

Copyright
by
Ruben J. Kubiak
2008

**Introduction of a Multi-Season Model Towards the Spread of
Infectious Diseases on Networks**

by

Ruben J. Kubiak

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF ARTS

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2008

**Introduction of a Multi-Season Model Towards the Spread of
Infectious Diseases on Networks**

APPROVED BY

SUPERVISING COMMITTEE:

Lauren Ancel Meyers

Manfred Fink

Für Angelika.

Acknowledgments

First of all, I would like to thank my girlfriend Angelika for all the support she gave me. It was a hard year for both of us, and I'm glad that she supported me to give me a chance to have both. A fantastic year in Austin and a wonderful relationship with a lovely women. :-*

Second, I want to thank my aunt Gudrun for my support. It is always enjoyable for me to visit you. And without your support, I wouldn't have had the chance to study in the USA. In addition, I want to thank my family. Thanks to my mother and my father for all the support in all the years since my birth.

Furthermore, I want to thank Lauren for the possibility to do my master in her great lab. It was a joy to work with you and I'm sure that I'll come back one day. Of course, I also want to thank the rest of Lauren's lab. Tom, you have always been a great officemate and helped me with programming issues, biological questions, and playing laser-tag. If I ever build a house myself, I will call you! Also, I want to thank Robert for his nice bacteria introduction and the rest of the lab for this nice and fun year.

But all this would not have been possible without Manfred Fink, who is in charge of this great Amerikaprogramm. Thank you for all your support and help.

At the very end, I would like to thank all I forgot above. So - thanks!

Introduction of a Multi-Season Model Towards the Spread of Infectious Diseases on Networks

Ruben J. Kubiak, M.A.

The University of Texas at Austin, 2008

Supervisor: Lauren Ancel Meyers

Outgoing from the known epidemiological percolation network model, a multi-season network model for infectious diseases was constructed. Cross-immunity effects through previous infections with other strains were taken into account. The cross-immunity function is arbitrary chosen, but the model has the capability to take every possible function into account, which only needs to fulfil very weak conditions. The multi-season process is modelled as a Markovian-chain, both in simulations and calculations. In the calculations, a Master-equation is used to determine the current cross-immunity for each vertex at each season. It can be demonstrated that all epidemic parameters go fast into an equilibrium. In addition, many epidemiological parameters are determined, such as the prevalence or probability of an epidemic, for multiple Poisson, scale-free and exponential networks. It is shown that the between season dynamics of the network have almost no influence on the networks and its epidemical behavior.

Table of Contents

Acknowledgments	v
Abstract	vi
Chapter 1. An Introduction into Classical Modeling of Infectious Disease Dynamics	1
1.1 The Standard SIR Model: Everybody Is in a State	1
1.2 Connection to the Real-World	2
Chapter 2. How to Predict an Epidemic with Network- and Percolation Theory	4
2.1 A Percolation Theory View on Epidemiological Network Models . . .	4
2.2 Properties of Network Degrees	6
2.2.1 Probability Generating Function of the Degree Distribution .	7
2.2.2 Moments of Degree Distributions	8
2.2.3 Powers of Degree Distributions	8
2.3 Connections between Vertices	9
2.3.1 Generating Function of Occupied Edges Attached to a Vertex	10
2.4 Finding the Component Sizes in a Network	11
2.5 Outbreak Sizes, Epidemic Sizes and Critical Transmissibility	13
2.5.1 Calculating the Average Outbreak Size in the Subcritical Phase	13
2.5.2 Epidemic Threshold - Going from the Subcritical Phase to the Supercritical Phase	14
2.5.3 Epidemic Sizes in the Supercritical Phase	14
2.6 Comparability of the Bond Percolation Network Model with the Standard SIR Model	16

Chapter 3.	Expansion of the Percolation Network Model to a Multi-Season Model with Cross-Immunity	17
3.1	Cross-Immunity as a Parameter of the Between-Host Dynamics . . .	18
3.2	Keeping Track of the Correct States	19
3.2.1	Definition of Markovian-Processes	19
3.2.2	Simplification of the Chapman-Kolmogorov-Equation to a Master-Equation in a Markovian-Process	21
3.2.3	Determination of the Correct Transition Probabilities	23
3.3	Comparison of the Between-Host Model with the Reality - The Case of an Finite Real-World	25
3.3.1	Limitations of the Transition Matrix Size	25
3.3.2	The Age Structure of the Population	26
3.3.3	The Real-World Degree Distribution	27
3.4	Modeling the Transmission Probability Between to Vertices	28
3.5	Three Different Models of Between-Season Mixture	31
3.5.1	The Non-Mixture Model	31
3.5.1.1	Transmissibility in the Non-Mixture Model	31
3.5.1.2	Calculation of Component Sizes in the Non-Mixture Model	34
3.5.2	The Half-Mixture Model	34
3.5.2.1	Transmissibility in the Half-Mixture Model	35
3.5.3	The Full-Mixture Model	35
3.5.3.1	The Transition Matrix and Transmissibility in the Full-Mixture Model	35
3.6	Two Different Ways to Calculate the Mean Epidemic Size	36
3.6.1	Calculate Every Possible Case	37
3.6.2	Calculate Only the Average Case	39
Chapter 4.	Simulating the Spread of Infectious Diseases in the Multi-Season Model	42
4.1	General Concerns for Network Modelling	42
4.1.1	Generating Random Numbers	42
4.1.1.1	Mersenne Twister Generator	42
4.1.1.2	Generating Poisson Distributed Random Numbers . .	43

4.1.1.3	Generating Scale-Free Distributed Random Numbers	44
4.1.1.4	Generating Exponential Distributed Random Numbers	45
4.1.2	Building the Network - an Algorithm for the Connection of Vertices	45
4.1.3	Spread of the Infection	48
Chapter 5.	Results of the Simulations and Calculations	50
5.1	Evaluating the Model	51
5.1.1	Epidemic Sizes	51
5.1.2	Degree-Specific Evolution of Epidemiological Risk	58
5.1.3	Distribution of Epidemic Sizes	66
5.2	The Impact of Social Mixing on Multi-Season Dynamics	78
5.3	The Impact of Network Structure on Multi-Seasonal Dynamics . . .	87
5.4	Comparison of the Different Calculation Methods	90
5.5	Implications of the Multi-Season Model	92
Vita		98

Chapter 1

An Introduction into Classical Modeling of Infectious Disease Dynamics

1.1 The Standard SIR Model: Everybody Is in a State

The classical approach to model a spread of an epidemic is the so called Suceptible-Infective-Recovered (SIR) model. It was first formulated by Lowell Reed and Wade Hampton Frost in the 1920s (Newman, 2002). The start point is a finite population of size N . This population can be divided into three states: susceptible (s), infective (i) and recovered (r) (Bailey, 1975). It is a full contact model - everybody in the population is connected to everybody else. The aim of the analysis is to determine the dynamics of a epidemic, therefore this static SIR case has to be changed introducing transition rates between the states. The average transition rate per unit time from suceptibles s to infectives i is called β . This is the rate with which infected individuals infect susceptibles. The average transition rate per unit time from infected i to recovered r is called γ . This is the rate with which infected individuals recover, acquire immunity or die. Now, it is possible to formulate this with three differential equations:

$$\frac{ds}{dt} = -\beta is, \quad \frac{di}{dt} = \beta is - \gamma i, \quad \frac{dr}{dt} = \gamma i \quad (1.1)$$

This standard SIR model is the most common. But there are many derivatives in use. For example, a less complex model is the Suceptible-Infective-Susceptible

Disease	R_0
Diphtheria	6-7
HIV	2-5
Influenza (1918 pandemic strain)	2-3
Measles	12-18
Mumps	4-7
Pertussis	12-17
Polio	5-7
Rubella	5-7
SARS	2-5
Smallpox	6-7

Table 1.1: Table of selected infectious diseases' R_0 values (CDC and WHO, 2007; Anderson, 1978; Wallinga and Teunis, 2004; Mills et al., 2004).

(SIS). It does not include the recovered state. Infected individuals recover into the susceptible state again. But most epidemiological models are more realistic and therefore more complex. It is possible to separate the three states into different sub-states, e.g. such as infectives, who can infect others, and infectives, who get infected but do not have a disease outbreak and do not transmit the disease. This leads to new transition rates and new differential equations, but the general idea remains the same.

1.2 Connection to the Real-World

Analyses of real-world epidemics do not provide these transition rates directly. Like done with SARS (Meyers et al., 2005), most researchers in epidemiology determine a so called basic reproductive rate R_0 . The basic reproductive rate (or sometimes called basic reproduction number) is the average number of secondary cases a typical infected case will cause in a population without immunity and with-

out a treatment from outside against the infection. Now, it is easy to determine the threshold for an epidemic. If the ratio of new infected individuals per old infected individual is $R_0 < 1$, the infection will cause an outbreak and die out on the long run. In the limit of an infinite population, the epidemic size would be $S = 0$. If $R_0 > 1$, the infection will probably cause an epidemic with an epidemic size of $S > 0$ in the limit of an infinite population. In the standard SIR model as shown in (1.1), for all R_0 values over 1 an epidemic will always develop. In reality, even with a basic reproductive rate above 1, an epidemic only occurs with a certain probability. In general, a higher R_0 value gives a higher epidemic probability. Table 1.1 shows R_0 values for selected infectious diseases.

Chapter 2

How to Predict an Epidemic with Network- and Percolation Theory

One of the weaknesses of the SIR model is the assumption of full spatial distribution. Spatial effects may have a great impact on epidemiological dynamics (Lloyd and May, 1996; Praham and Ferguson, 2006; Hagenaars et al., 2004). It is obvious that this is not true in reality. It is much more accurate to model the actual contacts of each person instead of assuming a full contact model. This is done in network models, where each person has only a well defined number of contacts.

2.1 A Percolation Theory View on Epidemiological Network Models

In general, percolation theory describes the behavior of connected clusters in a random graph. Its origin lies in graph theory, which describes mathematically the pairwise relations between objects. These objects are called *vertices* and the connections or relations are called *edges*. In this biological context, a graph is generally called a network. It can be *directed*, which means that the edge has a distinction between its two connecting vertices, or *undirected*, which means that there is no distinction between the two vertices associated with each edge. The concept of a graph is very general and not exclusive for inter-human networks. The first publication on graph theory was *Die Sieben Brücken von Königsberg* by

Leonhard Euler (Euler, 1741). Many other noticeable studies followed over the centuries, such as the *four color theorem* by Arthur Cayley (Cayley, 1879).

Percolation theory is specially relevant for the behavior of clusters and the influence of the graph's parameters, such as the mean number of vertices' connections, on the clusters. It is a common procedure to achieve cluster structure by giving the edges a probability of affiliation, also called transmission. This probability determines the character of an edge. In the real-world, the network could be an information network between individuals. Edges are connections between the individuals and with a certain probability, a special message could spread through this network through these edges. In this case the subgraph of all vertices (individuals), which got the message, would be a cluster.

According to the different types of graphs, percolation theory itself can be split up into different classes, such as:

- **bond percolation** unifies all random graphs at which the connection between vertices can be opened or closed
- **site percolation** unifies all random graphs at which the vertices itself can be opened or closed
- **directed percolation** is a special case of percolation classes, in which the transmission is directed in only one direction (like a directed network)

But all of these different classes have common properties. It exists a critical transmission probability p_c . This means if the value of the average transmission probability $\langle p \rangle = p$ reaches a critical value, the cluster size becomes infinite in the limit

of an infinite network, which is called *giant component* (gcc) of the network. This is called the supercritical phase. Below the threshold, a subcritical phase with clusters of finite sizes occur. It is necessary to point out that the supercritical phase is necessary to achieve gcc's, but not sufficient. A gcc can occur only with a certain probability dependent of p .

2.2 Properties of Network Degrees

In network-theoretical studies, each individual is represented by a vertex. Each vertex (or each individual) has its degree, which is nothing else as the number of connections to other vertices (or the number of connections to other individuals). If you have a network with N vertices and each vertex has a degree of $N - 1$, you will get the same situation as in SIR models - everybody is connected to everybody. But this is not always true. Nobody is connected to everybody in the real world. Also, the number of connections is different from individual to individual. Mathematically, this means that a vertex contains a probability p_k of possessing a distinguished degree k with $\sum_{k=0}^{\infty} p_k = 1$. For each network, these probabilities give a characteristic distribution, so that one can separate networks in different classes depending on the probability distributions. In principal, it is possible to construct an infinite number of probability distributions (and therefore network classes), but it was shown that real-world networks can be described by only few network classes (Newman, 2002; Watts and Strogatz, 1998; Williams and Martinez, 2000). Especially scale-free networks, Poisson networks, exponential networks and small-world networks can describe real-world networks with good approximation (May and Lloyd, 2001;

Watts and Strogatz, 1998; Newman, 2007).

Even if it was tried before (Pasor-Satorras and Vespignani, 2001), Mark E. J. Newman was the first to develop a comprehensive epidemiological percolation model in 2002 (Newman, 2002), which became well known in theoretical epidemiology and the basis of many epidemiological studies, as well as for this thesis. In the following, Newman’s model is described in more detail.

2.2.1 Probability Generating Function of the Degree Distribution

An important concept in network theory is to describe probability distributions by *probability generating functions* (pgfs). The expectation behind pgfs is to provide all different characteristics of a probability distribution within one function (Wilf, 1994). Therefore, a pgf is - as its name says - a function which generates a probability distribution of degrees thus:

$$P_0(x) = \sum_{k=0}^{\infty} p_k x^k \quad (2.1)$$

All information of the degree distribution are encapsulated by its pgf. Note that:

$$P_0(1) = \sum_{k=0}^{\infty} p_k = 1 \quad (2.2)$$

Given the pgf, the degree distribution can be reconstructed by repeated differentiation:

$$p_k = \frac{1}{k!} \left[\frac{d^k P_0}{dx^k} \right]_{x=0} \quad (2.3)$$

2.2.2 Moments of Degree Distributions

Using a pgf makes it easy to determine the moments of a probability distribution. The first moment, which is the mean, is given by the first derivative of the pgf with $x = 1$. For a degree distribution, it is the average degree in a network:

$$z = \langle k \rangle = \sum_{k=0}^{\infty} k p_k = \left[\frac{dP_0}{dx} \right]_{x=1} = P'_0(1) \quad (2.4)$$

In general, higher moments are given by higher derivatives. It is:

$$\langle k^n \rangle = \sum_{k=0}^{\infty} k^n p_k = \left[\left(x \frac{d}{dx} \right)^n P_0(x) \right]_{x=1} \quad (2.5)$$

2.2.3 Powers of Degree Distributions

The same is true for powers of a probability distributions, e.g. at the cumulative degree distribution of n vertices. It is generated by $[P_0(x)]^n$. This is correct in the general case, but it is taken $n = 2$ for illustration:

$$\begin{aligned} [P_0(x)]^2 &= \left[\sum_{k=0}^{\infty} p_k x^k \right]^2 \\ &= \sum_{i,j=0}^{\infty} p_i p_j x^{i+j} \\ &= p_0 p_0 x^0 + (p_0 p_1 + p_1 p_0) x^1 + (p_0 p_2 + p_1 p_1 + p_1 p_0) x^2 \\ &\quad + (p_0 p_3 + p_1 p_2 + p_2 p_1 + p_3 p_0) x^3 + \mathcal{O}(4) \end{aligned} \quad (2.6)$$

Now, it is possible to determine every moment and power of a degree distribution in a convenient way.

2.3 Connections between Vertices

Like individuals in the real-world, vertices do not have double connections in epidemiological network models. In general, not all connections are the same, which is also taken into account by network models. For example, consider a pair of vertices i and j (a pair of individuals), which is connected by an edge. Assume that vertex i is infected and vertex j is susceptible and the rate of contacts between them is r_{ij} . The infective i has the time τ in which i is infective to others. It is possible to define the probability T_{ij} that the disease will be transmitted from i to j . First, the probability that j will not get infected thru i is similar to Newman (2002):

$$\begin{aligned} 1 - T_{ij} &= \lim_{\delta t \rightarrow 0} (1 - r_{ij} \delta t)^{\frac{\tau}{\delta t}} \\ &= \lim_{\delta t \rightarrow \infty} \left(1 - \frac{r_{ij}}{\delta t}\right)^{\tau \delta t} \\ &= e^{-r_{ij} \tau} \end{aligned} \tag{2.7}$$

This gives the probability of transmission:

$$T_{ij} = 1 - e^{-r_{ij} \tau} \tag{2.8}$$

Unfortunately, r_{ij} is a specific variable for each individual and therefore, T_{ij} is varying between individuals. But in a network with a population size $N \rightarrow \infty$, the epidemic dynamic is only computable with an average transmissibility. For this, the assumptions are made that r_{ij} is an *independent identically distributed* (iid) random variable and τ is a constant value for all edges. Because r_{ij} is an iid random variable, T_{ij} is also an iid random variable.

It is now possible to calculate the average transmissibility T . The distribution for the interaction rates is given by $P(r)$ and the distribution of the interaction times is $P(\tau)$, then an average transmissibility is calculated by:

$$T = \langle T_{ij} \rangle = 1 - \int_0^\infty dr d\tau P(r) P(\tau) e^{-r\tau} \quad (2.9)$$

The transmissibility is always in the range $0 \leq T \leq 1$. With the average transmissibility, it is possible to relate the network approach to percolation theory.

2.3.1 Generating Function of Occupied Edges Attached to a Vertex

To solve the percolation problem, the propability distribution of occupied edges attached to a vertex has to be derived. As seen at section 2.2.1, a convenient way to do this is with a pgf.

The probability to have n occupied edges attached to a vertex with degree k is given by the binomial distribution $\binom{k}{n} T^n (1 - T)^{k-n}$. This leads the following generating function:

$$\begin{aligned} P_0(x; T) &= \sum_{k=0}^{\infty} \sum_{n=0}^k p_k \binom{k}{n} T^n (1 - T)^{k-n} x^n \\ &= \sum_{k=0}^{\infty} p_k \sum_{n=0}^k \binom{k}{n} T^n (1 - T)^{k-n} x^n \end{aligned}$$

which can be reformulated with the use of the binomial theorem:

$$\begin{aligned} &= \sum_{k=0}^{\infty} p_k (1 - T + xT)^k \\ &= P_0(1 + (x - 1)T) \end{aligned} \quad (2.10)$$

In addition, it is necessary to know the number of occupied egdes, leaving an infected vertex. This is requires different logic as calculated in equation (2.10).

$P_0(x; T)$ does not separate between infected or not infected vertices. To do so, first the distribution of edges leaving a vertex has to be determined by following a random edge. The probability to reach a vertex with degree k is proportional to kp_k . With the correct normalization, this leads to the pgf:

$$\frac{\sum_{k=0}^{\infty} kp_k x^k}{\sum_{k=0}^{\infty} kp_k} = x \frac{P'_0(x)}{P'_0(1)} \quad (2.11)$$

In an epidemiological context, each vertex needs to be infected through one of its edges by a neighbor before it can infect all of its neighbors. This is similar to the number of possible ways to leave a vertex excluding the edge I used to reach the vertex. The number of edges leaving the vertex is therefore the degree minus one. This is nothing else as equation (2.11) divided by one power of x and gives:

$$P_1(x) = \frac{P'_0(x)}{P'_0(1)} = \frac{1}{z} P'_0(x) \quad (2.12)$$

with z as the average degree.

The number of occupied edges leaving an infected vertex is similar to calculate as done in equation (2.10). This gives a pgf of:

$$P_1(x; T) = P_1(1 + (x - 1)T) \quad (2.13)$$

which generates the distribution of occupied edges, leaving an infected edge.

2.4 Finding the Component Sizes in a Network

Newman showed that it is possible to calculate the distribution of sizes of connected components in a network that are reached by choosing a random edge and

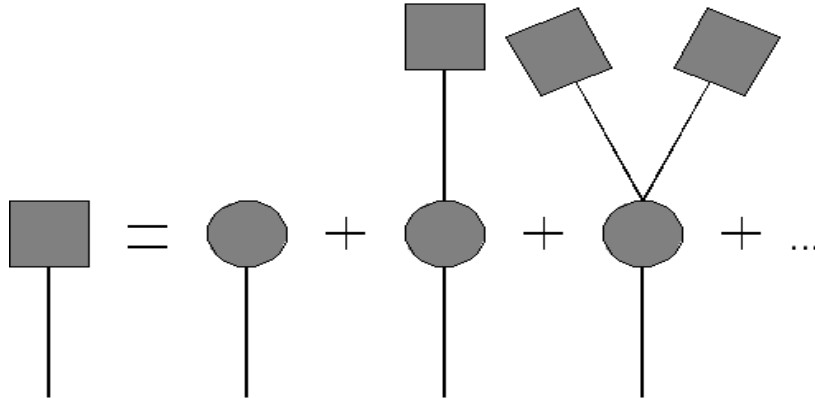


Figure 2.1: Schematic representatoin of the additive component of vertices reached by following a randomly choosen edge. The probability of each such component can be represented as the sum of the probability over all such components.

following it to its end (Newman, 2002). By using the pgf approach of section 2.2.1 again, it is possible to define $H_1(x; T)$ as the generating function of this distribution. Per definition, the giant component (if there is one or not) $H_1(x; T)$ is excluded so that component sizes of $H_1(x; T)$ are finite. In general, the probability to have a connection between vertex i and j goes as N^{-1} with N as network size. Therefore, the probability to get loops in $H_1(x; T)$ is $\lim_{N \rightarrow \infty} N^{-1} = 0$ in the limit of an infinite network. This has a very useful consequence: the structure of the component is tree-like (see figure 2.1). This makes it possible to calculate $H_1(x; T)$ with a self-consistent equation. Let g_k be the probability that the inital site has k edges coming out of it (note: we are excluding the edge, with wich we came in). This gives by

equation (2.6) the self-consistent equation:

$$\begin{aligned} H_1(x; T) &= xg_0 + xg_1H_1(x; T) + xg_2[H_1(x; T)]^2 + xq_3[H_1(x; T)]^3 + \mathcal{O}(4) \\ &= x \sum_{i=0}^{\infty} g_i [H_1(x; T)]^i \end{aligned} \quad (2.14)$$

According to equation (2.12), $P_1(x; T)$ is the pgf which describes the same aspect: the probabilities to leave an infected vertex without using the edge, with which the vertex got infected before. Therefore, $H_1(x; T)$ can be written as:

$$H_1(x; T) = xP_1(H_1(x; T); T) \quad (2.15)$$

Let $H_0(x; T)$ be the pgf for the size of a cluster reachable from a randomly chosen starting vertex. Hence these is given by:

$$H_0(x; T) = xP_0(H_1(x; T); T) \quad (2.16)$$

The derivation is analogous to equation (2.15).

