a comparison of the non-linear model (4.4) with the exact 2^N-state solution in Section 4.5.4 further quantifies the effect of the mean field approximation.

Finally, the mean field approximation also excludes information about the joint probability of states,

$$\Pr\left[X_1\left(t\right) = n_1, X_2\left(t\right) = n_2, \ldots, X_N\left(t\right) = n_N\right]$$

where all $n_j \in \{0, 1\}$, as in the 2^N- state Markov chain.

4.3   The time evolution of epidemics

Suppose that all $v_i\left(t\right)$ are sufficiently small to ignore the term $\mathrm{diag}(v_i\left(t\right))\beta A V\left(t\right)$ in (4.5), the time-dependent solution is

$$V\left(t\right) = e^{(\beta A - \delta I)t}V\left(0\right)$$

Since an adjacency matrix has the eigenvalue decomposition $A = U\Lambda U^T$, where $\Lambda = \mathrm{diag}(\lambda_j)$ and $\{\lambda_j\}_{1 \leq j \leq N}$ is the set of eigenvalues of A, and where the orthonormal matrix U has the eigenvectors of A as columnvectors (see e.g. [17, Appendix A]), we obtain

$$B = \beta A - \delta I = U\left(\beta\Lambda - \delta I\right)U^T$$

or $B = U\mathrm{diag}(\beta\lambda_j - \delta)U^T$. Thus,

$$V\left(t\right) = U\mathrm{diag}\left(e^{(\beta\lambda_j - \delta)t}\right)U^T V\left(0\right)$$

and, in order for $V(t)$ to be a probability vector, we must require that all eigenvalues $\beta\lambda_j - \delta \leq 0$ or that $\lambda_j \leq \frac{1}{\tau}$ for all j. This again leads to the requirement that $\tau \leq \frac{1}{\lambda_{\max}(A)}$. The analysis shows that, in the regime $\tau \leq \frac{1}{\lambda_{\max}(A)}$, the probability vector $V(t)$ tends exponentially fast to zero.

Ganesh *et al.* [31, Theorem 3.1] and Durrett [40] have bounded the probability that the virus spread process is not (yet) in the absorbing state as

$$\Pr\left[X(t) \neq 0\right] \leq \sqrt{N \left\|X(0)\right\|_1} \, e^{(\beta\lambda_{\max}(A) - \delta)t}$$

where the norm (see e.g. [17, Section A.3]) $\left\|X(0)\right\|_1 = \sum_{j=1}^N X_j(0)$. Since $\Pr\left[X(t) \neq 0\right]$ is related to $V(t)$ and the largest component of $V(t)$ precisely decays proportionally to $e^{(\beta\lambda_{\max}(A) - \delta)t}$, we may expect that the non-linear N-intertwined model is fairly accurate for $\tau \leq \tau_c = \frac{1}{\lambda_{\max}(A)}$, as also confirmed by simulations presented in Section 4.5.4.

4.4   The role of the spectrum of A

The sum $y(t) = \frac{1}{N}\sum_{i=1}^N v_i(t)$ gives the fraction of infected nodes in the network. Summing (4.3) over all i is equivalent to right multiplication of $V(t)$ by u^T because $\sum_{i=1}^N v_i(t) = u^T V(t)$. Then, we find from (4.5) that

$$\frac{du^T V(t)}{dt} = u^T \left(\text{diag}\left(1 - v_i(t)\right)\beta A - \delta I\right) V(t)$$

$$= \beta \left(u - V(t)\right)^T A V(t) - \delta u^T V(t)$$

Since $u^T A = D^T$ because $A = A^T$, we can write

$$\frac{d}{dt}\left(u^T V(t)\right) = (\beta D - \delta u)^T V - \beta V^T A V \qquad (4.17)$$

Invoking the eigenvalue decomposition $A = U\Lambda U^T$ of the symmetric adjacency matrix leads to

$$\frac{d}{dt}\left(u^T V(t)\right) = (\beta D - \delta u)^T V - \beta \left(U^T V\right)^T \Lambda \left(U^T V\right)$$

$$= (\beta D - \delta u)^T V - \beta \sum_{j=1}^N \lambda_j(A) z_j^2 \qquad (4.18)$$

where z_j is the j-th component of the vector $U^T V$: the scalar product $V.x_j$ or the projection of the vector V onto the j-th eigenvector x_j of A. We have that $0 \leq V^T A V = \sum_{j=1}^N \lambda_j(A) z_j^2$.

Equation (4.18) shows that the zero eigenvalues in the adjacency matrix of a graph do not contribute to the infected fraction $y(t) = \frac{u^T V}{N}$ of nodes. In general,

a matrix has a zero eigenvalue if its determinant is zero. A determinant is zero if two rows are identical or if some of the rows are linearly dependent. For example, two rows are identical if two distinct nodes are connected to a same set of nodes. Since the elements a_{ij} of an adjacency matrix A are only 0 or 1, linear dependence of rows here occurs every time the sum of a set of rows equals another row in the adjacency matrix. For example, consider the sum of two rows. If n_1 is connected to the set S_1 of nodes and n_2 is connected to the distinct set S_2, where $S_1 \cap S_2 = \varnothing$ and $n_1 \neq n_2$, then the graph has a zero eigenvalue if another node $n_3 \neq n_2 \neq n_1$ is connected to $S_1 \cup S_2$. These zero eigenvalues occur when a graph possesses a "local bi-partiteness". In real networks, this type of interconnection often occurs.

Lemma 13. *For any effective spreading rate* $\tau = \frac{\beta}{\delta} \geq 0$, *the components* $v_{i\infty}$ *of the steady-state infection probability vector obey*

$$\sum_{i=1}^{N} \left(d_i - \frac{1}{\tau} \right) v_{i\infty} = \sum_{j=1}^{N} \lambda_j (A) z_{j\infty}^2 \qquad (4.19)$$

from which

$$0 \leq \sum_{j=1}^{N} \lambda_j (A) z_{j\infty}^2 \leq \sum_{i=1}^{N} \frac{|\tau d_i - 1|}{\tau d_i + 1} d_i \leq 2L$$

where L *is the number of links.*

Proof. The equality (4.19) is an immediate consequence of (4.18). The first upper bound follows from (4.8). The second one from the basic equation of the degree $\sum_{i=1}^{N} d_i = 2L$. □

Since $\sum_{j=1}^{N} \lambda_j (A) = 0$ for any graph, the lower bound in Lemma 13 shows that the positive eigenvalues and their eigenvectors are more important than the negative ones. Because the left hand side of (4.19) is increasing in τ, the vector V_∞ is increasingly more aligned with eigenvectors of A belonging to positive eigenvalues. Lemma 9 shows that at $\tau = \tau_c + \varepsilon$, only the eigenvector of $\lambda_{\max} (A)$ plays a role. As τ increases, we now deduce that V_∞ is influenced by additional eigenvectors (proportional to $\lambda_j (A)$). The contribution of the eigenvector of $\lambda_{\max} (A)$ to $\sum_{j=1}^{N} \lambda_j (A) z_{j\infty}^2$ remains dominant, because it is the only eigenvector with all positive components and all eigenvectors in U are normalized, i.e. $x_j^T x_j = 1$. By combining (4.19) and (4.11), we have

$$\sum_{j=1}^{N} \lambda_j (A) z_{j\infty}^2 = \frac{1}{\tau} \sum_{i=1}^{N} \frac{v_{i\infty}^2}{1 - v_{i\infty}}$$

4.5 The N-intertwined model and other models

In this section, we will show that, by making additional approximations, our model can reproduce the differential equation (2.1) of the Kephart and White

model. Further, we present average cluster model that illustrate the averaging bias introduced by the application of the logistic equation for topology epidemics. Finally, we compare N-intertwined model with exact model for the reference.

4.5.1 Evaluation of the Kephart and White model

In a regular graph with degree k and adjacency matrix A_R, the degree vector $D = ku$ and the eigenvector belonging to the largest eigenvalue $\lambda_{\max}(A_R) = k$ is u such that (4.18) becomes

$$\frac{d}{dt}\left(u^T V(t)\right) = (\beta k - \delta)\, u^T V - \beta k \left(u^T V\right)^2 - \beta \sum_{j=2}^{N} \lambda_j(A_R)\, z_j^2$$

If we let $y(t) = \frac{u^T V}{N}$ and assume in the last sum that all eigenvalues and vectors are equal to the largest one, we again find the Kephart and White differential equation (2.1). Clearly, apart from the mean field approximation and the confinement to regular graphs (or nearly regular graphs), the Kephart and White model approximates the eigenvalue structure of a regular graph and only the largest eigenvalue and eigenvector are considered. Since $\sum_{j=1}^{N} \lambda_j(A_R) = 0$ implying that a non-negligible fraction of the eigenvalues are negative, the Kephart and White derivative $\frac{dy(t)}{dt}$ underestimates the actual rate of infection in the regular graph. Most likely, this underestimation is a general property of "homogeneous" virus spread models. A similar comment holds for the extended local models proposed by Pastor Satorras and Vespignani [22, Chapter 9].

For the simplest regular graph, the complete graph K_N, we observe that the equation (4.3) for each node i is identical. Thus, one might be led to put, $v_i = v_k$, for all $1 \le k \le N$ and for all t and such that $\sum_{j=1}^{N} a_{ij} v_j(t) = (N-1) v_i(t)$. In that case, the set of equations (4.4) reduces to a single equation,

$$\frac{dv_i(t)}{dt} = \beta(N-1)\, v_i(t)\left(1 - v_i(t)\right) - \delta v_i(t)$$

which is the Kephart and White differential equation (2.1). Although apparently correct, the assumption that $v_j = v_k$ (for all t) implies that all initial conditions also are the same. That full symmetry reduces the modeling of the network to that of a single node. Also, that local view of the single node is equivalent to ignoring all, but the largest eigenvalue in (4.18). In random attack strategies of computer viruses, where each node has equal probability to be infected initially, the full symmetry $v_i(0) = v_j(0)$ for any pair of nodes i and j is achieved.

4.5.2 Average cluster model and isoperimetric constant

As presented in Chapter 2, section 2.2.5, Ganesh *et al.* in [31] determine the two thresholds τ_c and τ_c^*. The first threshold, τ_c was examined in details in [41] as

well as in [3] and [32]. Below the first one, the lifetime of the epidemic is of the order $O(log(N))$, where N is the number of nodes in the network [31] and it is determined as reciprocal of the largest eigenvalue $\lambda_{\max}$ of the adjacency matrix i.e. $\tau_c = \frac{1}{\lambda_{\max}}$. Above the second threshold τ_c^*, the lifetime of the metastable state is of the order[2] $\Omega(e^{N^\alpha})$, where τ_c^* is such an effective spreading rate for which the infected fraction in the metastable state is $y_\infty(\tau) = O(N^{\alpha-1})$, for $\alpha > 0$ [31].

Let us divide an infected graph into two sets: a set $S(t)$ of infected nodes at time t and its complement $S^c(t)$, the set of non-infected nodes. The number of edges between the two clusters at time t is denoted as $e(S(t), S^c(t))$ and the number of nodes in each set as $|S(t)|, |S^c(t)|$. The birth of new infections is a sum of Poisson processes on each link connecting infected and non-infected clusters. The rate at which new infected nodes appear is thus $\beta e(S(t), S^c(t))$. On the other hand, the curing process is a sum of Poisson processes on all the infected nodes which leads to $\delta |S(t)|$. For large N, the increase/decrease of the fraction of infected nodes depends on these two processes and can be written as

$$N\frac{dy(t)}{dt} = \beta e(S(t), S^c(t)) - \delta |S(t)|$$

where $Ny(t)$ is the total number of infected nodes in the graph.

During the metastable state, individual nodes change their states from infected to susceptible and back, while the clusters $S(t)$ and $S^c(t)$ are constantly changing. When $\frac{dy(t)}{dt}$ is equal to zero, then the number of new nodes that are infected and the infected ones that are cured is equal, but the clusters may still be changing. For $\frac{dy(t)}{dt} = 0$, we find

$$\frac{e(S_\infty, S_\infty^c)}{|S_\infty|} = \frac{1}{\tau} \tag{4.20}$$

Because the infected cluster changes, it is possible to enclose the infection in a cluster that is poorly connected to the rest of the graph. In this case, the number of nodes in the infected cluster will decrease and not enough new nodes will be infected, resulting in infection extinction.

Let us define the minimum of the ratio (4.20) as the generalized isoperimetric constant for different maximal sizes of the infected cluster m,

$$\eta(G, m) = \inf_{S \subset \{1, .., n\}, |S| \leq m} \frac{e(S, S^c)}{|S|}, \qquad 0 < m \leqslant \lfloor N/2 \rfloor$$

Bounding the epidemic process with this worst case scenario leads to the second epidemic threshold 2.4.

Another important result [42] is the relation between the second smallest eigenvalue $\mu_{N-1}(Q)$ of the Laplacian matrix Q of a graph G to the standard isoperi-

---

[2]Ω is a lower asymptotic bound with definition: for a given real function $g(x)$, we denote $\Omega(g(x))$ as the set of functions such that $\Omega(g(x)) = \{f(x), \exists x_0, \exists c > 0 | 0 \leq cg(x) \leq f(x), \forall x \in , x > x_0\}$.

metric constant $\eta(G) = \eta(G, \lfloor \frac{N}{2} \rfloor)$

$$\eta(G) \geqslant \frac{\mu_{N-1}(Q)}{2} \tag{4.21}$$

The average infected cluster of a graph

Statistically, infection of some clusters is more probable than of others. Some nodes have higher probability of being infected because they are more connected or they are part of a highly connected cluster. Some examples are hubs in a star networks, or nodes in the middle of a lattice. The N-intertwined model takes into account interactions between pair of nodes and the position of a node in the network, but is not able to describe interactions of clusters. The ratio $\frac{e(S,S^c)}{|S|}$ is the best way to describe cluster interactions, but it is not a simple parameter to calculate. In this section, we show that averaging over the cluster connectivity leads to the logistic equation [1].

Not all the clusters are equally probable, but in the first approximation we will disregard this fact. For sufficiently large N, we can calculate the average increment $\frac{d\bar{y}(t)}{dt}$ in the number of infected nodes for a given cluster size of m nodes as a sum over all the clusters divided by the total number of combinations $\binom{N}{m}$.

$$N\frac{d\bar{y}(t)}{dt} = \frac{\beta}{\binom{N}{m}} \sum_{(\forall S)S \subset \{1,..,N\},|S|=m} e\left(S(t), S^c(t)\right) - \frac{\delta}{\binom{N}{m}} \sum_{(\forall S)S \subset \{1,..,N\},|S|=m} |S(t)|$$

For $\frac{d\bar{y}(t)}{dt} = 0$, we have

$$\tau = \frac{1}{\frac{1}{\binom{N}{m}} \sum_{(\forall S)S \subset \{1,..,N\},|S|=m} \frac{e(S(t),S^c(t))}{m}} \tag{4.22}$$

We will define the **average interconnection constant** of a graph as

$$\bar{\eta}(G, m) = \frac{1}{\binom{N}{m}} \sum_{(\forall S)S \subset \{1,..,N\},|S|=m} \frac{e(S, S^c)}{m}$$

Simply stated, we have averaged the ratio $\frac{e(S,S^c)}{|S|}$ over all combinations of clusters with m nodes, taking any of them as equiprobable and divided by the total number of combinations of m-sized clusters from the network with N nodes.

Lemma 14. *The average interconnection constant of a graph G obeys*

$$\bar{\eta}(G, m) = d_{av}(1 - \frac{m-1}{N-1}) \tag{4.23}$$

Proof. We denote the degree of a node i by d_i and the number of links from node i to the nodes inside set S as h_i. From the definition of the averaged cluster constant, we have

$$\bar{\eta}(G, m) = \frac{1}{\binom{N}{m}} \sum_{(\forall S)S=(i_1, i_2, \ldots i_m), S \subset \{1, \ldots, N\}, |S|=m} \frac{\sum_{p=1}^{m} (d_{i_p} - h_{i_p})}{m}$$

The number of clusters, in which a node i appears, is equal to $\binom{N-1}{m-1}$. Counting the number of combinations for different values of $h_{i_p} = 0, 1, 2 \ldots m$ we have

$$\bar{\eta}(G, m) = \frac{\binom{N-1}{m-1} \sum_{i=1}^{N} d_i}{m \binom{N}{m}} - \frac{1}{\binom{N}{m}} \sum_{(\forall S)S=(i_1, i_2, \ldots i_m), S \subset \{1, \ldots, N\}, |S|=m} \frac{\sum_{p=1}^{m} h_{i_p}}{m}$$

In order to count the number of clusters in which a node i appears with $h_{i_p} = k$ of its neighbors, we calculate all the combinations without a node i and all of its neighbors, which is $\binom{N-d_i-1}{m-k-1}$ and multiply that with all the combinations of a node i with k neighbors, which is $\binom{d_i}{k}$. If a node i has large degree, it is not always possible to include any number of its neighbors in the cluster S. The minimal number of neighbors k that can be in a cluster S is $k = \max(0, m + d - N)$ and the maximal k is $k = \min(d_i, m)$. Hence,

$$\bar{\eta}(G, m) = d_{av} - \frac{\sum_{i=1}^{N} \sum_{k=\max(0,m+d-N)}^{\min(d_i,m)} \binom{N-d_i-1}{m-k-1}\binom{d_i}{k}}{m\binom{N}{m}}$$

We can simplify the double sum and find (4.23). $\square$

In other words, the average interconnection constant $\bar{\eta}(G, m)$ is equal to the average degree of a node d_{av} excluding connections inside a cluster $\frac{m-1}{N-1} d_{av}$.

Assuming $\frac{d\tilde{y}(t)}{dt} = 0$ and using (4.22) and (4.23) for any graph, leads to

$$\tau = \frac{1}{d_{av}(1 - \frac{m-1}{N-1})}$$

from which follows the estimation of the average number of infected nodes in the graph as a function of τ:

$$m = 1 + (N-1)(1 - \frac{1}{\tau d_{av}}) \tag{4.24}$$

This result is equivalent to the Kephart and White result for k-regular ($k = d_{av}$) graphs [2]. The epidemic threshold (for $m = 0$) is

$$\tau_c = \frac{1}{d_{av}(1 + \frac{1}{N-1})} \approx \frac{1}{d_{av}}$$

The largest eigenvalue of an adjacency matrix is always larger then or equal to the average degree which implies $\frac{1}{d_{av}} \geqslant \frac{1}{\lambda_{\max}}$ [17] such that the threshold of the Kephart and White model is overestimating the real epidemic.

It is interesting to observe that not all the clusters have the same probability to appear as an infected cluster. If the probability distribution of infected clusters would be available, a more precise description of an epidemic process in a network could be derived. In Figure 4.3, the comparison between the N-intertwined model and the eq. (4.24) is given. The largest discrepancy can be observed for a power law graph.
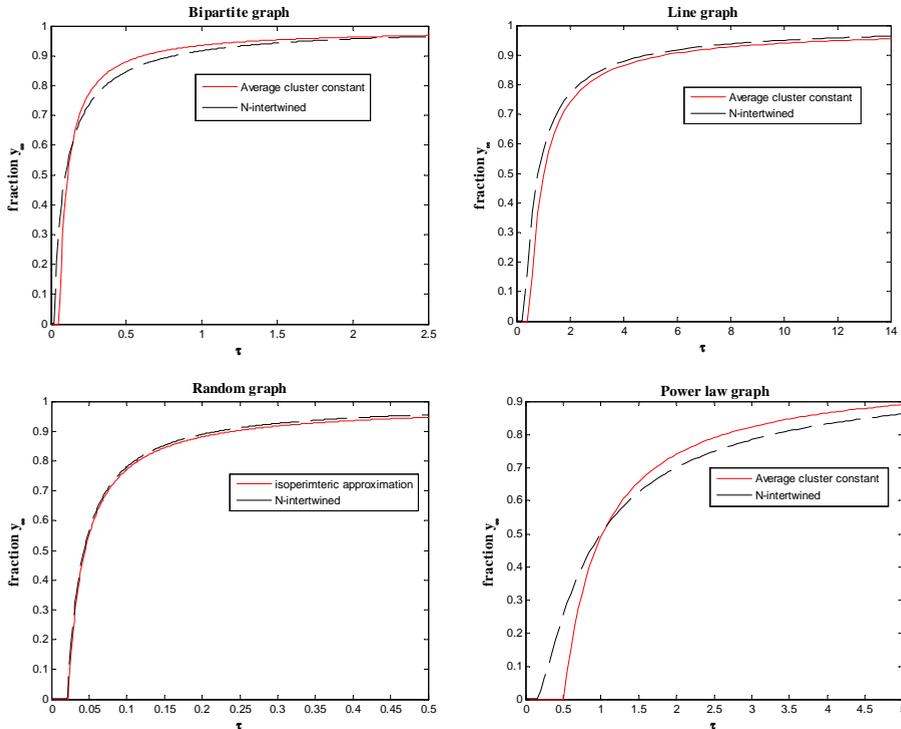

Figure 4.3: *The fraction of infected nodes as a function of τ. Comparison between the model given by eq. (4.24) and the N-intertwined model. All graphs consist of $N = 100$ nodes.*

4.5.3 The model of Wang *et al.*

In section 2.2.6, we have presented model of Wang *et al.*. In this section, we will compare the steady state of Wang model with the N-intertwined model. To help reader follow the arguments we will repeat the steady state equations of Wang *et al.* 2.5 and the N-intertwined model 4.6.

$$p_i = \frac{1 - \prod \left(1 - \beta_w p_{j,t-1}\right)}{1 - \left(1 - \frac{1}{2}\delta_w\right)\prod \left(1 - \beta_w p_{j,t-1}\right) + \frac{1}{2}\delta_w}$$

$$v_{i\infty} = \frac{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty}}{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty} + \delta}$$

Note that in the first equation (4.5.3) β and δ are probabilities, while in (4.5.3) they are rates. In order to distinguish between probability of transmission and rate of transmission, we denoted probability with β_w and δ_w.

The main difference in these two models lies in different application regions. Wang *et al.* is discrete model and can be used only for synchronous systems where there is specific time unit during which the infection will or will not be transmitted. In that sense, Wang *et al.* model tends to N-intertwined model in the limit $\beta \to 0$ and $\delta \to 0$ with $\frac{\beta_w}{\delta_w} = \tau$. If the process that is being modeled is in continuous time, Wang *et al.* model introduces an error which increases with increase of β_w and δ_w.

Further more, steady state solution in Wang *et al.* model does not depend only on ratio $\tau_w = \frac{\beta_w}{\delta_w}$, but also on a polynomial of β_w. Although completely natural for discrete time, this effect does not exist in continuous time. Therefore in the discrete model, the steady state is different for the same τ_w and different β_w.

The Wang *et al.* model as presented in [32] has one serious mistake. To cite: *[...] we assume that the probability that a curing event at node i takes place after infection from neighbors is roughly* 50%. First of all, it is not possible to have before and after in discrete time step and secon estimation of roughly 50% is not rigorous analytical modeling step.

Wang *et al.* and N-intertwined model are created with different systems in mined, namely discrete and continuous. While Wang *et al.* model can not be used for continuous time epidemic spreading, N-intertwined model is not appropriate for discrete time, sinhronous, epidemic spreading.

4.5.4 The exact 2^N Markov chain

Via simulations, we assess the accuracy of the N-intertwined Markov chain. Only small networks are simulated because we expect for small N the largest error. Fig. 4.4,4.5 and 4.6 present a typically view of the fraction $y(t)$ as function of time t in K_{11} for three different τ-regimes.
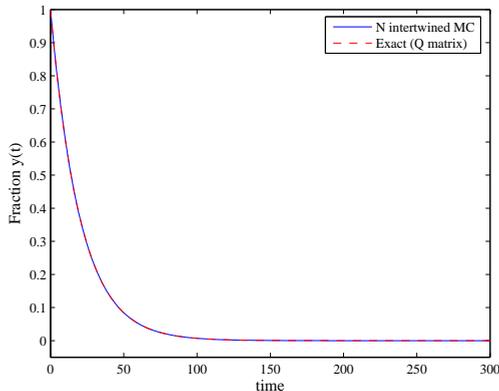
Figure 4.4: *The fraction $y(t)$ of infected nodes in K_{11} where $\tau = 10^{-3}$ as a function time computed exactly (via the Q-matrix) and with the N intertwined Markov chain model.*

Below the epidemic threshold $\tau_c = \frac{1}{10}$ ($N = 11$ in Fig. 4.4), the N-intertwined non-linear model is almost exact.

In a τ-region round τ_c, Fig. 4.5 illustrates that the deviations from the exact solution are substantial. But, sufficiently above τ_c as in Fig. 4.6, the accuracy of the N-intertwined non-linear model again improves. Since the N-intertwined non-linear model upperbounds the fraction of infected nodes as shown in Section 4.2, the relative small difference in Fig. 4.6 quantifies the effect of neglecting dependence in the mean field approximation.

In summary, for all graphs, if $\tau < \tau_c$, the N-intertwined Markov chain is very accurate. If $\tau > \tau_c$, the N-intertwined Markov chain differs from the exact solution, but the difference decreases with increasing network size N. The fact that the non-linear N-intertwined model and the exact 2^N-state Markov chain are close for large N is linked with a general property of Markov chains: A Markov chain can approximate any stochastic process arbitrarily close provided the number of states in the Markov chains is sufficiently large.

4.6 Special case - complete bi-partite graph

In this section we will consider complete bi-partite graphs. A complete bi-partite graph $K_{M,N}$ consists of two disjoint sets S_1 and S_2 containing respectively M and N nodes, such that all nodes in S_1 are connected to all nodes in S_2, while within each set no connections occur. Figure 4.7 gives an example of a complete bi-partite graph on 6 nodes.
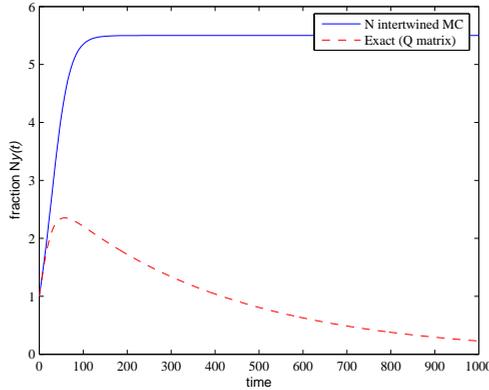
Figure 4.5: *The $N = 11$ times the fraction $y(t)$ of infected nodes in K_{11} where $\tau = 0.2 = 2\tau_c$ as a function time computed exactly (via the Q-matrix) and with the N intertwined Markov chain model.*

Notice that (core) telecommunication networks often can be modeled as a complete bi-partite topology. For instance, the so-called double-star topology (i.e. $K_{M,N}$ with $M = 2$) is quite commonly used because it offers a high level of robustness against link failures. For example, the Amsterdam Internet Exchange (see www.ams-ix.net), one of the largest public Internet exchanges in the world, uses this topology to connect its four locations in Amsterdam to two high-density Ethernet switches. Sensor networks are also often designed as complete bi-partite graphs.

