A Stochastic Model for Panic Behaviour in Disease Dynamics

Diplomarbeit
vorgelegt von

Rafael Brune

aus Göttingen

angefertigt
im Institut für Nichtlineare Dynamik
der Georg-August-Universität zu Göttingen

2008

Contents

1 Introduction

During the last few years the media paid great attention to different diseases that pose a threat to the modern human society. The public interest increased and with it the fear of a world wide pandemic. Disease outbreaks like the one of severe acute respiratory syndrome (SARS) or of the avian influenza (H5N1), which may mutate into a strain capable of fast human-human transmission, helped to bring broad attention to the field of epidemological research. Of course this kind of research is not new. The first so-called compartmental model to describe the time evolution of an outbreak in a small closed system, e.g. a school, was proposed by Kermack and McKendrick in 1927 [Ker27]. For quite some time their models did not get the attention they deserved. It was in 1979 when Anderson and May [And79] published about population dynamics in infectious diseases and picked up Kermack and McKendrick's work. Those compartmental models can be used to describe different kinds of diseases. The SIS model, see 2.5.2 on page 15, describes an endemic disease like the cold, whereas the SIR model, see 2.5.3 on page 16, can be used for epidemics like measles, mumps and rubella. Quickly those models where extended to include spatiality. Small subpopulations interact and the process of geographic disease spreading could be described. Other extensions include different human contact patterns to account for the great variability in human behaviour. More importantly, nowadays the travelling behaviour of humans is drastically different to what was possible e.g. in the Middel Age. Complex traffic networks such as public transport systems, highways and flight networks play an important role and really decrease the effective time one needs to travel to distant locations. The work of Brockmann et al. from 2006 [Bro06], which got massive media attention, shows how important alone the aspect of travel behaviour is for understanding disease spreading. All this research is done in hope for creating better predictions of outbreaks like the one of SARS which has been studied for example by Hufnagel et al. [Huf04]. Predictions help to organize countermeasures like vaccination and containment. In 2005 Ferguson et al. [Fer05] published an article proposing strategies for containing an emerging influenza pandemic in Southeast Asia based on a spatially explicit description of the area, commuting of individuals and random local contacts. As interesting and good those complex simulations are, they assume and disregard many different things from which we do not know their effect on the spreading of diseases. One of those pieces that need more research are the effects that the change of the human behaviour has if a disease is present in the population. Normally the models assume that an individual always follows a defined behaviour and that it would be independent from the location. In this thesis I introduce two models where local differences play a role and which model panic-like behaviour of the individuals when a disease is present in their vicinity.

This thesis has three main parts: The first (Chapter 2) is an introduction to stochastic descriptions of chemical reaction systems and how they are used in disease modelling. These descriptions can also be found in the standard literature for reaction-diffusion models [And79, Gar96, Ker27, Mur02]. The second (Chapter 3) is about numeric algorithms for stochastic simulations and summarizes the work of Gillespie and Gibson [Gib00, Gil76, Gil77] on stochastic algorithms that are important for this thesis. The third part (Chapters 4, 5 and 6) presents the new models, their simulation results and a discussion. The appendix contains excerpts from my actual implementations of the algorithms discussed in Chapter 3.

2 Approaches to Infectious Diseases

During the creation of a predictive model for a given system there are two problems one has to solve. First, breaking it down into a set of fundamental processes that describes the system's behavior, and second, finding a mathematical way to generate predictions based on this description. This chapter covers the first problem: Finding an appropriate mathematical description for the system of interest. Since the underlying dynamics involved in infectious diseases are similar to those used to describe chemical systems, this will be an introduction into chemical reaction diffusion models. The ideas introduced for these models are then transferred to the area of modeling infectious diseases.

2.1 Chemical Reactions

Consider a test tube filled with a system of different molecules A and B. These two chemical substances can react in a way that one molecule of kind A combines with one molecule of type B into a third molecule of kind C. This simple reaction can be described mathematically like this:

$$A + B \xrightarrow{k} C$$

Of course one can construct arbitrary complex reactions with many different reactants and products in the same way. But it is far more interesting to break the system down into *elementary* reactions with not more than two reactants and one to three products.

If all these reactions occur on large timescales in comparison to the normal non-reactive collisions between the molecules in our test tube, one can assume that the solution is *well mixed*. This means that the system can simply be represented by the number of each type of involved molecules. So instead of trying to calculate all movement and reaction data for every single molecule we can focus on how the number of molecules of each kind changes with time. This *well mixed* property allows us to write down the probability that a certain reaction n will occur in the next small time step $\mathrm{d}t$,

$$p_n = a_n \,\mathrm{d}t + O(\mathrm{d}t)$$

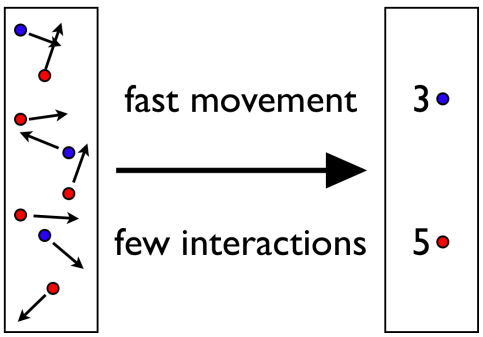

Figure 2.1: Ignoring the movement of individual molecules is valid, if all interactions happen on a large enough timescale. This allows a much simpler description of the system by variables which simply count the number of each molecule type.

where $O(\mathrm{d}t)$ are terms which are negligible for small $\mathrm{d}t$. The factors a_n are called the propensities of the particular reactions. These propensities[1] do not depend on $\mathrm{d}t$, but can be functions of variables such as volume, temperature, concentration and of course the reaction rate k. Their exact definition is not important at the moment, we will come back to them later.

We now extend our simple test system to involve a second reaction that reverses the first reaction.

$$A + B \xrightarrow{k_1} C$$

$$C \xrightarrow{k_2} A + B$$

Since we are only interested in the number of each kind of molecule we can simply describe the system's configuration with a simple vector S

$$S = (\#A,\, \#B,\, \#C)$$

Whenever one of the involved reactions takes place this status S instantly changes to a new status S'. For example the first reaction decreases the number of A and B molecules by one and increases the count for C

$$S' = (\#A - 1,\, \#B - 1,\, \#C + 1)$$

The probability of this transition can be written as

$$P(S', t + \mathrm{d}t \,|\, S, t) = a_1\, \mathrm{d}t + O(\mathrm{d}t) \tag{2.1}$$

As we can see here, the probability only depends on the actual state of the system and not on any of the past states. This means, this is a *Markov process*.

One approach to work with such stochastic systems is to define a probability variable $P(S)$ for every possible state S. With help of equation (2.1) one can write down a system of coupled differential equations, which describe the system with these probability variables. This set of equations is called the *Master Equation*.

$$\frac{\mathrm{d}P(S)}{\mathrm{d}t} = \sum_n \left(a'_n\, P(S'_n) - a_n\, P(S)\right) \tag{2.2}$$

The Master Equation describes the time evolution of the probabilities that the system is in a given state. Therefore equation (2.2) calculates the in- and outflux of probability from all states the system can come from and move into in one reaction step. Of course the amount of probability flowing from one state to another depends on the propensity of the respective reaction.

Figure 2.2 visualizes the flow of probabilities in a simple system consisting of these three reactions:

$$\begin{aligned} A + B &\xrightarrow{k_1} 2B \\ B &\xrightarrow{k_2} C \\ C &\xrightarrow{k_3} A \end{aligned} \tag{2.3}$$

To write down the master equation for this system in detail, the different transitions are added or substracted and weighted with their respective propensities.

[1] Propensities should never be mixed up with probabilities. As we learn later, propensities can be used to calculate actual probabilities, but they are something more raw. More information on this topic can be found in publications by Karl Popper[Pop82].