2.5 Outbreak Sizes, Epidemic Sizes and Critical Transmissibility

2.5.1 Calculating the Average Outbreak Size in the Subcritical Phase

The average outbreak size $s(T)$ is the size of an outbreak below the epidemic threshold or, if the transmissibility is above the threshold, the size of an outbreak if no epidemic occurs. It is given by:

$$\langle s(T) \rangle = H'_0(1; T) \quad (2.17)$$

By the use of equation (2.4) and the fact that all pgfs are equal to 1 for $x = 1$, $H'_0(1; T)$ can be re-written as follows:

$$H'_0(1; T) = 1 + P'_0(1; T)H'_1(1; T) \quad (2.18)$$

With equation (2.16), this gives:

$$H'_1(1; T) = 1 + P'_1(1; T)H'_1(1; T) = \frac{1}{1 - P'_1(1; T)} \quad (2.19)$$

and:

$$\langle s(T) \rangle = 1 + \frac{P'_0(1; T)}{1 - P'_0(1; T)} = 1 + \frac{TP'_0(1)}{1 - TP'_0(1)} \quad (2.20)$$

2.5.2 Epidemic Threshold - Going from the Subcritical Phase to the Supercritical Phase

The average outbreak size can reach values from $1 \leq \langle s(T) \rangle \leq \infty$. Because $P'_0(1) \neq 0$, T must be 0 to get an outbreak size of 1. On the other hand, $\langle s(T) \rangle$ diverges when $TP'_1(1) = 1$. This is the point at which the critical transition into the supercritical phase takes place and a giant component occur. Therefore, this gives a critical transmissibility T_c of:

$$\begin{aligned} T_c &= \frac{1}{P'_1(1)} = \frac{P'_0(1)}{P''_0(1)} \\ &= \frac{\sum_{k=0}^{\infty} kp_k}{\sum_{k=0}^{\infty} k(k-1)p_k} \\ &= \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \end{aligned} \quad (2.21)$$

As it is easy to see now, the critical transmissibility T_c is the reciprocal of the average degree of excising edges.

2.5.3 Epidemic Sizes in the Supercritical Phase

But above the epidemic threshold T_c , equation (2.15) is no longer valid. In the giant component, because of its infinite size, it is possible to have self-loops.

Therefore, H_0 needs to be redefined: H_0 is the pgf only for clusters smaller than the epidemic cluster (e.g. finite size clusters that are smaller than the infinite giant component). This gives an epidemic size $S(T)$ in the supercritical phase of:

$$S(T) = 1 - H_0(1; T) \quad (2.22)$$

Making use of equation (2.16), this can be written as:

$$\begin{aligned} S(T) &= 1 - P_0(u; T) \\ &= 1 - \sum_{k=0}^{\infty} k p_k (1 + (u - 1) T)^k \end{aligned} \quad (2.23)$$

where we substitute x with u . Hence, $u \equiv H_1(1; T)$ and therefore u is the solution of the self-consistency relation:

$$\begin{aligned} u &= P_1(u; T) \\ &= P_1(1 + (u - 1)T) \\ &= \frac{\sum_{k=0}^{\infty} k p_k (1 + (u - 1) T)^{k-1}}{\sum_{k=0}^{\infty} k p_k} \end{aligned} \quad (2.24)$$

Now, it is clear what u is the parameter for. It is the probability of following an edge with a non-infected vertex at its end and also the fraction of non-infected edges in a network. In general, the probability for a vertex with degree k to not get infected is:

$$\psi^k = (1 + (u - 1) T)^k \quad (2.25)$$

Another property of simple undirected semi-random network models is that the average epidemic size is also the probability of having an epidemic. Like typical for percolation models, it is not for sure that a gcc will occur in the supercritical

phase. This is in contrast to the standard SIR model (see chapter 1), where above the epidemic threshold, an epidemic always occurs.

2.6 Comparability of the Bond Percolation Network Model with the Standard SIR Model

The bond percolation network model and the standard SIR model are entirely different. In the SIR model, you have a full contact population. Everybody gets in contact with everybody else. Not so in the bond percolation model. In addition, the contact rates between individuals can be flexible. Not so in the SIR model.

But despite all these differences, it is possible to make both models comparable. Both models have a threshold factor for the transfer from the non-epidemic case to the epidemic case. And it is possible to connect both models with this property. In the SIR model, it is R_0 , and in the bond percolation model it is T_c . Now, it is possible to determine the connection of R_0 with T_c and equation (2.21):

$$T = R_0 \quad T_c = R_0 \left(\frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \right) \quad (2.26)$$

This is easy to understand if one remembers that T_c is the reciprocal value of the average excess degree (see page 14, equation (2.21)). R_0 only gets "re-normed" to the specific network properties. With this equation, it is possible to compare results of network models with real world studies and standard SIR models.

Chapter 3

Expansion of the Percolation Network Model to a Multi-Season Model with Cross-Immunity

In general, epidemic models, such as the standard SIR model and the percolation theory network model, are only good for one season of an epidemic. It is only possible to model one spread of the disease through the population with the existing models. But like for influenza, some diseases spread in waves around the world - the so called seasons. This is also true for many other infectious diseases. Individuals are exposed to the same disease many times in their live. They got infected, recover and are possibly infected again a few seasons later. But in general, previous infections have a big influence on the later infectibility. Our immune system is trained in recognizing known infections and very effective in fighting against these. On the other side, viruses and bacteria change through evolution. New strains evolve and our immune system is confronted with a new, evolved enemy. But because the disease often only changed in small parts, it is still possible that our immune system has cross-immunity. This is an immunity against a slightly different strain, and because of the similarity of the strains, the immune system has some immunity against the new strain. In parts, this problem is similar to the problem of modeling immunity by vaccination. But there are also huge differences. In the multi-season model, the immunity is not given through an (more or less) intelligent vaccination

strategy. It is determined by the last epidemics. In addition, cross-immunity is not an all-or-nothing model. It is determined by the phenotype difference of the strains and therefore more probabilistic.

3.1 Cross-Immunity as a Parameter of the Between-Host Dynamics

In most of the existing immunity models, the cross-immunity and immunity are parameters of the vertex (Nuno et al., 2008; Pasor-Satorras and Vespignani, 2002). This corresponds to a site percolation model (Newman, 2002). The immunity is part of the vertex, not of the edge which connect vertices (or individuals) and therefore part of the within-host dynamics. This immunity can be dependent on different parameters of the vertex, such as the degree. If the vertex has immunity, it will be impossible to infect it. In the case of full immunity, this makes no difference between the within-host and the between-host dynamics. In both cases, the vertices which got infected and have full immunity can not be infected again. It is the same as the full removing of all infected vertices and their respective edges from the network. Now, the infection in the next season is an infection in the remaining sub-network.

On the other hand, it makes a big difference in the case of partial immunity. In the within-host model, all immunity lies in the edges. This is like a bond percolation view on it. The cross-immunity has a direct influence on the transmissibility between vertices and therefore defines the between-host dynamics. This model is much more realistic than the within-host model. A high degree vertex is much more likely to get re-infected, even if it has cross-immunity from the last season, than

a low degree vertex, which has only very few contacts and hence the risk of an re-infection is much smaller.

In my multi-season model, the cross-immunity is only dependent on the time between the last infection and the current season. This assumption is due to the evolution of the strains. It is assumed that each season a new strain is introduced and that the differences between the strains is constant over time. This assumption is not a necessity for my model. It is also possible to model otherwise distributed or experimental observed differences without any problems.

3.2 Keeping Track of the Correct States

For the new multi-season model, it is necessary to keep track of the individual history of each vertex. Luckily, the vertices are only separated by one characteristic parameter - the degree. Therefore, it is possible to use a probabilistic model, which describes which fraction of degree k vertices is in which state. As we will see, the model fulfills the requirements for a Markovian-process and we could therefore use a Master-equation to describe the transition process of the probabilities.

3.2.1 Definition of Markovian-Processes

In general, a statistical process in a system is defined by the time-dependent stochastic random variable $r(t)$. According to (Honerkamp, 2002), the following definitions are necessary to define a stochastic process:

- $\rho_1(r, t)$ is the probability density to find the state r at time t
- $\sigma_2(r_2, t_2 | r_1, t_1)$ is the conditional probability density to find the state r_2 at

time t_2 , if the system was in r_1 at time t_1

- in general, $\sigma_n(r_n, t_n | r_{n-1}, t_{n-1}, \dots, r_1, t_1)$ is the conditional probability density to find the state r_n at time t_n , if the system was in $r_{n'}$ at times $t_{n'}$ with $n' = n - 1, \dots, 1$

Now, every process fulfilling:

$$\sigma_n(r_n, t_n | r_{n-1}, t_{n-1}, \dots, r_1, t_1) = \sigma_2(r_n, t_n | r_{n-1}, t_{n-1}) \quad (3.1)$$

is called a Markovian-process. The specialty of Markovian-processes is that they do not have a long-term history. For state r_n at time t_n , it is only necessary to know the state r_{n-1} at time t_{n-1} . This is exactly, what the multi-season model does. Therefore, it fulfills the requirements for a Markovian-process. It is not necessary to know the earlier states or how the system got into these states. This gives a very powerful formalism. Each probability density $\rho_n(r_n, t_n; r_{n-1}, t_{n-1}; \dots, r_1, t_1)$ can be described through $\rho_1(r, t)$ and $\sigma_2(r', t' | r, t)$ only. If $n = 3$ and $t_3 > t_2 > t_1$, it is under the use of the Markov-condition (3.1):

$$\begin{aligned} \rho_3(r_3, t_3; r_2, t_2; r_1, t_1) &= \sigma_3(r_3, t_3 | r_2, t_2; r_1, t_1) \rho_2(r_2, t_2; r_1, t_1) \\ &= \sigma_2(r_3, t_3 | r_2, t_2) \sigma_2(r_2, t_2 | r_1, t_1) \rho_1(r_1, t_1) \end{aligned} \quad (3.2)$$

It is possible to define every Markovian-process only with the conditional probabilities $\sigma_2(r', t' | r, t)$. These conditional probabilities are also called transition probabilities, because they transition a probability between two states at two different time points. If equation (3.2) is integrated about the whole scope of r_2 , this will give the

Chapman-Kolmogorov-equation:

$$\begin{aligned}
\rho_2(r_3, t_3; r_1, t_1) &= \sigma_2(r_3, t_3 | r_1, t_1) \rho_1(r_1, t_1) \\
&= \int dr_2 \rho_3(r_3, t_3; r_2, t_2; r_1, t_1) \\
&= \int dr_2 \sigma_2(r_3, t_3 | r_2, t_2) \sigma_2(r_2, t_2 | r_1, t_1) \rho_1(r_1, t_1)
\end{aligned} \tag{3.3}$$

According to equation (3.3), it is possible to write $\sigma_2(r_3, t_3 | r_1, t_1)$ as:

$$\sigma_2(r_3, t_3 | r_1, t_1) \equiv \int dr_2 \sigma_2(r_3, t_3 | r_2, t_2) \sigma_2(r_2, t_2 | r_1, t_1) \tag{3.4}$$

This means that the transition probability of going from state r_1 at time t_1 to state r_3 at time t_3 is given by the sum over all possible states r_2 at time t_2 .

To find a analogous relationship for $\rho_1(r_3, t_3)$, multipli equation (3.4) with $\rho_1(r_1, t_1)$ and integrate over r_1 :

$$\begin{aligned}
\rho_1(r_3, t_3) &= \int dr_1 \sigma_2(r_3, t_3 | r_1, t_1) \rho_1(r_1, t_1) \\
&= \int \int dr_2 dr_1 \sigma_2(r_3, t_3 | r_2, t_2) \sigma_2(r_2, t_2 | r_1, t_1) \rho_1(r_1, t_1) \\
&= \int dr_2 \sigma_2(r_3, t_3 | r_2, t_2) \rho_1(r_2, t_2)
\end{aligned} \tag{3.5}$$

This proofs that only the transition probabilities σ_n and the start probability $\rho_1(r_0, t_0)$ need to be known to describe the system at all possible times.

3.2.2 Simplification of the Chapman-Kolmogorov-Equation to a Master-Equation in a Markovian-Process

The Chapman-Kolmogorov-equation is sufficient to describe Markovian-processes, but it is not easy to handle. Luckily, it is possible to use an equivalent equation, the so called Master-equation, which is much easier to handle. Coming from stochastic physical processes (van Kampen, 2001), it is used for time differences $t_n - t_{n-1} = \tau$

with $\tau \rightarrow 0$ - but in addition, for all processes with $\tau = \text{const}$. Therefore, all transition probabilities $\sigma_2(r, t + \tau | r', t)$ must fulfill (Honerkamp, 2002):

$$\lim_{\tau \rightarrow 0} \sigma_2(r, t + \tau | r', t) = \delta(r - r') \quad (3.6)$$

with $\delta(r - r')$ as Dirac-delta-function.

Especially important is the first order term of the Taylor-development over small τ :

$$\sigma_2(r, t + \tau | r', t) = (1 - C(r', t)\tau) \delta(r - r') + \tau \omega(r, r', t) + \mathcal{O}(\tau^2) \quad (3.7)$$

with $\omega(r, r', t)$ as transition rate from state r to state r' at time t . $C(r', t)$ is the normalization of σ_2 and ensures that at time $t + \tau$ the system is at any state:

$$\begin{aligned} 1 &= \int dr \sigma_2(r, t + \tau | r', t) \\ &= (1 - C(r', t)\tau) + \tau \int dr \omega(r, r', t) + \mathcal{O}(\tau^2) \\ \Rightarrow C(r', t) &= \int dr \omega(r, r', t) \end{aligned} \quad (3.8)$$

This gives together with equation (3.3), (3.4) and (3.7):

$$\begin{aligned} \sigma_2(r, t + \tau | r', t) &= \int dr'' \sigma_2(r, t + \tau | r'', t) \sigma_2(r'', t | r', t') \\ &= \int dr'' [(1 - C(r'', t)\tau) \delta(r - r'') + \tau \omega(r, r'', t)] \sigma_2(r'', t | r', t') \\ &= (1 - C(r'', t)\tau) \sigma_2(r, t | r', t') \\ &\quad + \tau \int dr'' \omega(r, r'', t) \sigma_2(r'', t | r', t') + \mathcal{O}(\tau^2) \end{aligned} \quad (3.9)$$

Now, using equation (3.8) and (3.9) under division of $\tau \rightarrow 0$ gives:

$$\lim_{\tau \rightarrow 0} \frac{\sigma_2(r, t + \tau | r', t) - \sigma_2(r, t | r', t')}{\tau} = \frac{\partial \sigma_2(r, t | r', t')}{\partial t} \quad (3.10)$$

and:

$$\frac{\partial \sigma_2(r, t|r', t')}{\partial t} = \int dr'' [\omega(r, r'', t)\sigma_2(r'', t|r', t') - \omega(r'', r, t)\sigma_2(r, t|r', t')] \quad (3.11)$$

It is possible to define a similar equation for $\rho_1(r, t)$:

$$\begin{aligned} \frac{\partial \rho_1(r, t)}{\partial t} &= \int dr' \frac{\partial}{\partial t} \sigma_2(r, t|r', t') \rho_1(r', t') \\ &= \int dr'' dr' \omega(r, r'', t) \sigma_2(r'', t|r', t') \rho_1(r', t') \\ &\quad - \int dr'' dr' \omega(r'', r, t) \sigma_2(r, t|r', t') \rho_1(r', t') \\ &= \int dr'' [\omega(r, r'', t) \rho_1(r'', t) - \omega(r'', r, t) \rho_1(r, t)] \end{aligned} \quad (3.12)$$

In the epidemiological model, the time steps (in seasons) are discrete values. Therefore, the integral becomes a sum and it is:

$$\frac{\partial}{\partial t} \rho_n(s) = \sum_{n'=1}^{\infty} [\omega_{n'n}(s) \rho_{n'}(s) - \omega_{nn'}(s) \rho_n(s)] \quad (3.13)$$

Nevertheless, equation (3.11), (3.12) and (3.13) are all Master-equations. $\omega_{nn'}(s)$ is the transition rate from state n to n' at season s and $\omega_{n'n}(s)$ is the rate in the other direction. Both rates are weighted by the probability density $\rho_{n'}(s)$ resp. $\rho_n(s)$.

3.2.3 Determination of the Correct Transition Probabilities

As seen in equation (3.13), it is summed over all possible states N_{states} . The transition rates could therefore be expressed in a matrix notation. This transition matrix \mathcal{W} is two dimensional and consists out of $(N_{states}) \times (N_{states} + 1)$ elements. In addition, the transition matrix is dependent on the season. The transition matrix changes each season, because new possible states are occurring. N_{states} is increased

each new season by one. It is also a transition matrix for each degree class needed, because the transition rates are dependend of the degree k . Therefore, it is $N_{states} = N_{states}(s)$ and $\mathcal{W} = \mathcal{W}^k(s)$:

$$\mathcal{W}^k(s) = \begin{pmatrix} \omega_{00}^k(s) & \cdots & \omega_{0j}^k(s) & \cdots & \omega_{0N_{states+1}}^k(s) \\ \omega_{10}^k(s) & \cdots & \omega_{1j}^k(s) & \cdots & \omega_{1N_{states+1}}^k(s) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \omega_{i0}^k(s) & \cdots & \omega_{ij}^k(s) & \cdots & \omega_{iN_{states+1}}^k(s) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \omega_{N_{states}0}^k(s) & \cdots & \omega_{N_{states}j}^k(s) & \cdots & \omega_{N_{states}N_{states+1}}^k(s) \end{pmatrix} \quad (3.14)$$

for each season s and each degree k . Also one has to remember that 0 is a possible state. A vertex is in state 0, if it never got infected before. It is in state 1, if it got infected last season, in state 2, if the last infection was two seasons before the current season.

The transition rates $\omega_{nn'}^k(s)$ itself are constant for $n, n' \leq N_{states}(s)$. Also, it is only possible to stay in state 0 and not to be transitioned back, because it is impossible to go back to the "never infected state" after an infection. The probability of being not infected is for a degree k vertex $\psi_i^k(s) = (1 + (u(s) - 1)T(i))^k$ for all $i \in \{0, \dots, N_{states}(s)\}$, as seen in equation (2.25) on page 15. Therefore, it is for all seasons and degrees:

$$\omega_{00}^k(s) = \psi_0^k(s) \quad (3.15)$$

In addition, this leads to:

$$\omega_{i0}^k(s) = 0 \quad \forall i \in \{1, \dots, N_{states}(s)\} \quad (3.16)$$

And if a vertex in state 0 gets infected, it will only go into state 1:

$$\omega_{0j}^k(s) = \begin{cases} 1 - \psi_0^k(s) = 1 - \omega_{00}^k(s) & \text{if } j = 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.17)$$

Also, it is possible to go no more than one state up per season. If a vertex gets infected, it will go into state 1. For any possible i with $i \neq 0$, this leads to:

$$\omega_{ij}^k(s) = \begin{cases} \psi_i^k(s) & \text{if } j = i + 1 \\ 1 - \psi_i^k(s) & \text{if } j = 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.18)$$

Now, all possible cases are determined. The transition matrix of equation (3.14) is:

$$\mathcal{W}^k(s) = \begin{pmatrix} \psi_0^k(s) & 1 - \psi_0^k(s) & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 - \psi_1^k(s) & \psi_1^k(s) & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 - \psi_2^k(s) & 0 & \psi_2^k(s) & 0 & \dots & \dots & 0 \\ 0 & 1 - \psi_3^k(s) & 0 & 0 & \psi_3^k(s) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 - \psi_{N_{states}}^k(s) & 0 & \dots & \dots & \dots & 0 & \psi_{N_{states}}^k(s) \end{pmatrix} \quad (3.19)$$

With $\mathcal{W}^k(s)$ and $\rho_0(s = 0) = 1$ (because at the beginning nobody has immunity), it is possible to determine the state densities for all seasons with equation (3.13). This is what one would expect for a Markovian-process.