Due to the symmetry, for complete bi-partite graph the set of N equations for N-intertwined model reduces to only two [43], [41].

$$v_N = \frac{\tau^2 MN - 1}{N\tau(M\tau + 1)}; \; v_M = \frac{\tau^2 MN - 1}{M\tau(N\tau + 1)} \tag{4.25}$$

Because an epidemic steady state only exists if $V_{N\infty} > 0$ or $V_{M\infty} > 0$, Eq. (4.25) yields the epidemic threshold:

$$\tau = \frac{1}{\sqrt{MN}}. \tag{4.26}$$

This complies with [32] because according to [44] the spectral radius of the adjacency matrix of the graph $K_{M,N}$ is equal to $\sqrt{MN}$.

For effective spreading rates above the epidemic threshold the epidemic steady state y_∞ for the complete bi-partite graph $K_{M,N}$ satisfies

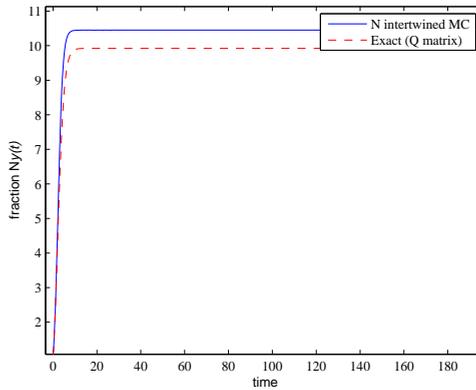$$y_\infty = \frac{Mv_{M\infty} + Nv_{N\infty}}{M + N}. \tag{4.27}$$

Figure 4.6: *The N times the fraction $y(t)$ of infected nodes in K_{11} where $\tau = 2$ as a function time computed exactly (via the Q-matrix) and with the N intertwined Markov chain model.*
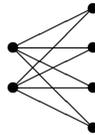
Figure 4.7: *Complete bi-partite graph $K_{2,4}$*

Substitution of Eq. (4.25) into Eq. (4.27) yields

$$y_\infty = \frac{(MN\beta^2 - \delta^2)(\beta N + \beta M + 2\delta)}{\beta(M + N)(\beta M + \delta)(\beta N + \delta)}. \tag{4.28}$$

The epidemic spreading is a stochastic process, and in the steady state, the system is taking a set of values around the mean epidemic steady state y_∞, see also Figures 4.9-4.8. Because the steady state probability of a node being infected does not depend on other nodes the steady state probability $\Pr[I_N, I_M]$ satisfies:

$$\Pr[I_N = x, I_M = w] = \binom{N}{x} v_{N\infty}^x (1 - v_{N\infty})^{N-x} \binom{M}{w} v_{M\infty}^w (1 - v_{M\infty})^{M-w} \tag{4.29}$$

4.6.1 The impact of infection delay

So far we have assumed that once a node is infected, it instantaneously becomes infectious. In reality, there may be a time lag between the arrival of a virus at

a node and the time this node itself starts to spread the virus. A virus could lie dormant on a host due to user inactivity or because the virus was designed in this manner for stealth reasons.

In [45], Wang and Wang have studied the impact of infection delay on the epidemic threshold and the epidemic steady state for regular graphs. In [45] the infection delay ϵ is defined as the length of time between the virus arrival at a node and the instant the node becomes infectious.

It is shown in [45] that the steady state for the fraction of infectious nodes satisfies

$$v_{N\infty} = \frac{\beta k - \delta e^{\delta\epsilon}}{\beta k},$$ (4.30)

which yields for the epidemic threshold:

$$\tau = \frac{e^{\delta\epsilon}}{k}.$$ (4.31)

Thus, the infection delay increases the epidemic threshold, which means that infection delay makes an epidemic die out more easily. Let us examine the impact of infection delay on virus spread on complete bi-partite graphs.

We can derive the following delay-differential equation for the evolution of $v_N(t)$, which as before, denotes the fraction of infected nodes in S_2 at time t:

$$\frac{dv_N(t)}{dt} = \beta M v_M(t - \epsilon)e^{-\delta\epsilon}(1 - v_N(t)) - \delta v_M(t),$$ (4.32)

where $v_M(t - \epsilon) = 0$ for $t < \epsilon$ and $v_M(t)$ denotes the fractions of nodes in S_1 that is infectious at time t. For $t \geq \epsilon$, the probability that a node in S_1 is infectious is the probability that the node was already infected at time $t - \epsilon$, since all nodes infected between $t - \epsilon$ and t are still being delayed. Curing a node during the infection delay period ϵ results in the $e^{-\delta\epsilon}$ factor.

Let us denote the steady state of Eq. (4.32) as $v_{N\infty}$. We solve for $v_{N\infty}$ by setting the right hand side of Eq. (4.32) equal to zero and $v_M(t - \epsilon) = v_{M\infty}$. Analogous to Eq. (4.30) we find for $v_{M\infty}$

$$v_{M\infty} = \frac{\beta N v_{N\infty} e^{-\delta\epsilon}}{\beta N v_{N\infty} e^{-\delta\epsilon} + \delta},$$ (4.33)

where the $e^{-\delta\epsilon}$ factor corresponds with the probability that a node is cured during the infection delay period ϵ.

Plugging Eq. (4.25) and $v_N = v_{N\infty}$ into Eq. (4.32) and solving the right hand side with respect to $v_{N\infty}$ we obtain the steady state solution for the fraction of infected nodes in S_2:

$$v_{N\infty} = \frac{MN\beta^2 - \delta^2 e^{2\delta\epsilon}}{\beta N(\beta M + \delta e^{\delta\epsilon})},$$ (4.34)

which yields for the epidemic threshold:

$$\tau = \frac{e^{\delta\epsilon}}{\sqrt{MN}}. \tag{4.35}$$

Analogous to the previous section it can be shown that for effective spreading rates above the epidemic threshold the epidemic steady state y_∞ for the complete bi-partite graph $B_{M,N}$ with infection delay ϵ satisfies

$$y_\infty = \frac{(MN\beta^2 - \delta^2 e^{2\delta\epsilon})(\beta N + \beta M + 2\delta e^{\delta\epsilon})}{\beta(M+N)(\beta M + \delta e^{\delta\epsilon})(\beta N + \delta e^{\delta\epsilon})}. \tag{4.36}$$

4.6.2 Simulation analysis

Virus spread without infection delay

In this section, we present a set of simulation results that will validate the mean field model for complete bi-partite graphs proposed in the previous section. We have considered complete bipartite graphs $K_{M,N}$ with $\{M = 10, N = 990\}, \{M = 500, N = 500\}$. Note that for $K_{10,990}$ and $K_{500,500}$ the epidemic threshold is $\tau_c = 0.0101$ and $\tau_c = 0.002$, respectively. The virus spread is a stochastic process, and it can be expected that during evolution some of the infections die out before reaching the steady state even though the effective spreading rate is above the threshold. These evolutions have been excluded from calculations of the expected number of infected nodes in the steady state.
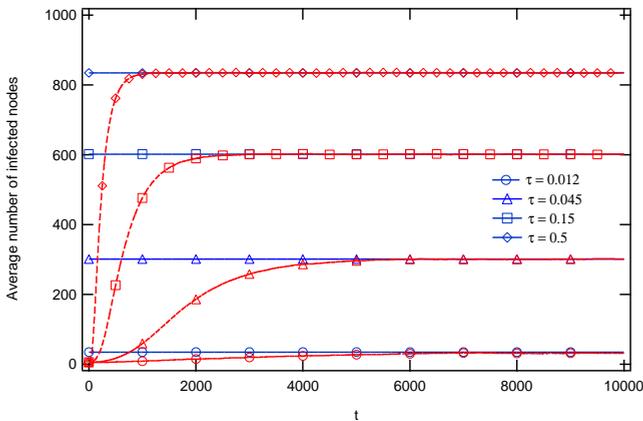

Figure 4.8: *Average number of infected nodes for $K_{10,990}$, excluding virus epidemics that died out.*

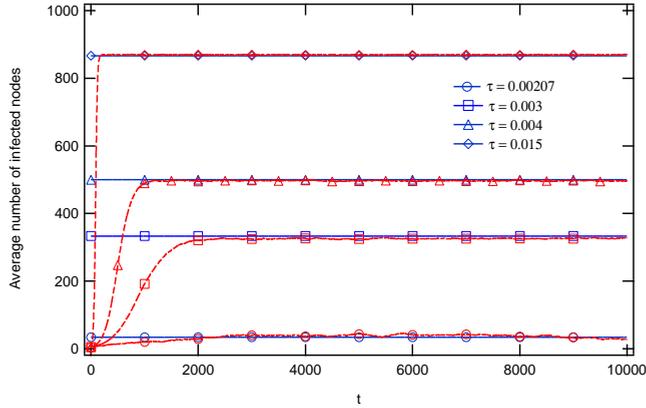Figures 4.8 and 4.9 show the average number of infected nodes for different values of τ.

Figure 4.9: *Average number of infected nodes for $K_{500,500}$, excluding virus epidemics that died out.*

Figures 4.10 and 4.11 show theoretical and simulated values for the mean number of infected nodes in steady state. Again, realizations of the system in which the virus died out during evolution are excluded in calculating the average. Simulation results also showed that below the threshold the virus dies out.

Figure 4.12 shows the standard deviation σ as a function of the effective spreading rate τ.

Note that our model, which is based upon mean field theory, fails to explain extinction of the virus before the steady state is reached for effective spreading rates above the threshold. We will deal with these issues in subsequent sections.

The impact of infection delay

We have conducted simulations for different values of the effective spreading rate $\tau = \frac{\beta}{\delta}$ on a complete bipartite graph $K_{M,N}$ with $\{M = 250, N = 750\}$ and for two values of the infection delay $\varepsilon \in \{10, 50\}$. Again the evolutions that died out are excluded in calculating the average number of infected nodes. Figures 4.13 and 4.14 show that our approximation Eq. (4.36) predicts the steady state well for the virus spread with infection delay.

Steady state probability distribution

We conducted simulations for the complete bi-partite graph $K_{M,N}$ with $M = 10, N = 990$ with the effective spreading rate $\tau \in \{0.045, 0.15, 0.5\}$. Note that the epidemic threshold for this case is $\tau_c = 0.0101$. Figure 4.15 also contains the probabilities that the virus dies out during system evolution.

We conclude from the simulation that Eq. (4.29) predicts the probability distribution of the number of infected nodes in steady state very well for large
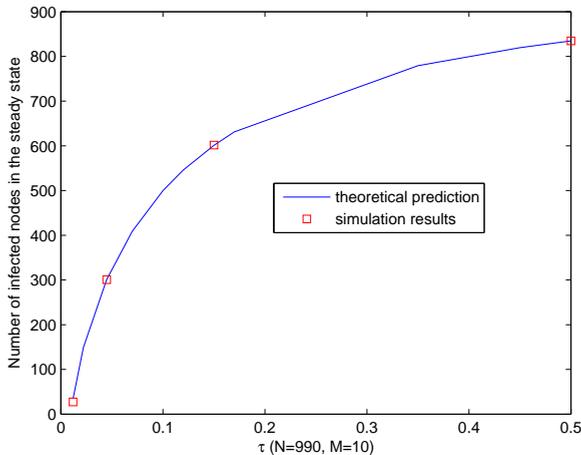
Figure 4.10: *Number of infected nodes in the steady state for $K_{10,990}$*

values of the effective spreading rate τ. For values of τ just over the threshold our model is less accurate in predicting the probability distribution. This confirms the result of section 4.2 that the N-intertwined model exhibits the largest deviation around $\tau = \tau_c$.

4.6.3   Extinction probability

Eventually, every epidemic on a finite population will die out. However, for effective spreading rates above the epidemic threshold, this will take an extremely long time in general [31].

   We approximate the probability that a virus dies out before reaching the steady state p_{ext} by the probability that all initially infected nodes are cured before they infect any other node. We initially infect N_0 nodes in the larger group of nodes S_2 (consisting of N nodes). Then p_{ext} equals the probability that all N_0 nodes are cured before they infect any of the susceptible M nodes to which they are attached, see Figure 4.16, where full and open circles denote infected and susceptible nodes, respectively. Let us first determine the probability p_M that one specific node will be cured before it has infected any of the susceptible M nodes, before time T. It is assumed that the infection process (over a link) and the node curing process are independent Poisson processes with rates β and δ, respectively. Furthermore, let T_β be a random variable that denotes the time it takes for a susceptible node to become infected over a link and T_δ denotes the time it takes for a node to cure. For the latter random variable, let $f_{T_\delta}(x)$ denote its corresponding probability density function. Suppose the infected node is cured at time x, with $0 \leq x \leq T$. This implies, that for all M susceptible nodes attached
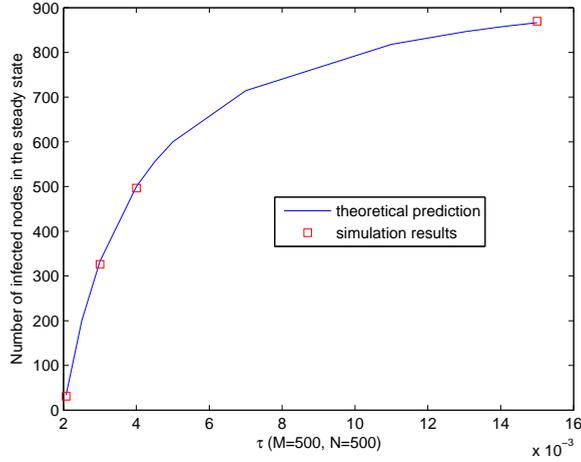
Figure 4.11: *Number of infected nodes in the steady state for $K_{500,500}$*

to the infected node, we require $T_\beta > x$. Applying the law of total probability we obtain:

$$p_M = \int_0^T [\Pr[T_\beta > x | T_\delta = x]]^M f_{T_\delta}(x) dx$$

$$= \int_0^T (e^{-\beta x})^M \delta e^{-\delta x} dx$$

$$= \frac{\delta}{\delta + M\beta}(1 - e^{-(\delta + M\beta)T}).$$

Because the curing processes of the N_0 infected nodes are independent, in order to obtain p_{ext}, we have to multiply the probabilities of each of them being cured before they infect other nodes, which leads to:

$$p_{ext} = \left(\frac{\delta}{\delta + M\beta}(1 - e^{-(\delta + M\beta)T}) \right)^{N_0} \tag{4.37}$$

In order to estimate how well Eq. (4.37) predicts extinction of a virus spread in the first phase, we have conducted simulations on the complete bi-partite graph $K_{M,N}$ with parameters $\{M = 10, N = 990, \tau = 0.045\}$. Figure 4.17 shows the probability of extinction evolving in time for the case of three initially infected nodes ($N_0 = 3$).
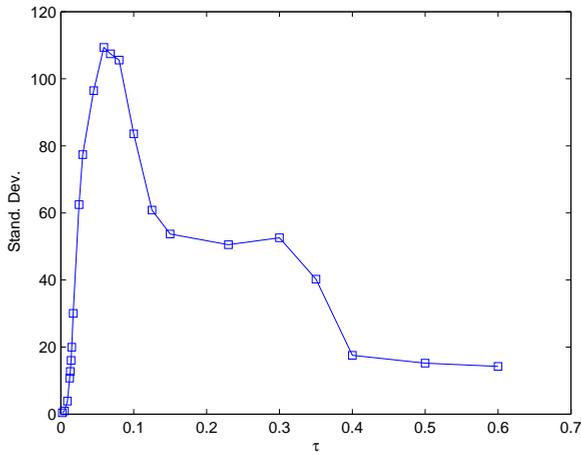
Figure 4.12: *Standard deviation around the steady state for $K_{10,990}$*

Figure 4.18 depicts p_{ext} for the number of initially infected nodes between 1 and 8. For a given graph, there is high probability that initially infected nodes are in the larger less connected set of nodes. If nodes in smaller well connected set are infected significantly different results can be expected.
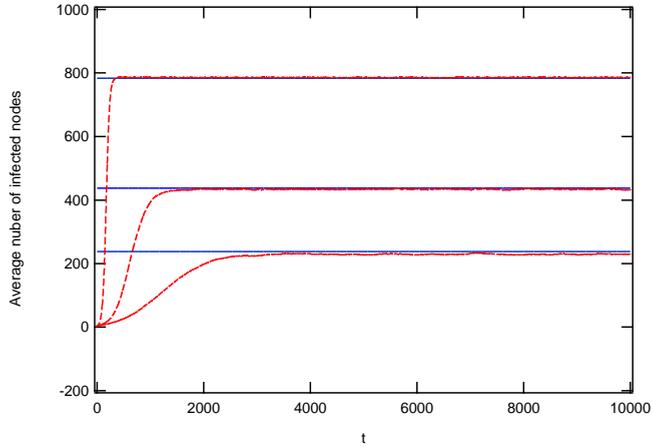
Figure 4.13: *Average number of infected nodes for $K_{250,750}$ with infection delay $\varepsilon = 10$, excluding virus epidemics that died out.*
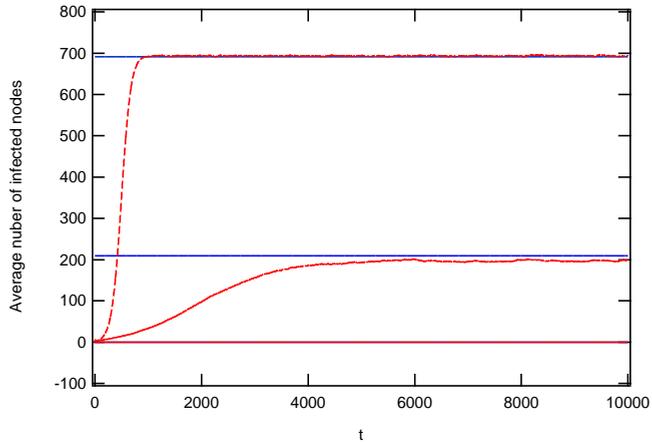
Figure 4.14: *Average number of infected nodes for $K_{250,750}$ with infection delay $\varepsilon = 50$, excluding virus epidemics that died out.*
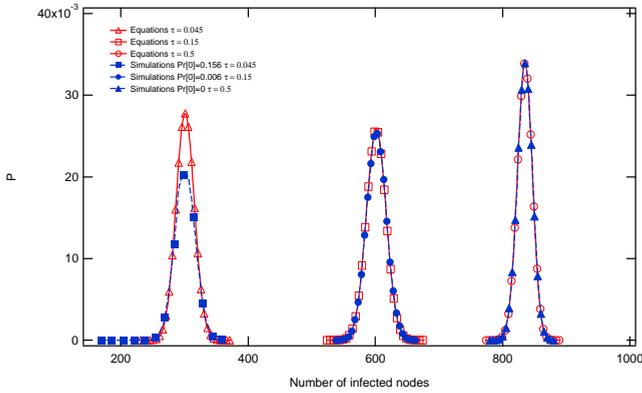
Figure 4.15: *Probability distribution of the number of infected nodes in the steady state for $K_{10,990}$*
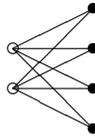
Figure 4.16: *Complete bi-partite graph K_{M,N_0}, with N_0 infected nodes*
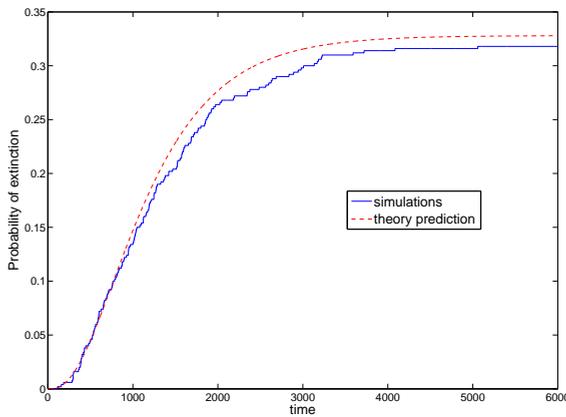
Figure 4.17: *Extinction of the virus as a function of time for $K_{10,990}$ with $\tau = 0.045$ for 3 initially infected nodes.*
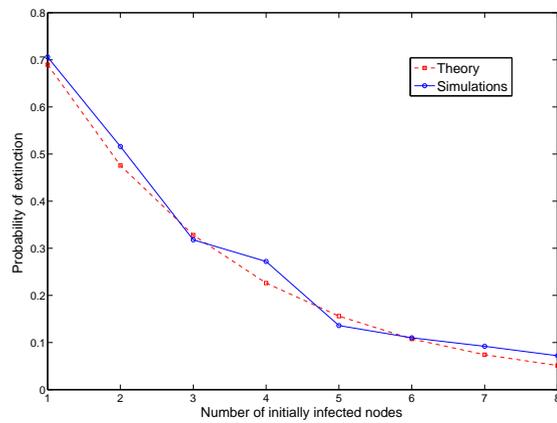
Figure 4.18: *Extinction of the virus after $T = 6000$ as a function of number of initially infected nodes, for $K_{10,990}$ with $\tau = 0.045$.*

Chapter 5

Heterogenous N-intertwined model

In this Chapter, we will study an extension of the N-intertwined model, where the curing and infection rates are node specific.

As in homogeneous case in Chapter 4, by separately observing each node, we will model the virus spread. In the case of heterogeneous model, the network is *bi-directional*, specified by asymmetric adjacency matrix A. Every node i at time t in the network has two states: infected with probability $\Pr[X_i(t) = 1]$ and healthy with probability $\Pr[X_i(t) = 0]$. If we apply Markov theory, the infinitesimal generator $Q_i(t)$ of this two-state continuous Markov chain is the same as in equation 4.1, with $q_{2;i} = \delta_i$ and

$$q_{1;i} = \sum_{j=1}^{N} \beta_j a_{ij} 1_{\{X_j(t)=1\}}$$

where the indicator function $1_x = 1$ if the event x is true else it is zero. The coupling of node i to the rest of the network is described by an infection rate $q_{1;i}$ that is a random variable, which essentially makes the process doubly stochastic.

Using the same analysis as in homogeneous case and denoting infection probability of node i with $v_i(t)$ and recalling that $\Pr[X_i(t) = 0] = 1 - v_i(t)$, the Markov differential equation [17, (10.11) on p. 182] for state $X_i(t) = 1$ turns out to be non-linear

$$\frac{dv_i(t)}{dt} = \sum_{j=1}^{N} \beta_j a_{ij} v_j(t) - v_i(t) \left(\sum_{j=1}^{N} \beta_j a_{ij} v_j(t) + \delta_i \right) \tag{5.1}$$

Written in matrix form, with

$$V(t) = \begin{bmatrix} v_1(t) & v_2(t) & \cdots & v_N(t) \end{bmatrix}^T$$

we arrive at

$$\frac{dV(t)}{dt} = A\text{diag}(\beta_j)V(t) - \text{diag}(v_i(t))(A\text{diag}(\beta_j)V(t) + C) \qquad (5.2)$$

where $\text{diag}(v_i(t))$ is the diagonal matrix with elements $v_1(t), v_2(t), \ldots, v_N(t)$ and the curing rate vector is $C = (\delta_1, \delta_2, \ldots, \delta_N)$.

We note that $A\text{diag}(\beta_i)$ is, in general and opposed to the homogeneous setting, not symmetric anymore, unless A and $\text{diag}(\beta_i)$ commute, in which case the eigenvalue $\lambda_i(A\text{diag}(\beta_i)) = \lambda_i(A)\beta_i$ and both β_i and $\lambda_i(A)$ have a same eigenvector x_i.

5.1   General heterogeneous steady-state

The metastable steady-state follows from (4.4) as

$$A\text{diag}(\beta_i)V_\infty - \text{diag}(v_{i\infty})(A\text{diag}(\beta_i)V_\infty + C) = 0$$

where $V_\infty = \lim_{t\to\infty} V(t)$. We define the vector

$$w = A\text{diag}(\beta_i)V_\infty + C \qquad (5.3)$$

and write the stead-state equation as

$$w - C = \text{diag}(v_{i\infty})w$$

or

$$(I - \text{diag}(v_{i\infty}))w = C$$

Ignoring extreme virus spread conditions (the absence of curing ($\delta_i = 0$) and an infinitely strong infection rate $\beta_i \to \infty$), then the infection probabilities $v_{i\infty}$ cannot be one such that the matrix $(I - \text{diag}(v_{i\infty})) = \text{diag}(1 - v_{i\infty})$ is invertible. Hence,

$$w = \text{diag}\left(\frac{1}{1 - v_{i\infty}}\right)C$$

Invoking the definition (5.3) of w, we obtain

$$A\text{diag}(\beta_i)V_\infty = \text{diag}\left(\frac{v_{i\infty}}{1 - v_{i\infty}}\right)C$$

$$= \text{diag}\left(\frac{\delta_i}{1 - v_{i\infty}}\right)V_\infty \qquad (5.4)$$

The i-th row of (5.4) yields the nodal steady state equation

$$\sum_{j=1}^{N} a_{ij}\beta_j v_{j\infty} = \frac{v_{i\infty}\delta_i}{1 - v_{i\infty}} \tag{5.5}$$

Let $\widetilde{V}_\infty = \mathrm{diag}(\beta_i)\, V_\infty$ and the effective spreading rate for node i, $\tau_i = \frac{\beta_i}{\delta_i}$, then we arrive at

$$\mathcal{Q}\left(\frac{1}{\tau_i\,(1 - v_{i\infty})}\right)\widetilde{V}_\infty = 0 \tag{5.6}$$

where the symmetric matrix

$$\begin{aligned}\mathcal{Q}\,(q_i) &= \mathrm{diag}\,(q_i) - A \\ &= \mathrm{diag}\,(q_i - d_i) + Q\end{aligned} \tag{5.7}$$

can be interpreted as a generalized Laplacian[1], because $\mathcal{Q}\,(d_i) = Q = \Delta - A$, where $\Delta = \mathrm{diag}(d_i)$. The observation that the non-linear set of steady-state equations can be written in terms of the generalized Laplacian $\mathcal{Q}\,(q_i)$ is fortunate, because, as will be shown in Section 5.1.1, the powerful theory of the "normal" Laplacian Q applies.

The modified steady-state vector $\widetilde{V}_\infty$ is orthogonal to each row (or, by symmetry, each column) vector of $\mathcal{Q}\left(\frac{1}{\tau_i(1-v_{i\infty})}\right)$. A non-zero modified steady-state vector $\widetilde{V}_\infty$ is thus only possible provided $\det \mathcal{Q}\left(\frac{1}{\tau_i(1-v_{i\infty})}\right) = 0$. In other words, the generalized Laplacian $\mathcal{Q}\left(\frac{1}{\tau_i(1-v_{i\infty})}\right)$ should have a zero eigenvalue with the modified steady-state vector $\widetilde{V}_\infty$ as corresponding eigenvector. Since the vectors $B = (\beta_1, \beta_2, \ldots, \beta_N)$ and $C = (\delta_1, \delta_2, \ldots, \delta_N)$ are given, the non-linear eigenvector problem (5.6) has, in general, a solution that cannot simply be recast to the homogeneous case where $B = \beta u$ and $C = \delta u$ (or $\beta_i = \beta$ and $\delta_i = \delta$ for all $1 \le i \le N$) in which the all-one vector $u = (1, 1, \ldots, 1)$.