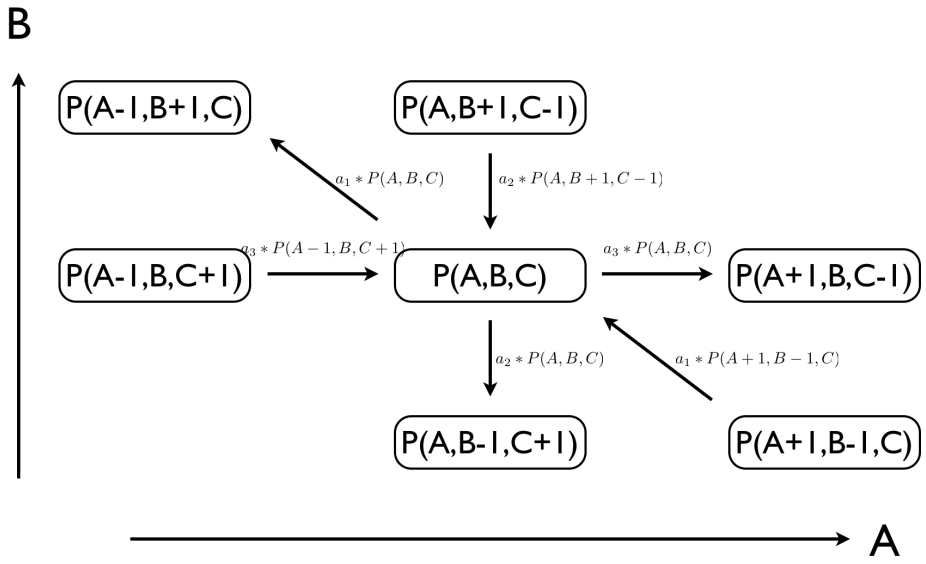

Figure 2.2: If the system has the configuration (A, B, C) there are three different states the system can evolve into, if the respective reaction occurs and also three different states where the actual state could have come from. The master equation now simply aggregates the different fluxes to describe the evolution of the probability to be in state (A, B, C) at time t.

$$\begin{aligned}\frac{\mathrm{d}P(A,B,C)}{\mathrm{d}t} &= -a_{1,(A,B,C)}\,P(A,B,C) + a_{1,(A+1,B-1,C)}\,P(A+1,B-1,C) \\ &\quad -a_{2,(A,B,C)}\,P(A,B,C) + a_{2,(A,B+1,C-1)}\,P(A,B+1,C-1) \\ &\quad -a_{3,(A,B,C)}\,P(A,B,C) + a_{3,(A-1,B,C-1)}\,P(A-1,B,C-1)\end{aligned}$$

Solving and integrating this equations can be fast and easy if the system has a reasonably small number of different possible states.

For more complex systems with large numbers of molecules one will, instead of calculating the probability for every state, switch to exact stochastic simulations to generate single trajectories of the system moving through the different system configurations. Different methods to do so are described in Chapter 3.

2.2 Meanfield Approximation

The evolution of the probabilities over time, which can be calculated from the master equation, are quite nice. With the help of some sums we can now have a look at the time evolution of mean values for our example system:

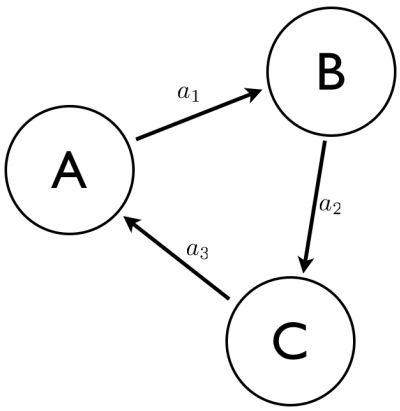

Figure 2.3: The example system has three different types of molecules and a reaction system in form of a loop. The flow from one molecule type to the next is described by the propensities of their reactions.

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle A\rangle = &\ \textstyle\sum_A \sum_B \sum_C & \frac{\mathrm{d}}{\mathrm{d}t} P(A,B,C)A \\
= &\ \textstyle\sum_A \sum_B \sum_C & \big(-a_{1,(A,B,C)}\, P(A,B,C)\, A + a_{1,(A+1,B-1,C)}\, P(A+1,B-1,C)\, A \\
& & -a_{2,(A,B,C)}\, P(A,B,C)\, A + a_{2,(A,B+1,C-1)}\, P(A,B+1,C-1)\, A \\
& & -a_{3,(A,B,C)}\, P(A,B,C)\, A + a_{3,(A-1,B,C-1)}\, P(A-1,B,C-1)\, A\big) \\
= &\ \textstyle\sum_A \sum_B \sum_C & \big(-a_{1,(A,B,C)}\, P(A,B,C) + a_{3,(A,B,C)}\, P(A,B,C)\big) \\
= & & \langle -a_1\rangle + \langle a_3\rangle
\end{aligned}
$$

The crucial and tricky part in this transformation is in between line two and three. After having spent some time looking at the terms most of them will simply cancel out.[2] The terms that stay are those that have a contribution to $\langle A\rangle$, which are the in- and outfluxes added over all different As and therefore the mean value of the propensities. This description is quite nice and somewhat intuitive but, most important, it allows to easy *see* the basic dynamics going on in the system. For our example system the meanfield differential equations[3] for all three variables are the following:

$$
\begin{aligned}
\mathrm{d}_t\langle A\rangle &= \langle -a_1\rangle + \langle a_3\rangle \\
\mathrm{d}_t\langle B\rangle &= \langle a_1\rangle + \langle -a_2\rangle \\
\mathrm{d}_t\langle C\rangle &= \langle a_2\rangle + \langle -a_3\rangle
\end{aligned}
$$

Figure 2.3 illustrates the meanfield equations. Since these meanfield equations are easy on the eyes they will be used quite often in this thesis.

Now, one may ask why this chapter is called meanfield *approximation,* if the calculations of the mean values did not include any approximations. The approximation comes into play when we take a closer look at the propensities. As already mentioned in the previous section, propensities describe the tendency of something to happen. In chemical reactions this tendency can be influenced by many different factors. For simplification and because we do not need them for later models we simply ignore factors like temperature and other physical properties of the

[2] A more explained calculation can be found in the thesis of Christian Thiemann [Thi08] or in the standard literature regarding master equations [Gar96].

[3] For simplification $\mathrm{d}/\mathrm{d}t$ is replaced by d_t from now on.

molecules. We see them moving around chaotically in our test tube and colliding with each other once in a while. So, how can we define those propensities? Let's have a look at the first reaction from our sample system:

$$A + B \xrightarrow{k_1} 2B$$

The first and obvious thing that comes to the interested reader's mind is, that the propensity a_1 for this reaction should be proportional to the reaction rate k_1:

$$a_1 \propto k_1$$

Looking at the reactants tells us that an increased number of A or B molecules should also increase the propensity. Since I prefer A molecules over type B molecules they will be in focus of attention first:

$$a_1 \propto k_1 \times A$$

Now each of these A molecules wants the chance to meet a B molecule. The probability to meet anyone of the B molecules out of the overall number of molecules in our system, N, naturally equals B/N:

$$a_1 \propto k_1 \times A \times \frac{B}{N}$$

One last thing we want to account for in this example is the volume Ω of the test tube. If the molecule density in the system decreases, for example due to a reaction that unites two molecules, the probability to meeting other molecules should also decrease. Since the density is simply the total number of molecules divided by the volume we finally come to this equation:

$$\begin{aligned} a_1 &= k_1 \times A \times \frac{B}{N} \times \frac{N}{\Omega} \\ &= k_1 \times A \times \frac{B}{\Omega} \end{aligned}$$

With the same procedure the propensities for reaction two and three are calculated:

$$\begin{aligned} a_2 &= k_2 \times B \\ a_3 &= k_3 \times C \end{aligned}$$

Since these two reactions resemble some kind of decay from B to C and from C to A they do not depend on meeting with another molecule or on the molecule density and have a rather simple form.