3.3 Comparison of the Between-Host Model with the Reality - The Case of an Finite Real-World

3.3.1 Limitations of the Transition Matrix Size

The between-host model can have unlimited seasons, $s_{max} \rightarrow \infty$, and therefore an infinite number of transition matrices. But in comparison with the real-world, this makes no sense. In the real world, nobody will live forever. It will be impossible that somebody goes from state $s = 200$ in state $s' = 201$, if we assume that the seasons correspond to the mean with respect years. To reflect this, it is possible to define a maximum number of states $N_{states}^{(max)}$. People die and this subtracts small constant factors from the state densities. Also, new individuals are

born, which gives a small constant term to the density of the first state, state $s = 0$. It is also possible to reflect birth and death effects in this model, but for complexity reasons I decided to neglect birth and death processes except the finite sizes of the transition matrices. As an assumption of this specific model, the population does not change over time. With this assumption, the average state is $\langle s \rangle$ is slightly different than in a more realistic model with birth and death. But this could only change the epidemic size itself, not the long-term dynamics of the infectious disease. Hence even in the infinite model, it is true that the probability of being in a very high state is zero:

$$\lim_{j \rightarrow \infty} \rho_j(s) = 0 \quad (3.20)$$

In this case, the transition matrix is infinite, but can be reduced to a finite size with an error $\epsilon \rightarrow 0$.

In addition, it is expected that the system will go into equilibrium and that the transition matrix will become homogeneous over the seasons. As a result, there will be a circulation of the states without a preferred state:

$$\lim_{i \rightarrow \infty} \rho_j(i) \neq 1 \quad \forall j \in \{0, \dots, N_{states}(s)\} \quad (3.21)$$

Of course, this does not mean that every season there will be an epidemic. But this transition matrix describes the average case for always getting an epidemic.

3.3.2 The Age Structure of the Population

As mentioned above, the age structure of a population has a finite size. This is obvious and leads to the finite size of the transition matrix. But a population

also has a structure to its age distribution. One of the assumptions is that there is no loss or addition of vertices in my model. A more realistic structure would give a bias to the transition matrix because vertices (individuals) are more likely to die between seasons - because they are old.

While this effect shouldn't change to much, another problem is the correlation between age and degree. The probability to have a high degree is certainly dependent on the age. A one year old baby, which is alone at home, has a much smaller degree than an adult who has a job and shares activities with other individuals. This bias is much harder to determine and in my approximation to the real world, I will neglect this correlation and assume no age structure and therefore no such correlation.

3.3.3 The Real-World Degree Distribution

Real-world degree distributions are hard to determine. It is possible to find several different distributions in the literature but in theory it is common to use only few network classes. A real-world network is a case between these borderline cases. Also, a specialty of real-world networks is its maximum degree $k^{(max)}$. Theoretically, it is possible to have a maximum degree of $k^{(max)} = N^{(max)} - 1$ with $N^{(max)}$ as network size. In this case, a vertex with degree $k^{(max)}$ is connected to all other vertices. But in reality, nobody has a connection to everyone else. This is something one needs to keep in mind, if a network model and simulation is built. There will be a difference in the results, if the finite size of the simulation gives a smaller cut-off value than used in the calculation. These is easy to understand, if one imagines a

simulation with one hundred vertices. It is not possible to get a degree with $k > 99$ but it could be possible in an infinite case - and therefore in the calculations. This fact needs to be considered in the calculations.

In addition, it is not very useful to allow degree 0 vertices in a epidemiological model. This are just individuals with no connections to others. Also, it is possible to define a network as a system of connections between vertices and the connected vertices themselves. In this case, all vertices have a degree of $k \geq 1$.

With all these arguments, it is clear that the number of transition matrices is as finite as the transition matrices themselves.

3.4 Modeling the Transmission Probability Between to Vertices

As seen in section 2.3 on page 9, the transmissibility between to connected vertices i and j without cross-immunity is determined by the rate of contacts and and the infection time. Hence the transmissibility with cross-immunity is different due to the possible cross-immunity of j - to whom the disease is spread. According to equation (2.7), it is:

$$\begin{aligned} 1 - T_{ij}^{without} &= \lim_{\delta t \rightarrow 0} (1 - r_{ij}\delta t)^{\frac{\tau}{\delta t}} \\ &= \lim_{\delta t \rightarrow \infty} \left(1 - \frac{r_{ij}}{\delta t}\right)^{\tau\delta t} \\ &= e^{-r_{ij}\tau} \end{aligned} \tag{3.22}$$

This gives:

$$T_{ij}^{without} = 1 - e^{-r_{ij}\tau} \tag{3.23}$$

This is the probability that the disease does not reach j from i . Therefore, $T_{ij}^{without}$ is the transmissibility without cross-immunity, or in another sense the only probability

that the disease is transmitted to the body of vertex j .

In general, the transmissibility with cross-immunity T is a function of the transmissibility without cross-immunity and the time between the last infection and the current season in season units:

$$T_{ij} = f(T_{ij}^{without}, \Delta s_j) \quad (3.24)$$

with Δs_j as difference between the actual season and the season at which j got infected the last time. The term for the probability of cross-immunity can be derived by experiments and is disease specific. In addition, it does not need to be linear, because it is highly dependent on the evolution of the disease strains (Adams and Sasaki, 2007). In this case, I will construct a theoretical cross-immunity function to show the model's dependence on it. Therefore, I propose:

$$T_{ij} = f(T_{ij}^{without}, \Delta s_j) = T_{ij}^{without} \gamma_j(\Delta s_j) \quad (3.25)$$

with $\gamma(\Delta s_j)$ as cross-immunity function.

With this knowledge, the transmissibility can be written in an analogous way as in section 2.3:

$$T(\Delta s) = \langle T_{ij}(\Delta s_j) \rangle = \left\langle T_{ij}^{without} \gamma(\Delta s_j) \right\rangle \quad (3.26)$$

As it is easy to see, it is no longer easy to find an average overall transmissibility. The transmissibility as a property of each edge is the probability of getting infected, not the probability to infect others. Therefore, the transmissibility can be divided in an incoming (defined by the vertex itself) and outgoing transmissibility (defined

by its neighbors). In general, it is possible to define the incoming transmissibility as:

$$T_{ij}^{incoming}(\Delta s_j) = T_{ij}^{without} \gamma(\Delta s_j) \quad (3.27)$$

with $T_{ij}^{without} = T_{ij}^{incoming}(\Delta s_j = 0)$ as average transmissibility in the first season without any immunity (which corresponds to state 0). In addition, this shows a binding property of $\gamma(\Delta s)$:

$$\gamma(\Delta s = 0) = 1 \quad (3.28)$$

which is easy to understand, because a vertex without immunity does not change its transmissibility.

But in general, the calculation of the outgoing transmissibility is not straightforward. It is highly dependent on the used kind of between-season network dynamics. High degree vertices spread the disease on average to a much higher fraction of its neighbors than low degree vertices. Now, if the neighbors do not change between the seasons, high degree vertices will have a much higher outgoing transmissibility than low degree vertices. This is a bias between the outgoing transmissibility and the degree, and hence these correlations destroy the symmetry of the transmissibility in the network. To be correct, it is a necessity to make an average incoming and outgoing transmissibility for each degree class of vertices. But, as we will see later, it is possible to find an easy and very accurate proximity for all outgoing transmissibilities.

3.5 Three Different Models of Between-Season Mixture

The above mentioned correlations only exist for a non-changing network. If every vertex changes all of its neighbors between the seasons, this correlations would not exist. Hence, it depends on the used model and therefore I use three different kinds of models to show the different possibilities. It is expected that the real-world is a mixture of these three models, and that therefore these models are like different borders of reality.

3.5.1 The Non-Mixture Model

In the non-mixture model, a between-season mixture does not exist. The network is generated at the beginning and it is kept constant for the entire simulation and calculation process. Like in most families or with peers in the world, the neighbors, with which a vertex has contacts, are kept the same. In this special case, it is not possible to model a correct average outgoing transmissibility for all vertices. But as mentioned above, it is possible to use a very good proximity.

3.5.1.1 Transmissibility in the Non-Mixture Model

The incoming transmissibility for each vertex is equal to (3.27). It is only dependent on Δs . Averaged over all states, it is:

$$\left\langle T_k^{incoming} \right\rangle (s') = \sum_{j=0}^{s'} T(j) \rho_j^k(s') \quad (3.29)$$

On the other hand, the outgoing transmissibility has the before mentioned correlations. It is dependent on the state of the vertex's neighbors, whose states are dependent on their neighbors. Hence every vertices' state depends on all other

vertices. As a result, it is not possible to model this in a $N \rightarrow \infty$ network. But it is possible to define the average outgoing transmission probability for a degree k vertex for a neighbor's finite size order. For example, if a vertex got infected at season s , its first order outgoing transmissibility will consist out of two parts - the transmissibility towards other infected neighbors and the uninfected neighbors. For the infected neighbors, it is at season $s' = s + 1$:

$$\left\langle T_k^{infected} \right\rangle (s') = T(1) \xi_s^k \quad (3.30)$$

with the average number of infected neighbors:

$$\xi_s^k = 1 + (k-1) \sum_{k'=0}^{\infty} \sum_{j=0}^s q_{k'} \rho_j^{k'}(s) (1 - \left[(1 + (u(s) - 1)T(j))^{k'-1} (1 - T(1)) \right]) \quad (3.31)$$

and $T(1)$ as transmission probability, ρ_j^k as fraction of degree k vertices in state j and $\psi_j^{k'}(s)$ as the probability for a degree k' vertex not to get infected being in state j at season s . The probability $q_{k'}$ is defined as the probability to reach a degree k' neighbor by following one of the vertex's edges:

$$q_{k'} = \frac{k' p_{k'}}{\sum_{k=0}^{\infty} k p_k} \quad (3.32)$$

The average outgoing transmissibility for uninfected neighbors is:

$$\begin{aligned} \left\langle T_k^{not\ infected} \right\rangle (s') = & (k - \xi_s^k) \sum_{k'=0}^{\infty} q_{k'} T(0) \rho_0^{k'}(s) (1 + (u(s) - 1)T(0))^{k'-1} \\ & + (k - \xi_s^k) \sum_{k'=0}^{\infty} \sum_{j=1}^s q_{k'} T(j+1) \rho_j^{k'}(s) (1 + (u(s) - 1)T(j))^{k'-1} \end{aligned} \quad (3.33)$$

Combined, equations (3.30) and (3.33) give the average outgoing transmissibility for an infected degree k vertex:

$$\left\langle T_k^{outgoing\ infected} \right\rangle (s') = \frac{1}{k} \left(\left\langle T_k^{infected} \right\rangle (s') + \left\langle T_k^{not\ infected} \right\rangle (s') \right) \quad (3.34)$$

Now, it is possible to calculate the first order of the outgoing transmissibility for each infected vertex at each season. But one has to keep in mind that the fraction of vertices in a special state is determined by the Master-equation (3.13). Hence it is only possible to calculate the next season after calculating the according outgoing transmissibility. And the next outgoing transmissibility is only calculable after using the Master-equation to determine the next season.

The average outgoing transmissibility for uninfected vertices is in parts similar to equation (3.33):

$$\left\langle T_k^{outgoing \text{ not infected}} \right\rangle (s') = \frac{1}{k} \sum_{k'=0}^{\infty} \sum_{j=0}^{s'} q_{k'} T(j) \rho_j^{k'}(s') \quad (3.35)$$

Obviously, the average transmissibility for a uninfected vertex does not depend on the degree k .

Now, it is possible to define the average outgoing transmissibility in first order for every degree k vertex as follows:

$$\begin{aligned} \left\langle T_k^{outgoing} \right\rangle (s') &= \left\langle T_k^{outgoing \text{ infected}} \right\rangle (s') \rho_1^k(s') \\ &+ \left\langle T_k^{outgoing \text{ not infected}} \right\rangle (s') \sum_{j=0, j \neq 1}^{s'} \rho_j^k(s') \end{aligned} \quad (3.36)$$

It is clear that the complexity of the average outgoing transmissibility grows with number of the used order. Luckily, the correlation strength is scaling with $\langle k \rangle^{-1}$ for each neighbor's order and the error for setting $\mathcal{O}(n) \sim \langle k \rangle^{-n} \approx 0$ for $n = 2, 3, 4, \dots, \infty$ is sufficiently small and approaches 0 fast for $n \rightarrow \infty$. Therefore, I propose that it is sufficient to use an easy approximation of the average outgoing transmissibility for all degree classes of vertices in this model. First, one needs to

determine the average incoming transmissibility for each degree class, like done in equation (3.29). In a second step, one can determine the average outgoing transmissibility as a probability of being connected to a degree k vertex. Hence it is:

$$\left\langle T_k^{outgoing} \right\rangle(s) = \sum_{k'=0}^{\infty} q_{k'} \left\langle T_k^{incoming} \right\rangle(s) \quad (3.37)$$

Under the use of equation (3.29) and (3.37), this reduces to:

$$\left\langle T_k^{outgoing} \right\rangle(s) = \sum_{k'=0}^{\infty} \sum_{j=0}^{s'} q_{k'} T(j) \rho_j^{k'}(s) \quad (3.38)$$

for any possible season s . One needs to remember that the states must be calculated before the outgoing transmissibility can be calculated. As a result, the limits of the sum over the states is correct.

3.5.1.2 Calculation of Component Sizes in the Non-Mixture Model

The difference between incoming and outgoing transmissibility needs a second look on the definition of the component sizes. As seen in section 2.4 on page 11, $H_1(x; T)$ defines the cluster reached by a random edge. This edge has an incoming transmissibility and therefore $P_1(x; T)$ needs also the incoming transmissibility as T (see also figure 2.1 again).

On the other hand, $H_0(x; T)$ defines the cluster reached by a random vertex. Therefore, the outgoing transmissibility is needed and hence $P_0(x; T)$ needs the outgoing transmissibility, too.

3.5.2 The Half-Mixture Model

In the half-mixture model, the mixing behavior between the seasons is defined by a random change of the neighbors of each vertex. But the vertex's degree

itself is kept the same. However, a difference still exists between the incoming and outgoing transmissibility. The definitions of the component sizes are the same as in 3.5.1.2.

3.5.2.1 Transmissibility in the Half-Mixture Model

The incoming transmissibility is the same as in the non-mixture model (see equation (3.29)), but the outgoing transmissibility is different. The random shuffle of the neighbors removes the correlations between the states of the vertices. Hence the outgoing transmissibility for a degree k at season s is the same as equation (3.38). The difference is that this time the outgoing transmissibility isn't an approximation anymore. In this case, it is the actual transmissibility.

3.5.3 The Full-Mixture Model

It is also possible to vary the degree of each vertex as well. This is the full-mixture model. The neighbors and the degree are shuffled by random after each season. But the underlying degree distribution, from which the degrees are drawn, is kept the same over the seasons. This has an important impact on the transition matrix.

3.5.3.1 The Transition Matrix and Transmissibility in the Full-Mixture Model

The transition matrix elements are different than in the other models. It makes no sense to make different matrices for different k values, because after each season the vertices get a new degree. Therefore, the transition matrix is independent

of k :

$$\mathcal{W}(s) = \begin{pmatrix} \psi_0(s) & 1 - \psi_0(s) & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 - \psi_1(s) & \psi_1(s) & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 - \psi_2(s) & 0 & \psi_2(s) & 0 & \cdots & \cdots & 0 \\ 0 & 1 - \psi_3(s) & 0 & 0 & \psi_3(s) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 - \psi_{N_{states}}(s) & 0 & \cdots & \cdots & \cdots & 0 & \psi_{N_{states}}(s) \end{pmatrix} \quad (3.39)$$

and hence the transition elements too. It is:

$$\psi_j(s) = \sum_{k=0}^{\infty} p_k \psi_j^k(s) = \sum_{k=0}^{\infty} p_k (1 + (u(s) - 1)T(j))^k \quad (3.40)$$

The average incoming transmissibility is analogous to equation (3.29), but without the dependency on k . In addition, the average outgoing transmissibility also has no dependency on k and is analogous to equation (3.38). If one compares these equations it is easy to see that, without any dependency of k , they are the same. Hence the average outgoing transmissibility is the average incoming transmissibility. It is:

$$\langle T \rangle(s) = \sum_{j=0}^s T(j) \rho_j(s) \quad (3.41)$$

3.6 Two Different Ways to Calculate the Mean Epidemic Size

The above explained way to calculate the correct states is for the case where one will have an epidemic every season. But as mentioned before, this is only correct with a certain probability. The epidemic size is also the probability of an epidemic. Therefore, one needs to calculate every possible case to determine the correct mean epidemic size per season.

3.6.1 Calculate Every Possible Case

In each season, it is possible to have an epidemic or not (if the average transmissibility is higher than the critical transmissibility - otherwise it is only possible to not have an epidemic). This means that there are up to 2^s possibilities to get a season s . All these cases need to be calculated and weighed, to find the correct mean epidemic size.

In general, one can calculate this by starting with the first season as usual. After defining the correct transmissibility, one can calculate the cluster reached by following a random edge with the self-consistent similar to equation (2.24):

$$\begin{aligned}
u(0) &= H_1(1; T(0)) \\
&= P_1(u(0); T(0)) \\
&= P_1(1 + (u(0) - 1)T(0)) \\
&= \frac{\sum_{k=0}^{\infty} kp_k \psi_0^{k-1}(0)}{\sum_{k=0}^{\infty} kp_k}
\end{aligned} \tag{3.42}$$

In the next step, it is possible to determine the average epidemic size as defined in equation (2.23):

$$S(0) = 1 - \sum_{k=0}^{\infty} p_k \psi^k(0) \tag{3.43}$$

Hence it will give an epidemic of size $S(0)$ with the probability $S(0)$, and no epidemic with the probability $1 - S(0)$. These are the two possible cases in this season.

In the next step, one can use the Master-equation (3.13) to determine the state of each vertex. If there was an epidemic in the last season, the transition

elements can be calculated as shown in the transition matrix (3.19) for the non-mixture, the half-mixture model and (3.39) for the full-mixture model. On the other hand, if there was no epidemic in the last season, the transition matrix will change to:

$$\mathcal{W}^k(s) = \mathcal{W} = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \quad (3.44)$$

because nobody got infected in the last season. With this, it is possible to determine u again. It is for all seasons s with $s > 0$ for the non- and half-mixture model:

$$u(s) = \frac{\sum_{k=0}^{\infty} kp_k \left(\sum_{k'=0}^{\infty} \sum_{j=0}^{n_{max}} q_{k'} \rho_j^{k'}(s) (1 - (1 - u(s))T(j-1)) \right)^{k-1}}{\sum_{k=0}^{\infty} kp_k} \quad (3.45)$$

And for the full-mixture model:

$$u(s) = \frac{\sum_{k=0}^{\infty} kp_k \left(\sum_{j=0}^{n_{max}} \rho_j(s) (1 - (1 - u(s))T(j-1)) \right)^{k-1}}{\sum_{k=0}^{\infty} kp_k} \quad (3.46)$$

Than, the epidemic sizes are for the non- and half-mixture model:

$$S(s) = 1 - \sum_{k=0}^{\infty} \sum_{j=0}^{n_{max}} p_k \rho_j^k(s) \psi_j^k(s) \quad (3.47)$$

And for the full-mixture one:

$$S(s) = 1 - \sum_{k=0}^{\infty} \sum_{j=0}^{n_{max}} p_k \rho_j(s) \psi_j^k(s) \quad (3.48)$$

To determine the average epidemic size, one needs to multiply the epidemic size its probability. Hence you can determine the probability of an epidemic of a

certain size at a certain season by multiplying the last epidemic sizes (or one minus the epidemic size - if there was no epidemic) and multiplying this with the epidemic size itself. This will give the correct probability $g_S(s)$ of an epidemic of size S at seasons s . In addition, one has to take care of the possibility that in one case the average transmissibility dropped under the critical transmissibility. In this case, you would get two of the same case, and therefore one case doesn't need to be considered.

After one determines the correct number of different cases N_{cases} , it is possible to determine the mean epidemic size at season s is:

$$\langle S \rangle(s) = \frac{\sum_{j=1}^{N_{cases}} g_{S_j}(s) S_j(s)}{\sum_{j=1}^{2^s} g_{S_j}} \quad (3.49)$$

which is the sum over all possible epidemics S_j at season s , weighted by their probabilities. In addition, this is normalized by the probabilities, because $\sum_{j=1}^{2^s} g_{S_j} \neq 1$ for every $T < 1$.

3.6.2 Calculate Only the Average Case

Another way to calculate the mean epidemic size is to calculate only the average case for every season. Of course, this approach has some weaknesses. For example, it is possible that in one of all cases at season s the average transmissibility drops below the critical transmissibility and that this network goes into the subcritical phase. But the average network does not go into this, because the average transmissibility of the averaged network does not drop below the critical transmissibility. In this case, the calculated mean epidemic size becomes also low. But on the other hand, this algorithm is much faster than the described one in section 3.6.1

(see figure 3.1 for a comparison). Actually, this calculation of the average case is the only one possible for a run over a large number of seasons; the algorithm of 3.6.1 would take too long.

This approach is mostly similar to 3.6.1. The only difference is that the transition matrix with $s' = s - 1$ is defined as follows:

$$\mathcal{W}^k(s) = \begin{pmatrix} (1 - S(s')) + S(s')\psi_0^k(s) & S(s') (1 - \psi_0^k(s)) & 0 & \cdots \\ 0 & S(s') (1 - \psi_1^k(s)) & (1 - S(s')) + S(s')\psi_1^k(s) & \cdots \\ 0 & S(s') (1 - \psi_2^k(s)) & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (3.50)$$

The difference between (3.50) and (3.19) is that the elements are multiplied by the probability that there was an epidemic (or not) in the last season. This is key to the "average behavior" of this calculation, and because of this, it is not necessary to calculate every possible case.