5.1.1   The generalized Laplacian $\mathcal{Q}\,(q_i)$

Since $\mathcal{Q}\,(q_i)$ is symmetric, all eigenvectors are orthogonal such that, with $\widetilde{V}_\infty = \mathrm{diag}(\beta_i)\, V_\infty$,

$$\sum_{j=1}^{N} \beta_j v_{j\infty} y_j = 0 \tag{5.8}$$

where y is the eigenvector belonging to $\lambda\,(\mathcal{Q}\,(q_i)) \ne 0$.

---

[1]All eigenvalues of the Laplacian $Q = \Delta - A$ in a connected graph are positive, except for the smallest one that is zero. Hence, Q is positive semi-definite. Much more properties of the Laplacian Q are found e.g. in [46] and [44].

Theorem 15. *If the network G is connected, all eigenvalues of $\mathcal{Q}(q_i)$ are positive, except for the smallest one $\lambda_N(\mathcal{Q}) = 0$.*

Proof. The theorem is a consequence of the Perron-Frobenius Theorem (see e.g. [47]) for a non-negative, irreducible matrix. Indeed, consider the non-negative matrix $q_{max}I - \mathcal{Q}(q_i)$, where $q_{max} = \max_{1 \le i \le N} q_i$, whose eigenvalues are $\xi_k = q_{max} - \lambda_k(\mathcal{Q})$ for $1 \le k \le N$. Since G is connected, then $q_{max}I - \mathcal{Q}(q_i)$ is irreducible and the Perron-Frobenius Theorem states that the largest eigenvalue $r = \max_{1 \le k \le N} \xi_k$ of $q_{max}I - \mathcal{Q}(q_i)$ is positive and simple and the corresponding eigenvector x_r has positive components. Hence, $\mathcal{Q}(q_i) x_r = (q_{max} - r) x_r$. Since eigenvectors of a symmetric matrix are orthogonal while $\widetilde{V}_\infty^T x_r > 0$, x_r must be proportional to $\widetilde{V}_\infty$, and thus $q_{max} = r$. Since there is only one such eigenvector x_r and since the eigenvalue $r > q_{max} - \lambda_k(\mathcal{Q})$ for all k (except that k for which $\lambda_k(\mathcal{Q}) = 0$, which is thus the smallest eigenvalue), all other eigenvalues of $\mathcal{Q}(q_i)$ must exceed zero. □

If the graph G is disconnected which means that A is reducible [17], the Theorem 15 still applies (see e.g. [47]), however, under the slightly weakened form that x_r has non-negative components (instead of positive, hence, zero components can occur) and that the largest eigenvalue r is non-zero (not necessarily strict positive). The consequence is that more than one zero eigenvalue can occur. From the point of virus spread, we may ignore disconnected graphs, because the theory can be applied to each connected component (cluster) of the network G. The symmetry of $\mathcal{Q}(q_i)$ implies that all eigenvalues are real and can be ordered. By Theorem 15, we have

$$0 = \lambda_N(\mathcal{Q}) \le \lambda_{N-1}(\mathcal{Q}) \le \ldots \le \lambda_1(\mathcal{Q})$$

Gerschgorin's theorem [48, p. 71-75] indicates that the eigenvalues of $\mathcal{Q}(q_i)$ are centered around q_i with radius equal to the degree d_i, i.e. an eigenvalue λ of $\mathcal{Q}(q_i)$ lies in an interval $|\lambda - q_k| \le d_k$ for some $1 \le k \le N$. Thus, there is an eigenvalue λ of $\mathcal{Q}(q_i)$ that obeys

$$q_k - d_k \le \lambda \le d_k + q_k$$

A solution of (5.6) requires that at least one eigenvalue of $\mathcal{Q}(q_i)$ is zero, while Theorem 15 states that there is only one zero eigenvalue. Hence, precisely one, say the j-th, of the Gerschgorin line segments that contain the eigenvalue $\lambda_N(\mathcal{Q}) = 0$, must obey $q_j \le d_j$ to have a non-zero solution of (5.6). However, more Gerschgorin segments may obey $q_k - d_k \le 0$. This couples $\frac{1}{\tau_j(1-v_{j\infty})} \le d_j$ for at least one j component and shows that, when $v_{j\infty} \to 1$, there must hold that $\tau_j \to \infty$. Hence, for at least one component j, there holds that

$$0 < v_{j\infty} \le 1 - \frac{1}{\tau_j d_j}$$

where the lower bound follows, by the Perron-Frobenius Theorem, from the fact that the network G is connected. This shows that there is a critical bound on $\tau_j > \frac{1}{d_j}$ for at least one component of τ. The critical threshold on the τ-vector is further explored in Section 5.1.2.

We also know that $\text{trace}(\mathcal{Q}(q_i)) = \sum_{k=1}^{N} \lambda_k(\mathcal{Q})$. Thus, with $\lambda_N(\mathcal{Q}) = 0$,

$$\sum_{k=1}^{N-1} \lambda_k(\mathcal{Q}) = \sum_{i=1}^{N} \frac{1}{\tau_i(1 - v_{i\infty})}$$

In addition, since

$$\text{trace}(\mathcal{Q}^2(q_i)) = \text{trace}(\text{diag}(q_i^2)) + \text{trace}(A^2)$$

$$= \sum_{i=1}^{N} \frac{1}{\tau_i^2(1 - v_{i\infty})^2} + 2L$$

we have that

$$\sum_{k=1}^{N-1} \lambda_k^2(\mathcal{Q}) = \sum_{i=1}^{N} \frac{1}{\tau_i^2(1 - v_{i\infty})^2} + 2L$$

Right multiplication of (5.4) by the all one-vector $u^T = (1, 1, \ldots, 1)$ yields

$$u^T A \text{diag}(\beta_i) V_\infty = u^T \text{diag}\left(\frac{\delta_i}{1 - v_{i\infty}}\right) V_\infty$$

With $u^T A = D^T = (d_1, d_2, \ldots, d_N)$, the degree vector, we have

$$\left(u^T \text{diag}\left(\frac{\delta_i}{1 - v_{i\infty}}\right) - D^T \text{diag}(\beta_i)\right) V_\infty = 0$$

or[2]

$$\sum_{j=1}^{N} \left(\frac{1}{\tau_j(1 - v_{j\infty})} - d_j\right) \beta_j v_{j\infty} = 0 \tag{5.9}$$

Similarly as deduced from Gershgorin's theorem, this sum shows that, at least one j term should be negative (because $\beta_j v_{j\infty} \geq 0$), i.e. $d_j \geq \frac{1}{\tau_j(1 - v_{j\infty})}$. Also, in view of (5.8), the vector y with components $y_j = \frac{1}{\tau_j(1 - v_{j\infty})} - d_j$ is a linear combination of eigenvectors of $\mathcal{Q}\left(\frac{1}{\tau_i(1 - v_{i\infty})}\right)$ belonging to a non-zero eigenvalue. In general, however, the vector y is not an eigenvector of $\mathcal{Q}\left(\frac{1}{\tau_i(1 - v_{i\infty})}\right)$.

---

[2] The result (5.9) also follows by adding all rows in (5.6)

$$\mathcal{Q}(q_i)\widetilde{V}_\infty = \text{diag}(q_i - d_i)\widetilde{V}_\infty + Q\widetilde{V}_\infty$$

and using the basic fact that the row sum of the Laplacian Q is zero.

5.1.2 The critical threshold

We known that the exact steady-state is $V_\infty = 0$, but the metastable steady-state (see [41] for a deeper discussion) is characterized by a second solution, the eigenvector of (5.6).

Theorem 16. *The critical threshold is determined by vector $\tau_c = (\tau_{1c}, \tau_{2c}, \ldots, \tau_{Nc})$ that obey $\lambda_{\max}(R) = 1$, where $\lambda_{\max}(R)$ is the largest eigenvalue of the symmetric matrix*

$$R = diag\left(\sqrt{\tau_i}\right) A \, diag\left(\sqrt{\tau_i}\right) \tag{5.10}$$

whose corresponding eigenvector has positive components if the graph G is connected.

Proof. At the critical threshold, the second, non-zero solution is $V_\infty = \varepsilon x$, where x is a vector with non-negative components and where ε is arbitrary small. This property allows us to approximate the generalized Laplacian $\mathcal{Q}(q)$ as

$$\mathcal{Q}\left(\frac{1}{\tau_i(1 - v_{i\infty})}\right) = \text{diag}\left(\frac{\delta_i}{\beta_i(1 - \varepsilon x_i)}\right) - A$$

$$= \text{diag}\left(\frac{\delta_i}{\beta_i}\right)(I - \varepsilon\text{diag}(x_i)) - A + O\left(\varepsilon^2\right)$$

such that (5.6) becomes to first order in ε

$$\mathcal{Q}\left(\frac{1}{\tau_i}\right)\text{diag}(\beta_i)\, x = 0$$

which can be rewriten as an eigenvalue equation for the adjacency matrix,

$$\text{diag}\left(\frac{1}{\delta_i}\right) A \, \text{diag}(\beta_i)\, x = x$$

Hence, x is the eigenvector of $\widetilde{A} = \text{diag}\left(\frac{1}{\delta_i}\right) A \, \text{diag}(\beta_i)$ belonging to the eigenvalue 1. Since $\widetilde{A}$ is a non-negative, irreducible matrix, the Perron-Frobenius Theorem [17, p. 451] states that $\widetilde{A}$ has a positive largest eigenvalue $\lambda_{\max}\left(\widetilde{A}\right)$ with a corresponding eigenvector whose elements are all positive and that there is only one eigenvector of $\widetilde{A}$ with non-negative components. Since any scaled vector $V_\infty = \varepsilon x$ must have non-negative components (because they represent scaled probabilities), we find that $\lambda_{\max}\left(\widetilde{A}\right) = 1$. Hence, for the given vectors $B = (\beta_1, \beta_2, \ldots, \beta_N)$ and $C = (\delta_1, \delta_2, \ldots, \delta_N)$, there are three possibilities:

$$\begin{cases} \lambda_{\max}\left(\widetilde{A}\right) < 1 & \text{not infected network} \\ \lambda_{\max}\left(\widetilde{A}\right) = 1 & \text{critical threshold} \\ \lambda_{\max}\left(\widetilde{A}\right) > 1 & \text{infected network} \end{cases} \tag{5.11}$$

where the inequalities sign are deduced by relating the largest eigenvalue to the norm of the matrix $\widetilde{A}$: higher eigenvalues correspond to a larger norm (see e.g. [17, Section A.3.1]). Of course, only in case $\lambda_{\max}\left(\widetilde{A}\right) = 1$, the eigenvector equation has a non-zero solution. If $\lambda_{\max}\left(\widetilde{A}\right) > 1$, then the first order expansion is inadequate and the full non-linear equation (5.6) needs to be solved.

The first order expansion process has caused $\widetilde{A}$ to be not symmetric, while $\mathcal{Q}\left(\frac{1}{\tau_i(1-v_{i\infty})}\right)$ is symmetric in general. Fortunately, there exist a similarity transform $H = \mathrm{diag}\left(\sqrt{\delta_i\beta_i}\right)$ which symmetrizes $\widetilde{A}$,

$$R = H\widetilde{A}H^{-1} = \mathrm{diag}\left(\sqrt{\frac{\beta_i}{\delta_i}}\right) A \mathrm{diag}\left(\sqrt{\frac{\beta_i}{\delta_i}}\right)$$

and $R = R^T$ has the same real eigenvalues as $\widetilde{A}$ (see [17, p.438]). The matrix R also demonstrates that only an effective rate per node, $\tau_i = \frac{\beta_i}{\delta_i}$, is needed. Thus, the equation that characterizes the critical threshold is

$$Ry = y$$

where $y = Hx$. The eigenvalue $\lambda_{\max}\left(\widetilde{A}\right) = \lambda_{\max}(R) = 1$ determines the critical vectors $\tau_c = (\tau_{1c}, \tau_{2c}, \ldots, \tau_{Nc})$. In general, there can be more than one critical vector because $\lambda_{\max}(R) = 1$ is a map of $\mathbb{R}^N \to \mathbb{R}$. $\square$

We remark that, since $\mathrm{trace}(R) = \mathrm{trace}(A) = 0$, that $\lambda_{\max}(R) = \lambda_1(R) = -\sum_{j=2}^{N} \lambda_j(R)$, where the eigenvalues are ordered as $\lambda_N \leq \lambda_{N-1} \leq \cdots \leq \lambda_1$.

Special cases

We illustrate that more than one critical vector obeys $\lambda_{\max}(R) = 1$.

1. The homogeneous threshold $\tau_{\mathrm{hom};c}$ is found when $\tau_i = \tau$, in which case $\lambda_{\max}(R) = 1$ reduces to $\frac{1}{\tau_{\mathrm{hom};c}} = \lambda_{\max}(A)$, a basic result in [41].

2. When $\frac{\delta_i}{\beta_i} = \frac{1}{\tau_i} = d_i$ for all $1 \leq i \leq N$, we observe that $\mathcal{Q}(d_i) = Q$ if $v_{i\infty} = \varepsilon > 0$, where ε is arbitrary small. In that case, the steady-state vector is $\widetilde{V}_\infty \to \varepsilon u$, thus $V_\infty = \varepsilon(\beta_1, \beta_2, \ldots, \beta_N)$ and the critical vector $\tau_c = \left(\frac{1}{d_1}, \frac{1}{d_2}, \ldots, \frac{1}{d_N}\right)$. In that case, $R = \mathrm{diag}\left(\sqrt{\frac{1}{d_i}}\right) A \mathrm{diag}\left(\sqrt{\frac{1}{d_i}}\right)$ and after a similarity transform $H_1 = \mathrm{diag}\left(\sqrt{\frac{1}{d_i}}\right)$, we obtain the stochastic matrix [17, p.484-486]

$$H_1 R H_1^{-1} = \Delta^{-1} A$$

whose largest eigenvalue is, indeed, equal to one.

5.1.3   Bounding $\lambda_{\max}(R)$

Applying the general Rayleigh formulation for any matrix M,

$$\lambda_{\max} = \sup_{x \neq 0} \frac{x^T M x}{x^T x}$$

and, knowing that all components of the eigenvector belonging to the largest eigenvalue are non-negative, we obtain

$$\lambda_{\max}(R) = \sup_{x \neq 0} \frac{x^T \mathrm{diag}\left(\sqrt{\tau_i}\right) A \mathrm{diag}\left(\sqrt{\tau_i}\right) x}{x^T x}$$

Let $z = \mathrm{diag}\left(\sqrt{\tau_i}\right) x$, then

$$\lambda_{\max}(R) = \sup_{z \neq 0} \frac{z^T A z}{z^T \mathrm{diag}\left(\frac{1}{\tau_i}\right) z} \tag{5.12}$$

In the sequel, we deduce several bounds from (5.12).

First, we rewrite (5.12) as

$$\lambda_{\max}(R) = \sup_{z \neq 0} \frac{z^T A z}{z^T z} \frac{z^T z}{z^T \mathrm{diag}\left(\frac{1}{\tau_i}\right) z}$$

$$\geq \sup_{z \neq 0} \frac{z^T A z}{z^T z} \frac{z^T z}{\inf_{z \neq 0} z^T \mathrm{diag}\left(\frac{1}{\tau_i}\right) z}$$

$$= \lambda_{\max}(A) \min_{1 \leq j \leq N} \tau_i$$

Thus,
$$\lambda_{\max}(A) \min_{1 \leq j \leq N} \tau_i \leq \lambda_{\max}(R) \leq \lambda_{\max}(A) \max_{1 \leq j \leq N} \tau_i \tag{5.13}$$

where the upper bound follows similarly from
$\sup_{z \neq 0} \frac{z^T A z}{z^T \mathrm{diag}\left(\frac{1}{\tau_i}\right) z} \leq \frac{\max_{z \neq 0} z^T A z}{\min_{z \neq 0} z^T \mathrm{diag}\left(\frac{1}{\tau_i}\right) z}$.

At the critical threshold where $\lambda_{\max}(R) = 1$, the bounds reduce, with $\tau_{\min} = \min_{1 \leq j \leq N} \tau_i$ and $\tau_{\max} = \max_{1 \leq j \leq N} \tau_i$, to the inequality for the minimum and maximum component of the critical τ-vector,

$$\tau_{\min;c} \leq \frac{1}{\lambda_{\max}(A)} \leq \tau_{\max;c}$$

Hence, there is always at least one τ-component below and one τ-component above the critical threshold in the homogeneous case $\tau_{\mathrm{hom};c} = \frac{1}{\lambda_{\max}(A)}$.

Next, a common lower bound (see e.g. [49],[50]) is obtained by letting $z = u$, the all-one vector, in (5.12). Equality in (5.12) is only achieved when z is the eigenvector such that, in all other cases,

$$\lambda_{\max}(R) \geq \frac{u^T A u}{u^T \operatorname{diag}\left(\frac{1}{\tau_i}\right) u} = \frac{2L}{\sum_{j=1}^{N} \frac{1}{\tau_j}} \tag{5.14}$$

For all regular graphs[3], the bound (5.14) is sharp, because u is the largest eigenvector of A belonging to $\lambda_{\max}(A) = d$. However, all eigenvectors of $\operatorname{diag}\left(\frac{1}{\tau_i}\right)$ are the basic vectors e_j with all components equal to zero, except for the j-th one that is equal to one. Written in terms of the average degree $E[D] = \frac{2L}{N}$ and the harmonic mean $E[\tau^{-1}] = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{\tau_j}$ yields

$$\lambda_{\max}(R) \geq \frac{E[D]}{E[\tau^{-1}]}$$

such that at the critical threshold, where $\lambda_{\max}(R) = 1$, there holds that $E\left[\tau_c^{-1}\right] \geq E[D]$. Unfortunately, the harmonic, geometric and arithmetic mean inequality[4], that leads to $\frac{1}{E[\tau^{-1}]} = N\left(\sum_{j=1}^{N}\frac{1}{\tau_j}\right)^{-1} \leq \frac{1}{N}\sum_{j=1}^{n}\tau_j = E[\tau]$, prevents us to clearly upper bound the average zero infection τ-region, $[0, E[\tau_c]]$. Approximative, by assuming $\frac{1}{E[\tau^{-1}]} \approx E[\tau]$, the average zero infection τ-region is upper bounded by the mean degree $E[D]$. Notice that, in the homogeneous case ($\tau_j = \tau$), the approximation is exact, leading to $\tau_{\text{hom;c}} \leq \frac{1}{E[D]}$.

There are several other interesting choices. A first alternative choice is $z = D$, where $D = (d_1, d_2, \ldots, d_N)$ is the degree vector. Or A second alternative choice is to choose the components of the vector z equal to a row vector of A, i.e. $z_j = a_{qj}$. However, these bounds are not very good.

5.1.4 Additional properties

We list here additional properties that have been proved in [41], and whose extension to the in-homogeneous setting is rather straightforward.

Lemma 17. *In a connected graph, either $v_{i\infty} = 0$ for all i nodes, or none of the components $v_{i\infty}$ is zero.*

Lemma 7 also follows from the Perron-Frobenius theorem as shown in the proof of Theorem 15.

---

[3]A regular graph is a graph where all degree $d_i = d$.
[4]For real positive numbers $a_1, a_2, \ldots, a_n$, the harmonic, geometric and arithmetic mean inequality is

$$\frac{n}{\sum_{j=1}^{n}\frac{1}{a_j}} \leq \sqrt[n]{\prod_{j=1}^{n} a_j} \leq \frac{1}{n}\sum_{j=1}^{n} a_j \tag{5.15}$$

Theorem 18. *The non-zero steady-state infection probability of any node i in the N-intertwined model can be expressed as a continued fraction*

$$v_{i\infty} = 1 - \cfrac{1}{1 + \frac{\phi_i}{\delta_i} - \delta_i^{-1}\sum_{j=1}^{N} \cfrac{\beta_j a_{ij}}{1 + \frac{\phi_j}{\delta_j} - \delta_j^{-1}\sum_{k=1}^{N} \cfrac{\beta_k a_{jk}}{1 + \frac{\phi_k}{\delta_k} - \delta_k^{-1}\sum_{q=1}^{N} \cfrac{a_{qk}\beta_q}{\ddots}}}} \tag{5.16}$$

where the total infection rate of node i, incurred by all neighbors towards node i, is

$$\phi_i = \sum_{j=1}^{N} a_{ij}\beta_j = \sum_{j \in \ neighbor(i)} \beta_j \tag{5.17}$$

Consequently, the exact steady-state infection probability of any node i is bounded by

$$0 \leq v_{i\infty} \leq 1 - \frac{1}{1 + \frac{\phi_i}{\delta_i}} \tag{5.18}$$

The continued fraction stopped at iteration k includes the effect of virus spread up to the $(k-1)$-hop neighbors of node i. In the homogeneous case where $\beta_j = \beta$ for all $1 \leq j \leq N$, we have that $\phi_i = \beta d_i$ is proportional to the degree of node i. The ratio $\tilde{\tau}_i = \frac{\phi_i}{\delta_i}$ is the total effective infection rate of node i.

Lemma 19. *In a connected graph G above the critical threshold, a lower bound of $v_{i\infty}$ for any node i equals*

$$v_{i\infty} \geq 1 - \frac{1}{\min_{1 \leq k \leq N} \frac{\phi_k}{\delta_k}} \tag{5.19}$$

Proof. Lemma 7 and Theorem 16 show that, for vectors τ above the critical threshold vector τ_c, there exists a non-zero minimum $v_{\min} = \min_{1 \leq i \leq N} v_{i\infty} > 0$ of the steady-state infection probabilities, which obeys (5.5). Assuming that this minimum $v_{\min}$ occurs at node i,

$$v_{\min} = 1 - \frac{1}{1 + \delta_i^{-1}\sum_{j=1}^{N} a_{ij}\beta_j v_{j\infty}} \geq 1 - \frac{1}{1 + \frac{\phi_i}{\delta_i}v_{\min}}$$

where we have used the definition (5.17). From the last inequality, it follows that

$$v_{\min} \geq 1 - \frac{\delta_i}{\phi_i} \tag{5.20}$$

such that (4.13) is proved. $\qquad\square$

By combining (4.8) and (4.13), the total fraction of infected nodes $y_\infty = \frac{1}{N}\sum_{k=1}^{N} v_{k\infty}$ in steady-state is bounded by

$$1 - \frac{1}{\min_{1 \leq k \leq N} \frac{\phi_k}{\delta_k}} \leq y_\infty \leq 1 - \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1 + \frac{\phi_i}{\delta_i}}$$

5.2 Special case - the regular graph with m curing rates

In this section, we derive the threshold for the spread of viruses on regular graphs with m different curing rates and the same infection rate per link β.

Assume that $n_1, n_2, .., n_m$ denotes the fraction of nodes with curing rate δ_1, δ_2 $,.., \delta_m$ ($\sum_{i=1}^{m} n_i = 1$). It is important to note that one of the assumptions is complete symmetry of the problem. For every node i, a fraction n_1 of neighbors has the curing rate δ_1, a fraction n_2 has curing rate δ_2 and so on.

Denote the number of infected nodes of type i in the population at time t by $X_i(t)$. The probability that a randomly chosen node of type i is infected is $v_i(t) \equiv \frac{X_i(t)}{N n_i}$. For every susceptible node the rate of infection is the product of the infection rate per node (β) and the probability that on a given link the susceptible node connects to an infected node ($\sum_{j=1}^{m} (n_j k) v_j$).

Therefore, we obtain the following differential equation describing the time evolution of $v_i(t)$:

$$\frac{dv_i}{dt} = \beta k \left(\sum_{j=1}^{m} n_j v_j \right)(1 - v_i) - \delta_i v_i, \; i = 1, .., m \tag{5.21}$$

For the general case with different curing rates, it is impossible to obtain an explicit solution for the system of equations (5.21). The standard approach for this type of system of nonlinear differential equations, is to study the qualitative behavior in the phase space.

Theorem 20. *Consider connected regular graphs where each node has exactly k neighbors. Assume that the infection rate along each link is β while the curing rate for each node is δ_i for a fraction n_i of the nodes, with $i = 1, .., m \leqslant k$ and $\sum_{i=1}^{m} n_i = 1$. Complete symmetry is assumed, where each node sees the same fraction of different curing rates. If we define the effective spreading rate as $\tau = \frac{\beta}{\delta^*}$, where δ^* is defined as the weighted harmonic mean of $\delta_1, ..., \delta_m$, i.e. $\delta^* = \left(\sum_{i=1}^{m} \frac{n_i}{\delta_i} \right)^{-1}$, then the epidemic threshold satisfies $\tau_c = \frac{1}{k}$.*

Proof. We denote the fraction of infected nodes of type i ($1 \leqslant i \leqslant m$) at time t as $v_i(t)$. This leads to a system of m differential equations (5.21).

We will use a Lyapunov function [51] to show that, under the condition $\beta \sum_{t=1}^{m} \frac{n_t}{\delta_t} - \frac{1}{k} \leq 0$, the origin is a global attractor for $\{v_1 \geq 0, v_2 \geq 0, .., v_m \geq 0\}$,

hence, that the virus dies out. Let $V(v_1, v_2, .., v_m) = \prod_{j=1}^{m} \delta_j \sum_{s=1}^{m} \frac{v_s}{\delta_s}$. Then, we have

$$\frac{dV}{dt} = -\left(\sum_{s=1}^{m} v_s\right)\left(\beta k V - \beta k \prod_{j=1}^{m} \delta_j \sum_{t=1}^{m} \frac{n_t}{\delta_t} + \prod_{j=1}^{m} \delta_j\right)$$

$$= -\left(\sum_{s=1}^{m} v_s\right)\left(\beta k V - k \prod_{j=1}^{m} \delta_j \left(\beta \sum_{t=1}^{m} \frac{n_t}{\delta_t} - \frac{1}{k}\right)\right).$$

The claim follows directly by applying Lyapunov's stability theorem. Next we consider the case $\beta \sum_{t=1}^{m} \frac{n_t}{\delta_t} - \frac{1}{k} > 0$. We first note that any trajectory of the system (5.21) can never leave the box $B = \{(v_1, ..., v_m)|0 \leq v_1 \leq 1, ..., 0 \leq v_m \leq 1\}$. This follows from $\frac{dv_1}{dt}|_{v_1=0} = \beta k (\sum_{j=1}^{m} n_j v_j) \geq 0$, and similar inequalities at the borders of the box B.

From the construction of the above Lyapunov function V, we can see that for $\beta \sum_{t=1}^{m} \frac{n_t}{\delta_t} - \frac{1}{k} > 0$, and for $(v_1, ..., v_m) \in B$ and sufficiently close to the origin, $\frac{dV}{dt} > 0$. This implies that the origin has an unstable manifold in B. Therefore, since any trajectory of system (5.21) can never leave the box B, system (5.21) has an attractor as the ω-limit set and, hence, the virus does survive. This finishes the proof of the theorem. $\square$

5.2.1 Virus spread on regular graphs with two curing rates

The two dimensional case ($m = 2$) of virus spread on a regular graph can be analyzed in more details. Applying Theorem 20, the spreading process has a threshold at $\tau = \frac{\beta}{\delta^*} = \frac{1}{k}$, where $\delta^* = \frac{\delta_1 \delta_2}{n_1 \delta_2 + n_2 \delta_1}$.