Now we can take a closer look at the mean field equations of our example system:

$$\begin{aligned} \mathrm{d}_t \langle A \rangle &= \left\langle -k_1 \times A \times \frac{B}{\Omega} \right\rangle + \langle k_3 \times C \rangle \\ &= -\frac{k_1}{\Omega} \langle AB \rangle + k_3 \langle C \rangle \\ \\ \mathrm{d}_t \langle B \rangle &= \left\langle k_1 \times A \times \frac{B}{\Omega} \right\rangle + \langle -k_2 \times B \rangle \\ &= \frac{k_1}{\Omega} \langle AB \rangle - k_2 \langle B \rangle \\ \\ \mathrm{d}_t \langle C \rangle &= \langle k_2 \times B \rangle + \langle -k_3 \times C \rangle \\ &= k_2 \langle B \rangle - k_3 \langle C \rangle \end{aligned}$$

It can be seen here, that inserting the propensity definition into the mean value description gives us the term $\langle AB \rangle$. A description that only uses the variables $\langle A \rangle$, $\langle B \rangle$ or $\langle C \rangle$ is much more interesting, since we only have access to these statistical features. Therefore we need to make a little assumption:

$$\langle AB \rangle \approx \langle A \rangle \langle B \rangle$$

This would be exact, if A and B were independent variables, but they are not. If A and B are large, the differences between these two terms become neglibible. Since I do not want to discuss meanfield calculations[4] in this thesis, we should just simply accept this approximation and go on. Therefore our example model ends up with this meanfield description:

$$\begin{aligned} \mathrm{d}_t \langle A \rangle &= -k_1 \frac{\langle A \rangle \langle B \rangle}{\Omega} + k_3 \langle C \rangle \\ \mathrm{d}_t \langle B \rangle &= k_1 \frac{\langle A \rangle \langle B \rangle}{\Omega} - k_2 \langle B \rangle \\ \mathrm{d}_t \langle C \rangle &= k_2 \langle B \rangle - k_3 \langle C \rangle \end{aligned}$$

Throughout the different chapters the differential meanfield equations will be used to describe the models discussed. They are quite handy, because one can see the dynamics going on and, at the same time, how the propensities are defined.

2.3 Reaction Diffusion Models

The single assumption of a *well mixed* property we made in the last section is also a major drawback. It is crucial for the behaviour of many interesting systems that they extend in different directions. This means that we must *not ignore diffusion* processes. For example if the test tube is so large, that the time one molecule needs to cover the distance by diffusion is greater than the time between two reaction events, diffusion cannot be neglected. Even further, in biological systems the reaction medium cannot be seen as homogeneous. The geometry of the cell needs to be accounted for. This means that we need to somehow model the involved diffusion processes.

Since we want to stay with in our approach to stochastic processes we need to somehow conserve the assumption of a well mixed media. This can be done by simply dividing the whole test volume into N smaller *subvolumes* in which the timescales are small enough so that the reaction processes are slower than the molecule's time to cross a subvolume by diffusion. The average distance a particle covers by diffusion depends on the time t and the local diffusion rate D:

$$\langle L^2 \rangle \propto Dt$$

This means, that by choosing the subvolumes as cubes of volume V, the following needs to be fulfilled.

$$\frac{V^{\frac{2}{3}}}{D} \ll \langle t \rangle$$

Where $\langle t \rangle$ denotes the average time between two reaction processes.

The system can now be described by a more complex vector, which stores the molecule counts for every subvolume.

$$S = (\#A_1, \ldots, \#A_N, \#B_1, \ldots, \#B_N, \#C_1, \ldots, \#C_N)$$

[4] See diploma thesis of Christian Thiemann instead - well not instead, better read both.

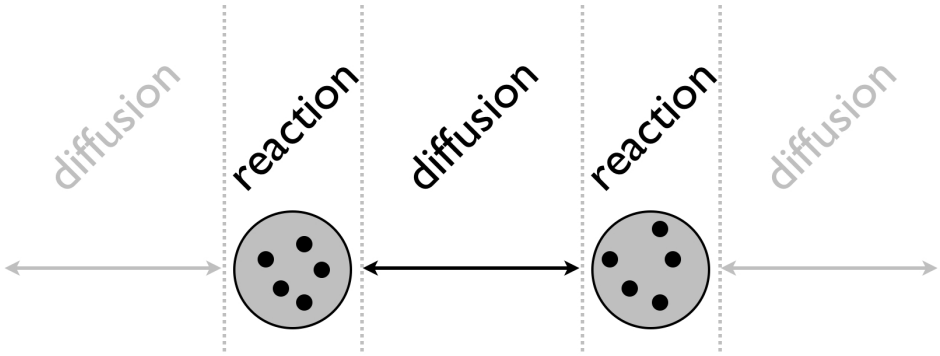

Figure 2.4: Reaction diffusion models allow to split up spatially extended systems into smaller subvolumes in which the *well mixed* assumption still holds.

This also means that reactions in one subvolume only affect the molecule counts of this one subvolume. But since we want to include some kind of diffusion process we need to add reactions that work across subvolumes. For a reaction that describes molecules moving from subvolume n to m one can write

$$\begin{aligned} A_n &\xrightarrow{d_i} A_m \\ B_n &\xrightarrow{d_i} B_m \\ C_n &\xrightarrow{d_i} C_m \end{aligned}$$

The index i is introduced to allow local different diffusion strengths for every subvolume. If we choose to have cubic subvolumes in a three dimensional test tube we need to add such diffusion processes to every side of each subvolume to correctly couple it with the neighbouring subvolumes.

The reactions inside each subvolume together with these diffusion reactions describe what is called a *reaction diffusion model*.

There are many different research fields in which reaction diffusion models are used to describe and later simulate complex systems. Chemistry and Biophysics for example rely heavily on such models. The advantages are that you can model complex geometry of spatially inhomogeneous systems; for example, biological cells have large amounts of particles involved in the reactions. Trying to solve the master equation for such a system would be not feasable in an acceptable amount of time.

As described in the previous section, basically we have small volumes which can be seen as well mixed and this is where the actual reactions are happening. To these uncoupled subvolumes we add reactions that allow particles to move from one subvolume to the next. In this way we can create arbitrarily complex systems the way we want them to be. A simple system that only extends in one direction could be sketched like this:

$$V_1 \overset{d}{\rightleftarrows} V_2 \overset{d}{\rightleftarrows} [\ldots] \overset{d}{\rightleftarrows} V_N$$

where every V_n is a subvolume. With such a *linear chain* one could simulate how fast a chemical reaction spreads spatially along the chain. Of course the observed velocity strongly correlates with the coupling strength d between the subvolumes. If we choose this coupling strength small enough we end up with only loosely coupled subvolumes. This contradicts our idea of choosing small enough subvolumes but opens up other interesting usages. On one hand you have many subvolumes with small numbers of particles and a high coupling. On the other hand we have smaller numbers of subvolumes with higher amounts of particles and a small coupling strength.

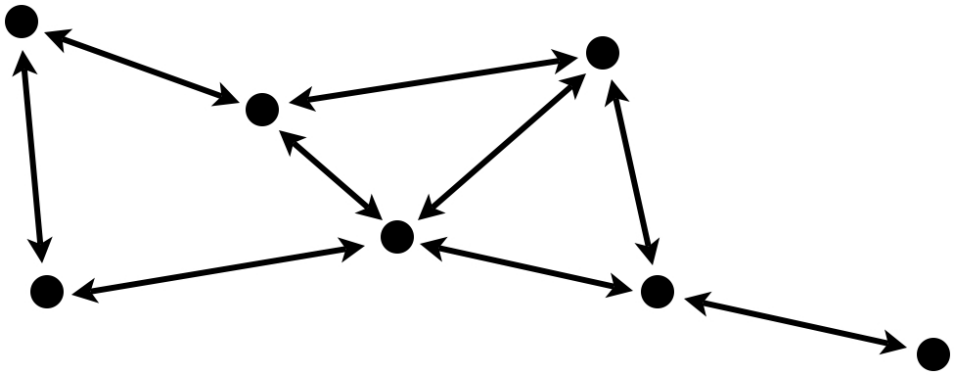

Figure 2.5: A simple network with nodes and edges. Each node represents a site where reactions take place. The edges show the coupling.

The first system is dominated by events of particles switching subvolumes, the latter by reactions inside each subvolume. In reality this system could be constructed as small well mixed testing tubes that are connected through a small hole in their walls.

2.4 Graphs and Networks

For spatial disease simulations these kind of stochastic models are used in the way that every subvolume is, for instance, a whole city, district or even country. The connections between these subvolumes are then seen as coupling strengths between two geographical areas. Mathematically this is nothing else but a weighted graph. The nodes in this graph represent reaction subvolumes and the edges with their respective weights represent how strong the corresponding nodes interact with each other. This effectively splits the description of this stochastic model into two parts. The first part is the description of the graph's structure. Often an adjacency matrix A_{ij} is used to mathematically formalise the connection between the nodes i and j.

This $n \times n$ matrix stores all combinations of the n nodes if they are connected $A_{ij} = 1$, if not $A_{ij} = 0$. The matrix entries can not only be used to store the adjacency, but also the coupling strength between two nodes i and j,

$$A_{ij} = \omega_{ij}$$

where ω_{ij} equals zero, if there is no connection from node j to node i. There should be no self loops in the graph, therefore all $A_{ii} = 0$. If the graph is undirected, this matrix is also symmetric.