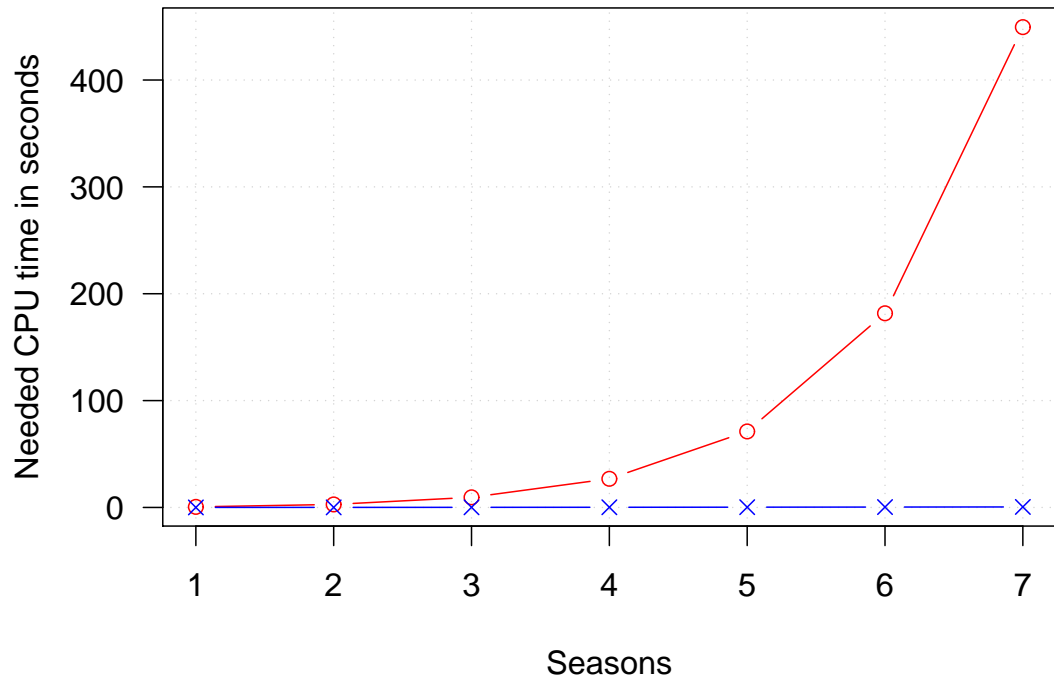


Figure 3.1: Comparison of the needed CPU time in seconds for a calculation of the mean epidemic size per season in *Mathematica* (Wolfram Research, Inc., 2007) on an Intel(R) Xeon(R) 5140. A Poisson network with $\lambda = 4$ was used. The red line and circles are the algorithm of 3.6.1. The blue line and crosses are the algorithm of 3.6.2.

Chapter 4

Simulating the Spread of Infectious Diseases in the Multi-Season Model

4.1 General Concerns for Network Modelling

As with every percolation model, there are some general concerns in writing a simulation. First of all, finite size effects should be kept as small as possible. Of course, an infinite network is not possible to realize in a computer simulation. But the error can be kept small by the use of a network with sufficiently large size. This leads to the need of fast code, because a large network causes many calculation units for the CPU. C++ is an especially fast, reliable program language, and therefore appropriate to write the source code in. Therefore, all simulations were written in C++. In the following, I will explain basic concepts of the programming in pseudo-code, because the full C++ code with explanations would be too long and would need a much more detailed view than is possible to describe in this thesis.

4.1.1 Generating Random Numbers

4.1.1.1 Mersenne Twister Generator

As for every Monte-Carlo simulation, it is necessary to have a reliable random number generator. In this case, I used the Mersenne Twister Generator (Matsumoto and Nishimura, 1998), which has a period length of $2^{19937} - 1$ and is equidistributed over 623 dimensions. In general, the Mersenne Twister cannot produce special

distributed random numbers. Its standard output are uniformly distributed doubles between 0 and 1.

4.1.1.2 Generating Poisson Distributed Random Numbers

To generate Poisson distributed random numbers, it is possible to use a characteristic of the Poisson distribution. The Poisson distribution is equivalent to the probability distribution of the needed time (in discrete steps) for an event to occur, which is independent of the needed time for the past events:

$$P(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (4.1)$$

with λ as the mean and k as discrete time. In physics, this characteristic is also known as *shot noise*. With this knowledge, it is possible to construct random numbers distributed by the Poisson distribution. Hence an algorithm in pseudo-code is:

```

set  $x = e^\lambda$ ,  $k = 0$  and  $p = 1$ ;
do {
    set  $r$  as random double between 0 and 1;
    set  $k = k + 1$ ;
    set  $p = p * r$ ;
}
while (  $p \geq x$  );
return  $k$ ;

```

The generation of r is done by the Mersenne Twister Generator.

4.1.1.3 Generating Scale-Free Distributed Random Numbers

Many real-world networks can not be described by Poisson distributed degrees. Scale-free networks are especially common and their degree distribution is a power-law distribution:

$$P(k; \alpha; \kappa) = \begin{cases} 0 & \forall k = 0 \\ Ck^{-\alpha}e^{-\frac{\alpha}{\kappa}k} & \forall k \geq 1 \end{cases} \quad (4.2)$$

with C as normalization constant and α and κ as specific parameters. In addition, it is not a true scale-free distribution, because the exponential tail produces a cut-off around degree $k = \kappa$. But this power-law with cut-off distribution is much more realistic than a true scale-free distribution. Another useful side effect is that this exponential tail makes the cumulative probability, which is the sum in respect of k over the distribution, finite and therefore the distribution normalizable, even if $k^{(max)} \rightarrow \infty$.

A fast algorithm to produce power-law distributed random number, which was also used in my simulations, is:

```
make a list  $pk$  of the degree  $k$  probabilities with size  $pk[k^{(max)}]$ ;
set  $k = 0$ ,  $i = 1$  and  $x = 0$ ;
set  $r$  as random double between 0 and 1;
do {
    set  $x = x + pk[i]$ ;
    set  $k = k + 1$ ;
    set  $i = i + 1$ ;
}
```

```

while (  $r \geq x$  );
return  $k$ ;

```

4.1.1.4 Generating Exponential Distributed Random Numbers

Also, exponential distributions are another class of degree distributions. An exponential distribution is:

$$P(k; \beta) = C e^{k\beta} \quad (4.3)$$

with C as normalization and β as specific parameter. The algorithm for the production of exponential distributed random numbers is equivalent to 4.1.1.3. Only the degree probabilities p_k are now taken from an exponential distribution.

4.1.2 Building the Network - an Algorithm for the Connection of Vertices

The building and connecting of randomly joint vertices is the most CPU intensive part of the simulation. While there are many different algorithms possible and also available in literature (Erdős and R enyi, 1959), I decided to write my own one. This gave me the ability to be flexible and find the best solution for the simulation’s task.

In general, the algorithm can be separated into three main parts. The first part builds the vertices with the correct degree distribution. The algorithms for the degree distributions have been described before.

Part 1

set the size N of the network;

make a list *vertices* of the vertices with size *vertices*[N];

```

do {
    fill vertices with random numbers smaller than  $N$  from the matching
    the degree distribution; }
while ( accumulative sum of vertices is odd );

```

Now, the vertices are constructed and have the correct degree. In the next step, the vertices are connected randomly to build a network:

Part 2

```

make a list net.old with size net.old[ $N$ ] and fill it with net.old[ $i$ ] =  $i$ ;
make a list net.new with size net.new[ $N$ ];
set size =  $N$ ;
do {
    set  $r$  as a random integer between 1 and size;
    take net.old[ $r$ ] and write it into net.new[size];
        for( start  $i$  at  $r$  and go to size ) {
            set net.old[ $i$ ] = net.old[ $i + 1$ ];
        }
    set size = size - 1;
}
while ( size > 1 );
destroy net.old and set size =  $N$ ;
make a list connections with size connections[ $N$ ];
for ( start  $i$  at 1 and go to size ) {
    for ( start  $j$  at 1 and go to vertices[net.new[ $i$ ]] ) {
        write net.new[ $i + j$ ] in connections[net.new[ $i$ ]];
    }
}

```

```

write net.new[i] in connections[net.new[i + j]];
set vertices[net.new[i + j]] = vertices[net.new[i + j]] - 1;
  if ( vertices[net.new[i + j]] = 0 ) {
    for ( start q at i + j and go to size ) {
      set net.new[q] = net.new[q + 1];
      set size = size - 1;
    }
  }
set net.new[i] = net.new[i + 1];
set size = size - 1;
}
destroy net.new;
destroy size;

```

After *Part 1* and *Part 2*, all vertices are connected randomly with the correct degree distribution. But it is still possible to find loops in the network. Per definition, a loop is a construct where one vertex has two ends of edges outgoing to the same vertex. As a result, it is possible to have loops between two vertices or self-loops within one vertex. Both cases are not allowed in my epidemiological model. Hence *Part 3* is needed to remove the loops:

Part 3

```

for (start i at 1 and go to N ) {
  if (check if two or more equal vertices are in connections[i]) {
    store this vertex in change;
  }
}

```



```

do {
  set  $r1$  as a random integer between 1 and  $size$ ;
  set  $r2$  as a random integer between 1 and the size of  $vertices[r1]$ ;
}
  while ( $i$  or  $change$  is in  $connections[r1]$  or  $r1 = i$  );
  overwrite  $change$  with  $vertices[r2]$  in  $connections[i]$ ;
  overwrite  $vertices[r2]$  with  $change$  in  $connections[r1]$ ;
}

```

After *Part 1*, *Part 2* and *Part 3*, you will have a correctly distributed network without loops.

4.1.3 Spread of the Infection

The spread of the infection is an easy process. First, one needs to choose and infect a vertex by random. In the second step, the disease needs to be spread through the network. This is done by a percolation process. The pseudo-code for this is:

```

set  $r$  as random integer between 1 and the size of the network  $N$ ;
do {
  take  $vertices[r]$  and write it into the list infected;
  get the neighbors of  $vertices[r]$ ;
  for (start  $i$  at 1 and go to the number of neighbors of  $r$ ) {
    get the incoming transmissibility  $t[i]$ ;
    set  $z$  as random number between 0 and 1;

```

```

if (is  $z < t[i]$ ) {
    write  $i$  into the list infected;
}
}
while (there are new infected);

```

After this, it is possible to determine the fraction of infected vertices. In addition, it is possible to get the average transmissibility - overall or sorted by degree classes - or the probability that a degree k vertex did not get infected. One needs to repeat the network construction algorithm or spreading algorithm after each season, if it is necessary due to the used mixing model.

Chapter 5

Results of the Simulations and Calculations

It is possible to determine many different specific values of an epidemic model. In the following, I will show the most informative, useful parameters and results of a multi-season model's simulation (see chapter 4 on page 42) . In addition, I will validate these results with calculations (see chapter 3 on page 17). I used *R* for data analysis and plotting of the graphs (R Development Core Team, 2007). The calculations were done using *Mathematica* (Wolfram Research, Inc., 2007).

In the following, all calculations and simulations were performed with $\gamma(\Delta s_j) = e^{\frac{-1}{\Delta s_j}}$ and:

$$T_{ij} = T_{ij}^{without} e^{\frac{-1}{\Delta s_j}} \quad (5.1)$$

if not stated otherwise. In addition, T_{ij} was determined by equation (2.26) with $R_0 = 2.5$. All simulations were performed on a 10,000 vertices network with 30 seasons. The simulations for each network type were averaged over 1,000 runs. The calculations were performed with the accurate, but slow algorithm (see section 3.6.1). The calculations were done until season 10, after this all following seasons were set as season 10. As shown in the following, this assumption is valid, because the networks are already in equilibrium at season 10. Also, there is no difference in calculations for the non-mixture and half-mixture model. Therefore, calculations

were made only for the half- and full-mixture model and the theoretical calculations of the non-mixture model were taken from the half-mixture model.

If not stated otherwise, Poisson networks with $\lambda = 4$, $\lambda = 8$ and $\lambda = 12$ were used. The exponential networks have β values of $\beta = \frac{1}{4}$, $\beta = \frac{1}{8}$ and $\beta = \frac{1}{12}$, where the scale-free networks have $\alpha = 1.4$, $\alpha = 1.5$ and $\alpha = 1.6$ with $\kappa = 30$.

5.1 Evaluating the Model

This section evaluates different parameters of the model, such as mean prevalence, distribution of epidemic sizes and the degree-specific evolution of epidemiological risk. The calculated predictions are compared to simulations for Poisson, exponential and scale-free networks.

5.1.1 Epidemic Sizes

The epidemic size of an epidemic model is the most important specific value. It is also the easiest one to compare. For example, it is not possible to design a network by asking every citizen about their contacts with other ones in the last two weeks. But it is possible to determine, how many people got ill and went to a hospital - just by counting it. Hence the epidemic size is the first value to evaluate the quality of an epidemic model. The model is not correct if it is not able to reproduce the experimental epidemic size. In addition, it makes different models comparable.

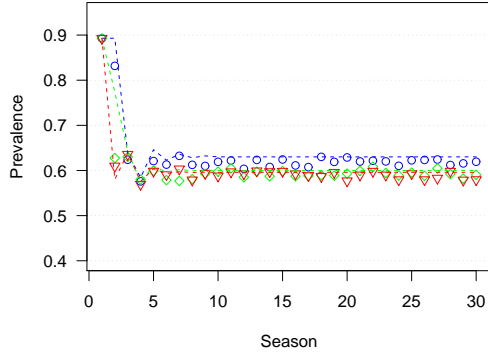
The mean epidemic sizes, also called prevalences, are shown in figures 5.1, 5.2 and 5.3. The simulated prevalence and the calculated prevalence are in very

good agreement for all network types. There are only small differences at season 2 and 3 between calculations and simulations, but the calculations predict the correct equilibrium's prevalence. Hence the calculations are capable of correctly describing the model's mean epidemic size.

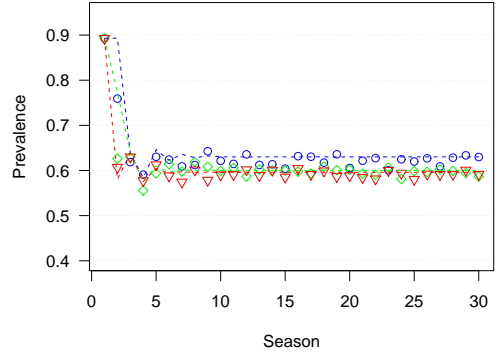
Also, the prediction of an equilibrium state is true. As simulations and calculations show, the prevalence goes into an equilibrium after approx. season $s > 6$ in the Poisson network and earlier in exponential and scale-free networks.

Furthermore, the mean epidemic size is the same for all three different networks in all mixture models at season 1 for Poisson networks. In contrast to the Poisson networks, the same R_0 value does not give the same epidemic size in exponential and scale-free networks. But the difference between the mean epidemic sizes keeps relatively constant in the equilibrium case (approx. for all seasons greater than 6) for all network types. This result is unexpected, especially because the epidemic size does not scale linearly with T in the percolation network model (Newman, 2002; Meyers et al., 2005). One explanation could be that these T values are in a quasi-linear branch of the T spectrum. Simulations with other T values could verify this behavior in the future.

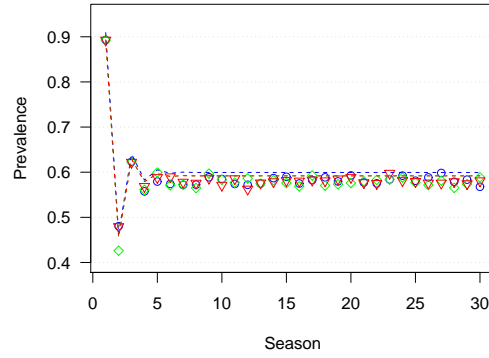
However, the mean epidemic size is only an average over all possible cases. And most of the time, the mean epidemic size is an impossible size to achieve. This can be seen by plotting all epidemic sizes with each probability or respectively frequency. For the simulations with $\lambda = 8$, I did this for various seasons in the histograms of figure 5.4. It shows that these networks do not form a single epidemic size, which would also be the mean epidemic size. As a point of fact, many different



(a) Prevalence for the non-mixture model

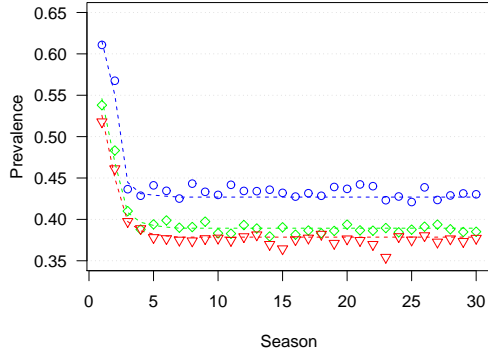


(b) Prevalence for the half-mixture model

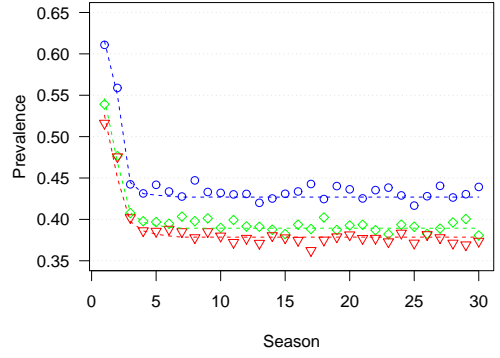


(c) Prevalence for the full-mixture model

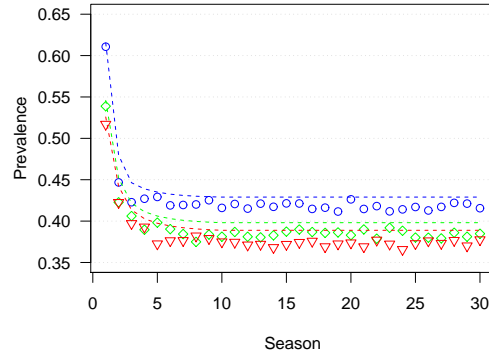
Figure 5.1: Prevalences for the different mixture models in Poisson networks. Blue circles are for the simulated data with $\lambda = 4$, green rhomboids are for $\lambda = 8$ and red triangles are for $\lambda = 12$. The colored lines are for the calculations respectively.



(a) Prevalence for the non-mixture model

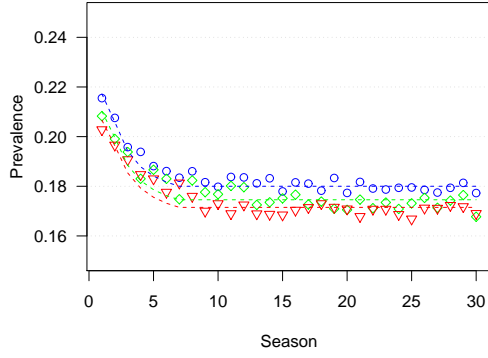


(b) Prevalence for the half-mixture model

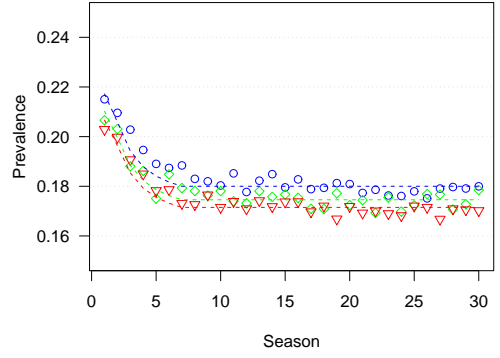


(c) Prevalence for the full-mixture model

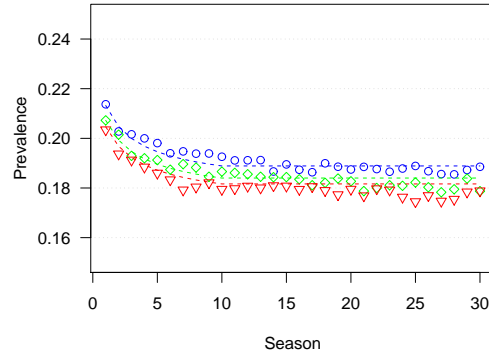
Figure 5.2: Prevalences for the different mixture models in exponential networks. Blue circles are for the simulated data with $\beta = \frac{1}{4}$, green rhomboids are for $\beta = \frac{1}{8}$ and red triangles are for $\beta = \frac{1}{12}$. The colored lines are for the calculations respectively.



(a) Prevalence for the non-mixture model

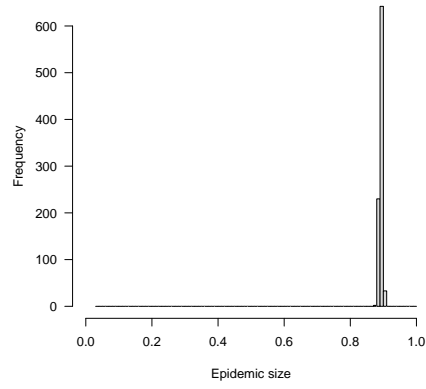


(b) Prevalence for the half-mixture model

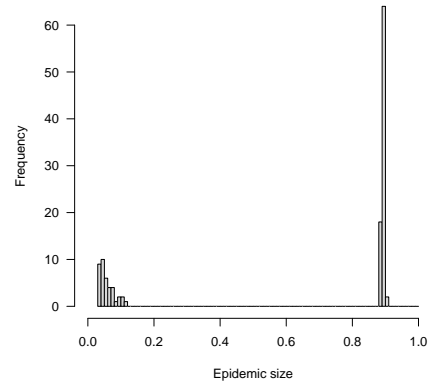


(c) Prevalence for the full-mixture model

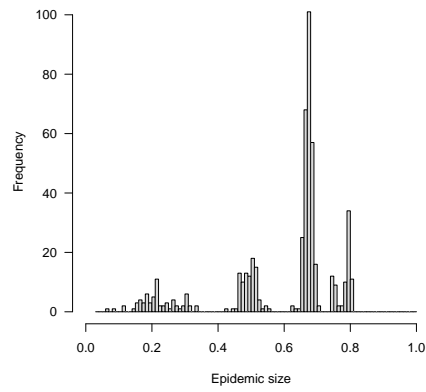
Figure 5.3: Prevalences for the different mixture models in scale-free networks. Blue circles are for the simulated data with $\alpha = 1.4$, green rhomboids are for $\alpha = 1.5$ and red triangles are for $\alpha = 1.6$. The colored lines are for the calculations respectively.