The phase portrait of two examples are depicted in Figure 5.1. The attractor for the case where virus survives is given by $(v_1, v_2) = (0.22, 0.17)$.

For system (5.21) where $m = 2$, it can be proven that the attractor is an equilibrium point of a nodal type, situated on a straight line L. It can also be shown that the system does not contain other equilibrium points in A or closed orbits. Therefore, in the case $m = 2$, this equilibrium point is a global attractor of system (5.21) in A.

Lemma 21. *The set of differential equations given by (5.21) for $m = 2$, has a straight line solution of the form $v_2 = \lambda v_1$.*

Proof. We have that

$$\left(\frac{dv_2}{dt} = \lambda \frac{dv_1}{dt}\right)_{v_2 = \lambda v_1}$$
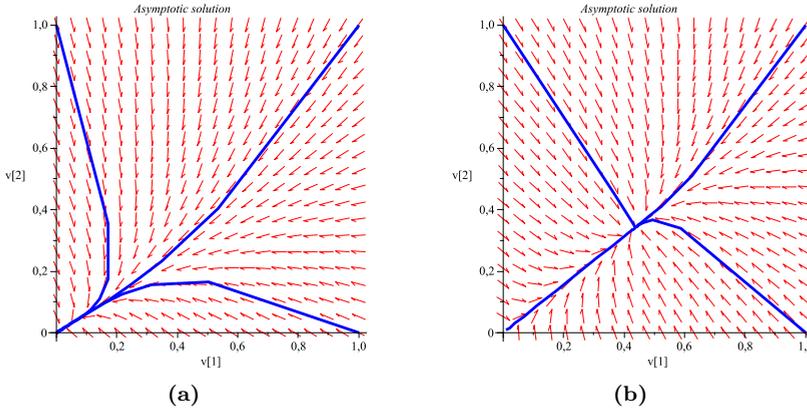
Figure 5.1: *Phase portrait for a regular graph with the two curing rates where a) virus dies out $\beta = 0.2$, $\delta_1 = 0.8$, $\delta_2 = 1.2$, $k = 4$, $n_1 = n_2 = 0.5$. b) virus survives $\beta = 0.4$, $\delta_1 = 0.8$, $\delta_2 = 1.2$, $k = 4$, $n_1 = n_2 = 0.5$.*

$$-v_1(\beta k n_1 \lambda^2 + (\beta k(n_1 - n_2) - \delta_1 + \delta_2)\lambda - \beta k n_2) \equiv -v_1 f(\lambda)$$

$f(\lambda)$ has got exactly one negative root and one positive root. The positive root λ_1 satisfies

$$\lambda_1 = \frac{\beta k(n_2 - n_1) + \delta_1 - \delta_2 + \sqrt{\Delta}}{2\beta k n_1},$$

where $\Delta = \beta^2 k^2 + 2\beta k(n_1 - n_2)(\delta_2 - \delta_1) + (\delta_1 - \delta_2)^2$. Therefore the straight line $L : v_2 = \lambda_1 v_1$ is a solution of system (5.21) for $m = 2$, which for $0 \le v_1 \le 1$ is situated in A.

By application of the Poincaré-Bendixson theorem [51] on A, the ω-limit set for the system (5.21) for $m = 2$, can be either an equilibrium point or an isolated periodic orbit. From the fact that there is a line solution through the equilibrium point, it follows that the ω-limit set is the equilibrium point. □

5.3 Special case - the complete bi-partite graph with two curing rates

We will now derive a model for virus spread on the complete bi-partite graph $K_{M,N}$ with two different curing rates and one spreading rate. The result is general in the sense that both curing rates can be distributed in both sets of nodes (S_1 and S_2).

Let us assume that a fraction p, with $p \in [0, 1]$, of nodes belonging to S_1 and a fraction q, with $q \in [0, 1]$, of nodes belonging to set S_2 have a curing rate δ_1, the rest have a curing rate δ_2. The total fraction of nodes with the curing rate δ_1 is $s = \frac{Mp + Nq}{M + N}$.

Denote the number of infected nodes of type 1 in the population of nodes from set S_1 at time t by $X_{i1}(t)$. The probability that a randomly chosen node of type 1 from set S_1 is infected is $v_{i1}(t) \equiv \frac{X_{i1}(t)}{Mp}$. Similarly, let v_{i2} denote the infection probability for nodes of type 2 from set S_1, (v_{j1} denotes type 1, set S_2; and v_{j2} denotes type 2, set S_2). Now, the rate at which the probability of infection for nodes of type 1, set S_1 changes is due to two processes: susceptible nodes becoming infected and infected nodes being cured. The curing rate for an infection probability v_{i1} for nodes of type 1, set S_1 is $\delta_1 v_{i1}$. The rate at which the probability v_{i1} grows is proportional to the probability of a node of type 1, set S_1 being susceptible, i.e. $1 - v_{i1}$. For every susceptible node the rate of infection is the product of the infection rate per node (β), the degree of the node (N) and the probability that on a given link the susceptible node connects to an infected node ($qv_{j1} + (1 - q)v_{j2}$).

Similarly, we obtain the differential equations for the other probabilities:

$$\begin{cases} \frac{dv_{i1}}{dt} = \beta N(qv_{j1} + (1-q)v_{j2})(1 - v_{i1}) - \delta_1 v_{i1}, \\ \frac{dv_{i2}}{dt} = \beta N(qv_{j1} + (1-q)v_{j2})(1 - v_{i2}) - \delta_2 v_{i2}, \\ \frac{dv_{j1}}{dt} = \beta M(pv_{i1} + (1-p)v_{i2})(1 - v_{j1}) - \delta_1 v_{j1}, \\ \frac{dv_{j2}}{dt} = \beta N(pv_{i1} + (1-p)v_{i2})(1 - v_{j2}) - \delta_2 v_{j2}, \end{cases} \quad (5.22)$$

In order to simplify the system of equations, we will substitute

$$i_1 = pv_{i1},\ i_2 = (1-p)v_{i2},\ j_1 = qv_{j1},\ j_2 = (1-q)v_{j2}$$

and

$$i = i_1 + i_2,\ j = j_1 + j_2$$

Therefore, we obtain the following differential equations for $i_1(t)$, $i_2(t)$, $j_1(t)$, $j_2(t)$:

$$\begin{cases} \frac{di_1}{dt} = p\beta Nj - \beta Nji_1 - \delta_1 i_1, \\ \frac{di_2}{dt} = (1-p)\beta Nj - \beta Nji_2 - \delta_2 i_2, \\ \frac{dj_1}{dt} = q\beta Mi - \beta Mij_1 - \delta_1 j_1, \\ \frac{dj_2}{dt} = (1-q)\beta Ni - \beta Nij_2 - \delta_2 j_2, \end{cases} \quad (5.23)$$

By solving the system of equations 5.23 for the steady state ($\frac{di_1}{dt} = \frac{di_2}{dt} = \frac{dj_1}{dt} = \frac{dj_2}{dt} = 0$) we can calculate the threshold:

$$\frac{\beta}{\delta^*} = \tau_c = \frac{1}{\sqrt{MN}} \quad (5.24)$$

$$\delta^* = \frac{\delta_1 \delta_2}{\sqrt{\delta_1^2(1-p)(1-q) + \delta_2^2 pq + \delta_1 \delta_2(p(1-q) + q(1-p))}} \quad (5.25)$$

Theorem 22. *Consider complete bi-partite graphs $K_{M,N}$ consisting of two disjoint sets S_1 and S_2 containing respectively M and N nodes. Assume that the*

infection rate along each link is β. For the nodes in S_1 a fraction p has curing rate δ_1 and in S_2 a fraction q of the nodes has curing rate δ_1, while the curing rate for a fraction $(1-p)((1-q))$ of the nodes is δ_2. If we define the effective spreading rate as $\tau = \frac{\beta}{\delta^*}$, where δ^* is defined as $\delta^* = \frac{\delta_1\delta_2}{\sqrt{(1-p)(1-q)\delta_1^2 + pq\delta_2^2 + \delta_1\delta_2(p(1-q)+q(1-p))}}$, then the epidemic threshold satisfies $\tau_c = \frac{1}{\sqrt{MN}}$.

Proof. First, we will show that if $\frac{\beta}{\delta^*} \leqslant \frac{1}{\sqrt{MN}}$, the virus dies out. $(0,0,0,0)$ is an equilibrium point for system (5.22). We will use a Lyapunov function to show that, under the condition $\frac{\beta}{\delta^*} \leqslant \frac{1}{\sqrt{MN}}$, the origin is a global attractor for $i_1 \geq 0, i_2 \geq 0, j_1 \geq 0, j_2 \geq 0$.

Let $V(i_1, i_2, j_1, j_2) = \delta_1\delta_2^2 i_1 + \delta_1^2\delta_2 i_2 + \beta N(p\delta_2 + (1-p)\delta_1)(\delta_2 j_1 + \delta_1 j_2)$. Then,

$$
\begin{aligned}
\frac{dV}{dt} = {} & (\beta^2 MN((1-p)(1-q)\delta_1^2 + pq\delta_2^2 + \\
& + \delta_1\delta_2((1-p)q + (1-q)p)) - \delta_1^2\delta_2^2)(i_1 + i_2) \\
& - \beta N\delta_2(\beta M(p\delta_2 + (1-p)\delta_1) + \delta_1\delta_2)i_1 j_1 \\
& - \beta N\delta_1(\beta M(p\delta_2 + (1-p)\delta_1) + \delta_2^2)i_1 j_2 \\
& - \beta N\delta_2(\beta M(p\delta_2 + (1-p)\delta_1) + \delta_1^2)i_2 j_1 \\
& - \beta N\delta_1(\beta M(p\delta_2 + (1-p)\delta_1) + \delta_1\delta_2)i_2 j_2.
\end{aligned}
$$

The extinction of the virus follows directly by applying Lyapunov's stability theorem. Next we will show that if $\frac{\beta}{\delta^*} > \frac{1}{\sqrt{MN}}$, the virus survives. We first note that any trajectory of the system (5.22) can never leave the box $B = \{(i_1, i_2, j_1, j_2) | 0 \leq i_1 \leq 1, \ 0 \leq i_2 \leq 1, 0 \leq j_1 \leq 1, 0 \leq j_2 \leq 1\}$. This follows from $\frac{di_1}{dt}|_{i_1=0} = p\beta N(j_1 + j_2) \geq 0$, and similar inequalities at the borders of the box B.

From the construction of the Lyapunov function, we can observe that for $\beta^2 MN((1-p)(1-q)\delta_1^2 + pq\delta_2^2 + \delta_1\delta_2((1-p)q + (1-q)p)) - \delta_1^2\delta_2^2) - \delta_1^2\delta_2^2 > 0$ and for $(i_1, i_2, j_1, j_2) \in B$ and sufficiently close to the origin, $\frac{dV}{dt} > 0$. This implies that the origin has an unstable manifold in B. Therefore, because any trajectory of system (5.22) can never leave the box B, system (5.22) has an attractor as the ω-limit set and hence the virus does survive. $\square$

The result from Theorem 22 holds for non-symmetric cases: a node from set S_1 sees different portion of nodes with curing rate δ_1 than a node from set S_2 ($p \neq q$). In the symmetric case ($p = q$), a more general result with m different curing rates can be derived, as in the case of the regular graph, described in Theorem 20.

Part II

Optimization of protection and Game theory

Chapter 6

Optimization of protection

Security and protection against malware spread are expensive luxury. In addition, security of the whole network depends on protection strategy of each node. What is the optimal protection in a network with N nodes? In an attempt to address these questions, we will study the global optimization problem for a network with N nodes in the case where spreading rate is the same for all the nodes and the sum of all curing rates is optimized. Because spreading rate per link is the same for all the nodes, without loss of generality, we will assume $\beta = 1$ and we will omit it from the equations.

6.1 Problem Formulation

Each node in a network has protection given by curing rate δ_i. Its level of security, the probability of infection influences neighbors. All the nodes in the network form network security. We will optimize overall security by strategic distribution of curing rates to nodes. As probabilities of infection are functions of curing rates of all nodes $v_{i\infty}(C)$, where $C = [\delta_1 \; \delta_2 \; .. \; \delta_N]^T$.

We formulate the following problem:

Problem 23.

$$(P_1): \begin{cases} min & f(C) = \sum_{j=1}^{N} v_{j\infty}(C) \\ s.t. & \sum_{j=1}^{N} \delta_j = 2L\alpha \\ & 0 \leq v_{i\infty} \leq 1; i = 1..N \\ & 0 \leq \delta_i \leq \delta_c; i = 1..N \end{cases}$$

Where $\alpha \in [0,1]$ and constant L refers to the number of links in the graph, and $\delta_c = 2L\alpha$.

In general, probability of infection $v_{i\infty}$ can not be found as explicit function of C, thus it is better to consider inverse problem of optimization of sum of curing rates such that probability of infection is bounded.

6.2 Inverse Optimization Problem

From equation 4.6, we have that

$$\delta_i = \frac{1 - v_{i\infty}}{v_{i\infty}} \sum_{j=1}^{N} a_{ij} v_{j\infty}$$

Let V denote the vector of infection probabilities $V = [v_{1\infty} \ v_{2\infty} \ .. \ v_{N\infty}]$. We formulate the following problem:

Problem 24.

$$(P_2) : \begin{cases} min & f(V) = \sum_{j=1}^{N} \delta_j(V) \\ s.t. & \sum_{j=1}^{N} v_{j\infty} = N\alpha \\ & 0 \le v_{i\infty} \le 1; i = 1..N \\ & 0 \le \delta_i \le \delta_c; i = 1..N \end{cases} \qquad (6.1)$$

where $\delta_c \le 2L$ and $\alpha \in [0,1]$, N is the number of nodes in the network. For $\alpha = 1$ optimization function is $f(V) = 0$. We derive in Theorem 25 that at the threshold for $\alpha = 0$, minimum of optimization function is $f(V) = 2L$.

For $0 < \alpha < 1$, the optimization problem (24) belongs to the class of sum of ratios fractional program [52].

We will separate the problem in two subproblems, namely $\alpha = 0$ and $\alpha \in (0, 1)$. This will allow us to treat the threshold point with different analytical tools. We will first consider the case for $\alpha = 0$, namely at the threshold, and than the case for $0 < \alpha < 1$ which is the sum of ratios fractional programming problem.

6.3 Optimization at the Threshold, $\alpha = 0$

In [53], Borgs *et al.* derive that minimum sum of curing rates applied on a network, such that contact process is at the threshold, is equal to the number of links in the network, $\sum_{j=1}^{N} \delta_j = 2L$, and it holds for all i, $\delta_i = d_i$.

We can derive the same result using N-intertwined model.

Theorem 25. *The minimum global price for a network at the threshold is*

$$f_V^{(\min)} = \sum_{j=1}^{N} d_j$$

and it is reached for each $\delta_i = d_i$.

Proof. Using the lower bound (5.14) on the largest eigenvalue of matrix A_δ, we have that if $\frac{2L}{\sum_{j=1}^{N} \delta_j} \geq 1$, the largest eigenvalue obeys $\lambda_{\max}(A_\delta) \geq \frac{2L}{\sum_{j=1}^{N} \delta_j} \geq 1$. If $\lambda_{\max}(A_\delta) \geq 1$ the network is above the threshold. Therefore, if $2L \geq \sum_{j=1}^{N} \delta_j$ which is equivalent to $\sum_{j=1}^{N} d_j \geq \sum_{j=1}^{N} \delta_j$ the network is infected. The equality $\lambda_{\max}(A_\delta) = \frac{2L}{\sum_{j=1}^{N} \delta_j} = 1$ holds if $\delta_i = d_i$ and the epidemic threshold is reached. $\square$

It is possible that other curing distributions satisfy $\sum_{j=1}^{N} d_j = \sum_{j=1}^{N} \delta_j$, however the minimum of the sum of curing rates cannot be lower than $\sum_{j=1}^{N} d_j = 2L$. Therefore for $\alpha = 0$, our minimization function is equal to 0. The optimal strategy is to distribute curing rates proportional to the degree of nodes.

The protection's efficiency depends on topological properties. In particular, for a complete graph, the minimum of the cost function is largest among all the graphs $f_V^{(\min)} = N(N-1)C$.

6.4 Sum of Ratios Fractional Programming, $\alpha \in (0,1)$

By rewriting equation (6.1)

$$
\begin{aligned}
f(V) &= \sum_{i=1}^{N} \delta_i(V) \\
&= \sum_{i=1}^{N} \frac{1 - v_{i\infty}}{v_{i\infty}} \sum_{j=1}^{N} a_{ij} v_{j\infty} \\
&= \sum_{i=1}^{N} \frac{\sum\limits_{j=1}^{N} a_{ij} v_{j\infty} - v_i^2 d_i}{v_{i\infty}}
\end{aligned}
$$

where d_i is the degree of i-th node.

Sum of Ratios generally is not a convex problem. Specifically, problem (24), Section 6.2 is not a convex problem which can be checked by restriction to a line.

For an overview of fractional programming, we refer to Schaible [52]. The sum of ratios fractional programming is a difficult problem, because many local minima can exist. Freund et al. [54] have shown that minimizing the sum of ratios is **NP** - complete problem. Several authors proposed algorithms for solving linear sum of ratios functional program [55], [56], [57], [58]. Only recently, several algorithms have been proposed for the case of nonlinear sum of ratios problem. Benson presented algorithm for globally solving equivalent convex maximization problem via branch-and-bound method in $2p + N$ space, where p is the number of fractions in the sum and N is the original size of the problem [59] and [60]. Wang et al. [61] generalized branch-and-bound method for the case where nominator, denominator and constrains are generalized multivariable polynomials and the feasible region is non convex.

6.4.1   Protection Proportional to the Node Degree for $\alpha \in (0, 1)$

At the threshold, the minimum of function $f(C)$ is reached for $\sum_{j=1}^{N} \delta_j = 2L$. The curing rate δ_i of a node i is equal to the degree d_i of that node. It is possible that the same strategy holds for the case where the threshold can not be reached due to the constrain $\sum_{j=1}^{N} v_j = \alpha N > 0$. This is equivalent to the constrained problem (23), Section 6.1, $\sum_{j=1}^{N} \delta_j < 2L$.

Consider problem (24), Section 6.2. If the curing rate is proportional to the degree with the same factor α, than all the infection probabilities are equal, which is formally stated in next Lemma.

Lemma 26. *If $\delta_i = (1-\alpha)d_i$, $\alpha < 1$, the probabilities of infection are all equal, namely $(\forall i)v_{i\infty} = \alpha$.*

Proof. If $\delta_i = (1-\alpha)d_i, \alpha < 1$, we have that

$$\sum_{j=1}^{N}\delta_j = (1-\alpha)\sum_{j=1}^{N}d_j \leq \sum_{j=1}^{N}d_j$$

we are certainly above the threshold. Let us assume that $v_{i\infty} = \alpha$, from 4.6 we have that $\delta_i = (1-\alpha)d_i$. $\square$

For the constrain $\sum_{j=1}^{N}v_j = \alpha N$ and $\alpha > 0$, is the protection which is proportional to the node degree the global optimum?

Theorem 27. *Consider problem (24), Section 6.2. For $\alpha > 0$ and for all ones vector u, solution $V = \alpha u$ is not a global minimum of function $f(V)$.*

Proof. Consider optimization problem (24) and let all nodes have the same infection probability $v_{i\infty} = \alpha$. The objective function is $f(V) = 2L(1-\alpha)$, L is the number of links. Now, let us take the node with the maximum degree d_{max} and the node with minimum degree d_{min} and change their infection probabilities as

$v_{d_{min}\infty} = \alpha - \gamma$

$v_{d_{max}\infty} = \alpha + \gamma,$

such that total sum of infection probabilities stays the same. We will assume for now that two nodes have no neighbors in common and that they are not connected. Any other case can be derived in similar manner.

The curing rates of these two nodes and their neighbors will change as follows

$$\delta_{d_{max}} = \alpha d_{max}(\frac{1}{\alpha + \gamma} - 1)$$

$$\delta_{d_{min}} = \alpha d_{min}(\frac{1}{\alpha - \gamma} - 1)$$

$$\delta_{d_{imax}} = (1-\alpha)d_{imax} - (\frac{1-\alpha}{\alpha}\gamma)$$

$$\delta_{d_{imin}} = (1-\alpha)d_{imin} + (\frac{1-\alpha}{\alpha}\gamma)$$

The new total sum of curing rates - objective function $f'(V)$ becomes

$$f'(V) = \sum_{j=1}^{N}\delta_j(V)$$

$$= 2L\alpha - S + \frac{S\alpha^2 - \alpha\gamma R}{\alpha^2 - \gamma^2} - R\frac{\gamma(1-\alpha)}{\alpha}$$
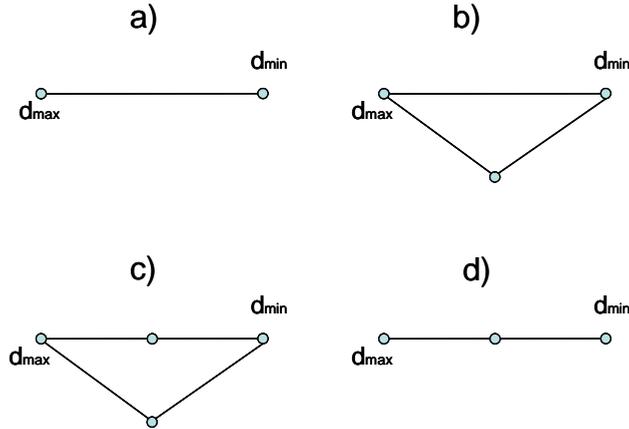
Figure 6.1: *Structures of connections between a node with maximum and minimum degree*

where $S = d_{max} + d_{min}$ and $R = d_{max} - d_{min}$. After some calculations it is possible to show that for any α, S, R there exist γ, such that $f'(V) < 2L\alpha$.

For other topological cases shown in Figure 6.1 similarly can be deduced that for any α, S, R there exist γ, such that $f'(V) < 2L\alpha$. We will skip details here. $\quad\square$

Corollary 28. *Consider problem (24), Section 6.2. For $\alpha > 0$ and for all ones vector u, vector $V = \alpha u$ is an upper bound on the global optimum.*

Proof. The Corollary is direct consequence of Theorem 27. $\quad\square$

It is interesting to note that by decreasing infection probability of highly connected node and increasing probability of a node with smallest degree for same γ we will always increase objective function. If the level of security is higher for highly connected node, the overall security will be worse.

In order to illustrate this fact, we have used simple graph depicted in Figure 6.2.

In Figure 6.3, the sum of infection probabilities as a function of α is depicted. Different strategies for distribution of δ have been considered. Curing vector C is represented as a vector which sum is equal to the number of links in a network multiplied by α. Different strategies are optimal in different regions of α.

What is the strategy to distribute the protection in a network? Should we reduce the probability of infection of the nodes with the highest degree, such that they do not spread infection to their neighbors? At the threshold, curing rates are proportional to the degree, but infection probabilities are all equal. If the same strategy is applied above the threshold, but with the constrain $\sum \delta = \alpha 2L, \alpha < 1$ infection probabilities of all nodes will be the same. The question is if the higher

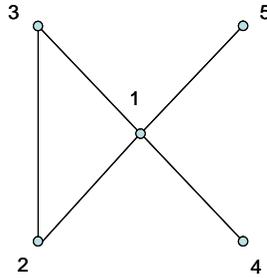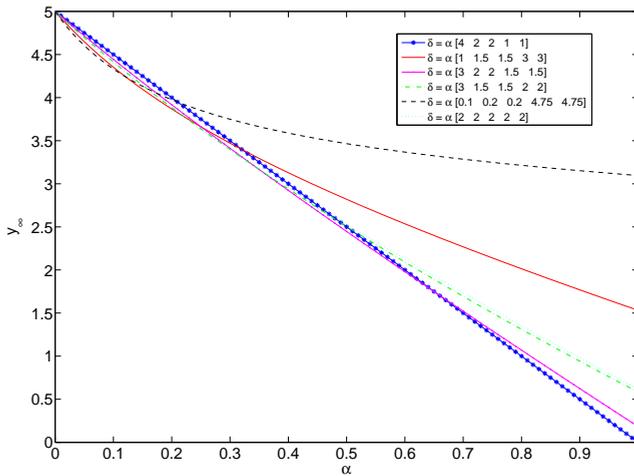**Figure 6.2:** *Example graph with 5 nodes.*

Figure 6.3: *Sum of infection probabilities as a function of α. Curing rate vector is represented by a vector which sum is equal to the number of links in the network multiplied by factor $\alpha \in [0,1]$.*

degree nodes have lower probability of infection does that increase overall security? The following theorem formulates non-intuitive answer.

Theorem 29. *Consider problem (24), Section 6.2. For $0 < \alpha < 1$, if high degree nodes have larger probability of infection than lower degree nodes, the overall probability of infection $\sum v_i$ is smaller.*

Proof. Let us assign factor x_i to each node, such that $\sum x_i = N$ and $0 < x_i < N$. Let us write each infection probability as $v_{i\infty} = \alpha x_i$. The equation $v_{i\infty} = \alpha x_i$ holds only for $\alpha < \frac{1}{x_i}$. Let the vector $X = [x_1 \ x_2 \ \dots \ x_N]$ be a strategy. We will fix strategy vector and compare for which α certain strategy is better. The sum of curing rates is

$$\sum \delta_i = \sum \sum a_{ij} \frac{x_i}{x_j} - \alpha \sum d_i x_j$$

In the case of $\forall i, \ x_i = 1$, we have $\sum \delta_i = 2L - \alpha 2L$. We will compare all the strategies with this one. Lets denote $Z = \sum \sum a_{ij} \frac{x_i}{x_j}$ and $W = \sum d_i x_j$. If we join contributions of two nodes we have a sum over all the links i, j.

$$
\begin{aligned}
Z &= \sum_{all\ links} \left(\frac{x_i}{x_j} + \frac{x_j}{x_i} \right) \\
&= \sum_{all\ links} \frac{x_i^2 + x_j^2}{x_i x_j}
\end{aligned}
$$

We have that $\frac{x_i^2 + x_j^2}{x_i x_j} \geq 2$, thus $Z \geq 2L$. Therefore, for $\alpha = 0$, $\sum \delta_i = Z$. Because Z is always larger than $2L$, W also has to be larger than $2L$, such that the sum of curing rates is smaller than the sum of curing rates in the case with all equal probabilities which is $2L - \alpha 2L$. Because $W = \sum d_i x_j$, it is always better to assign higher infection probability to high degree nodes. □

Investing in the protection of high degree such that they are the most protected nodes above the threshold is not the best strategy in the sense of global optimum.