2.5 Modelling Disease Dynamics

Now equipped with these mathematical fundament we can start thinking about modelling human diseases. We begin with looking at the basic processes involved.

If a single person gets infected and we observe the person in a small period of time, there are a few things that we could see. For example, the person can become healthy again. Or it simply stays ill for the time of observation. On the other hand, the person can also meet someone and, with a hopefully small probability, infect the other person.

In our description of molecules reacting with each other we now simply classify all involved individuals into *infected* (I) and *susceptible* (S) species. Our observation gives us this mathematical description:

$$S + I \xrightarrow{\alpha} 2I \tag{2.4}$$

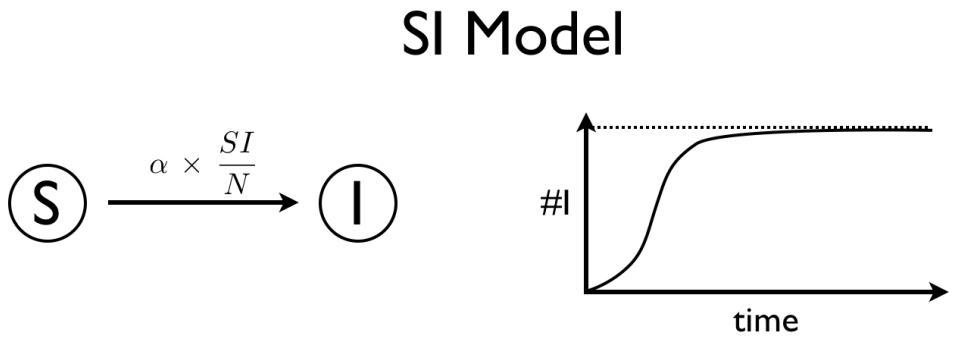

Figure 2.6: On the left is a small graph describing the flow of individuals in the SI model and on the right is a typical time evolution. The small dotted line equals the state where all individuals are infected $I = N$. After some time this state is reached inevitable.

This describes the process where an infected somehow infects a susceptible and makes it an infected individual. The rate α describes how effective an infected person is at infecting other individuals. The other described reaction was how an infected becomes healthy again. There are different models for different diseases that describe different diseases in which for example one infected individual becomes *resistant* (R) or one model where they become susceptible again. Of course these different models give different dynamics. One important thing to model here are the definitions for the reaction propensities. The idea is to strictly do the same as for the chemical reactions in Section 2.1. The test tube is the city we are looking at, the molecules are our individuals and the volume of the test tube is the size of our city, here set to the total number of individuals. Therefore we end up with this propensity a for reaction 2.4:

$$a = \alpha \times \frac{SI}{N}$$

2.5.1 SI Dynamics

The most simple model is the one which is only based on reaction (2.5). This model is called the SI model.

$$S + I \xrightarrow{\alpha} 2I \tag{2.5}$$

After some time everybody is infected and no susceptibles are left.

The meanfield equations for the SI model are defined as

$$\begin{aligned} \mathrm{d}_t S &= -\alpha \times S \times \frac{I}{N} \\ \mathrm{d}_t I &= +\alpha \times S \times \frac{I}{N} \end{aligned}$$

In this dynamics two stable points exist of which one is a metastable point. The metastable point exists where $S = N$ and $I = 0$. As soon as there is at least one infected individual in the population, the system will end up in the stable point at $S = 0$ and $I = N$.

2.5.2 SIS Dynamics

The SIS dynamics are a logical extension to the SI dynamics. Additionally to the infection reaction (2.5) that can occur in the SI dynamics, a second reaction (2.6) is added, which describes

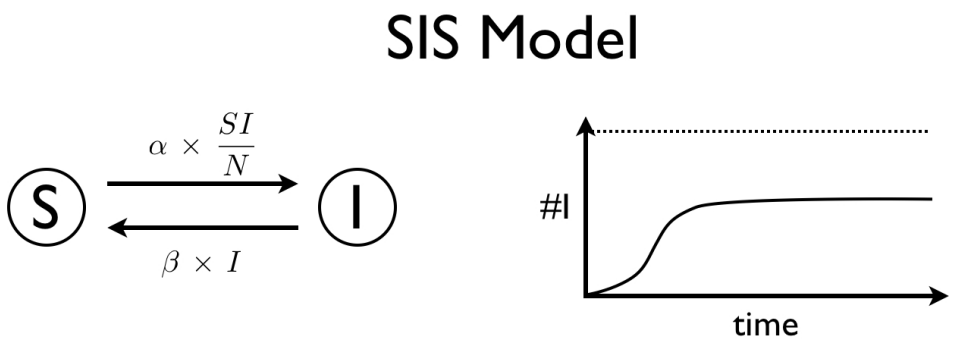

Figure 2.7: In the SIS model there is a flow of individuals back into class S. Therefore an equilibrial state exists, where in and outflux from I are equal.

the process that an individual becomes healthy again after it has been infected with the disease.

$$\begin{aligned} S + I &\xrightarrow{\alpha} 2I \\ I &\xrightarrow{\beta} S \end{aligned} \tag{2.6}$$

This reaction effectively creates a cycle which the individuals are going through. A susceptible individual can get infected by a contact with an infected individual and also becomes an infecteted individual. After some time, and maybe after the infected individual also infected other susceptible individuals, it will become healthy again. The cycle is closed.

For the SIS model the mean field equations are these:

$$\begin{aligned} \mathrm{d}_t S &= -\alpha \times S \times \frac{I}{N} + \beta \times I \\ \mathrm{d}_t I &= +\alpha \times S \times \frac{I}{N} - \beta \times I \end{aligned}$$

This model also exhibits stable points. Again a metastable point at $S = N$ and $I = 0$. The stable equilibrial point in this model lies at

$$\alpha \times S \times \frac{I}{N} - \beta \times I = 0 \quad \Leftrightarrow \quad S = \frac{\beta}{\alpha} \times N$$

which means that the infection level will stay around $I = N - {}^{\beta}/_{\alpha} \times N$. The SIS dynamics can therefore be seen as a model for an endemic disease that stays more or less at a constant level of infecteds in a population.

2.5.3 SIR Dynamics

The last model I want to discuss here is the SIR model. Here, a new kind of individuals is introduced. The R, which stands for removed or recovered, in the meaning that these individuals did recover from the disease and are immune or removed from the dynamic system because they did not survive the disease.

To include this new kind of individuals, the reaction (2.6) is changed such that, after a while, the infected individuals do not go back to the susceptibles S but to the new class R.

$$\begin{aligned} S + I &\xrightarrow{\alpha} 2I \\ I &\xrightarrow{\beta} R \end{aligned}$$

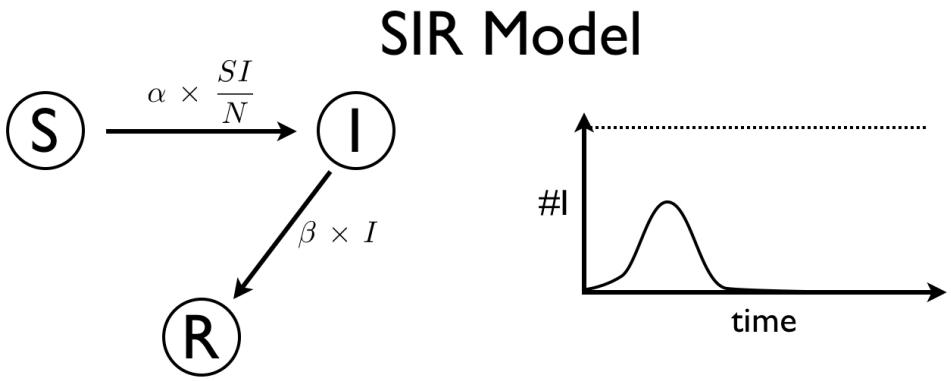

Figure 2.8: The SIR model adds a new kind of individuals: R. Instead of just becoming healthy again, individuals end up beeing resistant. In comparison to the SI and SIS model, we end up with a system, where no infected individuals are left. The disease died out.