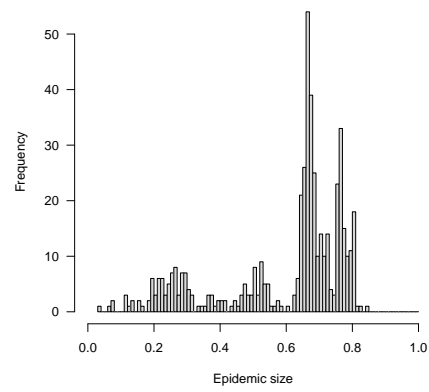
(a) Season 1



(b) Season 2

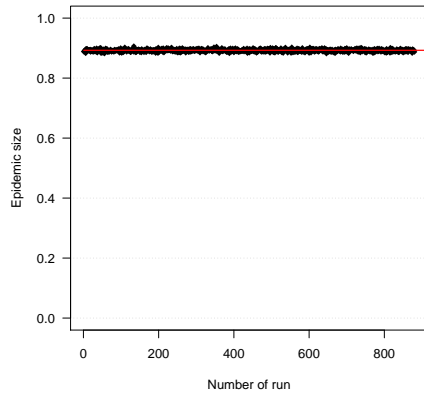


(c) Season 5

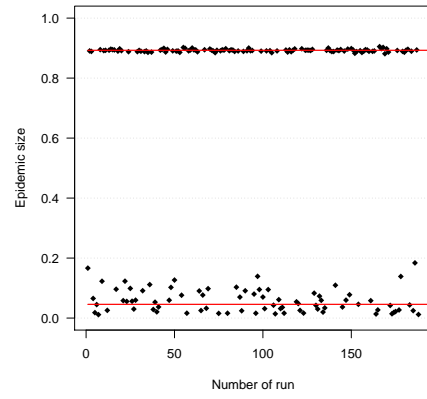


(d) Season 10

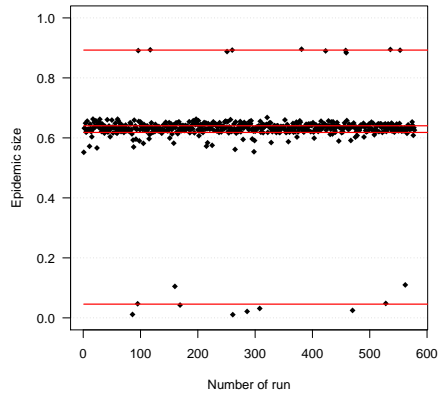
Figure 5.4: Histograms of the simulations' epidemic sizes of a Poisson $\lambda = 8$ network. The non-mixture model was used. Each bar has a width of 0.01.



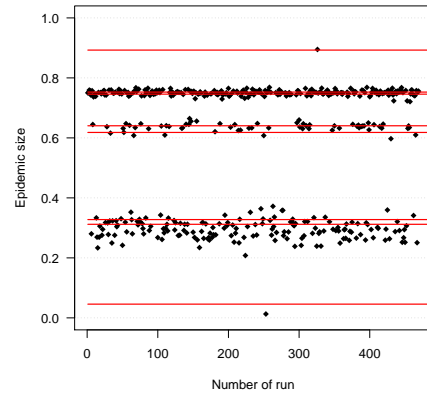
(a) Season 1



(b) Season 2



(c) Season 3



(d) Season 4

Figure 5.5: Epidemic sizes of a Poisson $\lambda = 10$ network. The half-mixture model was used. The black rhomboids are simulated data, the red lines are the calculated epidemic sizes.

epidemic sizes are produced at each season. But still, as seen in figure 5.1, the prevalences are correctly described by the calculations. This raises the question about the actual epidemic sizes. Figure 5.5 shows plots with all simulated epidemic sizes and the calculations. The calculations are able to correctly reproduce all epidemic sizes. The calculations are only a small bit off in the non-mixture model due to unconsidered correlations. Even if not shown, it is true for all network types. The axis with the number of runs changes, because there is only an epidemic with a certain probability each season. In the percolation network model, the probability of an epidemic is also the epidemic size (Newman, 2002). But this is no longer true for the multi-season model, because of the vertices history (see chapter 3). Another possibility is to look at the probability of infection for a single degree k vertex, as this determines the epidemic sizes (see equation (3.47) on page 38).

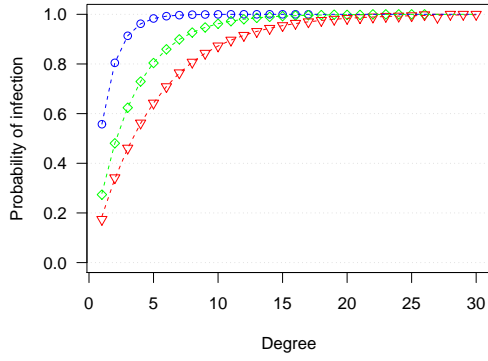
5.1.2 Degree-Specific Evolution of Epidemiological Risk

The probability of being infected is different for each degree class and each season. To get the average infection probability of a degree k vertex in season s , one has to average over all possible states in this degree class, as it is done in equation (3.47). Figure 5.6 shows the probabilities of infection, or epidemiological risk, for different degree classes in a Poisson network. Figure 5.7 and figure 5.8 show the same probabilities of infection for exponential and scale-free networks. Figures 5.6 - 5.8 show only the non-mixture model. One can see how the probabilities converge to a certain, network and season specific value. High degree vertices have a much higher chance to get infected than low degree vertices - a result which is expected.

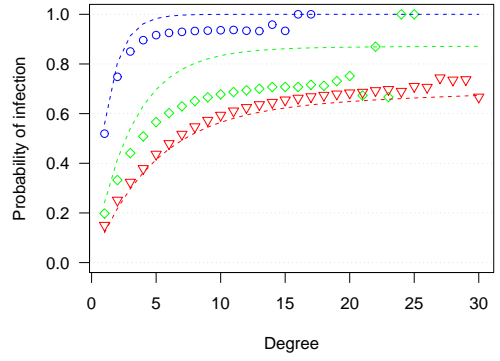
Also, one can see large fluctuations for high degrees or even missing values in Poisson networks. This is an effect of the small degree probability. The Poisson distribution has only a small deviation around its mean. Therefore, it is very unlikely to get high degree vertices. Hence, 1,000 runs are not enough to produce a network with this high degree. This is also the same explanation for the high fluctuations. Only very few networks have such high degree vertices. And as a result, there are not enough vertices to get the correct average probability of infection.

The calculations are able to reproduce the simulated data. Even if the calculations neglect the correlations of the non-mixture model, the agreement, except for season 2, is very good. This lack of correctness at season 2 is also visible in the calculations of the mean epidemic size. To see this behavior, one can also look at the probability of being infected for a fixed degree over several seasons. This was done in figures 5.9 - 5.11. Again, fluctuations and missing values in Poisson networks are due to the low degree probability (see above). But more can be explained with these Poisson plots. In season 2, one can see a negative peak for high degrees in Poisson networks. This peak gets even bigger with the degree. In addition, we have a lack of this peak for $k = 1$. The same kind of peak can be found in the mean epidemic size in season 2. If one compares both, it is obvious that the high degree vertices are responsible for this peak at season 2. The same behavior, but smaller, can also be observed for season 3 in some Poisson networks.

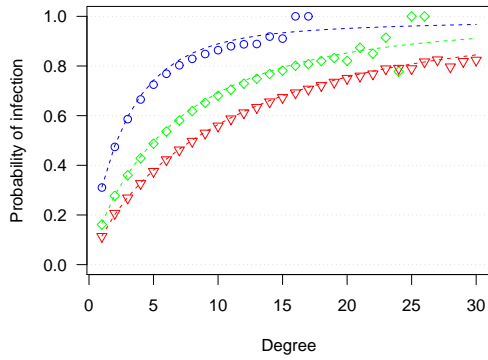
In exponential and scale-free networks, no such peak can be observed. But again, a comparison between the mean prevalence and the probability of infection shows that high-degree vertices determine the behavior of the networks.



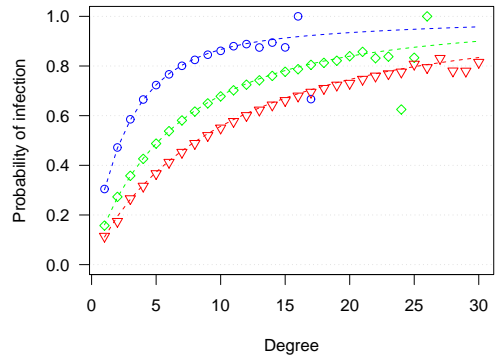
(a) Season 1



(b) Season 2

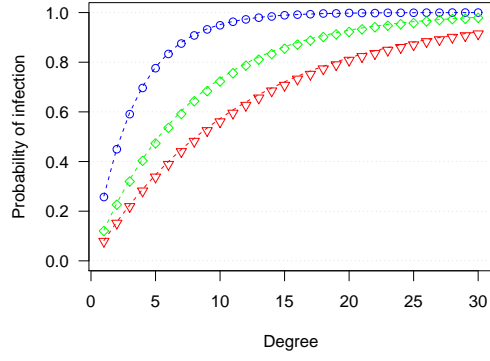


(c) Season 5

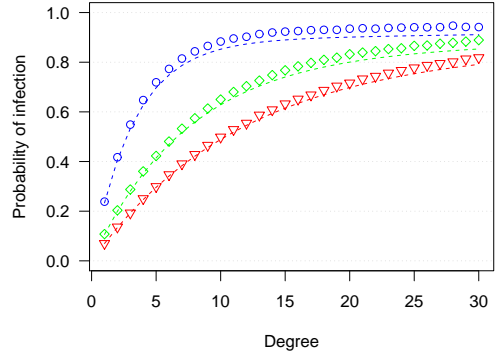


(d) Season 10

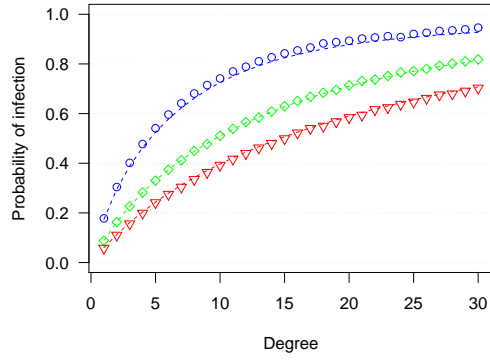
Figure 5.6: Probability of being infected for different degrees in Poisson networks. The non-mixture model was used. The blue circles are simulated data with $\lambda = 4$, the green rhomboids are simulated data with $\lambda = 8$ and the red triangles are simulated data with $\lambda = 12$. The lines are colored for the calculations respectively. The lack of simulated values is due to the small probability for this degree.



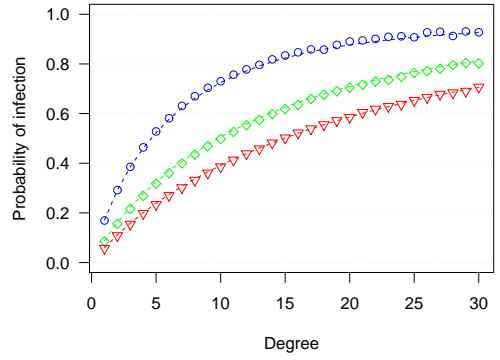
(a) Season 1



(b) Season 2

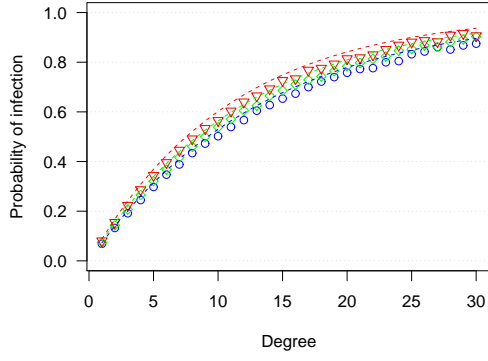


(c) Season 5

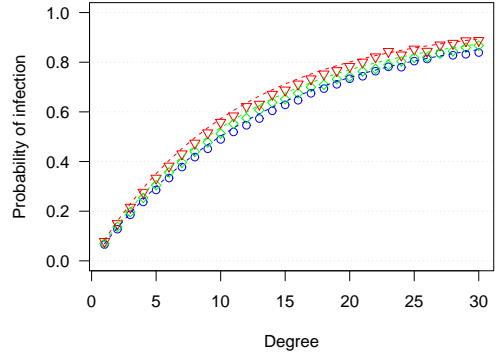


(d) Season 10

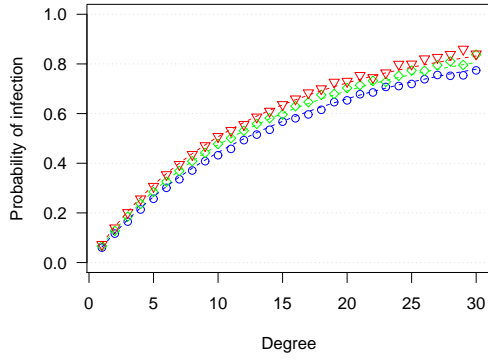
Figure 5.7: Probability of being infected for different degrees in exponential networks. The non-mixture model was used. The blue circles are simulated data with $\beta = \frac{1}{4}$, the green rhomboids are simulated data with $\beta = \frac{1}{8}$, the red triangles are simulated data with $\beta = \frac{1}{12}$. The lines are colored for the calculations respectively.



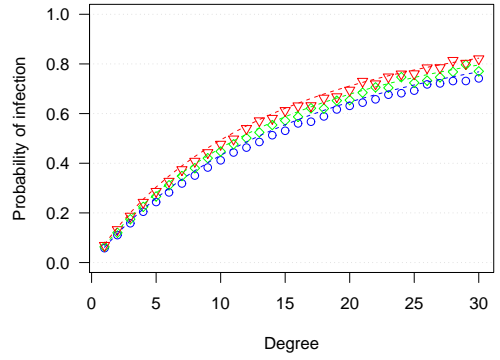
(a) Season 1



(b) Season 2



(c) Season 5



(d) Season 10

Figure 5.8: Probability of being infected for different degrees in scale-free networks. The non-mixture model was used. The blue circles are simulated data with $\alpha = 1.4$, the green rhomboids are simulated data with $\alpha = 1.5$, the red triangles are simulated data with $\alpha = 1.6$. The lines are colored for the calculations respectively.

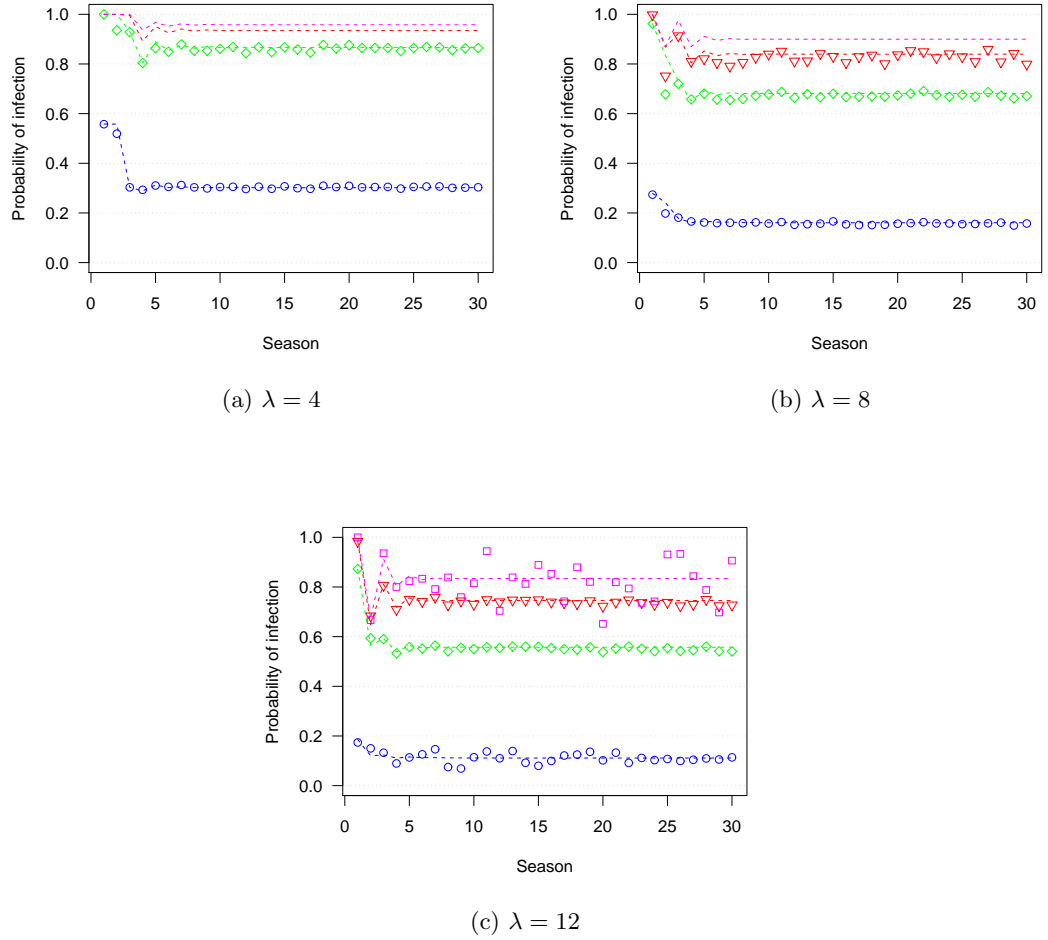
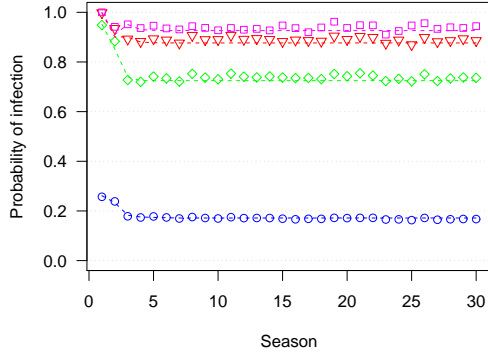
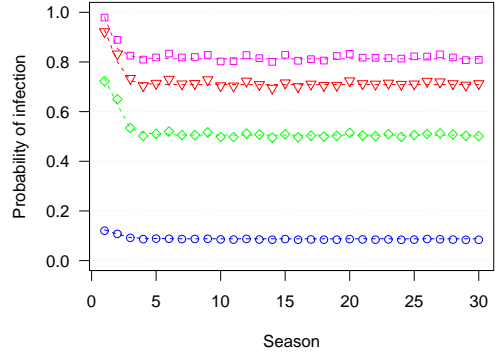


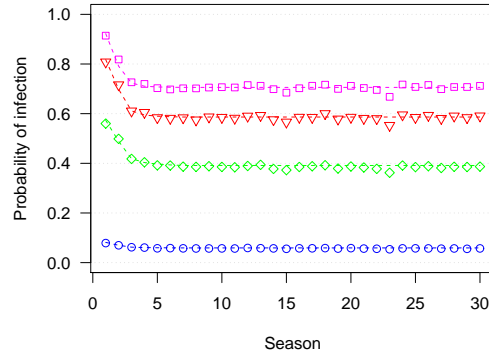
Figure 5.9: Probability of being infected for different degrees in Poisson networks. The non-mixture model was used. The blue circles are simulated data with $k = 1$, the green rhomboids are simulated data with $k = 10$, the red triangles are simulated data with $k = 20$ and the magenta squares are simulated data with $k = 30$. The lack of simulated values is due to the small probability for this degree.



(a) $\beta = \frac{1}{4}$

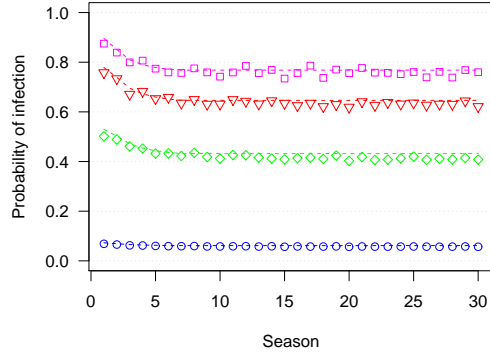


(b) $\beta = \frac{1}{8}$

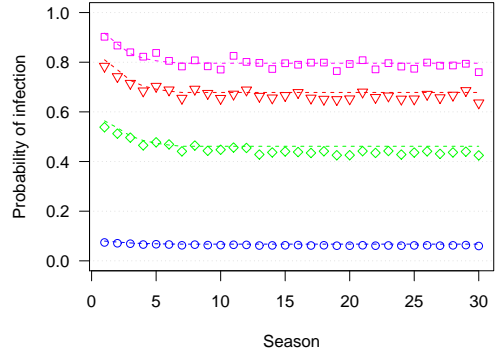


(c) $\beta = \frac{1}{12}$

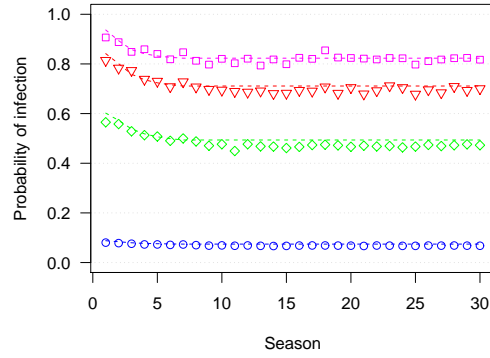
Figure 5.10: Probability of being infected for different degrees in exponential networks. The non-mixture model was used. The blue circles are simulated data with $k = 1$, the green rhomboids are simulated data with $k = 10$, the red triangles are simulated data with $k = 20$ and the magenta squares are simulated data with $k = 30$.