6.5   Special case - complete bi-partite graph, $\alpha \in (0, 1)$

In the case of complete bi-partite graph, the minimum threshold is reached when each node has protection proportional to its degree. However, there are other optimization problems such as the optimization with two fixed curing rates. Let us consider the heterogeneous case with two curing rates, presented in Chapter 5, section 5.3.

For any bi-partite graph, the threshold for the heterogeneous case is fixed and equal to $\delta^* = \sqrt{MN}$, where δ^* is defined in eq. (5.25). The threshold can be

reached for different values of δ_1, δ_2, p and q. For example, for $(\delta_1 = M, \delta_2 = N, p = 1, q = 0)$ the threshold is reached with δ_1 applied on nodes from set S_1, while for $(\delta_1 = M, \delta_2 = N, p = 1, q = 0)$ the threshold is also reached and the curing rate δ_1 is now applied on the nodes from the other set. The question is how can we decide which solution is better. One of the options is to minimize the total protection strategy applied on the network, while reaching the threshold. The total protection strategy, denoted by $f(p, q, \delta_1, \delta_2)$

$$f(p, q, \delta_1, \delta_2) = \sum_{l=1}^{M+N} \delta_l = Mp\delta_1 + M(1-p)\delta_2 + Nq\delta_1 + N(1-q)\delta_2 \qquad (6.2)$$

For the previous two cases, the total protection strategy is different. In case $(\delta_1 = M, \delta_2 = N, p = 1, q = 0)$, the total protection strategy is $f = (M^2 + N^2)$, and in the other case, $f = 2MN$, which is always smaller than or equal to the first case.

Let us formulate the general optimization problem for complete bi-partite graph:

Problem 30. Minimize

$$f(p, q, \delta_1, \delta_2) = Mp\delta_1 + M(1-p)\delta_2 + Nq\delta_1 + N(1-q)\delta_2 \qquad (6.3)$$

subject to the conditions

$$\sqrt{MN} = \frac{\delta_1\delta_2}{\sqrt{(1-p)(1-q)\delta_1^2 + pq\delta_2^2 + \delta_1\delta_2(p(1-q) + q(1-p))}} \qquad (6.4)$$
$$0 \leq p, q \leq 1$$
$$0 < \delta_1, \delta_2$$

The optimization problem is non-linear with non-linear conditions. However, the minimum of the function f for any graph and any set of curing rates is equal to twice the number of links L in the network

$$f_{\min} = 2L.$$

In the case of the complete bi-partite graph, the minimum is $f_{\min} = 2MN$ and it is reached for $(\delta_1 = M, \delta_2 = N, p = 1, q = 0)$ or $(\delta_1 = N, \delta_2 = M, p = 0, q = 1)$. This means that for $N > M$, the larger curing rate proportional to the number of links in set S_1 will be assigned to the nodes from that set.

Further, we can have a situation, where curing rates (δ_1, δ_2) are fixed and we will optimize the parameters (p, q). This optimization problem can be formulated as follows.

Problem 31. For two fixed curing rates δ_1, δ_2 , minimize function (6.2), subject to the conditions (6.4).

From the threshold condition we can determine one of the variables p or q. We will derive equations for variable q (the case with p is analogue),

$$q = \frac{\delta_1(MN\delta_1(1-p) + MN\delta_2 p - \delta_1 \delta_2^2)}{MN(\delta_1^2(1-p) + \delta_1\delta_2(2p-1) + \delta_1^2)} \tag{6.5}$$

By substituting q in f, the total sum of curing rates becomes a function of parameter p only and optimization is simplified. The function is of the form $f(p) = \frac{P_2(p)}{P_1(p)}$ where $P_1(p)$ is a polynomial of the first order in p and $P_2(p)$ is a polynomial in the second order in p.

Lemma 32. *For any fixed δ_1, δ_2, the optimal solution of minimization problem (31) is on the boundary of the region ($p = 0$ or $p = 1$ or $q = 0$ or $q = 1$).*

Proof. The function $f(p)$ is not defined for $P_1(p) = 0$, which holds for $p = \frac{\delta_1}{\delta_1 - \delta_2}$. The value $\frac{\delta_1}{\delta_1 - \delta_2}$ does not belong to the interval $[0, 1]$. The second derivative of $f(p)$ is strictly negative in the interval $q \in [0, 1]$.

$$\frac{d^2 f(p)}{dp^2} = -\frac{2\delta_1^2 \delta_2^2(\delta_1 - \delta_2)^2}{(\delta_1(1-q) + \delta_2 q)} < 0, q \in [0, 1]$$

Therefore, $f(p)$ is concave in the interval of interest and minimum is on the boundaries of the interval.   □   □

For given δ_1, δ_2, it is not always possible to reach the threshold. In the case $\delta_1, \delta_2 < \sqrt{MN}$, the threshold cannot be reached and the network is in the state of permanent infection. For example, if $\delta_1, \delta_2 < \sqrt{MN}$ and $\delta_1 > \delta_2$, if we take only the larger curing rate for the whole network, we have $\frac{1}{\delta_1} < \frac{1}{\sqrt{MN}}$.

If the threshold can be reached, Lemma 32 shows that either set S_1 or set S_2 is completely protected with only one curing rate. In order to minimize the sum of curing rates we are interested how many times we can apply smaller curing rate. Without loss of generality, let $\delta_1 < \sqrt{MN} < \delta_2$ and $N > M$. Firstly, we will assign δ_2 to all the nodes from larger set with N nodes and δ_1 to the smaller set. If the effective spreading rate obeys $\frac{1}{\delta^*} > \frac{1}{\sqrt{MN}}$, than $p = 1$, and q can be calculated from equation (6.5). In the case $\frac{1}{\delta^*} < \frac{1}{\sqrt{MN}}$, the network is cured and below the threshold. Then $q = 0$ and p can be calculated from the condition for the threshold.

With the previous result, the curing rate space for problem (23), Section 6.1 is bounded in a box $\delta_i \leq \delta_c$, $\delta_c = 2L$, which makes the set space non-empty, compact and convex.

Chapter 7

Protecting against network infections: a game theoretic perspective

Network security has become one of the major challenges of communication networking. Security breaches come in many forms, such as the spread of viruses and worms in the Internet, as well as social engineering compromises and direct exploitation of a hosts vulnerability. In such a breach, an exposed (infected) host becomes a new source of infection, which attacks other unprotected machines. We shall generalize any such breach and model it as an infection process.

In order to overcome such breaches and their implied damage, network users and nodes can be equipped with protection and curing tools, to which we shall refer as protection strategies (or curing strategies). For example, a protection strategy is an antivirus software, with its signature quality and the speed of response to new virus strains.

A major source of complication in network security is the typically autonomous nature of decision making in the network, most notably in the Internet. Indeed, administration and policy enforcement are not possible at the inter-networking level (as opposed to intra-networking within a company), hence a majority of users is left to make independent decisions, including the choice of the protection strategy. Clearly, while such decisions are made autonomously by users and nodes, they do influence other users, through the potential infection processes. This gives rise to a noncooperative game.

Consider an internetwork that includes a company network that has servers with vital data, as well as hosts of individual users that are divided into subnetworks. Suppose that each machine is administrated by an independent decision maker. The companys servers will seek a higher level of protection, due to the importance of the information they contain and the fact that many users hosts

will be able to connect to them. On the other hand, for individual users, the price of tools such as antivirus software and hosts firewall will often be too high compared to the value of the security they provide. Moreover, a user host often has just a small number of neighbors, namely other hosts that can connect to them hence potentially endangering them. Therefore, these hosts would compromise with a lower level of protection, hence decreasing the level of security of the whole network, and, in turn, putting a higher burden on the companys servers. Investigating such a network security game requires a proper model, which captures both the process of infection spread as well as the games structure. We obtain such a model by combining game theoretic principles with epidemic theory.

With the rapid growth of Internet and decentralization of services, the game theoretical framework has become an important tool for network modeling. Game theoretic models have been employed in various networking contexts, such as flow control [62], [63], routing [64], [65], and bandwidth allocation [66]. These studies mainly investigated the structure of the network operating points i.e., the Nash equilibria of the respective games. Such equilibria are inherently inefficient [67] and, in general, exhibit suboptimal network performance. As a result, the question of how much worse the quality of a Nash equilibrium is with respect to a centrally enforced optimum has received considerably attention e.g., [68, 69, 70]. In order to quantify this inefficiency, several conceptual measures have been proposed in the literature. Most notably, the price of anarchy [71] corresponds to a worst-case analysis and it is the ratio between the worst Nash equilibrium and the social optimum.

Recently, network security under a game theoretical setting was considered in [72]. That study addressed the interplay between protection and infection and noted the influence of the underlying topology, however it focused on the case of just two simple strategies, namely being fully protected or totally unprotected. In particular, if a node chooses the "fully protected" strategy, its security level does not depend on those of its neighbors. Somewhat similar work appears in [73, 74], where Lelarge *et al.* generalize game settings to incorporate weak security solutions. However, the problem is tractable only for sparse random graphs and trees. In [75], Jiang *et al.* consider a network security game, where the level of security is determined by weights assigned to a topology and the infection process is not modeled. A framework that is closer to the present study is that of *IDS* (Interdependent security games) [76, 77]. As opposed to [72], in *IDS* games security levels of agents are interdependent even when they choose the "protected strategy". However, the *IDS* framework does not consider the influence of the underlying topology, as it restricts its attention to the case of a complete graph.

7.1   The virus protection game

Consider a network with N nodes defined by an adjacency matrix A. This is an underlying topology over which a virus can spread with an infection rate $\beta = 1$ per

link. Each node i chooses its curing rate among an infinite number of strategies from the interval $\delta_i \in [0, \infty]$, so as to minimize its *cost function* $J^i = c_i \delta_i + v_{i\infty}$, where c_i is a positive value that stands for the *relative price of protection* and quantifies the trade-off of the user between the money (and any overhead) invested in protection and the penalty of being infected. For example, a firm may give much importance to security, hence its relative price of protection would be smaller than that of a private Internet user. Thus, the utility function of a node i is a weighted sum of the curing rate per node, δ_i, and the probability of infection in the steady-state, $v_{i\infty}$.

To sum up, the game has N players, corresponding to the nodes of a graph. Each node i chooses a curing strategy δ_i so as to minimize its cost function $J^{(i)}$. The strategies chosen by all nodes result in a certain steady-sate infection probability for each node, $v_{i\infty}$. The latter is also the percentage of time that the node is in the infected state. We term this game as *the virus protection game*.

A *Nash equilibrium point* (*NEP*) is a strategy profile such that no user can benefit from unilaterally changing its strategy. We shall denote an NEP by a vector $\overrightarrow{\delta} = [\ \delta_1 \quad \delta_2 \quad .. \quad \delta_N\ ]^T$ and a corresponding vector of individual probabilities of infection $V_\infty = [v_{1\infty}, v_{2\infty}, ..v_{N\infty}]^T$. The probability of infection $v_{i\infty}$ depends on the states of other nodes as in equation (4.6) and, therefore, the utility function $J^i(\overrightarrow{\delta}, A)$ depends on the vector of curing strategies and the system (network) parameters.

In section 7.1.3, a simple case with just two nodes and one link illustrates the utility function behavior and the optimization process of individual nodes. An example of a utility function for a network with two nodes is given in Figure 7.1. The utility function of the second node $J^{(2)}$ is calculated for different values of the constants c_1 and c_2. In the case $c_1 > 1, c_2 > 1$, the utility function only increases due to the fact that the protection price is larger than the corresponding security it offers.

In the case $c_1 < 1, c_2 < 1$, the utility function decreases due to the decrease of the infection probability. A network is clean of viruses if the curing rates of all nodes satisfy the threshold relation (5.11). Whether the network is able to reach the threshold depends on the price each node is prepared to pay.

Clearly, it is of interest to establish the existence of an NEP and characterize it. We shall show that a Nash equilibrium always exists. We shall also show that the NEP's quality, in terms of overall network security and protection against viruses, largely depends on the properties of the underlying topology.

7.1.1 Nash Equilibrium

First we indicate that the individual probabilities of infection $v_{i\infty}$ are strictly convex in δ_i. This will be later used to establish the quasi-convexity of the cost function, with which we shall prove the existence of a Nash equilibrium.
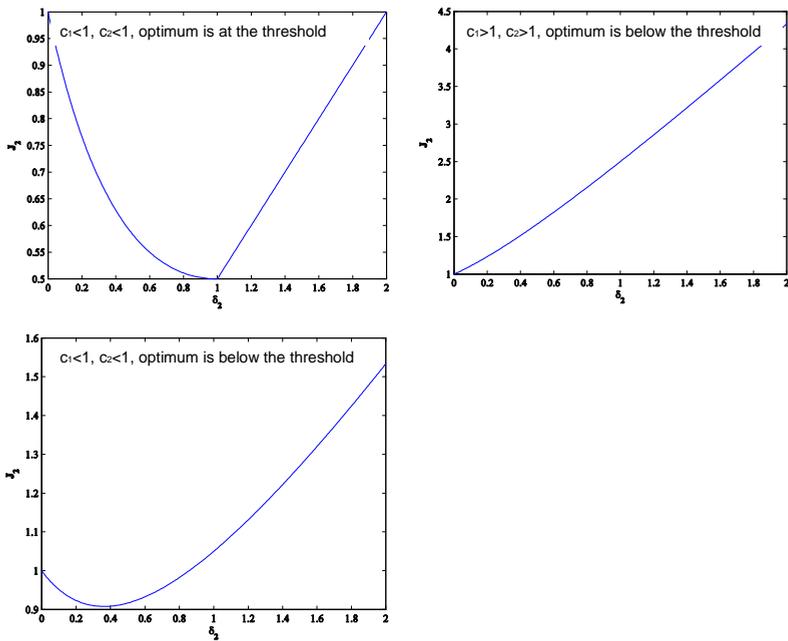
The following result is taken from [78].

Figure 7.1: *Utility function for a network with two nodes and different parameters c_1 and c_2. The curing rate δ_1 of the first node is optimal.*

Lemma 33. *For fixed curing rates of other nodes, the probability of infection* $v_{i\infty}(\delta_i)$ *is a strictly convex function in* δ_i.

Lemma 34. *For the utility function defined as* $J^{(i)}(\delta_i, \delta_{-i}) = c_i\delta_i + v_{i\infty}(\delta_i, \delta_{-i})$, $c_i > 0$, *the function is quasi-convex in each* δ_j, $j = 1..N$.

Proof. Let us first show that $J^{(i)}$ is quasi-convex in δ_j. For any $j \neq i$, the utility function $J^{(i)}$ is quasi-convex in j

$$c_i\delta_i + v_{i\infty}(\delta_i, \lambda\delta_j + (1-\lambda)\delta'_j, \delta_{-i,-j})$$
$$\leq \max\{c_i\delta_i + v_{i\infty}(\delta_i, \delta'_j, \delta_{-i,-j}), c_i\delta_i + v_{i\infty}(\delta_i, \delta_j, \delta_{-i,-j})\}$$

which holds for any c_i. The probability of a node being infected $v_{i\infty}(\delta_i, \delta_{-i})$ is convex function in δ_i. When δ_i reaches the threshold value for the curing rate δ_{ic}, the infection probability becomes zero. The utility function $J^{(i)}$ is a sum of a convex function and a linear - strictly increasing function and, therefore, it is quasi-convex in the domain of interest. $\square$

Theorem 35. *For a set of strategies* $\forall i$ $\delta_i \in [0, \delta_{max}]$ *which is non-empty, compact and convex, and for the continuous and quasi-convex cost function in each* δ_i, *the game has at least one Nash equilibrium.*

Proof. The set of minimizers of a quasi-convex function on a convex set is convex. Continuity of the cost function implies upper-hemicontinuity of the point to set correspondence [79]. $\square$

The existence of an NEP means that the protection game has at least one stable point. We proceed to explore the properties of the Nash equilibria, which indicate the ability of a network to protect itself from epidemics.

7.1.2 Characterization of equilibrium

An NEP can be in two very different regions, namely above or at the threshold (5.11). The NEP does not exist below the threshold: in order to realize why, consider the following example. A node i has a curing rate δ_i and the curing rates of other nodes are fixed. If inequality holds $\lambda_{\max}(A_{\delta_c}) < 1$, the cost function of a node i is $J^{(i)} = c_i\delta_i$. Therefore, a node i is able to reduce the curing rate such that its cost function decreases, because the probability of infection $v_{i\infty} = 0$ will not change.

If the optimum of the cost function is reached at the threshold point, we can have multiple Nash equilibria. At the critical point $\overrightarrow{\delta}_c$, we have that

$$\lambda_{\max}(A_{\delta_c}) = 1 \tag{7.1}$$

This can be easily shown on a simple example of a two-nodes network. This infinite set of NEPs is bounded and we will establish the worst case scenario.

When the network is in the regime above the threshold, numerical calculations suggest that only one equilibrium exists. However, this was not established formally, and the uniqueness of the NEP above the threshold remains an open problem.

In the case of two nodes (see section 7.1.3), a unique Nash equilibrium exists if $\sqrt[3]{c_1 c_2^2} + \sqrt[3]{c_1^2 c_2} > 1$, $(\delta_1 \delta_2 < 1)$. For $\sqrt[3]{c_1 c_2^2} + \sqrt[3]{c_1^2 c_2} \leq 1$, $(\delta_1 \delta_2 \geq 1)$, the example with two nodes shows multiple NEPs.

Next, we determine the influence of the relative price of protection vector $\overrightarrow{c}$ on the Nash equilibria. This shall be later used to bound the equilibrium value of the cost function.

In some cases, all the nodes of a network decide not to protect themselves against infection, leaving the overall network unsecured. If a node is unprotected, i.e., $\delta_i = 0$, the infection probability is always equal to 1 and it does not depend on the curing rates of other nodes.

The next theorem makes a distinction between networks with a vector $\overrightarrow{c}$ such that every node chooses not to be protected at all and networks where the equilibrium point is reached with curing rates larger than 0.

Theorem 36. *In a virus protection game, for a network with N nodes and with cost function for a node i defined as*

$$J^{(i)} = c_i \delta_i + v_{i\infty}$$

the following holds:

1. *If $\forall i \ c_i \geq 1$, the only Nash equilibrium is defined by the curing rate vector $\overrightarrow{\delta} = [\ 0 \quad 0 \quad .. \quad 0 \ ]^T$.*

2. *If $c_i < \frac{1}{d_i}$, where d_i is the degree of a node i, the curing rate of a node i in the Nash equilibrium is different from zero, $\delta_i \neq 0$.*

Proof. Point 1.

Consider any two nodes in the network i, j. Since the network is a connected network with N nodes, at least one node from the pair will be connected to at least one other node. For the same δ_i, δ_j, we can compare the infection probability of neighboring nodes in a network with N nodes $(v_{i\infty}, v_{j\infty})$, with the case of a network with only two connected nodes $v_{i\infty}, v_{j\infty}^{(2)}$. The infection probability of a connected node will increase due to possible connections to infectious nodes and its neighbor will also feel this effect. It holds that

$$v_{i\infty} \geq v_{i\infty}^{(2)}, v_{i\infty} \geq \frac{1 - \delta_i \delta_j}{1 + \delta_i} \tag{7.2}$$

Similarly, we have

$$v_{j\infty} \geq v_{j\infty}^{(2)}, v_{j\infty} \geq \frac{1 - \delta_i \delta_j}{1 + \delta_j}$$

Equality holds for $N = 2$.

For a node i to increase δ_i from zero it has to hold that for some $\delta_i > 0$

$$
\begin{aligned}
J^{(i,\delta_i=0)} &> J^{(i,\delta_i>0)} \\
1 &> c_i\delta_i + v_{i\infty} \\
v_{i\infty} &< 1 - c_i\delta_i
\end{aligned}
\tag{7.3}
$$

and similarly for node j.

From (7.2) and (7.3) we have

$$
\delta_j > c_i\delta_i + c_i - 1
\tag{7.4}
$$

and similarly, for node j it holds

$$
\delta_i > c_j\delta_j + c_j - 1
\tag{7.5}
$$

from (7.4) and (7.5) we have

$$
(1 - c_i c_j)\delta_j > -1 + c_i c_j
$$

which gives for positive c_i, c_j, δ_j

$$
\begin{aligned}
\delta_j &> -1, \ c_i c_j < 1 \\
\delta_j &< -1, \ c_i c_j > 1
\end{aligned}
$$

And similarly we have for δ_i. We can conclude that for $c_i > 1, c_j > 1$ for nodes i and j there is no other solution than $\delta_i = \delta_j = 0$. We can continue the process for any other two nodes in the network concluding that the only solution is $\vec{\delta} = [\ 0 \quad 0 \quad .. \quad 0\ ]$, which proves the first point of the theorem.

Point 2.

For scaled rates such that $\beta = 1$, the infection probability of a node i is

$$
v_{i\infty} = \frac{\displaystyle\sum_{j=1}^{N} a_{ij}v_{j\infty}}{\displaystyle\sum_{j=1}^{N} a_{ij}v_{j\infty} + \delta_i}
$$

The first derivative of the cost function for a node i for $\delta_i = 0$ is

$$
\begin{aligned}
\left.\frac{dJ^{(i)}}{d\delta_i}\right|_{\delta_i=0} &= c_i + \left.\frac{\delta_i \displaystyle\sum_{j=1}^{N} a_{ij}\frac{\partial v_{j\infty}}{\partial \delta_i} - \displaystyle\sum_{j=1}^{N} a_{ij}v_{j\infty}}{\left(\displaystyle\sum_{j=1}^{N} a_{ij}v_{j\infty} + \delta_i\right)^2}\right|_{\delta_i=0} \\
&= c_i - \frac{1}{\displaystyle\sum_{j=1}^{N} a_{ij}v_{j\infty}}
\end{aligned}
\tag{7.6}
$$

which achieves its maximum for $\sum_{j=1}^{N} a_{ij} v_{ji\infty} = d_i$. If $c_i < \frac{1}{d_i}$, the first derivative of the cost function is smaller than zero for any set of curing rates of other nodes. This proves the second point of the theorem. $\square$

Theorem 36 shows that if antivirus software or other means of protection against viruses are too expensive, such that $\forall i \; c_i \geq 1$, the NEP is unique and the network will end up in the completely infected state. In order to steer a decision maker i to chose protection over infection, the relative price should satisfy the inequality $c_i < \frac{1}{d_i}$. The higher the degree of a node, the more it is exposed to infection, hence the required relative price is lower. For example, a large firm typically has many interactions over the Internet and thus its degree is higher. Therefore, its required relative price of antivirus software is lower than that of a smaller firm, which has less opportunities to get infected.

In order to determine the global optimum and the worst case scenario that can happen in a virus protection game, we establish an upper bound on the minimum of the cost function $J_{\min}^{(i)}$.

Lemma 37. *The minimum of the cost function $J_{\min}^{(i)}$ is bounded by $J_{\min}^{(i)} \leq 1$.*

Proof. For curing rate $\delta_i > 0$, $v_{i\infty}$ is bounded. The cost function for $\delta_i = 0$ is $J^{(i)}(\delta_i = 0) = 1$ and the minimum cannot be larger than this value. Therefore, we have

$$
\begin{aligned}
J_{\min}^{(i)} &= c_i \delta_{iopt} + v_{iopt\infty} \leq 1 \\
\delta_{iopt} &\leq \frac{1 - v_{iopt\infty}}{c_i} \leq \frac{1}{c_i}
\end{aligned}
\tag{7.7}
$$

In the case of a network above the threshold, inequality (7.7) holds because the function's minimum cannot be larger than 1 ($J^{(i)}(\delta_i = 0) = 1$). We have

$$
\begin{aligned}
J^{(i)} &= c_i \delta_{iopt} \leq 1 \\
\delta_{iopt} &\leq \frac{1}{c_i}
\end{aligned}
$$

$\square$

If multiple Nash equilibria exist, the curing vector $\overrightarrow{\delta_c}$ is bounded as in Lemma 37. If the relative price of a protection strategy for a node i is too high, the other nodes in the network will have to pay more for the security of the whole network.

7.1.3   Unconstrained case with 2 nodes

For a network with two nodes and one link, each node chooses its strategy out of the interval $\delta_i \in [0, \infty]$. The utility function is defined as $J^{(i)} = c_i \delta_i + v_{i\infty}$. The probabilities of infection follow from (4.6) as $v_{1\infty} = \frac{1-\delta_1\delta_2}{1+\delta_1}$; $v_{2\infty} = \frac{1-\delta_1\delta_2}{1+\delta_2}$. The

Nash equilibrium point (NEP) is reached for $\delta_{1opt} = \sqrt[3]{\frac{1}{c_2 c_1^2}} - 1, \delta_{2opt} = \sqrt[3]{\frac{1}{c_2^2 c_1}} - 1$. If optimal solutions $\delta_{1opt}, \delta_{2opt}$ satisfy $\sqrt{\delta_{1opt} \delta_{2opt}} > 1$, the network NEP will be at the threshold and the cost functions reduce to $J^{(j)} = c_2 \delta_2, J^{(j)} = c_2 \delta_2$. In this case, both nodes will choose smaller curing rates than $\delta_{1opt}, \delta_{2opt}$ such that new values $\delta'_{1opt} \delta'_{2opt}$ satisfy $\sqrt{\delta'_{1opt} \delta'_{2opt}} = 1$. All the solutions that satisfy $\delta'_{1opt} < \sqrt[3]{\frac{1}{c_2 c_1^2}} - 1, \delta'_{2opt} < \sqrt[3]{\frac{1}{c_2^2 c_1}} - 1$ and $\sqrt{\delta'_{1opt} \delta'_{2opt}} = 1$ are optimal and nodes will not change their curing rates. This yields an infinite number of Nash equilibrium points.

In Figure 7.1, for a network with two nodes, the cost function of the second node $J^{(2)}$ is calculated for different values of constants c_1 and c_2.

7.2 Price of anarchy

Clearly, if we could dictate the security strategy of the whole network, we would be able to obtain a better solution. However, as mentioned, the Internet is a decentralized system, and it is challenged by persistent virus infections. Therefore, security of the whole network depends on the decisions of independent users. Yet, is it possible and feasible to completely cure the Internet? How far is the Internet from the global optimal point in the presence of a virus protection game?

In a noncooperative networking game, it is important to know the social welfare attained at the operating points, namely the Nash equilibria. Social welfare is defined as the well-being of the community as a whole. In this specific case of protection game, it is the cost of all the protection in the network. A Nash equilibrium typically exhibits non optimal social welfare. This penalty of selfish behavior is quantified by the price of anarchy (PoA), which is defined as:

$$PoA = \frac{Cost\ of\ worst\ NEP}{Social\ optimum}$$

7.2.1 Social optimum

Social optimum is the solution to the global optimization problem. In Chapter 6, we discussed constrained optimization problem where protection costs are not taken into account. Unconstrained global optimization problem is significantly more complex and we will consider optimization of upper bound on optimal solution presented in Chapter 6. Assume that a "network manager" has the same relative price of security C for all the nodes. In order to compare global optimum, with the NEP we will assume that $C = \frac{1}{N} \sum_{j=1}^{N} c_j$. The corresponding (global) optimization problem is

Minimize

$$J_M = \sum_{j=1}^{N} v_{j\infty} + C \sum_{j=1}^{N} \delta_j$$

For some C, a network will be in the regime at the threshold and $\sum_{j=1}^{N} v_{j\infty} = 0$. The global cost function becomes $J_M = C \sum_{j=1}^{N} \delta_j$. A manager can be interested in optimizing the overall protection, such that $\sum_{j=1}^{N} v_{j\infty} = 0$, which reduces the problem to $J_M = C \sum_{j=1}^{N} \delta_j$.