This gives us the following meanfield equations for the three individual types S, I and R:

$$\begin{aligned} \mathrm{d}_t S &= -\alpha \times S \times \frac{I}{N} \\ \mathrm{d}_t I &= +\alpha \times S \times \frac{I}{N} - \beta \times I \\ \mathrm{d}_t R &= +\beta \times I \end{aligned}$$

Given a system where every individual is from the susceptible kind and there happens to appear one single infected individual which starts to spread the infection to other individuals, we can observe how the count of infected people raises and at the same time decreases the number of susceptibles. During all this, the number of infected individuals is constantly reduced and increases the count of recovered individuals. The total number of infecteds reaches its maximum when the supply of susceptibles is running low. After some time all infected individuals become recovered and the disease dies out.

Therefore we again have one stable point where $I = 0$ and $S = 0$ and one metastable point where $I = 0$ and $S > 0$.

3 Stochastic Simulation Methods

This chapter discusses different methods and algorithms that can be used to examine complex stochastic systems. It concentrates on exact trajectory generation and does not talk about deterministic meanfield approximations of the stochastic system. More insight in meanfield approximations and algorithms than the few hints given in Section 2.2can be found in Christian Thiemann's diploma thesis, which I strongly suggest to read.

3.1 Monte Carlo Algorithms

The Monte Carlo method relies on the idea of repeadetly using random number input to generate a result. Since this is such a repetitive task, Monte Carlo methods are implemented in computer algorithms. The next subsections describe a set of different algorithms that can be used to generate valid trajectory data for stochastic systems. This stands in contrast to trying to solve the master equation (Section 2.1).

In 1976 Gillespie [Gil76, Gil77] developed two exact stochastic simulation algorithms. The first is called the Direct Method and is a straight forward approach to generate a valid single trajectory. The second algorithm is called the First Reaction Method and works different than the Direct Method but is provably equivalent. The First Reaction Method has the advantage of being a little bit faster, because it can be easily optimized. Later Gibson et al. [Gib00] published their Next Reaction Method which extends Gillespie's First Reaction Method and uses some intelligent ideas to optimize the amount of recalculations needed during each iteration which can decrease the running time dramatically. The next sections show how these algorithms work mathematically and explain how to actually implement them.

3.1.1 Direct Method

Consider a system of r different reactions. All these reactions also have a rate constant k_n. When the system is prepared in a defined state, different transitions can happen. A transition is an execution of one of the possible reactions. At most there are r transitions possible. After a transition the system again has at most r different transitions. A random number generator can be used to choose which transitions really happen. The tricky part is to actually generate the random numbers with the correct distribution at every timestep to generate a statistically valid trajectory.

Gillespie's Direct Method calculates explicitly which reaction μ should happen and at which time τ. So these are the two open questions: How can we decide which reaction should occur and when?

This means, we need the probability distributions for reactions and for times. For this purpose a new function, the *Reaction Probability Density Function* $P(\tau, \mu)$, has been introduced by Gillespie. Here $P(\tau, \mu)\mathrm{d}\tau$ is the probability that, with the system in a given state at time t, the next reaction will happen in the intervall $(t + \tau, t + \tau + \mathrm{d}\tau)$ and that this reaction will be a reaction of type μ. Since we are interested in these μ and τ we need to find an analytical expression for $P(\tau, \mu)$. First, we remember that in Section 2.1 we already found that $a_\mu \mathrm{d}\tau$ is the probability that reaction μ with propensity a_μ will occur in the time intervall $(t + \tau, t + \tau + \mathrm{d}\tau)$. This leads

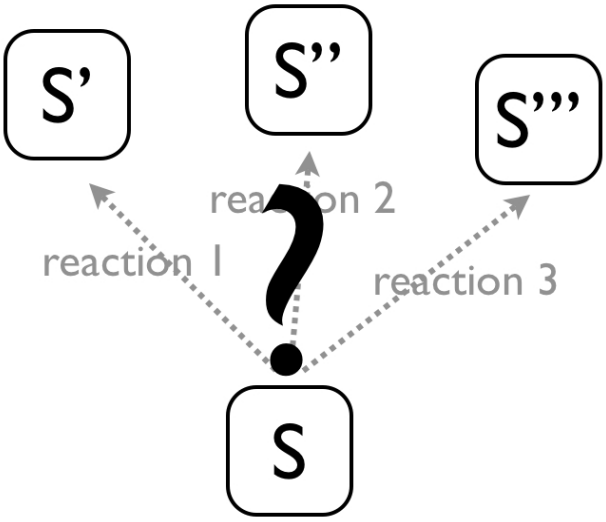

Figure 3.1: The big questions: Which reaction should occur? When should it occur?

to the idea to define a $P_0(\tau)$ which should be the probability that no reactions occurs in the intervall $(t, t+\tau)$. These two probabilities multiplied will also give us a description for $P(\tau, \mu)$:

$$P(\tau, \mu)\mathrm{d}\tau = \mathrm{P}_0(\tau) \times \mathrm{a}_\mu \mathrm{d}\tau \tag{3.1}$$

Since $[1 - \sum_\nu a_\nu \mathrm{d}\tau']$ is the probability that no reaction occurs in the infinitesimal interval $\mathrm{d}\tau'$ we can construct a selfreferential definition of $P_0(\tau)$:

$$P_0(\tau' + \mathrm{d}\tau') = P_0(\tau') \times [1 - \sum_\nu a_\nu \mathrm{d}\tau']$$

This leads directly to the analytical expression

$$P_0(\tau) = \exp(-\sum_\nu a_\nu \tau).$$

For easier reading we now denote

$$a_0 = \sum_\nu a_\nu,$$

the sum of all reaction channel propensities. Together with (3.1) we get

$$P(\tau, \mu) = a_\mu \exp(-a_0 \tau). \tag{3.2}$$

This equation allows us to correctly generate probability distributions for μ and τ. At first we integrate $P(\tau, \mu)$ from 0 to ∞,

$$\begin{aligned} P(\mu) &= \int_0^\infty P(\mu, \tau')\mathrm{d}\tau' \\ &= \int_0^\infty a_\mu \exp(-a_0 \tau')\mathrm{d}\tau' \\ &= \frac{a_\mu}{a_0}, \end{aligned} \tag{3.3}$$

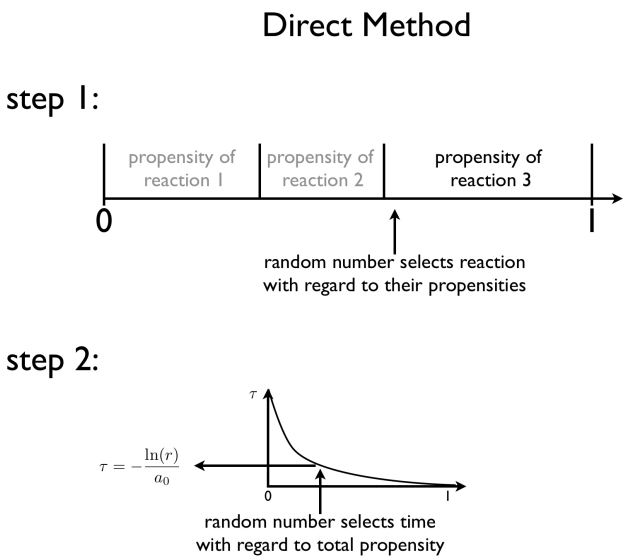

Figure 3.2: Gillespie's Direct Method uses two random numbers per iteration. One to select the reaction and the other to generate the reaction time.

and now we sum up $P_0(\tau, \mu)$ over all μ

$$\begin{aligned} P(\tau) &= \sum_{\mu'} P(\mu', \tau) \\ &= \sum_{\mu'} a_{\mu'} \exp(-a_0 \tau) \\ &= a_0 \exp(-a_0 \tau). \end{aligned} \tag{3.4}$$

Finally with these two probability distributions (3.3), (3.4) we are able to generate random numbers to correctly choose a reaction μ and its time τ. This leads us directly (well, maybe the name *'Direct Method'* has something to do with it) to Gillespie's algorithm:

Direct Method Algorithm (Gillespie)
1. *Prepare initial state (time, molecules, etc).*
2. *Calculate propensities a_μ for all reactions.*
3. *Choose μ according to probability distribution in* (3.3).
4. *Choose τ according to exponential distribution in* (3.4).
5. *Change system state according to selected reaction μ and increase t by τ.*
6. *Repeat from step 2.*

This simple and short algorithm can simulate exactly even complex processes with lots of reaction channels where solving the master equation would be an impossible task.