(a) $\alpha = 1.4$



(b) $\alpha = 1.5$



(c) $\alpha = 1.6$

Figure 5.11: Probability of being infected for different degrees in scale-free networks. The non-mixture model was used. The blue circles are simulated data with $k = 1$, the green rhomboids are simulated data with $k = 10$, the red triangles are simulated data with $k = 20$ and the magenta squares are simulated data with $k = 30$.

5.1.3 Distribution of Epidemic Sizes

It is still true that a certain epidemic occurs with the same probability as its certain epidemic size, but the probability to get this special epidemic size depends only on the history of the network. Therefore, it is possible to determine the correct probabilities for each season by multiplying the single probabilities for each season. This has been done for the half-mixture model in figure 5.12 for a Poisson network. No simulated data was used, because the probabilities decline very fast. Much more than 1,000 runs would be needed to show these probabilities with simulated data. But in fact, this is not needed to validate the correctness of the calculations. It was shown before that the calculations are capable to reproduce the correct mean epidemic size. In addition, the calculations can determine all epidemic sizes correctly. Now, only if it is also possible to calculate the correct probability of a certain epidemic size, it would be possible to calculate the correct mean epidemic size (see section 3.6.1). As this is possible, it is clear that the probabilities must be correct, too. It is also possible to compare the histogram of the simulated data with the calculated probabilities. This is done in figure 5.13 for a Poisson network and one can see that both plots are in good agreement. Again, the distribution of epidemic sizes goes into an equilibrium, like the mean epidemic sizes and the epidemiological risk.

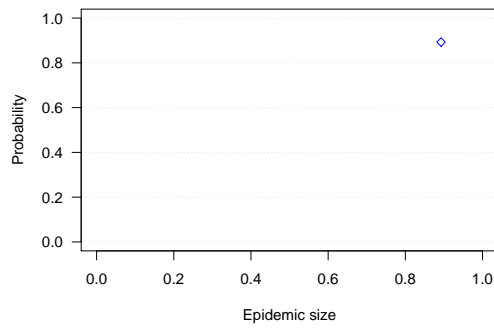
To visualize the probabilities in a better way, it is possible to plot the accumulative probability of getting a epidemic smaller than a certain size. This is done for the half-mixture model in figures 5.14 - 5.16. If one compares figure 5.14 with figure 5.5, one can see that the branch with the steepest slope in the accumula-

tive probability plot corresponds to the epidemic size with the most simulated data points in it. This is also true for exponential and scale-free networks. Again, this supports the validation of the calculations.

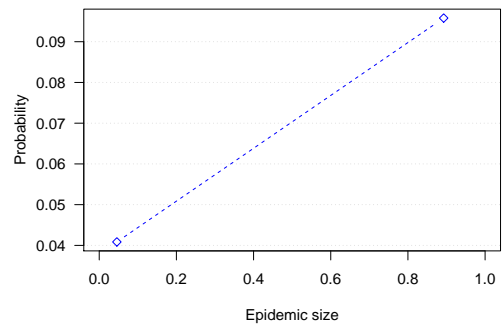
The probabilities for the full-mixture model are almost the same. Figure 5.15 illustrates the accumulative probabilities of getting an epidemic smaller than a certain size in a full-mixture model. Figures 5.14 and 5.15 show both very similar accumulative probability functions. Hence, this result is in good agreement with the findings regarding the mean epidemic size. For Poisson networks, season 2 is the one with the biggest prevalence difference in the mixture models as well (see also tables 5.1 and 5.2).

Another possibility is to look at the probabilities of having an epidemic at all, which is the accumulative probability over all epidemic sizes. Furthermore, it is possible to verify this by simulated data. 1,000 runs are not enough to give a good distribution of epidemic sizes, but it is enough to correctly determine the probability of getting an epidemic for each season. This was done in figure 5.23 for a Poisson network.

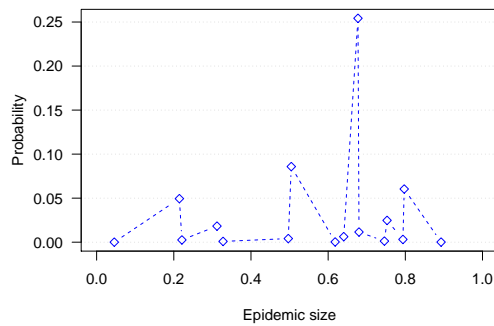
Again, the calculations match the simulations. But the mean epidemic size not longer corresponds with the probability of an epidemic. In fact, the probability of having an epidemic is slightly smaller than the mean epidemic size. Also, the difference between the mixture models is the greatest at season 2. This can also be seen in table 5.1 and 5.2, where the probabilities of having an epidemic for the first 10 seasons are shown. Every simulated probability is in the 95% confidence interval of the calculations. This is also shown in tables 5.3 and 5.4 for exponential networks



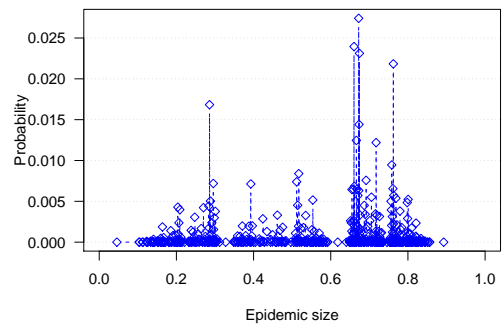
(a) Season 1



(b) Season 2

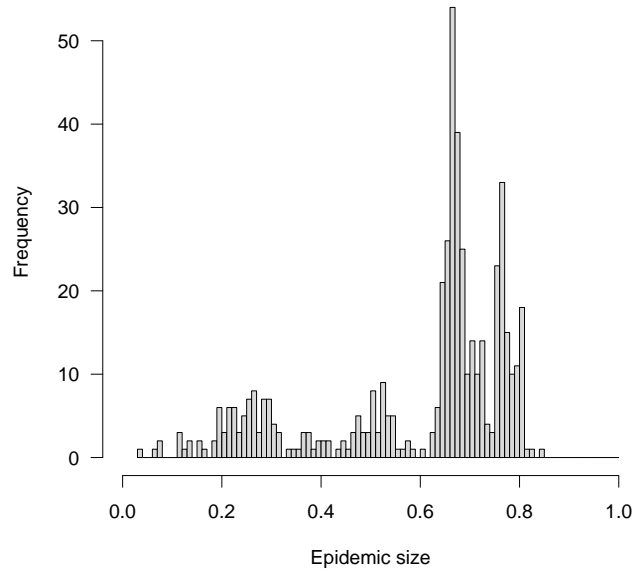


(c) Season 5

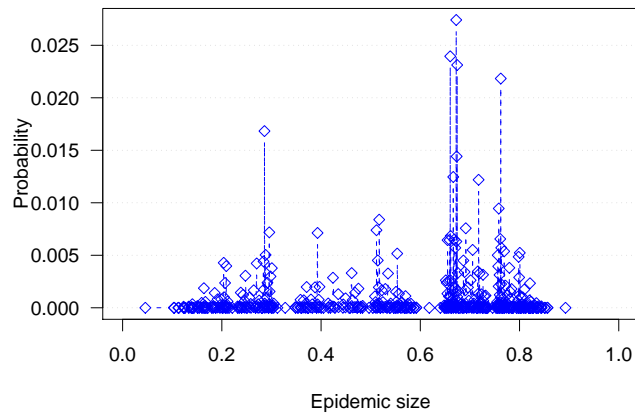


(d) Season 10

Figure 5.12: Probabilities of epidemic sizes for a Poisson $\lambda = 10$ network. The half-mixture model was used. The blue circles are calculated data, the blue line is only for better visibility. The probability to have an epidemic size other than shown by a circle is 0.

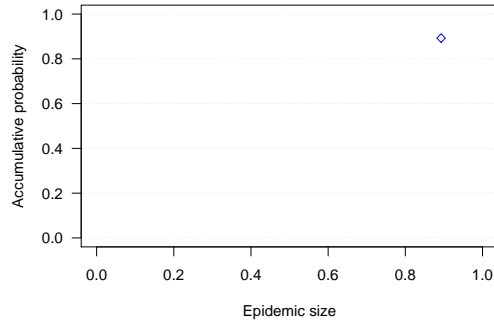


(a) Season 10 - Simulated data

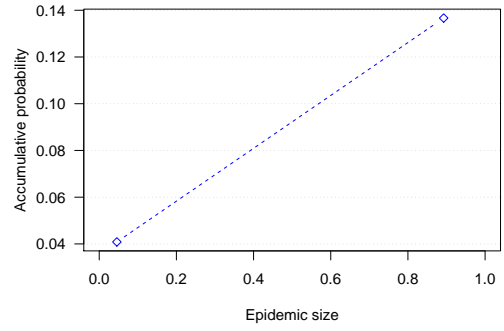


(b) Season 10 - Calculated data

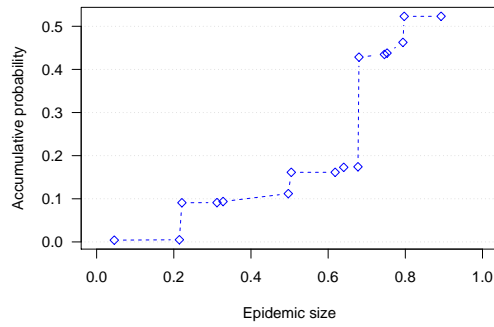
Figure 5.13: Probabilities and frequencies of epidemic sizes compared for a Poisson $\lambda = 8$ network. The half-mixture model was used. The blue circles are calculated data, the blue line is only for better visibility. The probability to have an epidemic size other than shown by a circle is 0.



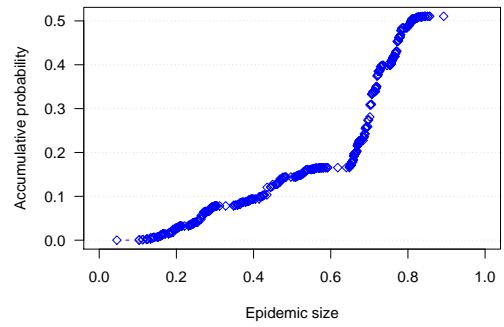
(a) Season 1



(b) Season 2

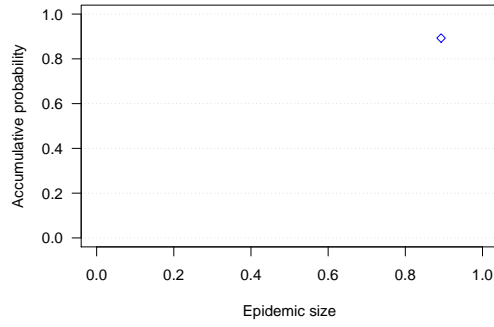


(c) Season 5

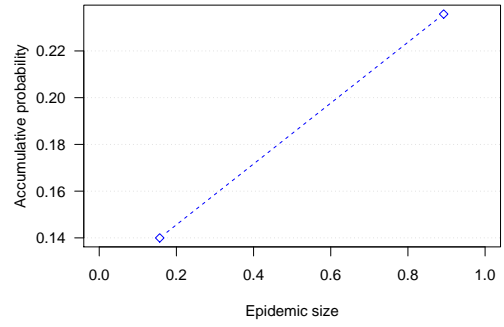


(d) Season 10

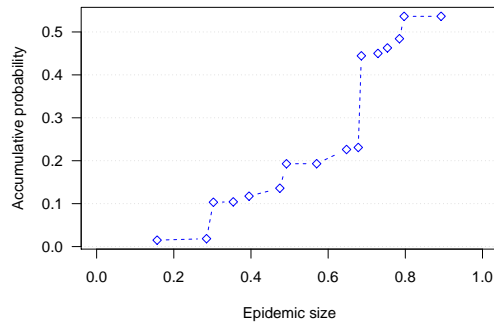
Figure 5.14: Accumulative probability of having a epidemic size smaller than a certain epidemic size for a Poisson $\lambda = 10$ network. The half-mixture model was used. The blue circles are calculated data, the blue line is only for better visibility. The probability to have an epidemic size other than shown by a circle is 0.



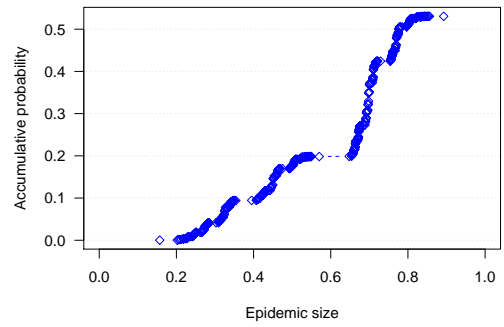
(a) Season 1



(b) Season 2

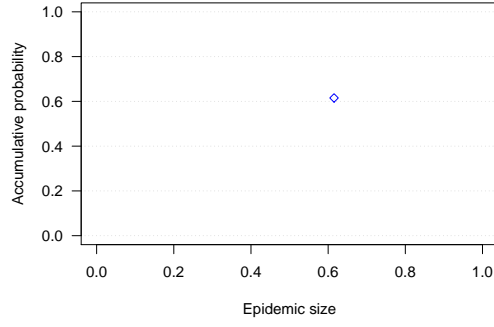


(c) Season 5

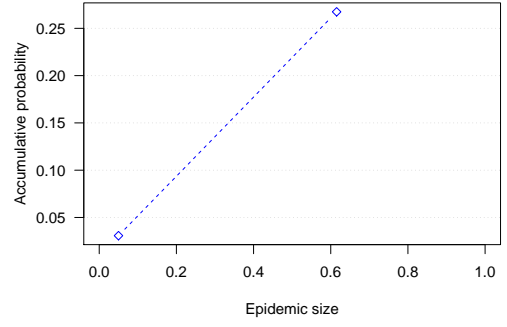


(d) Season 10

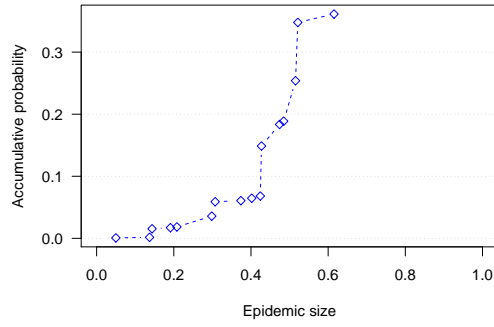
Figure 5.15: Accumulative probability of having a epidemic size smaller than a certain epidemic size for a Poisson $\lambda = 10$ network. The full-mixture model was used. The blue circles are calculated data, the blue line is only for better visibility. The probability to have an epidemic size other than shown by a circle is 0.



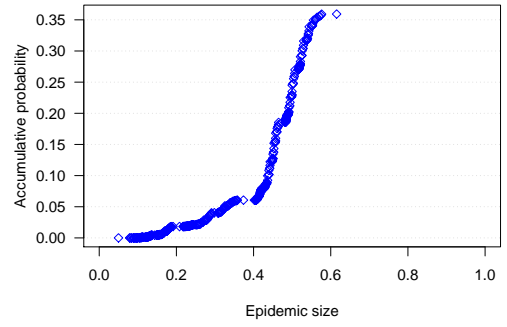
(a) Season 1



(b) Season 2

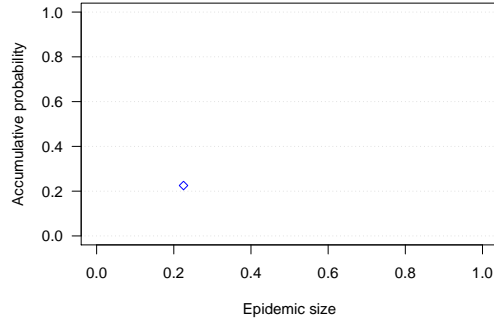


(c) Season 5

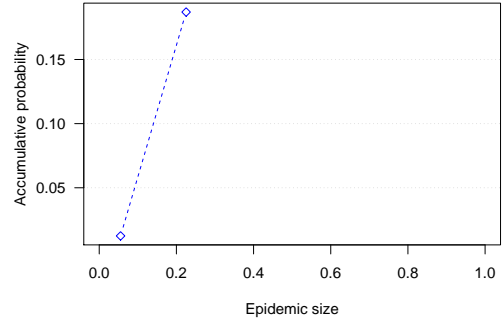


(d) Season 10

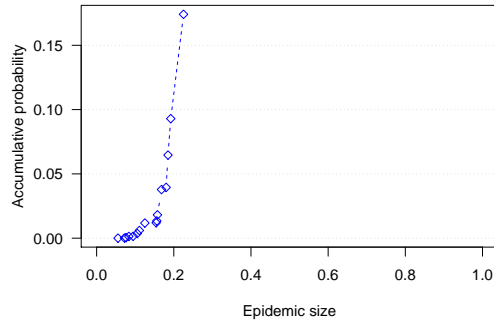
Figure 5.16: Accumulative probability of having a epidemic size smaller than a certain epidemic size for an exponential $\beta = \frac{1}{4}$ network. The half-mixture model was used. The blue circles are calculated data, the blue line is only for better visibility. The probability to have an epidemic size other than shown by a circle is 0.



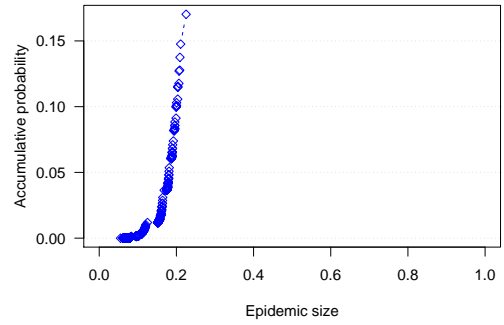
(a) Season 1



(b) Season 2



(c) Season 5



(d) Season 10

Figure 5.17: Accumulative probability of having a epidemic size smaller than a certain epidemic size for a scale-free $\alpha = 1.4$ network. The half-mixture model was used. The blue circles are calculated data, the blue line is only for better visibility. The probability to have an epidemic size other than shown by a circle is 0.

Season	Calc. half	Sim. non	Sim. half
1	0.893 (0.873 - 0.911)	0.904	0.880
2	0.137 (0.117 - 0.160)	0.155	0.139
3	0.585 (0.554 - 0.615)	0.593	0.573
4	0.483 (0.452 - 0.514)	0.465	0.468
5	0.523 (0.492 - 0.554)	0.528	0.530
6	0.510 (0.479 - 0.541)	0.518	0.517
7	0.513 (0.482 - 0.544)	0.524	0.517
8	0.510 (0.479 - 0.541)	0.523	0.523
9	0.511 (0.480 - 0.542)	0.529	0.500
10	0.511 (0.480 - 0.542)	0.522	0.517

Table 5.1: Table of probabilities of having an epidemic for each season. The probabilities are separated into simulated and calculated data for the non- and half-mixture models. The values in the brackett are the lower and upper limits of the 95% confidence interval for a proportion (Wilson, 1927). A Poisson $\lambda = 10$ network was used.

Season	Calc. full	Sim. full
1	0.893 (0.873 - 0.911)	0.890
2	0.236 (0.211 - 0.263)	0.234
3	0.597 (0.566 - 0.627)	0.609
4	0.511 (0.480 - 0.542)	0.528
5	0.536 (0.505 - 0.567)	0.533
6	0.528 (0.497 - 0.559)	0.536
7	0.531 (0.500 - 0.562)	0.533
8	0.530 (0.499 - 0.561)	0.522
9	0.530 (0.499 - 0.561)	0.555
10	0.531 (0.500 - 0.562)	0.526

Table 5.2: Table of probabilities of having an epidemic for each season. The probabilities are separated into simulated and calculated data for the full-mixture model. The values in the brackett are the lower and upper limits of the 95% confidence interval for a proportion (Wilson, 1927). A Poisson $\lambda = 10$ network was used.

Season	Calc. half	Sim. non	Sim. half
1	0.615 (0.585 - 0.645)	0.609	0.605
2	0.267 (0.241 - 0.295)	0.261	0.269
3	0.360 (0.331 - 0.390)	0.336	0.320
4	0.361 (0.332 - 0.391)	0.351	0.329
5	0.361 (0.332 - 0.391)	0.343	0.341
6	0.360 (0.331 - 0.390)	0.333	0.345
7	0.360 (0.331 - 0.390)	0.323	0.318
8	0.359 (0.330 - 0.389)	0.319	0.337
9	0.359 (0.330 - 0.389)	0.355	0.353
10	0.359 (0.330 - 0.389)	0.302	0.326

Table 5.3: Table of probabilities of having an epidemic for each season. The probabilities are separated into simulated and calculated data for the non- and half-mixture models. The values in the brackett are the lower and upper limits of the 95% confidence interval for a proportion (Wilson, 1927). An exponential $\beta = \frac{1}{4}$ network was used.