A network can be in two significantly different states, namely above or at the threshold. Due to the fact that unconstrained global optimization problem is significantly more complex, these two states have to be discussed separately. Thus, we split the optimization problem into two different problems, namely: optimization of the network at the threshold and above the threshold. We assume that a manager optimizes in the regime where the network NEP is, i.e.: if a network NEP reaches the threshold, the manager will optimize with the constraint $\sum_{j=1}^{N} v_{j\infty} = 0$ while if a network NEP is above the threshold, the network manager will optimize the function $J_M = \sum_{j=1}^{N} v_{j\infty} + C \sum_{j=1}^{N} \delta_j$ with the constraint $\sum_{j=1}^{N} v_{j\infty} > 0$. In the case of multiple NEPs, where some are above and others are at the threshold, the network manager optimizes at the threshold.

The network is below the threshold if the curing rates of individual nodes satisfy the inequality

$$\lambda_{\max}(diag(\frac{1}{\delta_i})A) \leq 1$$

If the strict inequality $\lambda_{\max}(diag(\frac{1}{\delta_i})A) < 1$ holds, the vector $\overrightarrow{\delta}$ cannot be a Nash equilibrium point, because there is a point $\delta_j^* < \delta_j$ such that $\lambda_{\max}(diag(\frac{1}{\delta_i^*})A) = 1$. The equality $\lambda_{\max}(diag(\frac{1}{\delta_i})A) = 1$ can be a Nash equilibrium point, if the condition $(\forall i)\frac{\partial J^{(i)}}{\partial \delta_i} = 0$ is satisfied.

Above the critical threshold, the probabilities of infection $v_{i\infty}$ are larger than zero and interesting parameters for the optimization are the sum of infection probabilities and the sum of curing rates $J_M = \sum_{j=1}^{N} v_{j\infty} + C \sum_{j=1}^{N} \delta_j$.

7.2.2 The worst NEP at the threshold

As shown in section 7.1.2, there can be an infinite number of NEPs in this regime. The set of Nash equilibria is bounded in this regime as $(\forall i)\delta_i < \frac{1}{c_i}$, thus the worst NEP is also bounded.

We proceed to determine the worst possible case of an NEP and the global optimal point.

Lemma 38. *The worst case Nash equilibrium, when the network is at the threshold, is bounded by*

$$J_M < C\sum_{j=1}^{N}\frac{1}{c_j} \tag{7.8}$$

where $C = \frac{1}{N}\sum_{j=1}^{N}c_j$.

Proof. Each curing rate is bounded by the constant c_i as in Lemma 37 and the set of Nash equilibria is therefore bounded as in (7.8). $\qquad\square$

The minimum price that has to be paid for a network which is clean of viruses is determined by the number of links, as shown in Theorem 25.

At the threshold, the minimum of the global cost function is a linear function of the number of links in the network $J_M^{(\mathrm{min})} = 2LC$.

For an NEP such that $\delta_1 = \delta_2 = ..\delta_N = \delta_c$ (homogeneous case), the price is $J_M = N\lambda_{\max}(A)C$.

For example, the largest eigenvalue of a line graph is $\lambda_{\max}(A) \simeq 2 = \delta_c$, while that of a star topology is $\lambda_{\max}(A) = \sqrt{N-1} = \delta_c$. These two graphs are interesting examples, as both have the same number of links $L = N - 1$. Thus, in the homogeneous case, the level of protection required for a star topology is significantly higher than for a line topology with the same number of nodes and links. The minimal global price is the same for these two topologies; however, in the homogeneous case, the NEP price is significantly higher for a star topology $(J_M = CN\sqrt{N-1})$, than for a line topology $(J_M = 2CN)$.

At the threshold, the cost of the social optimum is $J_M^{(\mathrm{min})} = C\sum_{j=1}^{N}d_j$. The cost of the worst Nash equilibrium is upper bounded as in lemma 38, under the constraint that the network's NEP is at the threshold, which depends on the vector $\overrightarrow{c}$.

Theorem 39. *The price of anarchy for a network that reaches an NEP at the threshold is bounded by*

$$PoA \leq \frac{\sum_{j=1}^{N}\frac{1}{c_j}}{C\sum_{j=1}^{N}d_j}$$

where $C = \frac{1}{N}\sum\limits_{j=1}^{N} c_j$.

Proof. Follows from Lemma 38 and Theorem 25.   □

It is interesting to note that if nodes regard security as an important issue ($c_j \ll 1$, $\sum\limits_{j=1}^{N} \frac{1}{c_j}$ is large), the price of anarchy can be very high. It is necessary to help the network reach a more efficient NEP, by starting the system from a point close to the optimal.

7.2.3   The worst NEP above the threshold

The network is above the threshold if the curing rates satisfy the inequality $\lambda_{\max}(diag(\frac{1}{\delta_i})A) > 1$. In general, the optimization function is $J_M = C\sum\limits_{j=1}^{N}\delta_j + \sum\limits_{j=1}^{N} v_{j\infty}$. Due to the complexity of the general problem, we will consider the upperbound strategy where $\delta_i = (1-\alpha)d_i, \alpha < 1$. The upperbound was stated in Theorem 28, Chapter 6.

The optimization function reduces to $J_M = C(1-\alpha)\sum\limits_{j=1}^{N} d_j + N\alpha$. This function shows threshold behavior around the point $C = \frac{N}{\sum\limits_{i} d_i}$.

$$
J_M = \begin{cases} N, & C \geq \frac{N}{\sum\limits_{i} d_i} \\ C\sum\limits_{i} d_i & C < \frac{N}{\sum\limits_{i} d_i} \end{cases} \tag{7.9}
$$

For $C < \frac{N}{\sum\limits_{i} d_i}$, the threshold is reached and the cost function is equal to the sum of degrees $J_M = C\sum\limits_{i} d_i < N$. In the case $C \geq \frac{N}{\sum\limits_{i} d_i}$, the optimum is reached for curing rates equal to zero and $J_M = N$.

Compared with the optimization at the threshold, where the cost function minimum can be $J_M^{(\min)} = O(N^2)$ for the complete graph, the cost function minimum cannot be larger than the size of the network $J_M^{(\min)} = O(N)$.

If a network is above the threshold ($v_{i\infty} > 0$), we considered a special case where curing rates are proportional to degrees with the same factor $1-\alpha$ and we can estimate the price of anarchy.

Theorem 40. *For global optimum calculated for curing rates proportional to the*

degrees, the price of anarchy above the threshold is bounded by

$$PoA \leq \begin{cases} \dfrac{\sum\limits_{j=1}^{N} \frac{1}{c_j}}{N}, & C \geq \frac{N}{\sum_i d_i} \\[4ex] \dfrac{\sum\limits_{j=1}^{N} \frac{1}{c_j}}{C \sum\limits_{j=1}^{N} d_j}, & C < \frac{N}{\sum_i d_i} \end{cases}$$

where $C = \frac{1}{N} \sum\limits_{j=1}^{N} c_j$.

Proof. Follows from Lemma 38 and equation (7.9). ☐

7.3 Managing a network by constraining the infection probabilities

We proceed to discuss how a manager can influence and control the Nash equilibria of the virus protection game. In section 7.1.2, we have shown how a Nash equilibrium depends on the relative price of protection vector $\overrightarrow{c}$. If a network is at the epidemic threshold, more Nash equilibrium points exist. By varying the relative price of protection vector $\overrightarrow{c}$ a manager can influence the network equilibrium point. A manager may be able to do that by determining (or affecting, e.g., through subsidies) the cost of protection means, e.g., antivirus software, hence indirectly influencing c_i. Here, the "manager" may be an antivirus supplier, which gives cheaper (per unit) antivirus to entities that have many Internet interactions and are densely connected to other nodes.

In section 7.1.2, Theorem 36, some conditions are introduced that can give guidance to the choice of the relative price of protection. If all $c_i > 1$, there is only an unprotected state, and no one will buy antivirus protection. If $c_i < \frac{1}{d_i}$, a node will always invest some money in protecting itself. These results make it possible for an antivirus supplier to estimate what price will make a network more secure. In Theorem 39, we have seen that too low relative prices can lead a network further away from the global optimum. If large firms invest in expensive security, other nodes can buy cheaper antivirus software such that the network reaches the threshold.

The other option for a manager is to set up upper bounds on infection probabilities, for all relative prices $c_i \geq 1$, which will determine the Nash point as presented in Theorem 41.

Theorem 41. *If $\forall i \; v_{i\infty} \leq B_i$, $\forall i \; c_i > 1$, the only Nash equilibrium is reached for*

$$(\forall i) \; \delta_{i\,\min} = \frac{(1 - B_i) \sum\limits_{j} a_{ij} B_j}{B_i}$$

Proof. The result for the unconstrained case with N nodes shows that a node will tend to decrease its curing rates till they all become terminally infected (Theorem 36). The only NEP is out of the bounded region, thus the feasible minimum will be on the bound such that $\forall i \; v_{i\infty} \leq B_j$. Nodes that are above the bound will tend to decrease their curing rates, which draws other nodes to do the same till they all reach the constraint of infection probability B_j. Thus, the minimum is reached for $\forall i \; v_{i\infty} = B_j$. The minimum point for all the nodes exists, and the corresponding curing rate can be calculated from Equation (4.6), for $v_{j\infty} = B_j$

$$B_i = \frac{\sum\limits_{j} a_{ij} B_j}{\sum\limits_{j} a_{ij} B_j + \delta_i}$$

Now, the curing rates are

$$(\forall i) \; \delta_{i\,\min} = \frac{(1 - B_i)\sum\limits_{j} a_{ij} B_j}{B_i}$$

$\square$

For $B_i \rightarrow 0$ and B_j finite for $j \neq i$, the curing rate of node i will tend to infinity $\delta_i \rightarrow \infty$. For $B_j = B$, $\delta_i = d_i(1 - B)$, where d_i is the degree of a node i, we have the vector of curing rates $\overrightarrow{\delta}$

$$\overrightarrow{\delta_{\min}} = \begin{bmatrix} \beta d_1(1 - B) & \beta d_2(1 - B) & .. & \beta d_N(1 - B) \end{bmatrix}^T$$

However, this is not a stable point. If there is an unfair player in the game, which reduces its security against the rules $v_{i\infty} > B$, it can cause other players to pay more than what was planned. The security of the whole network is harmed.

This result suggests a strategy for steering autonomous systems (ASs), or Internet service providers, to invest money in their own security, which is proportional to the number of "links", that is, interactions they have with other ASs. The way to "force security" upon ASs is by asking a certain fixed probability of infection $v_{i\infty} < B$, for all relative prices $c_i > 1$. Together with the fact that the cheapest threshold, in terms of the total security ($\sum \delta_i$), is reached when the nodes are protected proportionally to their own degrees, this seems to be a very fair way to provide overall security. Bigger ASs with more connections towards other ASs will have to protect themselves more, in order to provide the same level of security, while smaller ASs will invest proportionally to their sizes and profits.

Part III

Influence of quarantine on epidemic spread

Chapter 8

Virus spread in social networks

The protection of important networks is not fast enough in practice, and the infection easily reaches all the segments of network. In this chapter, we propose and analyze a fast method to stop or reduce epidemic spreading on networks. When an epidemic is detected, a network cut is performed by removing links leading to several disconnected clusters of nodes. This clustering allows limited intercommunity communication between nodes to continue, while possibly quarantining the rest of the network. Many real world networks from on-line social networks to airline transport networks and Internet AS network typically show a strong community structure [80], [12]. Depending on the speed of the epidemic reaction, it is possible to totally prevent any risk of infection for a number of disconnected clusters. Even with very delayed reaction, the amount of protection, that has to be applied in the network in order to stop the spreading, can be reduced. Thus clustering can be used in addition to other protection methods.

The removal of links as protection against epidemics was proposed in mathematical epidemiology. The Equal Graph Partitioning (EGP) method uses immunization to remove specific nodes that cut the graph into clusters [81]. However, the immunization takes time, while individual nodes can stop communicating with other nodes immediately after receiving the news about the epidemic. Several authors have studied the reduction of disease spreading using air line restrictions. Goedecke *et al.* [10] and Epstein *et al.* [11] used the Susceptible Exposed Infected Recovered ($SIER$) model and dynamic in time travel restrictions. Marcelino *et al.* [12] used the Susceptible Infected (SI) model together with edge betweenness and Jaccard coefficient to increase in spreading time [12] by 81% by removing 25% of the links. Due to the multicommunity structure of the network with most connected nodes not being the most central, the optimal strategy for flight cancellation is not the removal of nodes (cities), but the removal of intercommunity

flights, which introduced an increase in spreading time [12]. We are interested in specific link removal such that intra-communty communication is preserved. We are not interested in optimizing of the clustering algorithm, but instead in the general improvement of protection that is possible by using a well-defined clustering algorithm.

Several algorithms have been proposed to find network communities. Modularity maximization is the most popular method. Modularity is a quantitative criterion to evaluate how good a graph partition is [82]. It maximizes links within communities, while minimizing the links between them. Modularity maximization is an NP problem, given the exponential number of existing partitions. We use a greedy heuristic proposed by Clauset *et al.* [83] to find an optimal modularity clustering.

In order to quantify the improvements of the network clustering in terms of epidemics, we use the epidemic threshold concept and the N-intertwined model on a large set of networks.

8.1   Quarantine model and networks

The protection method of dividing the network into clusters by removing links will be referred to as *clustering* or *quarantining*. The moment when a network is quarantined determines how many nodes are completely protected, since the virus is not able to infect nodes outside its cluster. In the first case, if we are able to quarantine a network into clusters faster than the virus is spreading, only a single cluster will contain infected nodes. On the other hand, if the virus infects all the clusters before a quarantine takes place there are still benefits, which are discussed in more details in section 8.3. Usually, the effective speed of clustering the network will be somewhere in between.

We discuss the two boundary cases separately. In the first case we determine the size of the clusters, which provides an estimate of how many nodes will never get infected. The size of the clusters also affects the performance of the network. Larger clusters mean that a larger part of the network can continue exchanging information. Second, we show that the epidemic threshold that divides non-infected from infected networks improves in networks that display clustering features.

If the infection is spreading very fast and all the clusters get infected, the number of infected nodes in the metastable state is reduced. We discuss the improvement with the respect to the number of removed links.

To show influence of clustering on epidemic spreading, we use several real-world networks. First, the Internet AS level topology obtained by Route View in 2006 and posted by the University of Oregon is used to illustrate the influence of cutting and virus spread on large infrastructural networks. The network consists of $N = 22,963$ ASs and $L = 48,436$ connections or links. Further, we used an example of a social network between weblogs on US politics recorded in 2005 by Adamic and Glance [84]. The political blog network is shown in Fig. 8.1, with
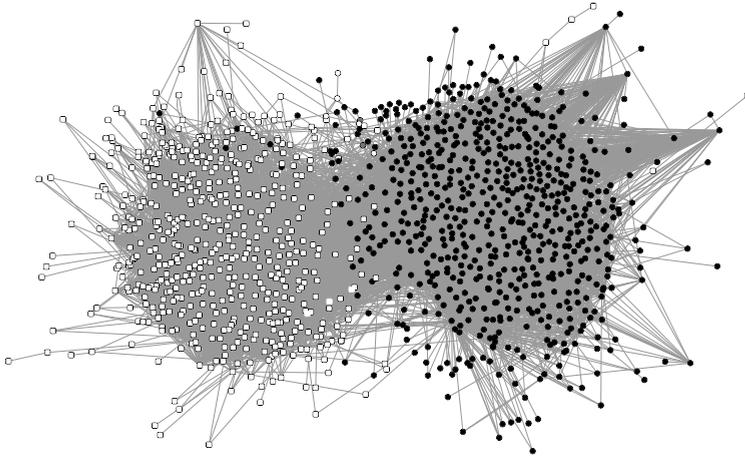
Figure 8.1: *Network of weblogs on US politics clustered network using modularity maximization. Nodes belonging to different clusters are differently colored.*

nodes belonging to different clusters colored in different colors. The network has $N = 1,222$ users and $L = 19,021$ relations between them. Finally, we examine online social network of friends from www.digg.com, gathered by the Delft University of Technology. The digg network has $N = 281,471$ friends and $L = 4,354,174$ connections.

In disease modeling, transport networks are frequently used. To illustrate the influence of traveling patterns on virus spread, we investigate the direct airport-to-airport American traffic network maintained by the U.S. Bureau of Transportation Statistics and European direct airport-to-airport traffic network obtained from European commission for statistics Eurostat. The USA network consists of $N = 2,188$ airports and $L = 31,331$ connections or links. The European network consists of $N = 1,247$ airports and $L = 22,830$ of connections or links.

In order to extend our understanding of the effects of clustering on the network robustness against virus spread, we include several artificial networks with $N = 1,000$ nodes.

We consider three Erdös-Rényi (ER) random graphs with a different number of links. Each node in ER random graph is connected to every other node with probability p. The probability p determines the number of links in the network [85]. We model power law networks using the Barabási-Albert model (BA) of preferential attachment for different number of links [86]. Finally, we use an artificial model of clustered networks [87]. The network is constructed in a similar manner as the ER random graph with two probabilities of link existence, one

for inter-community connections and the other for intra-community connections. We have generated several different networks with $N = 1,000$, two clusters and different modularity. The modularity was set to $Q = 0.17, 0.28, 0.49$ for networks with $L = 10,000$ links and $Q = 0.004, 0.16, 0.49$ for networks with $L = 100,000$. Further, we have considered networks with $4, 6, 8, 10$ clusters. We chose to generate a greater number of networks with two clusters because most of the real-world networks consist of mainly two big clusters.

We additionally consider the square lattice, line, ring and tree topologies.

8.2   Early clustering

In this section, we examine the case of instant clustering where a network is clustered faster than the worm is spreading, resulting in a single infected cluster. Defending the network and performing quarantines provides important advantages. First of all, if a network is cut on time and the infection is limited to one cluster, only a percentage of nodes will eventually be exposed to infection. Second, from the interlacing theorem of graph theory [44], the largest eigenvalue of a subgraph is always smaller than that of the graph. Thus, the thresholds $\tau_c = 1/\lambda_{max}$ will always increase for any subgraph, making the subgraphs more robust to epidemic spreading. The case that all the clusters are initially infected is discussed in section 8.3. Finally, the lifetime of the metastable state depends on the number of nodes [31] as $\Omega(e^{N^\alpha})$, for $\alpha > 0$.

One of the improvements introduced by clustering is a reduction of the largest eigenvalue $\lambda_{\max}$ of the smaller clusters with respect to the original graph. This increases the threshold τ_c, the border between infected and non-infected networks. The ration between the largest eigenvalue of a cluster and the largest eigenvalue of the whole network versus the modularity Q for several networks is shown in Fig. 8.2 and 8.3.

The behavior of $\lambda_{\max Cluster}$ for the different network is diverse. For networks with high modularity, such as the lattice and tree topologies, improvement $(\frac{\lambda_{\max Cluster}}{\lambda_{\max G}})$ is not so significant. For the same type of networks e.g. BA or ER with different number of links, a reduced modularity results in a reduced $\lambda_{\max}$, which is an improvement. For both cases, the modularity is reduced by generating topologies with a larger number of links (by respectively increasing parameter m in BA model and parameter p in ER model). In addition, the difference between the two largest eigenvalues of different clusters is greater for BA than for ER. The effect can be caused by the homogeneity of the degree distribution of clusters in the ER case, while BA shows a significantly heterogeneous cluster degree distribution.

The threshold τ_c $(\frac{1}{\lambda_{\max Cluster}})$ increases as a function of the number of links removed between a cluster and the rest of the network, as shown in Fig. 8.4 and 8.5. In order to preserve as much network communication as possible upon link removal, a small number of links should be removed during the quarantine. On
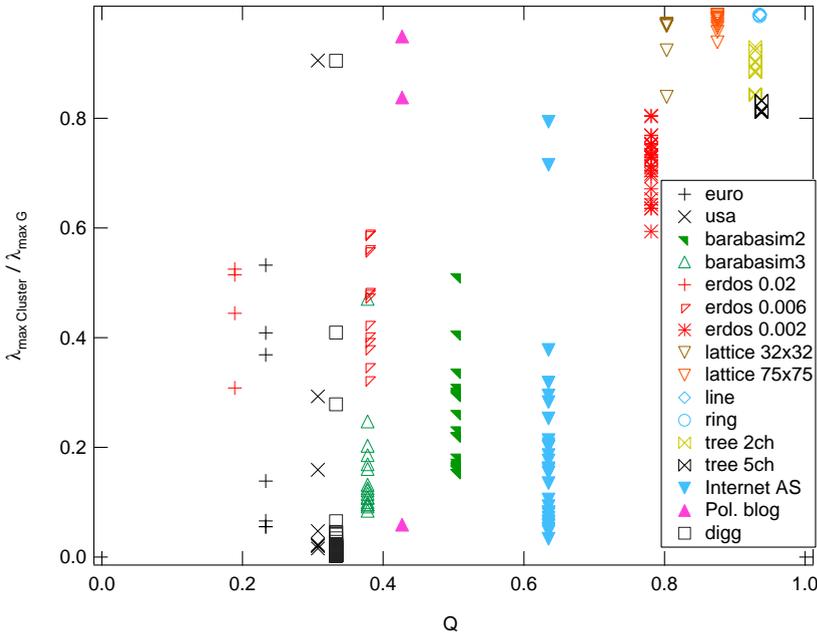
Figure 8.2: *Relative largest eigenvalue of the each cluster $\lambda_{max\,Cluster}/\lambda_{max\,G}$ as a function of the modularity Q for real-world networks and real world models.*

the other hand, τ_c is inversely proportional to $\lambda_{\max Cluster}$. Hence the networks with best performance show clusters with both low $\lambda_{\max Cluster}$ and low L_{out}, close to the point $(0,0)$ in the figures. Real-world networks such as the airline networks and AS network perform well, while artificial networks perform much better the smaller the number of clusters the graph has.

For individual graphs, the dependency of threshold improvement versus the number of links removed is close to linear, which is indicated by change in lower bound on largest eigenvalue $\lambda_{\max} \geq \frac{2L}{N}$. Sparse ER graphs are clustered easily, with small number of removed links, but show no significant improvement of τ_c. The artificial clustered graph with modularity 0.04 shows the worst performance in the number of removed links.

The size of the clusters after cutting is an important variable for the performance of the network. Large clusters will allow for node communication after a quarantine. But on the other hand smaller clusters will be more robust to virus spread. The size of the clusters is decided by the modularity algorithm.

Another parameter to consider is the size of the largest cluster after the quarantine. The distribution of number of cluster sizes is shown in Fig. 8.6. In the case of early clustering, network is cut into clusters before the virus can reach any other cluster except for the one it starts to spread in. The worst case scenario
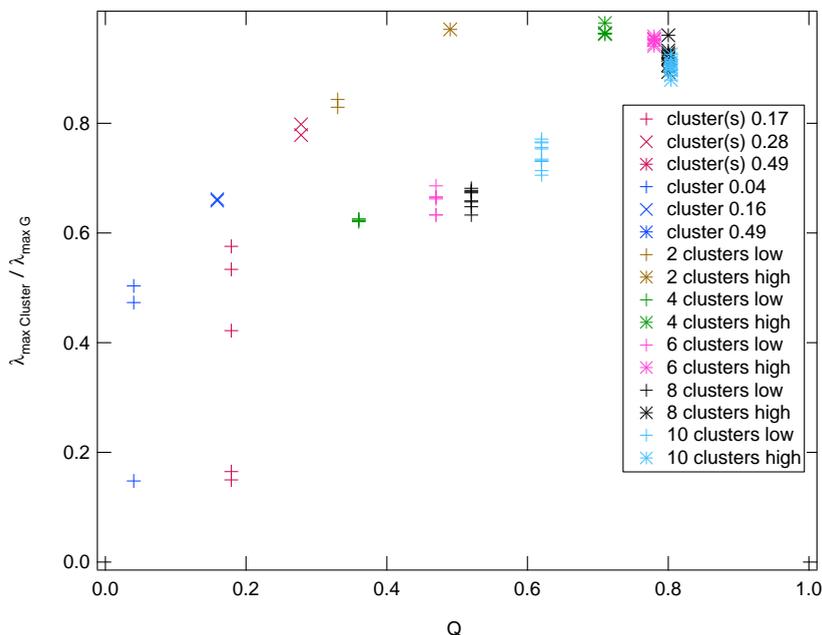
Figure 8.3: *Relative largest eigenvalue of the each cluster $\lambda_{max\,Cluster}/\lambda_{max\,G}$ as a function of the modularity Q for cluster network models.*

is when the virus starts to spread in the largest cluster. Most of the networks have one cluster that contains half of the nodes. In the case of the European air network, three clusters pop up, thus leaving more than two thirds of network protected. A BA graph has many small clusters of the size one fifth of network, which leaves four fifths of network protected, Fig.8.6.

The Digg network has one large cluster which covers the half of the network and many significantly smaller ones. The USA air network and the political blog network have 2 large clusters, while the European air network has 3 large clusters and several small ones. The Internet AS topology is more differentiated. There are 8 clusters with $1,000 - 1,500$ nodes and two larger ones with $3,000$ and $6,000$. Artificial networks show different behavior. ER and BA have a lot of smaller clusters comparable in size. In Fig. 8.7, the number of nodes in the cluster is given as a function of the number of removed links between the cluster and the rest of the network. The air network of USA airports has the largest cluster with the smallest number of deleted links, while the European air network has 3 clusters.

In Fig. 8.8 and 8.9, for the same network, larger clusters tend to have a larger $\lambda_{\max\,Cluster}$ than the smaller clusters. This is, however, not true for any graph: by comparing the line graph of any size with the complete graph of any smaller
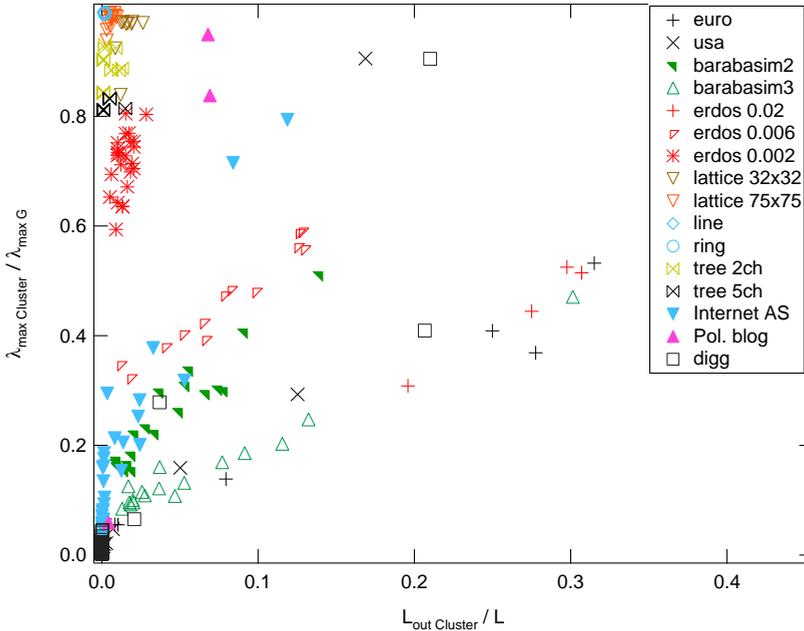
Figure 8.4: *Relative largest eigenvalue of the cluster $\lambda_{max\,Cluster}/\lambda_{max\,G}$ as a function of relative number of links leaving the cluster $L_{out\,Cluster}/L$ real world networks and real world models.*

size.