Since most operating systems only provide a (pseudo) random number generator, which delivers uniformly distributed random numbers from the unit interval, step 3 and 4 in the algorithm need a little more attention. To get random numbers according to the exponential distribution (3.4) a method called *inverse transform sampling*[1] is used which gives us, for an uniformly distributed

[1] This method only works for a small set of simple probability distributions. Since Monte Carlo algorithms heavily rely on sampling from arbitrary complex probability distributions, often *cumulative distribution functions (CDF)* are used. For implementation and getting the idea have a look at, e.g., the book series *Numerical Recipes [Pre07]* or for a short introduction *Wikipedia*.

number r,

$$\tau = -\frac{\ln(r)}{a_0}$$

Choosing the right reaction μ according to (3.3) is straightforward. Simply divide the unit intervall into smaller intervalls with sizes according to their respective a_μ. Therefore we choose μ such that the uniformly distributed r fulfills

$$\sum_{\nu=1}^{\mu-1} \frac{a_\nu}{a_0} < r \leq \sum_{\nu=1}^{\mu} \frac{a_\nu}{a_0}$$

The algorithm using these transformations is illustrated in Figure 3.2.

My commented sample implementation of the Direct Method in C can be found in Section A.1 on page 43.

3.1.2 First Reaction Method

Gillespie also published a second algorithm which, instead of generating μ and τ directly, calculates a time τ_i for each reaction channel first and then simply chooses the channel with the earliest time for execution.

The probability for a given reaction μ to occur during the time interval $(t+\tau, t+\tau+\mathrm{d}\tau)$ is the following:

$$P_\mu(\tau)\mathrm{d}\tau = a_\mu \exp(-a_\mu \tau)\mathrm{d}\tau \tag{3.5}$$

But this is only valid if there is no change in μ's reactant numbers. This means, if there is a reaction that happens earlier and which influences this reaction, the probability is not valid anymore. Having this in mind one can now generate reaction times τ_μ for each reaction channel according to distribution (3.5),

$$\tau_\mu = -\frac{\ln(r)}{a_\mu}.$$

Where r is again a uniformly distributed random number from the unit interval.

From this set of times we now choose the smallest one to actually occur. We combine this distribution and selection into one probability $P'(\tau,\mu)\mathrm{d}\tau$ which denotes the probability that reaction μ will happen during the time interval $(t+\tau, t+\tau+\mathrm{d}\tau)$,

$$P'(\tau,\mu)\mathrm{d}\tau = P(\tau < \tau_\mu < \tau + \mathrm{d}\tau) \times P(\tau_\nu > \tau \quad \forall \nu \neq \mu).$$

The first term describes the time interval in which it should happ en and the second term denotes that the time is the smallest of all times. Together with distribution (3.5) we can write

$$P(\tau < \tau_\mu < \tau + \mathrm{d}\tau) = a_\mu \exp(-a_\mu r)\mathrm{d}\tau,$$

whereas the second term can be written as

$$\begin{aligned} P(\tau_\nu > \tau, \forall \nu \neq \mu) &= P(\frac{\ln(r_\nu)}{a_\nu} > \tau, \forall \nu \neq \mu) \\ &= \prod_{\forall \nu \neq \mu} P(r_\nu < \exp(-a_\nu \tau)) \\ &= \prod_{\forall \nu \neq \mu} \exp(-a_\nu \tau) \end{aligned}$$

This last step is possible since the random numbers r_ν are uniformly distributed in the unit interval and the probability that it will be less than a given number between 0 and 1 equals just that number.

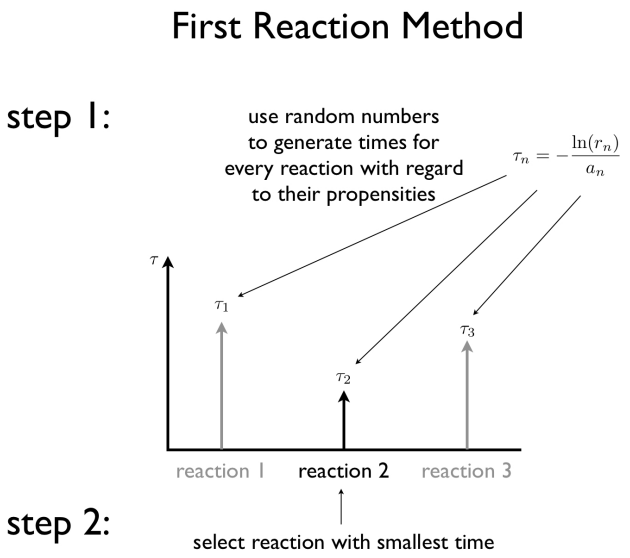

Figure 3.3: Gillespie's First Reaction Method needs random numbers for every reaction to generate their times. Then the smallest time is selected and the reaction executed.

Putting everything together again gives us

$$\begin{aligned} P'(\tau,\mu)\mathrm{d}\tau &= a_\mu \exp(-a_\mu r)\mathrm{d}\tau \times \prod_{\forall\nu\neq\mu} \exp(-a_\nu\tau) \\ &= a_\mu \mathrm{d}\tau \times \prod_{\forall\nu} \exp(-a_\nu\tau) \\ &= a_\mu \exp(-\sum_{\nu} a_\nu\tau)\mathrm{d}\tau \end{aligned}$$

This is exactly the distribution $P(\tau,\mu)\mathrm{d}\tau$ we already found in equation (3.2)

$$P'(\tau,\mu)\mathrm{d}\tau = P(\tau,\mu)\mathrm{d}\tau$$

Therefore the First Reaction Method is as valid as the Direct Method described before. This leads us to the following algorithm description:

First Reaction Method Algorithm (Gillespie)

1. Prepare initial state (time, molecules, etc).

2. Calculate propensities a_μ for all reactions.

3. For each μ generate a time τ_μ according to an exponential probability distribution with parameter a_μ like (3.4)*.*

4. Choose μ to have the smallest time τ_μ.

5. Set τ to τ_μ.

6. Change system state according to selected reaction μ and increase t by τ.

7. Repeat from step 2.

This algorithm has no real advantages performance-wise over the Direct Method, but as we will see in the following subsection it can be easily optimized and enhanced to allow the simulation of very complex systems without increasing computing time in an unacceptable way.

3.1.3 Next Reaction Method

Later, in 2000, Gibson [Gib00] published an extended version of the First Reaction algorithm which addresses its major bottlenecks. All these bottlenecks take time proportional to the number of reaction channels. First there is the propensity calculation of a_μ, second, we have to generate all times τ_μ and finally we have to sort all times to find out which is the first one to happen.

The first thing Gibson proposes is to reuse as much data as possible from the previous iteration. This means not recalculating all a_μ if nothing changed for this reaction channel, not regenerating the times τ_μ if the respective a_μ did not change, and using an ordered data structure to quickly retrieve the first time. If we want to reuse τ_μ from previous iterations we need to change from relative times to absolute values τ_μ^a since all relative times are invalid after increasing the time in the last iteration. This change is quite simple, we just add the actual time during τ_μ generation and store it instead.

$$\tau_\mu^a = t_n + \tau_\mu$$

This gives us the following distribution for τ_μ^a which depends on the actual time t_n at iteration n:

$$P(\tau_\mu^a > \tau) = \begin{cases} \exp(-a_\mu(\tau - t_n)) & \text{if } \tau > t_n \\ 1 \end{cases} \tag{3.6}$$

Not so intuitive is that we are really able to reuse τ_μ at all. Normally reusing values generated from random numbers during Monte Carlo simulations is not allowed.

For all reaction channels, that were not influenced by the last executed reaction step and therefore their a_μ did not change the distribution Eq. (3.6), we do not need to draw a new random number and therefore do not need any recalculation for their τ_μ^a. This saves us a lot of recalculations during every timestep of the algorithm execution, but also introduces the need of a *dependency graph*. This graph stores which reaction channels need to be recalculated depending on the actual executed reaction. Consider this simple example:

$$\begin{aligned} A &\rightarrow B \\ B &\rightarrow C \end{aligned}$$

Particles of type A decay to type B which then decay into type C. If the first reaction is executed the number of A and B particles are changed, so every reaction that depends on either A or B particles needs to have their τ_μ^a recalculated. In this case reaction two needs to be updated, because it depends on the B particle count. If on the other hand reaction two is executed all reactions that depend on B or C particles need an update, which in this case are none. As an implementation one stores a list for every reaction channel, which holds all other reaction channels, which need to be updated after execution.