Season	Calc. full	Sim. full
1	0.615 (0.585 - 0.645)	0.633
2	0.429 (0.399 - 0.460)	0.443
3	0.421 (0.391 - 0.452)	0.433
4	0.415 (0.385 - 0.446)	0.405
5	0.415 (0.385 - 0.446)	0.400
6	0.413 (0.383 - 0.444)	0.414
7	0.412 (0.382 - 0.443)	0.394
8	0.411 (0.381 - 0.442)	0.407
9	0.411 (0.381 - 0.442)	0.408
10	0.411 (0.381 - 0.442)	0.400

Table 5.4: Table of probabilities of having an epidemic for each season. The probabilities are separated into simulated and calculated data for the full-mixture model. The values in the brackett are the lower and upper limits of the 95% confidence interval for a proportion (Wilson, 1927). An exponential $\beta = \frac{1}{4}$ network was used.

Season	Calc. half	Sim. non	Sim. half
1	0.225 (0.200 - 0.252)	0.225	0.215
2	0.187 (0.164 - 0.212)	0.193	0.201
3	0.179 (0.157 - 0.204)	0.180	0.201
4	0.176 (0.154 - 0.201)	0.177	0.176
5	0.174 (0.152 - 0.200)	0.171	0.185
6	0.173 (0.151 - 0.198)	0.174	0.180
7	0.172 (0.150 - 0.197)	0.174	0.169
8	0.171 (0.149 - 0.196)	0.171	0.184
9	0.171 (0.149 - 0.196)	0.172	0.171
10	0.170 (0.148 - 0.195)	0.172	0.172

Table 5.5: Table of probabilities of having an epidemic for each season. The probabilities are separated into simulated and calculated data for the non- and half-mixture models. The values in the brackett are the lower and upper limits of the 95% confidence interval for a proportion (Wilson, 1927). A scale-free $\alpha = 1.4$ network was used.

Season	Calc. full	Sim. full
1	0.225 (0.200 - 0.252)	0.217
2	0.215 (0.191 - 0.242)	0.206
3	0.210 (0.186 - 0.236)	0.196
4	0.207 (0.183 - 0.233)	0.214
5	0.205 (0.181 - 0.231)	0.203
6	0.203 (0.179 - 0.229)	0.200
7	0.202 (0.178 - 0.228)	0.180
8	0.201 (0.177 - 0.227)	0.193
9	0.200 (0.176 - 0.226)	0.180
10	0.199 (0.175 - 0.225)	0.209

Table 5.6: Table of probabilities of having an epidemic for each season. The probabilities are separated into simulated and calculated data for the full-mixture model. The values in the brackett are the lower and upper limits of the 95% confidence interval for a proportion (Wilson, 1927). A scale-free $\alpha = 1.4$ network was used.

and in tables 5.5 and 5.6 for scale-free networks, even if the difference is much smaller in season 2 compared with Poisson networks. In addition, the simulated probability of getting an epidemic for the non-mixture model in exponential networks is a few times slightly out of the confidence interval for the calculations of the half-mixture model. Probably, this is because of the only in the non-mixture model existing correlations.

In addition, there is almost no difference in the mixture models in Poisson networks, which is not so in exponential networks. One can see a clear difference in the mixture models. The separation is between the non- and half-mixture model on one side and the full-mixture model on the other side. The non- and half-mixture model have only a small difference in the epidemic size. But the full-mixture model has a difference of about 5% compared to the others. That means that an epidemic is 5% more likely in the full-mixture model than in the non- and half-mixture models. This is in agreement with the trend seen in Poisson networks.

As for Poisson and exponential networks before, the full-mixture model has a bigger probability of getting an epidemic in scale-free networks. Hence, this behavior is consistent throughout all network types, but the difference is not as big as in exponential networks. In addition, the probability of having an epidemic corresponds to the prevalence for the non- and half-mixture model, but not for the full-mixture model. The question remains open, if this is just an artifact of $\alpha = 1.4$ networks or if it is a general behavior of scale-free networks.

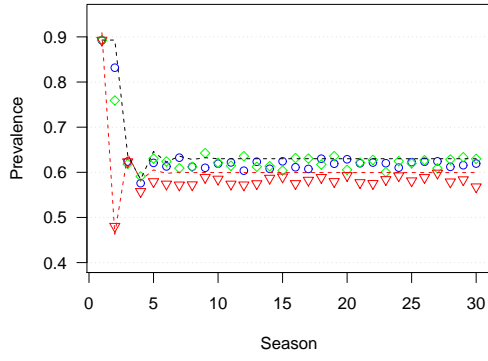
Nevertheless, one can see that the mixture has only a small impact on certain epidemiological parameters, which is surprising.

5.2 The Impact of Social Mixing on Multi-Season Dynamics

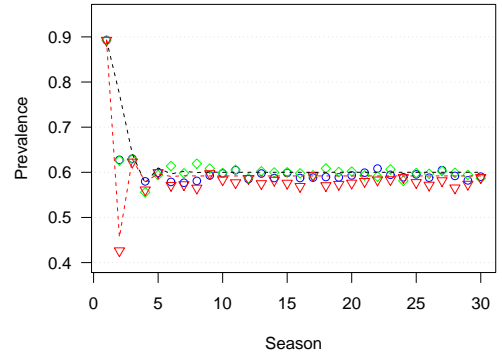
As it was shown before, the difference in the mixture models is very small for the probability of epidemic sizes for almost all networks. Therefore, it is consistent to look at the mean epidemic size for different mixture models. This was done in figures 5.18 - 5.20. The mean epidemic sizes follow the same behavior as the other network parameters. In all networks, the mixture model has only a small influence on the mean epidemic size. The difference in the prevalence stays nearly the same over all seasons. For Poisson networks, the difference gets smaller with the size of the mean degree λ . In addition, the full-mixture model shows the biggest peak at season 2.

This can also be seen in a pure difference plot. Figure 5.21 shows corresponding plots of the differences in the mixture models in a Poisson network. The difference is the biggest for $\lambda = 4$ and gets smaller for $\lambda = 8$ to its smallest values for $\lambda = 12$. This is also true for the difference at equilibrium as for the peak difference at season 2. In addition, the difference between the non- and half-mixture models is much smaller than the difference between both of them and the full-mixture model. Again, this finding is surprising. In fact, it is very unexpected that a static network shows a similar behavior as a completely flexible one. Furthermore, the mean epidemic size is almost the same. Combined with the findings in figure 5.1, it appears that the mean degree and the mixture model only have a minor impact on the mean epidemic size.

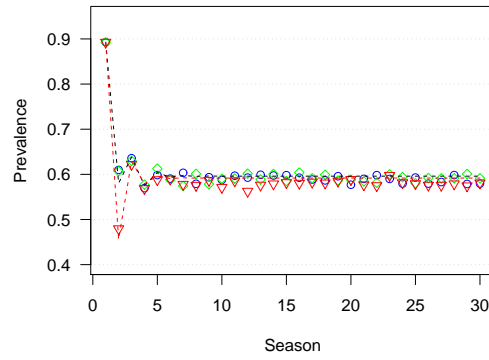
Furthermore, it is possible to compare the epidemiological risk in different mixture models. This is done in figure 5.22 for a Poisson network. The results are



(a) $\lambda = 4$



(b) $\lambda = 8$



(c) $\lambda = 12$

Figure 5.18: Prevalences of Poisson networks with different mixture models. The blue circles are simulated data for the non-mixture model, the green rhomboids are simulated data for the half-mixture model and the red triangles are simulated data for the full-mixture model. The black line is the calculated data for the non- and half-mixture model, the red line is for the full-mixture model.

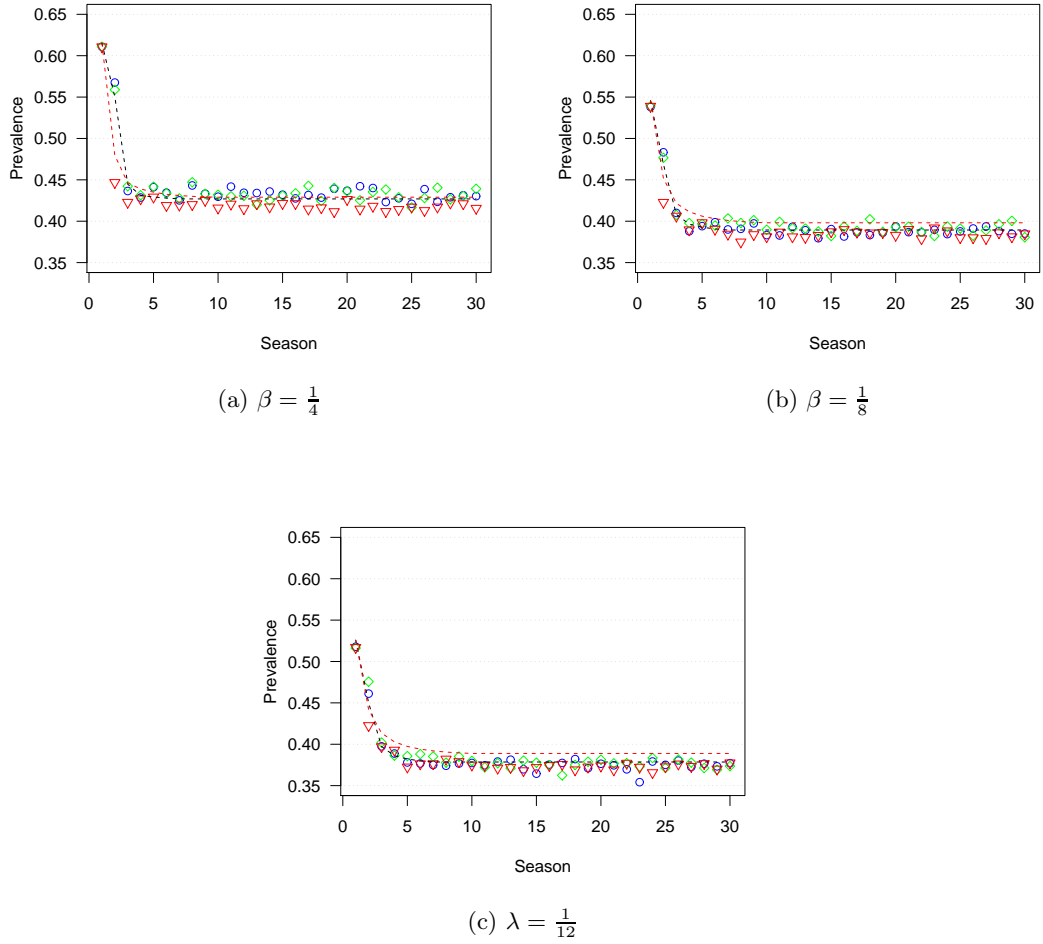
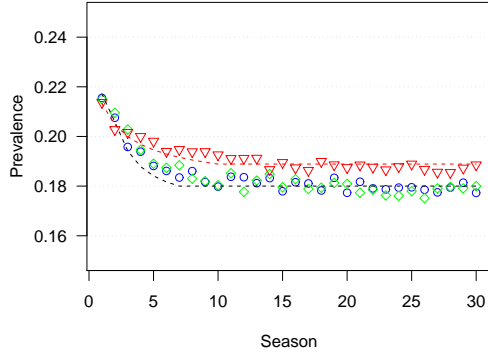
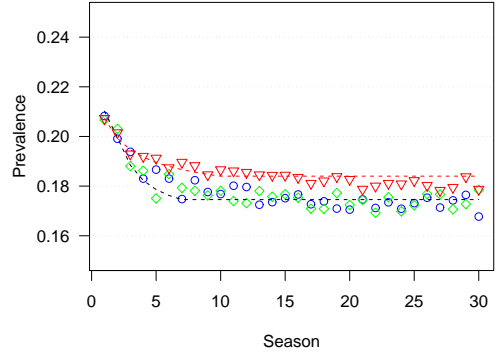


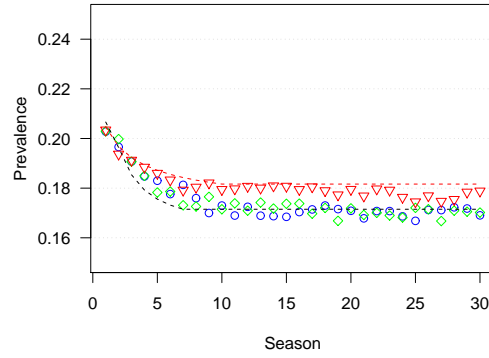
Figure 5.19: Prevalences of exponential networks with different mixture models. The blue circles are simulated data for the non-mixture model, the green rhomboids are simulated data for the half-mixture model and the red triangles are simulated data for the full-mixture model. The black line is the calculated data for the non- and half-mixture model, the red line is for the full-mixture model.



(a) $\alpha = 1.4$



(b) $\alpha = 1.5$



(c) $\alpha = 1.6$

Figure 5.20: Prevalences of scale-free networks with different mixture models. The blue circles are simulated data for the non-mixture model, the green rhomboids are simulated data for the half-mixture model and the red triangles are simulated data for the full-mixture model. The black line is the calculated data for the non- and half-mixture model, the red line is for the full-mixture model.

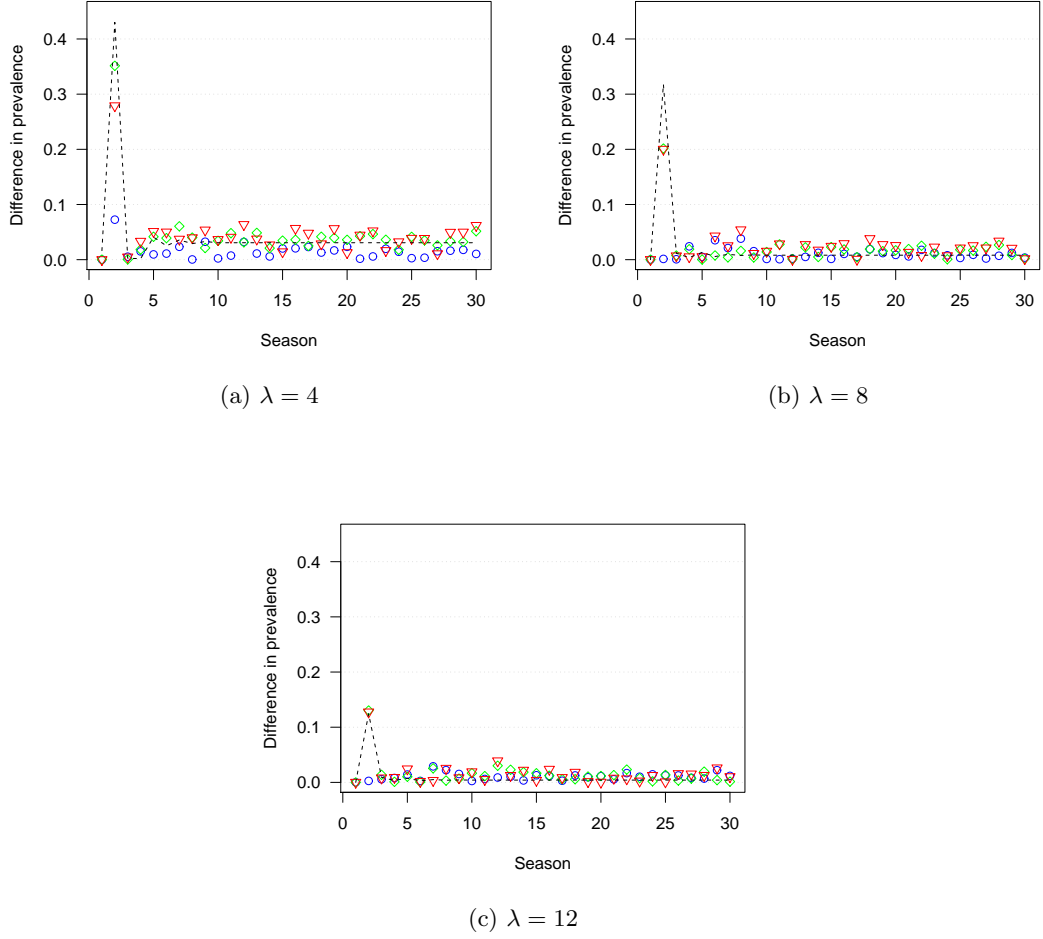
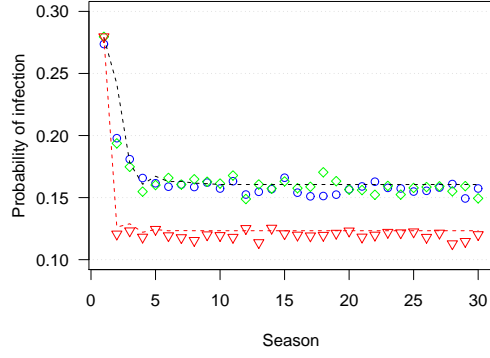
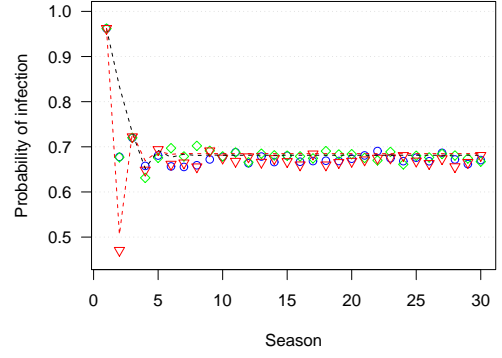


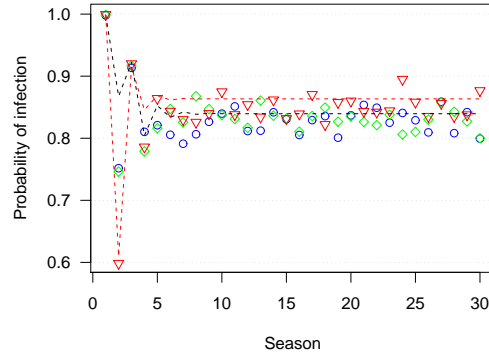
Figure 5.21: Differences in prevalence of Poisson networks with different mixture models. The blue circles are simulated data of the difference between the non- and the half-mixture model, the green rhomboids are simulated data of the difference between the non- and full mixture model and the red triangles are simulated data of the difference between the half- and full mixture model. The black line is the calculated data of the difference between the half- and full-mixture model.



(a) $k = 1$



(b) $k = 10$



(c) $k = 20$

Figure 5.22: Probability of being infected for different degrees in a Poisson $\lambda = 8$ network in different mixture models. The blue circles are simulated data for the non-mixture model, the green rhomboids are simulated data for the half-mixture model and the red triangles are simulated data for the full-mixture model. The black line is the calculated data for the non- and half-mixture model and the red line is for the full-mixture model.

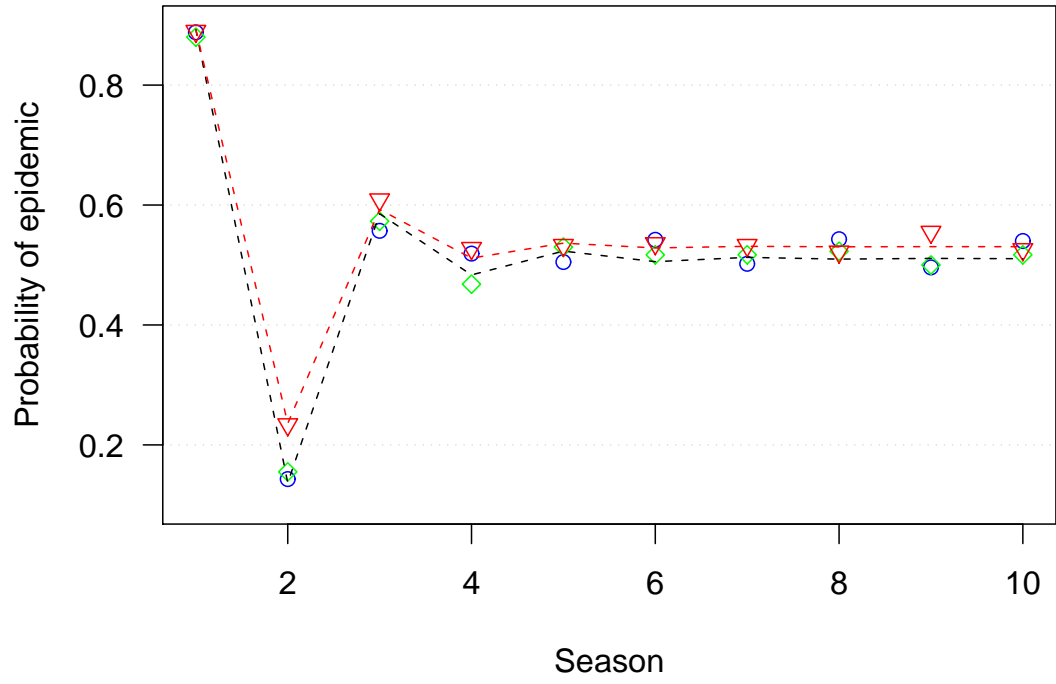


Figure 5.23: Probabilities of having an epidemic for each season in a Poisson $\lambda = 10$ network. The blue circles are for the simulated non-mixture model, the green rhomboids for the half-mixture model and the red triangles for the full-mixture model. The black line is the calculated probability for the non- and half-mixture model and the red line for the full-mixture model.