8.3   Delayed clustering

This section assumes that all the clusters are infected before the quarantine process clustered the network. We examine the performance using the N-intertwined model. In order to clean the infected network, it is necessary to apply a protection/cleaning rate δ such that the effective spreading rate $\tau = \frac{\beta}{\delta}$ is below the threshold $\frac{1}{\lambda_{max}}$. If the network is completely infected and then clustered, the amount of cleaning that is reduced because $\lambda_{\max\,Cluster} \leq \lambda_{\max\,G}$, therefore $\tau_c(G) \geq \tau_c(Cluster)$. Therefore, if the network is clustered, it will be easier to clean the network from infection.

Fig. 8.10 presents the percentage of infected nodes as a function of effective spreading rate τ for different clusters in the network Cluster 0.004.

We calculate the fraction of infected nodes in the clustered network y_{clust} for the effective spreading rate τ for which the number of infected nodes in the original network $y_{tot,50\%}, y_{tot,80\%}$ reaches 50% and 80%. Then, we calculate the
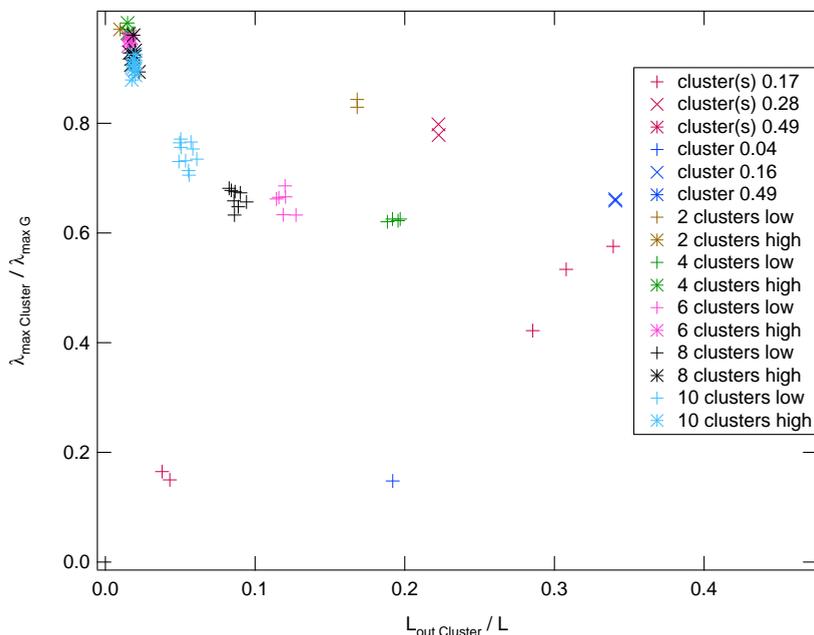
Figure 8.5: *Relative largest eigenvalue of the cluster $\lambda_{max\,Cluster}/\lambda_{max\,G}$ as a function of relative number of links leaving the cluster $L_{out\,Cluster}/L$ for cluster network models.*

difference between the original value and improved one:

$$i_{50\%} = y_{tot,50\%} - y_{clust},\, i_{80\%} = y_{tot,80\%} - y_{clust}$$

We calculate the fraction of infected nodes for several networks. Larger networks as the Internet AS and the Digg network are more computationally demanding and are left out of the analysis. In Fig. 8.11, the upper bound on reduction of infected nodes exhibits the tendency to decrease with the modularity of the graph. The improvement is different when there are 50% and 80% of infected nodes in the original network. Air travel networks and ER networks with small average degree do not show significant difference between improvements and have generaly small improvements.

The number of infected nodes decreases with the increase of the number of removed links in the whole network, shown in Fig. 8.12. This is not surprising because the power of spreading in a network decreases with links removal. Real-world networks do not show a significant reduction in number of infected nodes.
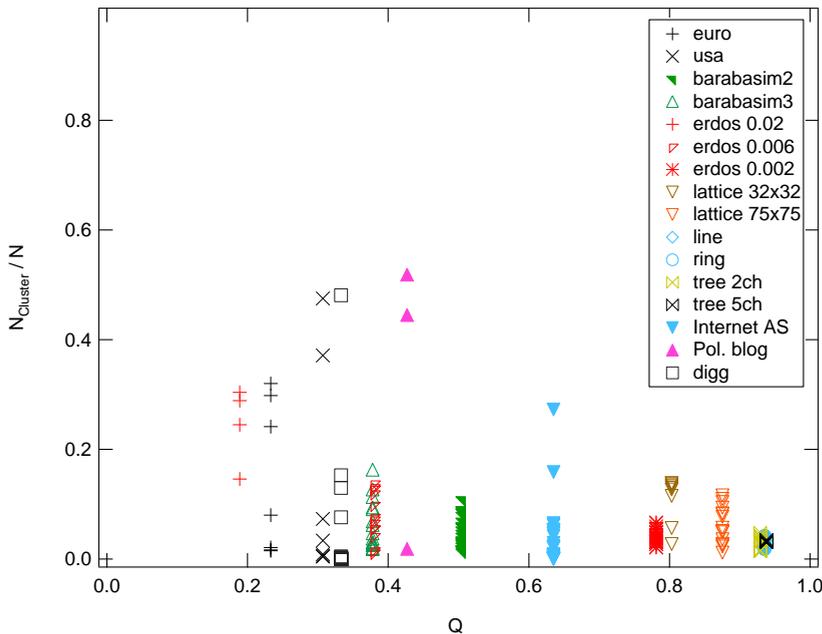
Figure 8.6: *Relative number of nodes in the cluster $N_{Cluster}/N$ as a function of modularity Q real world networks and real world models.*

8.3.1 Random removal of nodes

In this section, we compare the threshold τ_c between quarantined networks with networks where the same number of links has been randomly removed. We give the largest eigenvalue of the original graph $\lambda_{\max G}$, the size of the giant connected component $\frac{N_{rand.big.comp}}{N_G}\%$, its largest eigenvalue $\lambda_{\max rand}$, the size $\frac{N_{big.clust}}{N_G}\%$ and the largest eigenvalue $\lambda_{\max l.Clust}$ of the largest cluster in the clustered network in Table 8.1. Links are removed at random and the average over many simulations of the largest eigenvalue of the largest connected component is calculated together with the variance of the largest eigenvalue.

The results are presented in Table 8.1. A large part of the network remains connected and can transmit infection, which is an expected result of random link removal. Between 80% and 90% of the network can be affected compared with at most 50% in case of clustering. Further, the largest eigenvalue of the largest cluster is still smaller than that of the large component in the case of random link removal.

In USA airlines network, ER with $p = 0.002$ and $p = 0.006$ some smaller cluster have larger $\lambda_{\max Cluster}$. In ER graphs and the political blog, two or more components similar in size have the same or close largest eigenvalue. In the case of political blog the advantage of clustering over random link removal lies in the
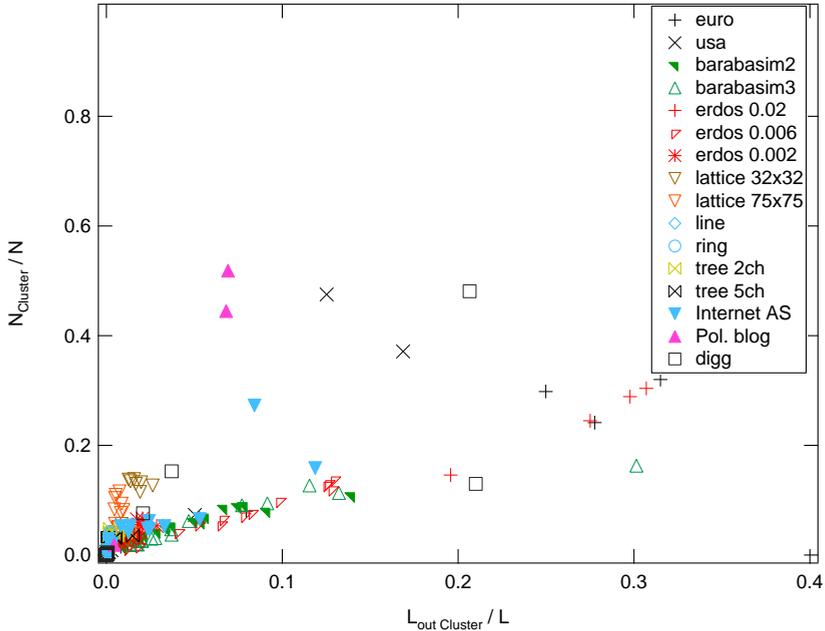
Figure 8.7: *Relative number of nodes in the cluster $N_{Cluster}/N$ as a function of relative number of links leaving the cluster $L_{out\,Cluster}/L$ for real world networks and real world models.*

fact that the other half of the nodes will not get infected if the clustering is performed before the virus has spread. In the case of AS Internet topology the smaller cluster of $N = 3,600$ nodes also has a larger $\lambda_{\max\,Cluster}$ than the largest cluster of $6,200$ nodes. The Digg network also has smaller cluster of $N = 36,491$ nodes with the largest eigen value $\lambda_{\max\,Cluster} = 701.61$, while all the rest of the network has significantly smaller largest eigenvalue. In the case of cluster 28s and 49s, two disconnected components have the same largest eigenvalue, which is the same as for random removal.

The variance of largest eigenvalue for different simulations of random link removal is less than 0.2 in all cases.

8.4 Discussion of results

When dividing the network into clusters, a virus can be stopped and annihilated faster. However, protection comes with a cost. Shutting down links from the network reduces the communication and reachability of nodes in the network. Assuming that the graph is disconnected only temporally, we calculate the price of quarantine as the number of links that are removed from the graph as a result
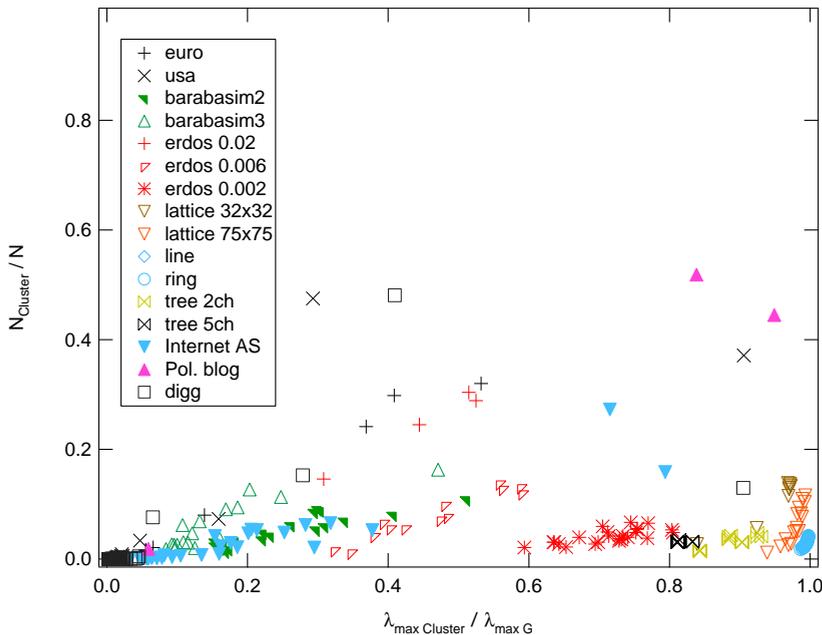
Figure 8.8: *Relative number of nodes in the cluster $N_{Cluster}/N$ as a function of relative largest eigenvalue of the cluster $\lambda_{max\,Cluster}/\lambda_{max\,G}$ for real world networks and real world models.*

of a modularity clustering.

The number of removed links varies from 0.4% to 60%. Most of the considered networks have around 50% of removed links which is significant. In networks where a small number of links is removed, no significant improvement in largest eigenvalue and number of infected nodes is found in the steady-state.

Although the modularity maximization algorithm is popular [82], it has not passed a rigorous theoretical examination. The question is also how good is its resulting clustering. We have not examined other algorithms that may perform differently, because we have concentrated on keeping the communities intact.

The largest eigenvalue improvement using the modularity algorithm is comparable with random links removal for several networks; however, in this case the worm can spread to 90% of the network.
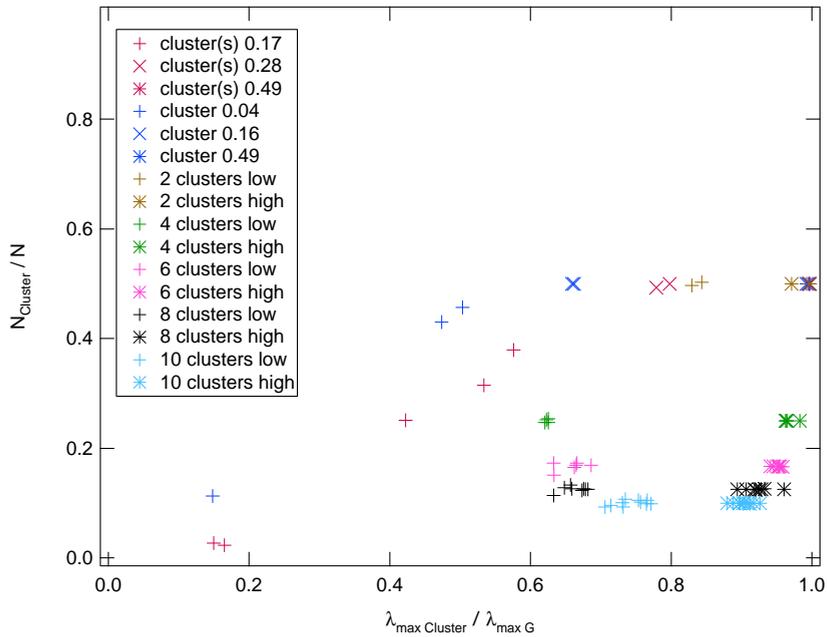
Figure 8.9: *Relative number of nodes in the cluster $N_{Cluster}/N$ as a function of relative largest eigenvalue of the cluster $\lambda_{max\,Cluster}/\lambda_{max\,G}$ for cluster network models.*

Network	$\lambda_{\max G}$	$\frac{N_{rand.big.comp}}{N_G}\%$	$\lambda_{\max rand}$	$\frac{N_{big.clust}}{N_G}\%$	$\lambda_{\max l.Clust}$
Euro	80.92	83.23	53.48	31.99	43.07
USA	144.61	96.51	118.67	47.54	42.36
BA 2m	16.09	85.50	12.01	10.08	8.22
BA 3m	28.11	88.40	20.41	16.30	13.24
Cluster 0.17s	22.88	100	11.77	37.9	13.17
Cluster 0.28s	23.51	100	18.41	50.0	18.77
Cluster 0.49s	25.32	100	25.23	50.00	25.26
ER 0.002	3.59	83.41	3.29	6.68	2.67
ER 0.006	7.23	93.2	4.29	13.7	4.03
ER 0.02	20.93	100	10.05	30.4	10.77
AS '06	71.61	90.59	58.49	27.27	51.22
Pol. Blog	74.08	99.01	69.88	51.88	62.11
Digg	775.33	92.7	582.11	48, 13	317.32

Table 8.1: *Comparison of the random links removal strategy with clustering strategy - largest eigenvalue of largest connected component and largest cluster.*
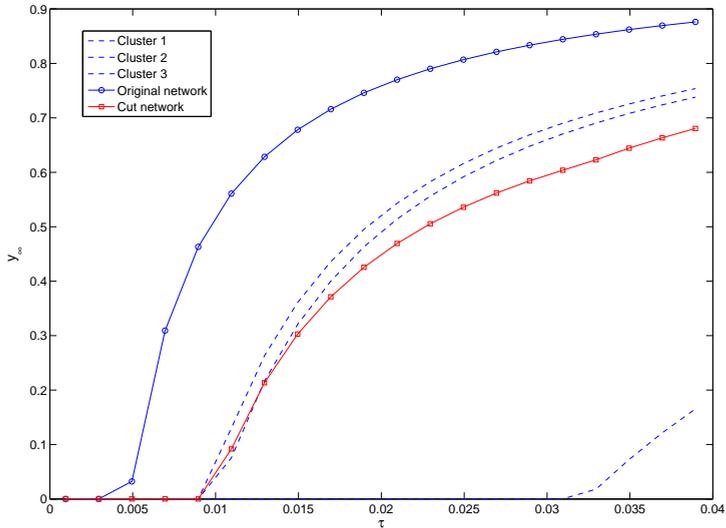
Figure 8.10: *Percentage of infected nodes y_∞ as a function of effective spreading rate τ for original network and clustered network.*

Network	N	L_{tot}	$L_{removed}\%$
Euro	$1,247$	$14,952$	47.27%
USA	$2,179$	$31,326$	18.11%
BA 2m	1000	$1,971$	42.46%
BA 3m	1000	$2,673$	58.88%
ER 0.002	808	980	17.34%
ER 0.006	1000	$3,054$	51.27%
ER 0.02	1000	$9,938$	55.02%
AS '06	$22,963$	$48,436$	20.62%
Pol. Blog	$1,222$	$19,021$	7.16%
Digg	$281,471$	$4,354,174$	25.02%

Table 8.2: *Network cost, the number of removed links.*

Figure 8.11: *The difference between number of infected nodes in the original network and clustered network as a function of modularity in the case when 50% and 80% of nodes are infected in original network.*
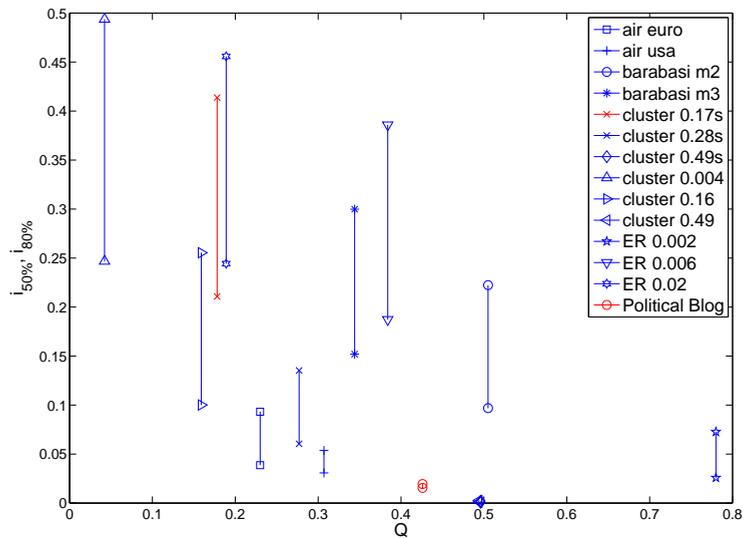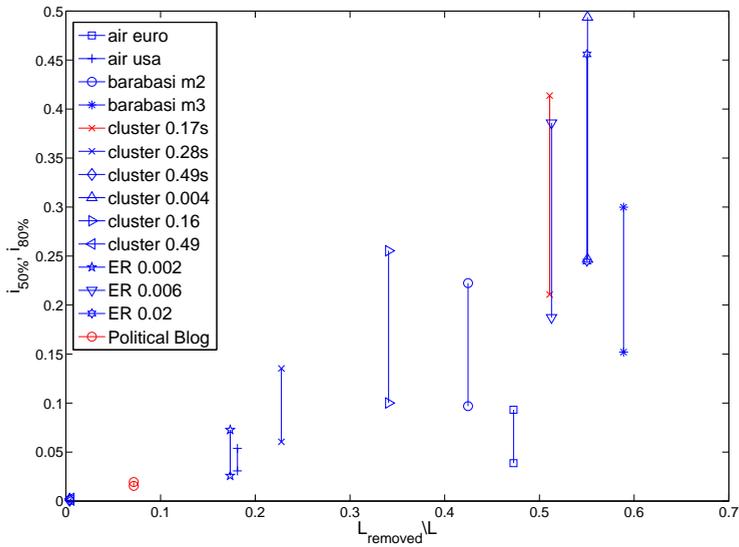
Figure 8.12: *The difference between number of infected nodes in the original network and clustered network as a function of relative number of removed links in the case when 50% and 80% of nodes are infected in original network.*

Chapter 9

Conclusions

The robustness of the infrastructure against failures and attacks has motivated the analysis of an epidemic spreading process in a given, fixed network, represented by the adjacency matrix A. Individual interactions are not homogeneous, but dictated by the structure of the network. Models of spreading processes should take the network topology into account.

We studied the N-intertwined SIS epidemic model. The exact 2^N state Markov chain model was compared with introduced N-intertwined model, whose only approximation lies in the application of mean field theory. The exact Markov chain provides insight into the virus spread process (the time of convergence to the absorbing state) for two boundary cases – the line graph and the complete graph. The N-intertwined model relates network topology parameters to the spreading process (largest eigenvalue and degrees of the nodes). The influence of the mean field approximation is quantified. Several upper bounds for the steady-state infection probabilities are presented.

N-intertwined model reduces for regular graphs to the basic Kephart and White epidemiological model after additional simplifications. We have explored the phase transition phenomenon and shown that, for a fixed graph, the epidemic threshold τ_c is consequence of the mean field approximation. We have presented the relation between spreading rate τ and convergence time towards the extinction of epidemics for two extreme cases (full mesh and line graph). This is especially important for smaller epidemics where τ is close to the epidemic threshold and where the lifetime of an epidemic varies significantly. The largest eigenvalue of the adjacency matrix of the graph is rigorously shown to define an epidemic threshold of the N-intertwined model (as well as of other mean field models).

As a special case, we have studied the spread of viruses on the complete bipartite graph $K_{M,N}$. Using N-intertwined model we have calculated the average number of infected nodes in the steady state and confirmed these results by means of simulations. In addition, the model was improved by introduction of infection delay. We also presented a heuristic for the prediction of the extinction probability

in the first phase of the infection. Simulations show that for the case without infection delay this time dependent heuristic is quite accurate.

The heterogeneous N-intertwined virus spread model has been described and analyzed in the steady-state. Since it applies to any network and any combination of node infections and curing vectors, we believe that the heterogeneous N-intertwined virus spread model is useful for a wide range of practical infection scenarios in networks, from computer viruses to epidemics in social networks and in nature. The critical threshold regime is investigated using generalized Laplacian. The largest eigen value of non-symmetric infection matrix was bounded. An upper and lower bounds for the steady-state infection probabilities are presented.

We considered two special cases in more details, namely the regular graph and the complete bi-partite graph. Using Lyapunov's stability theorem, we have shown that for regular graphs, the epidemic threshold is inversely proportional to the degree of nodes and directly proportional to the harmonic mean of curing rates. Further, we have considered the heterogeneous case with 2 curing rates for the complete bi-partite graph.

We discussed global protection optimization problem. We formulated inverse problem which is a sum of ratios fractional programming problem in the case above the threshold. We determined the global optimum for the case at the threshold. An upper bound on the global optimum above the threshold is presented. For the case above the threshold, it was shown that highly connected nodes should have higher or equal probability of infection compared to low degree nodes if the network reaches the global optimum. At the threshold, the minimum of the social cost function is $O(L)$, where L is the number of links in the network.

We studied special case of optimization problem in complete-bipartite graph. For the case of fixed curing rates, the distribution is convex function in fractions p, q.

In Chapter 7, a novel framework for network security under the presence of autonomous decision makers was presented. We have established the existence of a Nash equilibrium point (NEP) and investigated its properties. In particular, we showed that, when the price of protection is relatively high (namely, $\forall i \ c_i \geq 1$), the only equilibrium point is that of a completely unprotected network; while if this price is sufficiently low for a node (namely, $c_i < \frac{1}{d_i}$), it will always invest in protecting itself.

A network can be in two significantly different regimes, namely above or at the threshold. If a network reaches Nash equilibrium at the threshold, multiple equilibria may exist. The question of uniqueness of the Nash equilibrium above the threshold remains an open question.

Although the optimal value of the social cost is the same for networks with the same number of links L, the non-optimal distribution of curing rates at an NEP results in much worse social welfare in some topologies (e.g., a star graph) than

in other topologies (e.g., a line graph). When optimizing above the threshold, we considered a specific case, for which we showed that the global utility function is always smaller than the number of nodes in the network. This specific case provides some insight on the social performance in the general case.

We have also proposed two methods for steering the network equilibrium, namely by influencing the relative prices and by imposing an upper bound on infection probabilities.

Chapter 8 we studied the influence of protection measures on epidemic spread. We studied how clustering/quarantining influence different network from artificially generated to several real world examples. We have found that real-world networks tend to show a better epidemic threshold τ_c after clustering than artificially generated graphs.

For all the networks under study, the curing rate can improve between 29% an 83% for the largest connected component with respect to the original graph. This wide range of values demonstrates the effect of the network topology on the virus spread. Regarding the network clustering features, an easily clustered graph does not guarantee a slower epidemic threshold, but the way the links intertwine between inter- and intra-communities are key.

Overall, network protection against cascading failures can be improved for any kind of graph. However, the number of removed links in a graph in order to apply a quarantine has been shown to be in the range of 7% to 58% of the links. These values are, in practice, unacceptably high. The advantages of early quarantine are shadowed by the fact that up to half of the links must be shut down for the quarantine to take effect.

The real-world networks have typically two or three big clusters and several smaller ones, while BA and ER graphs have several smaller ones comparable in size. BA and ER graphs are assumed to model the real-world complex networks. However, in respect of the size of the clusters, BA and ER fail to match real world networks.

Additional to the epidemic spread analysis, this diversity in results appears valuable to create a general classification of types of networks. The degree distribution of the graph has been so far widely used for this purpose. For instance, a network classification could be generated by taking the largest eigenvalue of the adjacency matrix of clusters $\lambda_{\max Cluster}$ vs. L_{out} distributions as an input.

The clustering with random removal of links has led us to conclude that the largest eigenvalue of the largest cluster can be less or comparable to the largest eigenvalue of the biggest component generated by random links removal. However, other clusters have a significantly smaller largest eigenvalue, which leads to a smaller amount of cleaning necessary to completely remove the worm from the network. Furthermore, if only the largest cluster is infected only up to 50% of the network will need cleaning.

9.1   Future Work

The results in this thesis provide a strong foundation for future work. Each chapter provides several research directions.