Now that we know what to do with reactions that do not need a recalculation, we need to take a look at reactions that need the update. Since we do not want to draw a new random number for a time value we have not used yet, we simply rescale their τ_μ^a. Since distribution (3.6) has to be fulfilled after timestep n and the change in the a_μ, the correct rescaling for timestep $n+1$ is this:

$$\tau_{\mu,n+1}^a = \frac{a_{\mu,n}}{a_{\mu,n+1}}(\tau_{\mu,n}^a - t_{n+1}) + t_{n+1} \tag{3.7}$$

This allows us to save drawing a new random number and should also save computing time due to its simple math.

With these changes to the First Reaction method we reduced the number of random numbers used during each step to one and lowered the amount of recalculations to parts where it can not be avoided. The last thing Gibson introduced is a data structure to store all τ_μ^a in an ordered fashion. Since sorting algorithms are an old topic in applied mathematics, the right algorithm

is available in an algorithm book of your choice[2]. In this case Gibson chose an *Indexed Priority Queue* to get $O(1)$ access time to the smallest value and $O(\log n)$ running time for element updates. His modified *update* procedure avoids removing and adding elements from and back to the queue and should save some computing time for reactions with many temporarly inactive reaction channels whose τ equals infinity and would therefore be moved many times through the whole tree structure.

Now with all these fancy additions to Gillespie's First Reaction method, Gibson's Next Reaction algorithm has these steps:

Next Reaction Method Algorithm (Gibson)
1. Prepare initial state (time, molecules, etc).
2. Generate Dependency Graph.
3. Calculate propensities a_μ for all reactions.
4. For each μ generate a time τ_μ^a according to an exponential probability distribution with parameter a_μ^a like (3.6) *and store them in the indexed priority queue.*
5. Choose μ to have the smallest time τ_μ^a in the priority queue.
6. Set t to τ_μ^a.
7. Change system state according to selected reaction μ.
8. Update all reaction times τ_μ^a, according to (3.7)*, which are listed in μ's dependency list.*
9. Generate a new τ_μ^a for reaction μ according to distribution (3.6) *and update the value in the priority queue.*
10. Repeat from step 5.

This algorithm offers some great performance gains when compared to Gillespie's initial First Reaction and Direct Method. Especially in systems with many loosely coupled reaction channels the storing of a_μ and τ_μ^a saves a great amount of computing time. In cases with many sparsely populated reaction channels the fast sorting of times is helpful.

A small excerpt from my implementation source code that implements the above description can be found in Section A.2 on page 44 together with some small comments.

[2]For example *Introduction to Algorithms* [Cor01] gives a good overview over the algorithms needed here.

4 Modelling Panic Behaviour

Equipped with fancy stochastic algorithms and interesting disease dynamics models it is now time to put everything together. Since we want to look at disease spreading and especially at how the spreading is influenced by different human behaviour, we need to create a testbed in which we can play around and measure the relevant observables. At the same time, the simulation should be as simple as possible to avoid any unforeseen consequences. After introducing a simulation system that fulfills these requirements, different ideas of panic behaviour are discussed and added to the framework. Thus, this chapter also contains mostly theoretical background, whereas the corresponding results are shown and discussed in the later chapters.

4.1 Generic Linear Chain Model

In the early beginnings of disease modelling, researchers concentrated on the fundamentals. For example: *What is the time evolution of a disease in a closed population*? This is really important, since all these models, like SIR, answer that question and we rely on the research that has been done with these *compartmental models.* But our question is more of the kind *What is the time evolution of a disease spreading in a spatially extended population?* The spatial extent is what makes it trickier, as we have already seen in Section 2.3.

The compartmental models describe a single point, which has different properties like the amounts of different particles. Quantities like length, area or even volume have no relevance, until we introduce dimensions. In Section 2.1 we intentionally removed all dimensions and reduced the system effectively to this dimensionless point. In disease dynamics this point correspond to, e.g., a city, district or a closed group of individuals. Now, when there are two populations which somehow interact and these interactions include that individuals switch from one population to the other, it can happen that one of these individuals carries a disease and effectively initializes the infection of the healthy population. As soon as there is the possibility of an interaction between our two *points,* these points have a relation, a link, to each other. If we add more and more *points* to our system and add links between them we end up with the description from Section 2.4. But a complex network with lots of nodes and many links is very high-dimensional and is not what I would call a simple test system. What we want is a spatially extended system so that we can easily quantify the speed of disease spreading.

For speed measurements we need the location x of the wavefront at every time t during the simulation and from this we calculate the distance the disease has travelled over time. Since we only have nodes as locations, it is unclear how to define distance. The easiest approach is to stay in our nodes-and-links description and construct a linear chain of nodes. Every node in this chain has exactly two neighbours with an exception of the start and end node, which have only one. With this rather simple structure the definition of distance can be very intuitive and simple. The distance between node i and node j is the amount of links between these two nodes. Or, for simplification, all nodes get a number starting from the first node so that the distance equals $|i - j|$. Therefore all variables for each node get an additional index i which identifies the node and location in the linear chain.

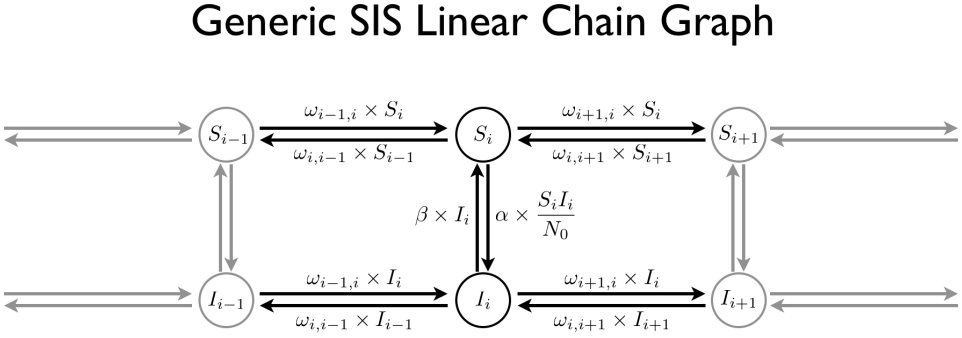

Figure 4.1: Every node in the linear chain consists of the two SIS model's classes S and I. Between neighboring nodes jumping events can occur, which move one individual from one class of a node to the same class on the other node.

For the SIS model this leads us to the following fundamental reactions:

$$\begin{aligned} S_i + I_i &\xrightarrow{\alpha} 2I_i \\ I_i &\xrightarrow{\beta} S_i \\ X_i &\xrightarrow{\omega_{ji}^X} X_j \end{aligned}$$

where X_i stands for all different classes, in this case S_i and I_i. For a linear chain the ω_{ji} are defined as

$$\omega_{ji}^X = \begin{cases} \omega_0 & \text{where } |j-i| = 1 \\ 0 & \text{else.} \end{cases}$$

The meanfield rate equations now have the form

$$\begin{aligned} \mathrm{d}_t S_i &= -\alpha \times \frac{S_i I_i}{N_0} + \beta \times I_i + \sum_j (\omega_{ij}^S S_j - \omega_{ji}^S S_i) \\ \mathrm{d}_t I_i &= +\alpha \times \frac{S_i I_i}{N_0} - \beta \times I_i + \sum_j (\omega_{ij}^I I_j - \omega_{ji}^I I_i) \end{aligned}$$

There are a few important things to note about this model which are quite important for the later discussion and modifications of it. The infection reaction propensity in this model is defined as $\alpha \times {}^{S_i I_i}/_{N_0}$. The important difference to the previous models is the use of an N_0. We want to model cities that can be nearly empty and cities that are crowded with people. So having equally sized cities with N_0 individuals each, at initialisation time, allows us to have a high or low density. If somehow a great amount of people left one city for another, the disease has a higher probability to spread inside the overcrowded city and a lower probability in the nearly empty city. This is quite intuitive: two persons accidently meeting in a deserted New York is quite improbable, but during normal days there are many people meeting every second. If it is still unclear where the N or N_0 comes from, try to read Section 2.2 again.