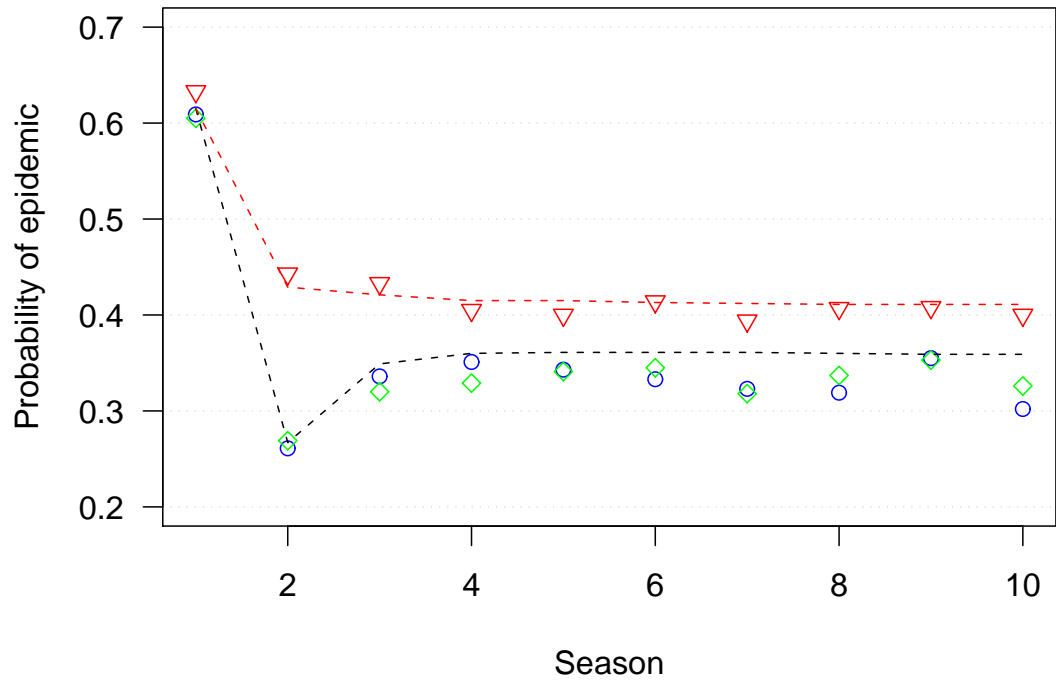


Figure 5.24: Probabilities of having an epidemic for each season in an exponential $\beta = \frac{1}{4}$ network. The blue circles are for the simulated non-mixture model, the green rhomboids for the half-mixture model and the red triangles for the full-mixture model. The black line is the calculated probability for the non- and half-mixture model and the red line for the full-mixture model.

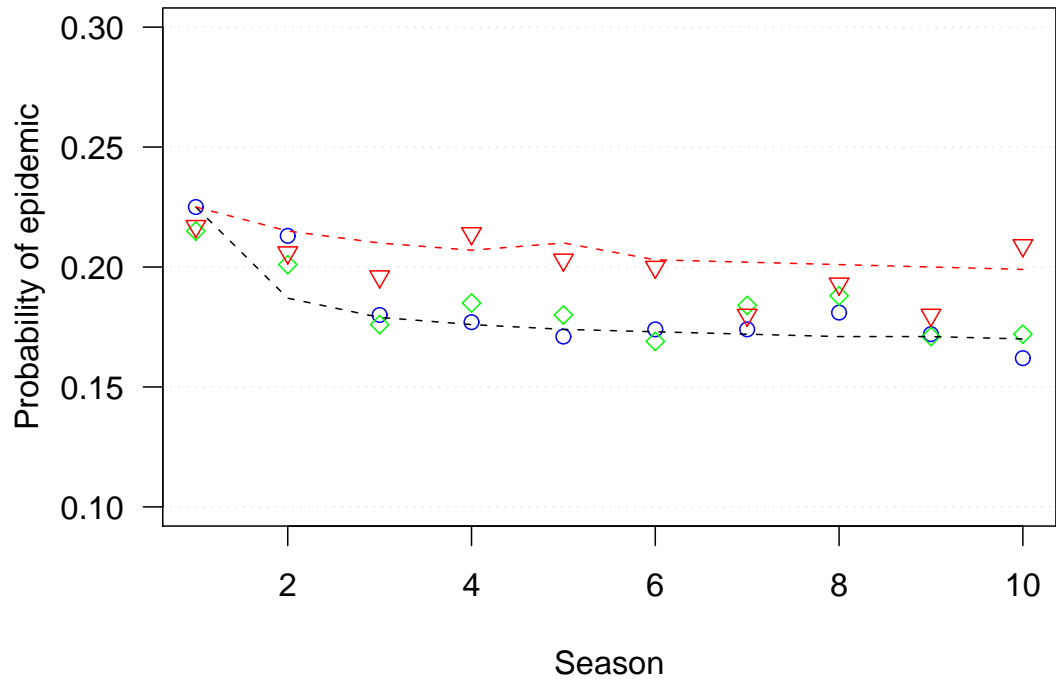


Figure 5.25: Probabilities of having an epidemic for each season in a scale-free $\alpha = 1.4$ network. The blue circles are for the simulated non-mixture model, the green rhomboids for the half-mixture model and the red triangles for the full-mixture model. The black line is the calculated probability for the non- and half-mixture model and the red line for the full-mixture model.

in good agreement with the findings for the mean epidemic size. Regardless of the mixing model, the behavior of the network is determined by its high-degree vertices. This is true for all three analyzed network types.

Figures 5.23, 5.24 and 5.25 show the probabilities for epidemics in different mixture models. Again, one can see that the mixture models only have a minor influence. In addition, the biggest difference is between the non- and half-mixture models and the full-mixture model. This is consistent for all network types.

5.3 The Impact of Network Structure on Multi-Seasonal Dynamics

As shown before, different network types have different mean epidemic sizes for the same $R_0 = 2.5$ value in general. But it is still a question how a Poisson, exponential and scale-free network differ from each other. To determine this the approx. same mean degree of $\langle k \rangle = 10$ for all three network types is taken. This leads to a Poisson network with $\lambda = 10$, an exponential network with $\beta = 0.105$ and a scale-free network with $\alpha = 0.9$ and $\kappa = 30$.

The mean epidemic size is plotted in figure 5.26, which shows that even the same mean degree does not give the same epidemic sizes. In addition, one can see that the prevalence is highest for Poisson networks and lowest for scale-free networks at the same R_0 values and mean degrees. Figure 5.26 also shows that the impact of cross-immunity is highest for Poisson networks and lowest for scale-free networks. The question still exists, if this is due to the higher epidemic size or because of the network structure.

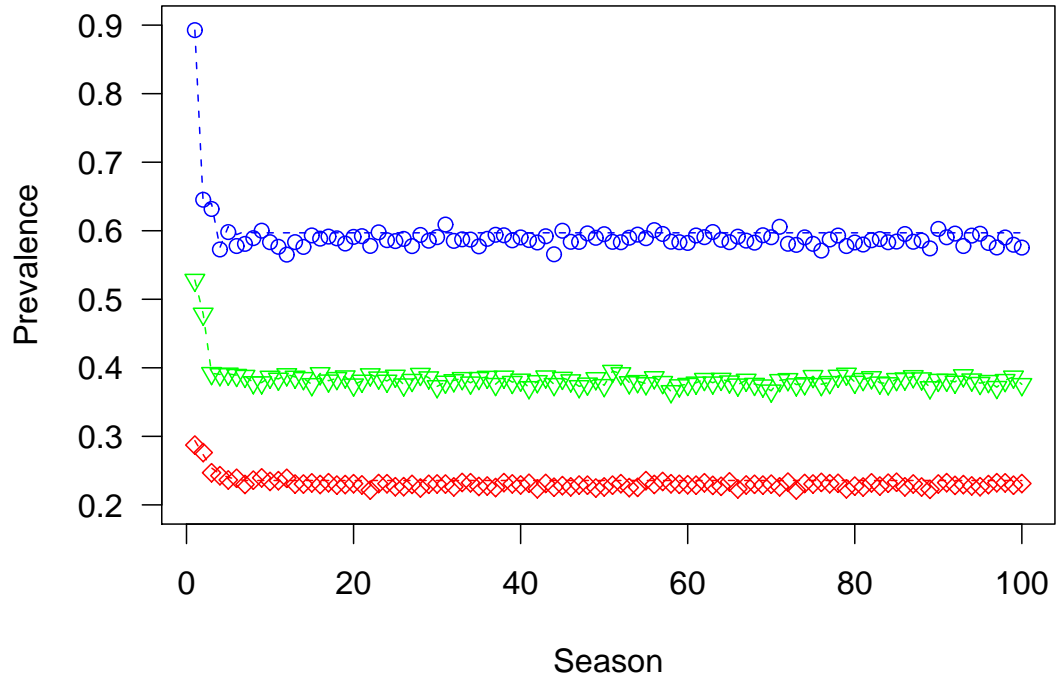
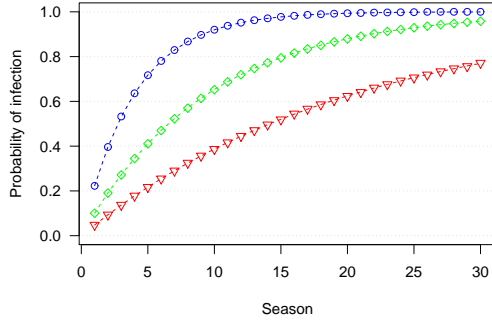
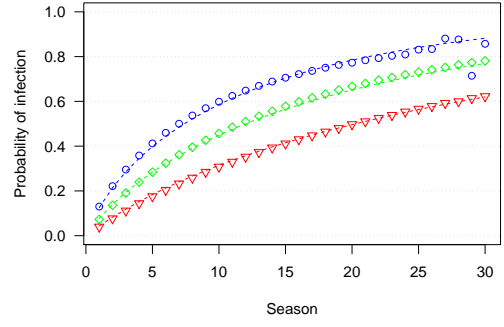


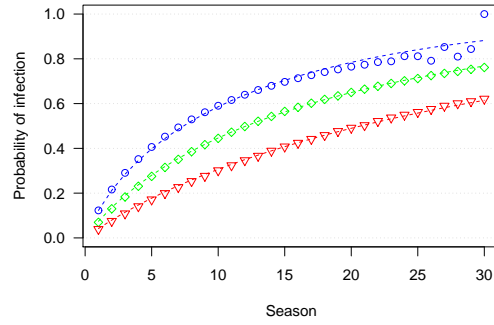
Figure 5.26: Probabilities of getting an epidemic for different network types. The non-mixture model was used. The blue circles are a simulated Poisson $\lambda = 10$ network, the green triangles are a simulated exponential $\beta = 0.105$ network and the red rhomboids are a simulated scale-free $\alpha = 0.9$ and $\kappa = 30$ network. The colored lines are the calculated values respectively.



(a) Season 1



(b) Season 10



(c) Season 100

Figure 5.27: Probabilities of getting infected for degree $k = 10$ vertices for different network types. The non-mixture model was used. The blue circles are a simulated Poisson $\lambda = 10$ network, the green triangles are a simulated exponential $\beta = 0.105$ network and the red rhomboids are a simulated scale-free $\alpha = 0.9$ and $\kappa = 30$ network. The colored lines are the calculated values respectively

5.4 Comparison of the Different Calculation Methods

As described in section 3.6 on page 36, two different ways of calculating the spread of epidemics are possible. In previous results, only one method is used. To see the difference in the accuracy of both, both methods compared to simulated data from a Poisson $\lambda = 8$ network in figure 5.28 have been plotted using the non-mixture model.

As said before, the calculate-everything-algorithm is much more accurate. Especially in season 2, this algorithm is much better than the faster one, which only calculates the mean network. Hence it is only the correct decision to take the mean-network-algorithm, if one does not have the time to use the better but slower one.

Another aspect of figure 5.26 is time. The mean epidemic size was plotted for 100 seasons. This shows that it was clearly sufficient in the results before to look only at the first 30 seasons, because the system is approx. at season 10 in its equilibrium state.

Furthermore, it is possible to determine the probability of being infected. This was done in figure 5.27. It shows that the system is in an equilibrium state at approx. season 10 as well. Also, one can see that the slope of the Poisson network becomes more similar to the other network types.

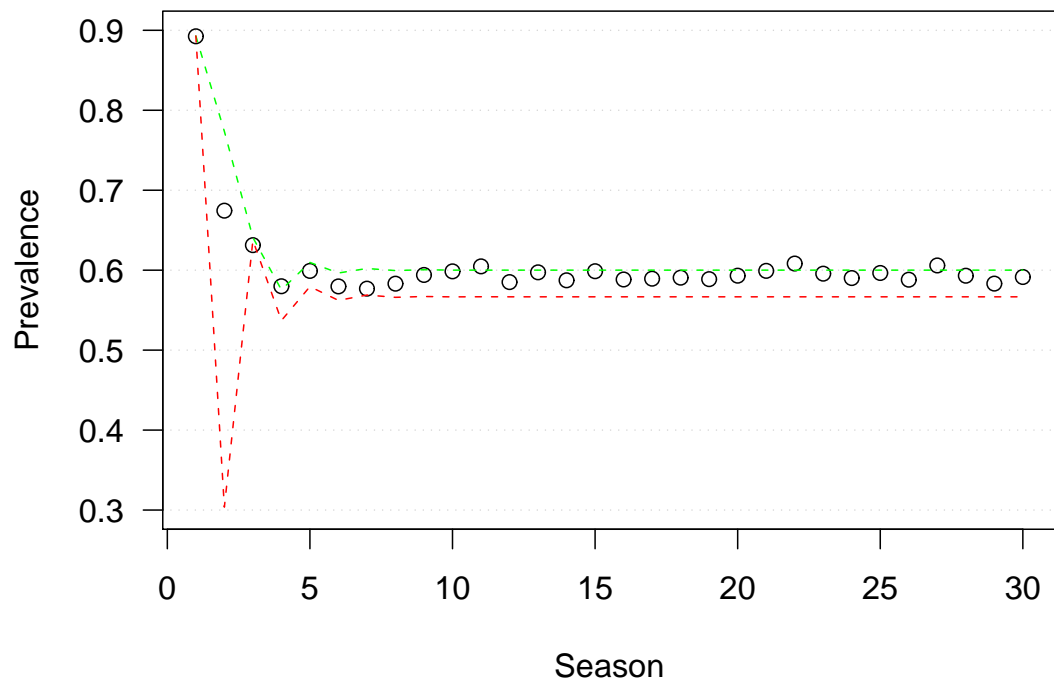


Figure 5.28: Prevalence for a Poisson $\lambda = 8$ network. The non-mixture model was used. The black circles are the simulated data. The green line is the accurate, calculate-everything-algorithm, the red line is the faster, mean-network-algorithm.

5.5 Implications of the Multi-Season Model

Outgoing from Newman's percolation model (Newman, 2002), a robust multi-season model for the spread of epidemic diseases could be defined. It was shown that a state model, based on a Markovian-process, is capable to describe the epidemic behavior over many seasons. In addition, this model is capable to use any possible experimental validated or just user-defined function of cross-immunity. This is especially interesting in predicting future disease dynamics.

The most impressive finding is the missing impact of social mixture dynamics. The between season mixture has almost no influence on the behavior of the network. Of course, further tests with other network types and different cross-immunity functions are necessary. But if this result is valid, the impact on epidemic modeling should be big. First of all, it is not only totally against one's first intuition. It also has big issues on the modeling of networks, and it is not necessary to try to model the fluctuations of networks. Once somebody has a good network to describe the real-world, nothing more needs to be considered in a model over multiple seasons - of course with the restriction that the underlying degree distribution does not change. In addition, this allows less complex models and hence it is faster to compute.

Second, the prediction of an equilibrium state is true for every network type and every mixture model, and only the relaxation time is changes between the systems. In the future, it would be interesting to compare this with cross-immunity SIR models or other differential equations that have the capability to show deterministic chaos.

Another finding is that the epidemic behavior is almost only determined by high-degree vertices. This result has relevance for vaccination strategies. An example of this would be the 1918 flu pandemic (Mills et al., 2004). Like in 1918, it is possible that a fraction of the network has been exposed to a similar strain before. Therefore, these individuals have a strong cross-immunity against a new pandemic strain, and only a small fraction of people do not. Even if the fraction of immune people would be enough to give herd immunity in full contact models, it could be that that this unimmune fraction is also the one with the highest degree. Hence, this could lead to an epidemic similar to the flu of 1918, where a totally atypical epidemical behavior was shown: neither elder nor young children had been the main group of infected, but instead young persons in their best years (Mills et al., 2004). Probably this age group of people was also the one with the highest degrees. Hence, it would have been the most promising strategy, if one could have vaccinated these high-degree individuals (vertices). In addition, ideas of vaccination strategies according to degree-classes has been proposed before (Ferrari et al., 2006). The results of this thesis give another argument to this strategy.

Also one key finding is that the probability of an epidemic is not equal to its mean epidemic size, like it is in the epidemiological percolation model (Newman, 2002). The difference between both values is different for each network, but in general valid. Another task for the future would be to control this finding in the real-world. Diseases such as flu have this seasonal behavior and should therefore show this prevalence-probability difference. In addition, this probability difference could be a test for the validity of this given model.

Bibliography

- B. Adams and A. Sasaki. Cross-immunity, invasion and coexistence of pathogen strains in epidemiological models with one-dimensional antigenic space. *Math. Biosciences*, 210(2), 2007.
- R. M. Anderson. Population biology of infectious diseases: Part i. *Nature*, 280: 361–367, 1978.
- N. T. J. Bailey. *The mathematical theory of infectious diseases and its applications*. Hafner Press, New York, 1975.
- A. Cayley. On the colourings of maps. *Proc. Royal Geographical Society*, 1:259–261, 1879.
- CDC and WHO. Smallpox: Disease, prevention, and intervention. Training course title, 2007.
- P. Erdős and A. R nyi. On random graphs. *Publicationes Mathematicae*, 6, 1959.
- L. P. Euler. Solv tio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741.
- M. J. Ferrari, S. Bansal, and L. A. Meyers. Network frailty and the geometry of herd immunity. *Proc. R. Soc. B*, 273, 2006.
- T. J. Hagenaars, C. A. Donnelly, and N. M. Ferguson. Spatial heterogeneity and the persistence of infectious diseases. *J. Theor. Biol.*, 229, 2004.

- J. Honerkamp. *Statistical Physics*. Springer-Verlag, Heidelberg, 2002.
- A. L. Lloyd and R. M. May. Spatial heterogeneity in epidemic models. *J. Theor. Biol.*, 179, 1996.
- M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1), 1998.
- R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Phys. Rev. E*, 64(2), 2001.
- L. A. Meyers, B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham. Network theory and SARS: predicting outbreak diversity. *Jour. Theo. Bio.*, 232(1):71–81, 2005.
- C. E. Mills, J. M. Robins, and M. Lipsitch. Transmissibility of 1918 pandemic influenza. *Nature*, 432:904–906, 2004.
- M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1), 2002.
- M. E. J. Newman. Component sizes in networks with arbitrary degree distributions. *Phys. Rev. E*, 76(4), 2007.
- M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applicationa. *Phys. Rev. E*, 64(2), 2001.

- M. Nuno, C. Castillo-Chavez, Z. Feng, and M. Martcheva. *Lecture notes in mathematics: Mathematical models of influenza: The role of cross-immunity quarantine and age-structure*, pages 349 – 364. Springer, Berlin, 2008.
- R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86, 2001.
- R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65(3), 2002.
- P. E. Praham and N. M Ferguson. Space and contact networks: capturing the locality of disease transmission. *J. R. Soc. Interface*, 3, 2006.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, Amsterdam, 2001.
- J. Wallinga and P. Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.*, 160(6): 509–516, 2004.
- D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
- H. Wilf. *Generatingfunctionology*. Academic Press, London, 1994.

R. J. Williams and N. D. Martinez. Simple rules yield complex food webs. *Nature*, 404:180–183, 2000.

E. B. Wilson. Probable inference, the law of succession, and statistical inference. *JASA*, 22:209–212, 1927.

Wolfram Research, Inc. Mathematica Edition: Version 6.0, Champaign, Illinois, 2007.

Vita

Ruben was born on a rainy 21st of November 1984 in Bremen, Germany. After a happy childhood in Neerstedt, Germany, with two older siblings and a yellow sandpit, Ruben moved with his family to Wildeshausen, Germany. He grew up in a nice and friendly environment, only disturbed by something called primary school, also known as St.-Peter-Schule Wildeshausen, Germany. After finishing the fourth grade, he went on to Orientierungsstufe Wildeshausen, Germany, where his thirst of knowledge was stopped by ultra-motivated teachers. Luckily two years later, Ruben went to the local high school, whose name was Gymnasium Wildeshausen, Germany - which strangely enough has nothing to do with a sport gym. He is still known for his unbroken school-wide record in unfinished homeworks. Finishing the Gymnasium with his Abitur (high school diploma) in 2004, he decided to study physics at the Bayerische-Julius-Maximilians-Universität Würzburg, Germany. After one year of study, he believed that it was a good idea to study psychology, too. Maybe it wasn't, but he studied it for his second degree until he went to the University of Texas at Austin in August 2007 to study physics as a graduate student.

In addition, Ruben likes to write of himself in third person in *vitae*.

Permanent address: 2616 Salado
Apt. 212
Austin, TX, 78705

This thesis was typeset with \LaTeX 2 ϵ by the author.