The appearance of several worms that use social networking Web sites to spread introduces the need of accurate relation graph in computer worms modeling. The N-intertwined model can be applied to social website worms and compared with the data. However, gathering data on computer malware is bounded by legal issues.

The N-intertwined model is derived under the mean-field approximation and the complete view of the process is only possible using the exact 2^N-state model. The size of the matrix is an unavoidable obstacle, however the specific structure of the infinitesimal generator matrix can give insights into the time evolution of epidemics.

Another line of research is related to heterogeneous N-intertwined model, which models a wide range of practical infection scenarios in networks, from computer viruses to epidemics in social networks and in nature. The interplay between topology and curing and protection parameters can be explored in more details. For example, two structurally different networks can have the same infection probability vector V by applying adequate infection and curing rates. It should be examined whether any graph structure can be simulated by only applying different curing and protection rates on a complete graph.

Game theoretical framework of epidemic spreading opens more questions than it answers. By fixing all relative price to $c_i = c$, the topology influence becomes dominant and reflects on the game outcome. Further more, topology can be diluted i.e. links and nodes. For example the game can be steered by deleting a limited number K of links or nodes. The nodes play the usual game, namely: each node i buys immunity in a way that minimizes its cost function J_i. The goal is to identify a set of K such links or nodes, which maximizes the global security at the Nash equilibrium.

Another method to steer the game outcome is to set bounds on the risk from infection. For example, players can agree on the upper bound on the probability of each player getting infection. This can lead to a construction of a distributed algorithm for $P2P$ network security. Other algorithms or rules which can improve network security in distributed manner should be explored.

Chapter 6 considers the constrained optimization problem above the threshold. Numerical simulation on a set of real world graphs can add new insights and approaches to the problem. Another question is if two graphs have the same number of links L, nodes N and the same total curing rates $\sum \delta_i = H_1$, is there a specific structure such that it has the most optimal (minimal) average number of infected $\sum v_i$. Such structures can be considered as more robust against virus spread. Starting point would be the fact that on the threshold point two graphs with the same number of links and nodes have the same total curing rate $\sum \delta_i = 2L$, (Theorem 25 of this thesis).

In chapter 8, we considered modularity to be the partitioning algorithm, but there exist a large number of partitioning algorithms that try to optimize different variables. The investigation of how different clustering algorithms affect the epidemic dynamics is an open question. This section also poses the question what other protection algorithms can be applied to specific computer networks.

All the above mentioned research lines are directly derived from the thesis results. A completely new direction would be the application of the N-intertwined model to general risk analysis framework. Contemporary computer networks are very complex and interdependent. System (computer or program) is modeled as a black box with interfaces that connect it with the other components and interactions between components is modeled with Markov intertwined theory. Protection of these boxes can be viewed as 'curing' of malfunctions. It is important to capture service/system dependencies that are present in contemporary complex computer networks.

Bibliography

[1] D. J. Daley and J. Gani, *Epidemic modelling: An Introduction*. Cambridge University Press, 1999.

[2] J. O. Kephart and S. R. White, "Measuring and modeling computer virus prevalence," *Security and Privacy, IEEE Symposium on*, 1993.

[3] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, no. 14, pp. 3200–3203, Apr 2001.

[4] L. Briesemeister, P. Lincoln, and P. Porras, "Epidemic profiles and defense of scale-free networks," in *WORM '03: Proceedings of the 2003 ACM workshop on Rapid malcode*. New York, NY, USA: ACM, 2003, pp. 67–75.

[5] E. M. Jin, M. Girvan, and M. E. J. Newman, "Structure of growing social networks," *Physical Review E*, vol. 64, no. 4, pp. 046 132+, 2001.

[6] M. E. J. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," *Phys. Rev. E*, vol. 66, no. 3, p. 035101, Sep 2002.

[7] P. T. Eugster, R. Guerraoui, A. Kermarrec, and L. Massouli, "From epidemics to distributed computing," *IEEE Computer*, vol. 37, pp. 60–67, 2004.

[8] D. Chakrabarti, J. Leskovec, C. Faloutsos, S. Madden, C. Guestrin, and M. Faloutsos, "Information survival threshold in sensor and p2p networks." in *INFOCOM*. IEEE, 2007, pp. 1316–1324.

[9] E. C. Jr., Z. Ge, V. Misra, and D. Towsley, "Network resilience: Exploring cascading failures within bgp," in *Allerton Conference on Communication, Control and Computing*, October 2002.

[10] D. M. Goedecke, G. V. Bobashev, and F. Yu, "A stochastic equation-based model of the value of international air-travel restrictions for controlling pandemic flu," in *WSC '07: Proceedings of the 39th conference on Winter simulation*. Piscataway, NJ, USA: IEEE Press, 2007, pp. 1538–1542.

[11] J. M. Epstein, D. M. Goedecke, F. Yu, R. J.Morris, D. K. Wagener, and G. V.Bobashev, "Controlling pandemic flu: The value of international air travel restrictions," *PLoS ONE*, vol. 2, no. 5, p. e401, 2007.

[12] J. Marcelino and M. Kaiser, "Reducing influenza spreading over the airline network," *PLoS Currents Influenza*, Aug 2009.

[13] W. O. Kermack and A. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London*, vol. A, no. 115, p. 700721, Aug 1927.

[14] P. Whittle, "The outcome of a stochastic epidemic — a note on bailey's paper," *Biometrika*, vol. 42, p. 116122, 1955.

[15] D. D. J. and G. J., "A deterministic general epidemic model in a stratified population," *Probability, Statistics and Optimisation: A Tribute to Peter Whittle*, p. 11132, 1994.

[16] N. N. B. (Jr.) and D. P. Sankovich, "N.n.bogolyubov and statistical mechanics," *Russian Mathematical Surveys*, vol. 49, no. 5, p. 19, 1994.

[17] P. V. Mieghem, *Performance Analysis of Communication Systems and Networks*. Cambridge University Press, 2006.

[18] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, "On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations," *Journal of Mathematical Biology*, vol. 28, no. 4, pp. 365–382, June 1990.

[19] B. Reid, "Reflections on some recent widespread computer break-ins," *Computers under attack: intruders, worms, and viruses*, pp. 145–149, 1990.

[20] B. Krebs, "A short history of computer viruses and attacks," Washington Post, Tech. Rep., 2003.

[21] M. Landesman, "A may 2007 / may 2008 state of the web comparative," Scansafe Security Brief, Tech. Rep., 2008.

[22] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York, NY, USA: Cambridge University Press, 2004.

[23] C. C. Zou, D. Towsley, and W. Gong, "On the performance of internet worm scanning strategies," *Elsevier Journal of Performance Evaluation*, vol. 63, pp. 700–723, 2003.

[24] S. Kamkar, "Technical explanation of the myspace worm," http://namb.la, Tech. Rep., 2005.

[25] M. Ciubotariu, "Worm js.yamanneram," Symantec, Tech. Rep., 2006.

[26] A. Lelli, "Worm js.twettir," symantec, Tech. Rep., 2009.

[27] M. M. P. Center, "Technical details on worm:win32/koobface.gen!d," Microsoft, Tech. Rep., 2009.

[28] M. Keeling, "The implications of network structure for epidemic dynamics," *Theoretical Population Biology*, vol. 67, no. 1, pp. 1 – 8, 2005.

[29] A. Barbour and D. Mollison, "Epidemics and random graphs," *Gabriel, J.P., Lefevre, C., Picard, P. (Eds.), Stochastic Processes in Epidemic Theory*, p. 8689, 1990.

[30] C. Moore and M. E. J. Newman, "Epidemics and percolation in small-world networks," *Phys. Rev. E*, vol. 61, no. 5, pp. 5678–5682, May 2000.

[31] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proceedings of IEEE INFOCOM*, 2005, pp. 1455–1466.

[32] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: An eigenvalue viewpoint," in *Reliable Distributed Systems, IEEE Symposium on.* Los Alamitos, CA, USA: IEEE Computer Society, 2003.

[33] M. Draief, A. Ganesh, and L. Massoulié, "Thresholds for virus spread on networks," in *valuetools '06: Proceedings of the 1st international conference on Performance evaluation methodolgies and tools.* New York, NY, USA: ACM, 2006, p. 51.

[34] M. Garetto and W. Gong, "Modeling malware spreading dynamics," in *Proceedings of IEEE INFOCOM*, 2003, pp. 1869–1879.

[35] H. Hermanns, *Interactive Markov Chains: The Quest for Quantified Quality*, ser. Lecture Notes in Computer Science. Springer, 2002, vol. 2428.

[36] C. A. Massachusetts and C. Verghese, "The influence model: A tractable representation for the dynamics of networked markov chains," in Dept. of EECS. 2000, Tech. Rep., 2000.

[37] P. L. Nikoloski Z., Deo N., "Correlation model of worm propagation on scale - free networks," *Complexus Network Modelling*, vol. 3, pp. 169–182, 2006.

[38] G. F. D. Duff and D. Naylor, *Differentiation Equations of Applied Mathematics.* New York: Wiley, 1966.

[39] G. Canright and K. Engoe-Monsen, "Spreading on networks: a topographic view," in *Proceedings of the European Conference on Complex Systems*, nov 2005.

[40] R. Durrett and X.-F. Liu, "The contact process on a finite set," in *The Annals of Probability*, vol. 16, no. 3, 1988.

[41] P. Van Mieghem, J. Omic, and R. Kooij, "Epidemic spreading in networks," *IEEE Transactions on Networking*, pp. 1–14, Feb 2009.

[42] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.

[43] J. Omic, R. Kooij, and P. Van Mieghem, "Virus spread in complete bi-partite graphs," in *Proceedings of BIONETICS*, 2007.

[44] D. M. Cvetkovic, M. Doob, and H. Sachs, *Spectra of Graphs, Theory and Applications, third edition*. Johann Ambrosius Barth Verlag, Heidelberg, 1995.

[45] Y. Wang and C. Wang, "Modeling the effects of timing parameters on virus propagation," in *WORM '03: Proceedings of the 2003 ACM workshop on Rapid malcode*. New York, NY, USA: ACM, 2003, pp. 61–66.

[46] N. Biggs, *Algebraic graph theory*. Cambridge University Press, 1974.

[47] F. R. Gantmacher, *The Theory of Matrices*. New York, NY, USA: Chelsea, 1959, vol. 2.

[48] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, UK: Claredon-Press, 1965.

[49] P. V. Mieghem, "A new type of lower bound for the largest eigenvalue of a symmetric matrix," *Linear Algebra and its Applications*, vol. 427, no. 1, pp. 119 – 129, 2007.

[50] S. Walker and P. V. Mieghem, "On lower bounds for the largest eigenvalue of a symmetric matrix," *Linear Algebra and its Applications*, vol. 429, no. 2-3, pp. 519 – 526, 2008.

[51] P. H. J. Guckenheimer, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer-Verlag, New York, 1983.

[52] H. Frenk and S. Schaible, "Fractional programming," in *Encyclopedia of Optimization*, C. A. Floudas and P. M. Pardalos, Eds. Springer, 2009, pp. 1080–1091.

[53] C. Borgs, J. Chayes, A. Ganesh, and A. Saberi, "How to distribute antidote to control epidemics," *to be published*, 2009.

[54] R. W. Freund and F. Jarre, "Solving the sum-of-ratios problem by an interior-point method," *Journal of Global Optimization*, vol. 19, no. 1, pp. 83–102, 2001.

[55] J. E. Falk and S. W. Palocsay, "Image space analysis of generalized fractional programs," *Journal of Global Optimization*, vol. 4, no. 1, pp. 63–88, 1994.

[56] H. Konno and N. Abe, "Minimization of the sum of three linear fractional functions," *Journal of Global Optimization*, vol. 15, no. 4, pp. 419–432, 1999.

[57] H. Konno and K. Fukaishi, "A branch and bound algorithm for solving low rank linear multiplicative and fractional programming problems," *Journal of Global Optimization*, vol. 18, no. 3, pp. 283–299, 2000.

[58] H. Konno, Y. Yajima, and T. Matsui, "Parametric simplex algorithms for solving a special class of nonconvex minimization problems," *Journal of Global Optimization*, vol. 1, no. 1, pp. 65–81, 1991.

[59] H. P. Benson, "Using concave envelopes to globally solve the nonlinear sum of ratios problem," *Journal of Global Optimization*, vol. 22, no. 1-4, pp. 343–364, 2002.

[60] ——, "Solving sum of ratios fractional programs via concave minimization," *Journal of Optimization Theory and Applications*, vol. 135, no. 1, pp. 1–17, 2007.

[61] Y.-J. Wang and K.-C. Zhang, "Global optimization of nonlinear sum of ratios problem," *Applied Mathematics and Computation*, vol. 158, no. 2, pp. 319–330, 2004.

[62] E. Altman, "Flow control using the theory of zero sum markov games," *IEEE Transactions on Automatic Control*, vol. 39, pp. 814–818, 1992.

[63] Y. A. Korilis and A. A. Lazar, "On the existence of equilibria in noncooperative optimal flow control," *JOURNAL OF THE ACM*, vol. 42, pp. 584–613, 1995.

[64] E. Altman, T. Basar, T. Jimnez, and N. Shimkin, "Competitive routing in networks with polynomial cost," *IEEE Transactions on Automatic Control*, vol. 47, pp. 92–96, 2000.

[65] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multi-user communication networks," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 510–521, 1993.

[66] A. A. Lazar, A. Orda, and D. E. Pendarakis, "Virtual path bandwidth allocation in multiuser networks," *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 861–871, 1997.

[67] G. Debreu, "A social equilibrium existence theorem," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 38, no. 10, pp. 886–893, October 1952.

[68] E. Koutsoupias and C. H. Papadimitriou, "Worst-case equilibria." *Computer Science Review*, vol. 3, no. 2, pp. 65–69, 2009.

[69] T. Roughgarden, "Designing networks for selfish users is hard," in *In Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, 2001, pp. 472–481.

[70] T. Roughgarden, E. Tardos, and va Tardos, "How bad is selfish routing?" *Journal of the ACM*, vol. 49, pp. 236–259, 2001.

[71] C. Papadimitriou, "Algorithms, games, and the internet," in *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing.* New York, NY, USA: ACM, 2001, pp. 749–753.

[72] J. Aspnes, K. Chang, and A. Yampolskiy, "Inoculation strategies for victims of viruses and the sum-of-squares partition problem," in *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms.* Society for Industrial and Applied Mathematics, 2005, pp. 43–52.

[73] M. Lelarge and J. Bolot, "Network externalities and the deployment of security features and protocols in the internet," in *SIGMETRICS '08: Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems.* New York, NY, USA: ACM, 2008, pp. 37–48.

[74] ——, "A local mean field analysis of security investments in networks," in *NetEcon '08: Proceedings of the 3rd international workshop on Economics of networked systems.* New York, NY, USA: ACM, 2008, pp. 25–30.

[75] L. Jiang, V. Anantharam, and J. Walrand, "Efficiency of selfish investments in network security," in *NetEcon '08: Proceedings of the 3rd international workshop on Economics of networked systems.* New York, NY, USA: ACM, 2008, pp. 31–36.

[76] H. Kunreuther and G. Heal, "Interdependent security," *Journal of Risk and Uncertainty*, vol. 26, no. 2, pp. 231–249, March 2003.

[77] M. Kearns and L. E. Ortiz, "Algorithms for interdependent security games," in *Advances in Neural Information Processing Systems.* MIT Press, 2004.

[78] P. V. Mieghem and J. Omic, "In-homogeneous virus spread in networks," Technical report 20080801, Delft University of Technology, Tech. Rep., 2008.

[79] I. B. Rosen, "Existence and uniqueness of equilibrium point for concave n-person games," *Econometrica*, vol. 33, no. 3, p. 520534, 1965.

[80] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.

[81] Y. Chen, G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley, "Finding a better immunization strategy," *P.R.L.*, vol. 101, no. 5, p. 058701, 2008.

[82] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, June 2006.

[83] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, p. 066111, 2004.

[84] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 u.s. election: divided they blog," in *LinkKDD '05: Proceedings of the 3rd int. workshop on Link discovery.* ACM Press, 2005, pp. 36–43.

[85] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, pp. 17–61, 1960.

[86] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, p. 509, 1999.

[87] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, p. 026113, Feb 2004.

List of Abbreviations

KW	Kephart and White model
P2P	Peer to peer network
BGP	Boarder gateway routing protocol
SI	Susceptible Infected model
SIS	Susceptible Infected Susceptible model
SIR	Susceptible Infected Recovered model
NEP	Nash Equilibrium Point
PoA	Price of Anarchy
KM	Kermack and McKendrick model
BD	Birth and death process
VP	Vespignani and Pastor-Satorras model
IP	Internet protocol
AJAX	asynchronous JavaScript and XML
DNS	Domain Name service
IMC	Interactive Markov chain

Acknowledgments

I would like to dedicate this dissertation to my family. I owe endless gratitude to my parents for bringing me to this world with their unconditional love and to my sisters who never let me feel alone. Without my partners love and support and our beautiful daughter and her smiles in the morning I would never be able to complete this work.

This thesis would not be possible without the help as well as physical and remote presence of many people.

I owe deepest gratitude to Prof. Piet Van Mieghem, my promoter, who was alway enthusiastic and inspiring teacher and protective and caring as a parent. Special thanks to my supervisor Rob Kooij, who's teachings always included much more than science. I am indebted to all the committee members for their time and effort spent on my thesis and all the interesting conversations we had. In particular, I would like to thank Prof. Ariel Orda for introducing me to the inspiring field of game theory and dr. Ayalvadi Ganesh for open discussions and sharing ideas and knowledge.

Working on the 19th floor of EWI building was one of the most enjoyable working experiences I had. I would like to thank my dear friends, colleges and Huijuan and Tom for shared secrets, uncountable favours and working time filled with fun. Many thanks to the enthusiastic researchers with whom I had the opportunity to work: Javier, Siyu and Yue. My heartfelt thanks go to my old officemates Alemrima for her many original translations of Dutch to our own language, Santpal, for silent winds, dark earth and Monday morning rain and Bingjie, for friendship. I will never forget lunches in Aula with Anthena, Edgar, Daije, Antonio, Rogir, Stojan, Katja, Jing, Fernando and Christian. My gratitude goes to Wynand for his kindness and constructive comments on the propositions accompanying this thesis.

Laura, Marjon, Wendy and Stefanie have helped a lot with all kinds of documents and procedures.

Now I know that I am not complete without my friends here in The Netherlands. Thank you Aleksandar, Veronika, Ivan, Filip, Marko, Milos and Jelena. Special thanks go to my basketball team members and to our captain and my dear comrade Elena.

I would never become what I am today without my childhood friends: Milja, Nevena, Dragica, Fea, Andrej and Ugljesa. I owe my deepest gratitude to Zoja and Minja for bringing Semsa and Juzni Vetar into my life and to Darko for his park. I also thank most sincerely to Natasa, Ljiljana, Mirjana and Ivana for accepting me for who I am.

Finally, I would like to thank Marija for her support.

Jasmina Omić
Delft, August 2010

Curriculum Vitae

Jasmina Omić was born in Zagreb, Croatia on March 4, 1981. After graduating, in October 2005 from Electrotechnical faculty of Belgrade University, Serbia, with a M.Sc. degree in Electronics and Telecommunications, she joined NIL d.o.o, where she worked as a certified network engineer in the field of network configuration and security. In May 2006, she became a researcher and a PhD candidate at Network Architecture and Services (NAS) Group of Delft University of Technology. Under the supervision of the NAS group chair Prof. Van Mieghem, she performed research in the field of epidemics on networks, focusing on the influence of network structure on the process.

During her PhD studies, she assisted in lecturing the Data Communications Networking and Performance Analysis, as well as guided two students in their effort towards a M Sc degree in Electrical Engineering. She also served as a reviewer for various international conferences and journals in the field of telecommunications.

Publications by the Author

- Omic, J. S., R. E. Kooij and P. Van Mieghem, *Virus spread in complete bipartite graphs*, Bionetics 2007, December 10-13, Budapest, Hungary, 2007. **- Ch. 4**

- Van Mieghem, P. and J. Omic, *In-homogeneous Virus Spread in Networks*, Delft University of Technology, report2008081, 2008. **- Ch. 5**

- Van Mieghem, P., J. S. Omic and R. E. Kooij, *Virus Spread in Networks*, IEEE/ACM Transaction on Networking, Vol. 17, No. 1, February, pp. 1-14, 2009. **- Ch. 3, 4**

- Omic, J., A. Orda and P. Van Mieghem, *Protecting against network infections: A game theoretic perspective*, IEEE INFOCOM 2009, April 19-25, Rio de Janeiro, Brazil, 2009. **- Ch. 6, 7**

- Omic, J., R. E. Kooij and P. Van Mieghem, *Heterogeneous Protection in Regular and Complete Bi-partite Networks*, Networking 2009, May 11-15, Aachen, Germany, 2009. **- Ch. 5**

- Omic, J., Hernandez, J. M. and Van Mieghem, P., *Network protection against worms and cascading failures using modularity partitioning*, International Teletraffic Congress, September 2010, Amsterdam, Netherlands, 2010. **- Ch. 8**

- Omic, J., Van Mieghem, P., *Pandemics and Networks: the case of the Mexican flu*, ICMHA'10, October 2010, San Francisco, USA, 2010.

NGInfra PhD Thesis Series

1. Strategic behavior and regulatory styles in the Netherlands energy industry Martijn Kuit, 2002, Delft University of Technology, the Netherlands.

2. Securing the public interest in electricity generation markets, The myths of the invisible hand and the copper plate Laurens de Vries, 2004, Delft University of Technology, the Netherlands.

3. Quality of service routing in the internet: theory, complexity and algorithms Fernando Kuipers, 2004, Delft University of Technology, the Netherlands.

4. The role of power exchanges for the creation of a single European electricity market: market design and market regulation Franois Boisseleau, 2004, Delft University of Technology, the Netherlands, and University of Paris IX Dauphine, France.

5. The ecology of metals Ewoud Verhoef, 2004, Delft University of Technology, the Netherlands.

6. MEDUSA, Survivable information security in critical infrastructures Semir Daskapan, 2005,Delft University of Technology, the Netherlands.

7. Transport infrastructure slot allocation Kaspar Koolstra, 2005, Delft University of Technology, the Netherlands.

8. Understanding open source communities: an organizational perspective Ruben van Wendel de Joode, 2005, Delft University of Technology, the Netherlands.

9. Regulating beyond price, integrated price-quality regulation for electricity distribution networks Viren Ajodhia, 2006, Delft University of Technology, the Netherlands.

10. Networked Reliability, Institutional fragmentation and the reliability of service provision in critical infrastructures Mark de Bruijne, 2006, Delft University of Technology, the Netherlands.

11. Regional regulation as a new form of telecom sector governance: the inter-actions with technological socio-economic systems and market performance Andrew Barendse, 2006, Delft University of Technology, the Netherlands.

12. The Internet bubble - the impact on the development path of the telecom-munications sector Wolter Lemstra, 2006, Delft University of Technology, the Netherlands.

13. Multi-agent model predictive control with applications to power networks Rudy Negenborn, 2007, Delft University of Technology, the Netherlands.

14. Dynamic bi-level optimal toll design approach for dynamic traffic networks Dusica Joksimovic, 2007, Delft University of Technology, the Netherlands.

15. Intertwining uncertainty analysis and decision-making about drinking water infrastructure Machtelt Meijer, 2007, Delft University of Technology, the Netherlands.

16. The new EU approach to sector regulation in the network infrastructure in-dustries Richard Cawley, 2007, Delft University of Technology, the Nether-lands.

17. A functional legal design for reliable electricity supply, How technology af-fects law Hamilcar Knops, 2008, Delft University of Technology, the Nether-lands and Leiden University, the Netherlands.

18. Improving real-rime train dispatching: models, algorithms and applications Andrea D'Ariano, 2008, Delft University of Technology, the Netherlands.

19. Exploratory modeling and analysis: A promising method to deal with deep uncertainty Datu Buyung Agusdinata, 2008, Delft University of Technology, the Netherlands.

20. Characterization of complex networks: application to robustness analy-sis Almerima Jamaković, 2008, Delft University of Technology, Delft, the Netherlands.

21. Shedding light on the black hole, The roll-out of broadband access net-works by private operators Marieke Fijnvandraat, 2008, Delft University of Technology, Delft, the Netherlands.

22. On stackelberg and inverse stackelberg games and their applications in the optimal toll design problem, the energy markets liberalization problem, and in the theory of incentives Kateřina Staňková, 2009, Delft University of Technology, Delft, the Netherlands.

23. On the conceptual design of large-scale process and energy infrastructure systems: integrating flexibility, reliability, availability,maintainability and economics (FRAME) performance metrics Austine Ajah, 2009, Delft University of Technology, Delft, the Netherlands.

24. Comprehensive models for security analysis of critical infrastructure as complex systems Fei Xue, 2009, Politecnico di Torino, Torino, Italy.

25. Towards a single European electricity market, A structured approach for regulatory mode decision-making Hanneke de Jong, 2009, Delft University of Technology, the Netherlands.

26. Co-evolutionary process for modeling large scale socio-technical systems evolution Igor Nikolić, 2009, Delft University of Technology, the Netherlands.

27. Regulation in splendid isolation: A framework to promote effective and efficient performance of the electricity industry in small isolated monopoly systems Steven Martina, 2009, Delft University of Technology, the Netherlands.

28. Reliability-based dynamic network design with stochastic networks Hao Li, 2009, Delft University of Technology, the Netherlands.

29. Competing public values Bauke Steenhuisen, 2009, Delft University of Technology, the Netherlands.

30. Innovative contracting practices in the road sector: cross-national lessons in dealing with opportunistic behaviour Mnica Altamirano, 2009, Delft University of Technology, the Netherlands.

31. Reliability in urban public transport network assessment and design Shahram Tahmasseby, 2009, Delft University of Technology, the Netherlands.

32. Capturing socio-technical systems with agent-based modelling Koen van Dam, 2009, Delft University of Technology, the Netherlands.

33. Road incidents and network dynamics, Effects on driving behaviour and traffic congestion Victor Knoop, 2009, Delft University of Technology, the Netherlands.

34. Governing mobile service innovation in co-evolving value networks Mark de Reuver, 2009, Delft University of Technology, the Netherlands.

35. Modelling risk control measures in railways Jaap van den Top, 2009, Delft University of Technology, the Netherlands.

36. Smart heat and power: Utilizing the flexibility of micro cogeneration Michiel Houwing, 2010, Delft University of Technology, the Netherlands.

37. Architecture-driven integration of modeling languages for the design of software intensive systems Michel dos Santos Soares, 2010, Delft University of Technology, the Netherlands.

38. Modernization of electricity networks: Exploring the interrelations between institutions and technology Martijn Jonker, 2010, Delft University of Technology, the Netherlands.

39. Experiencing complexity: A gaming approach for understanding infrastructure systems. Geertje Bekebrede, 2010, Delft University of Technology, the Netherlands.

40. Epidemics in networks; Modeling, optimization and security games. Jasmina Omić, 2010, Delft University of Technology, the Netherlands.

Order information: info@nextgenerationinfrastructures.eu