The other thing to note here is that the definition of the ω_{ji}^X sets every link to ω_0 and thus also to the same for every class of individuals in the system. This means that everybody, infected or not, has the same travelling probabilities. In the modifications of the model we introduce in the next sections, we play around with this assumption and let different classes have different probabilities.

Since we also discuss variants of an SIR linear chain model, it should be said that it works the same way as the SIS linear chain model. Every class of individuals can travel to neighboring nodes, whereas in the nodes the normal SIR dynamics take place.

4.2 Increased Diffusion Model

Instead of setting the diffusion rate on all links to the same value, we introduce a modification of the generic linear chain model, where the ω_{ji}^X is modified based on the local infection level,

$$\omega_{ji}^X = \omega_0(1 + \nu\frac{I_i}{N_0})$$

where ν is a factor which can be choosen freely to change how strong the additional diffusion should be. The density of infected individuals increases the normal diffusion rate additive, such that without any infecteds the diffusion rate stays the same as in the normal model.

In the simulations we look at two variants of this. The first has the additional diffusion rate on links for all different classes of individuals. The second model has it only for the susceptible class.

The idea behind this is to model a panic like behaviour of individuals. In this model the individuals sense that there is a disease outbreak in their city and as a reaction try to move away from it. Therefore they increase the local diffusion rate. Since it is interesting to consider that infected and recovered individuals could know about their current health status, we also look at the variant where only susceptibles show this panic behaviour and increase their diffusion rate.

4.3 Directed Flight Model

The increased diffusion model simulates that the individuals just try to get away from the disease outbreaks. They choose the direction they want to go randomly. *That's really stupid.* Maybe mankind is a little bit brighter and thinks before travelling away. It is a lot wiser to choose the direction based on the status of the city one would end up in. So an individual has to compare the local city i to the remote cities j:

$$\Delta_{ji}^I = \frac{I_i}{N_0} - \frac{I_j}{N_0}$$

This difference is positive, if the density of infecteds in the local city i is higher than in the remote city j. We can use this value to modify the diffusion rate of the corresponding link ω_{ji}. Since we do not want to lower the diffusion rate below our base rate ω_0 we have to consider the two cases $\Delta_{ji}^I < 0$ and $\Delta_{ji}^I \geq 0$,

$$\omega_{ji}^X = \begin{cases} \omega_0(1 + \mu\Delta_{ji}^I) & \text{where } \Delta_{ji}^I \geq 0 \\ \omega_0 & \text{else.} \end{cases}$$

We can also simply use the Heaviside function $\Theta(x)$ to replace the cases such that we have a single-line description and make it easier to read:

$$\omega_{ji}^X = \omega_0(1 + \mu\Delta_{ji}^I\Theta(\Delta_{ji}^I))$$

The μ is used as a factor to adjust the force strength forcing the people to move. This model has been simulated with the two different variants again. One where everybody is affected by the urge to travel away, if an outbreak happens in their city, and the other, where only the susceptibles see the need to bring them selves to a safer place.

5 Simulation, Results and Discussion

5.1 Simulated System

The linear chain model systems that have been introduced in the previous chapter have been simulated with an implementation of the next reaction method that I implemented[1] in *C*. Since these models have a large parameter space and we want to take a look at the effects the new increased diffusion and directed flight models have on the disease spreading, some of the parameters are set to a constant throughout all of the simulations. In Figure 5.1 the value of α has been varied to show the effect on the normal SIR and SIS dynamics. Later on we only look at $\alpha = 3.0$, $\beta = 1.0$ and $\omega_0 = 1.0$. This makes it easier to concentrate on the parameters of the new models which are ν in the additional diffusion model and μ in the directed flight model. With these parameters the effect can be gradually increased from not present, $\nu = \mu = 0$, to any desired strength, $\nu > 0$, $\mu > 0$.

The simulation setup is such that a linear chain for the desired base model, SIS or SIR, is created with a length of 500 nodes and no periodic boundary conditions. Then every node is populated with an initial number of $N_0 = 500$ susceptibles. To initialise an outbreak 10 infected individuals are added to the first node in the linear chain. The simulation has two abort conditions: there are no infected individuals left anywhere in the system or the wave front reaches node number 400.

During each simulation different measurements are taken: If the wave front reaches node number 200 and later also node 400, the time between these two events is used to calculate the speed of the wave front. This is a fairly accurate speed measurement since the wave front speed is very stable as the Figures 5.8 and 5.9 show.

If the wave front reaches the 100th, node a snapshot of the actual system state is taken which can be used to have a look at the wave front shape and the distribution of the other individual classes.

In a periodic simulation time interval the total number of individuals from each class in the whole linear chain is recorded to get an integral over all nodes.

And as a last quantity we can distinguish between simulations that finished due to reaching the 400th node or due to lack of infected individuals. This allows to see if the new models have an effect on the stability of the spreading wave.

All this data has been recorded for the different framework parameters: SIS or SIR model, additional diffusion oder directed flight model, all individual classes affected or only the susceptibles and of course many different values for ν and μ. For each of those parameter sets 100 simulations were started and recorded resulting in around 32000 individual simulation runs. Since, on average, 4 minutes are needed per simulation, on a modern processor this results in 89 days[2] pure simulation time and 2 GB of valuable data. A small *C* program was then used to read in the data, calculate averages and standard deviations and to finally generate appropiate input files for *gnuplot*. The processed and refined results are presented on the next pages together with descriptions how to read them, their meaning and some possible explanations.

[1] The implementation details can be found in the appendix of this thesis.

[2] On good days, meaning that nobody else tried to work, e.g. christmas holidays, the MPIDS cluster system can do it in 3 to 4 days. As one can guess now, without the optimized algorithms this would not be feasible in finite time.

5.2 Results

The idea behind the additional diffusion and directed flight models is to investigate the effect of a dependency between diffusion and reaction processes in the standard SIS and SIR models. In normal reaction diffusion systems the diffusion process only depends on the number of individuals that are affected by the diffusin itself and not, like in the new introduced models, also by the number of other individuals. Therefore the diffusion channel for each class can behave independently from the others and, most important, the diffusion behaviour can be very different in each node of the system. If we go back to the chemical particle description one can see the idea behind the new models as particles that influence the behaviour of other particles with a force. In the additional diffusion model we have the infected particles pushing other particles out of the subvolume. We investigated two variants, one where only susceptibles are affected by the pushing force of the infecteds and the other where every particle is influenced by the force. Since a pushing force inside a subvolume increases the number of particles leaving that subvolume, we have a locally increased diffusion. Since we are in the subject of human diseases, this model can be seen as a panic model: the people of a city or area get informed that a disease spreads in their region and they panic and simply try to run in a random direction in the hope to avoid an infection.

Since we are talking about humans, which should be more intelligent in choosing an appropiate reaction to a disease outbreak, we came up with the directed flight model to simulate a saner behaviour: Instead of choosing a random direction for their evasion maneuver, the neighbours are compared with the local area and evaluated if the risk to get an infection decreases if travelled to a neighbouring site. On the particle level this could mean that the force reaches over the length of one subvolume into the next subvolume and if there is the same amount of infecteds in both subvolumes the forces cancel each other out. Therefore the increase in the local diffusion is directed away from subvolumes with higher levels of infected individuals. Important here is that the diffusion process is not symmetric anymore. More appropriate would be to use the name “drift”. Again, this drift can be different at each node, depending on the local amount of infected individuals *and* their next neighbours.

With those descriptions in mind it should be clear that additional diffusion and directed flight are *toy* models. We use them as prototypes to get some hints what effects can appear if there are additional dynamics affecting the particles node switching behaviour. Therefore the simulations are made as simple as possible. There are the basic reaction diffusion models, SIS and SIR, which have been studied for a long time and whose dynamics are well understood and we can now carefully add the new models and slowly increase their influence via the parametes μ and ν. If we want to understand what happens in more realistic and also more complex systems we first need to understand the effect of each of the small parts that make up the whole system.

So, what happens if there are people that try to run away?

Let us first have a look on the plain SIS and SIR dynamics, which can be seen in Figure 5.2 on page 34. In the SIS model we have a wave front and a tail that stays around a constant level of infecteds. If the diffusion moves an infected individual from the front line into a not yet infected node the reaction dynamics will create more and more infected individuals in this node until the critical level, where infection and recovery reaction have equal rates, is reached. In the SIR model there is no equilibrium level of infecteds. As long as there is enough susceptible supply, the number of infecteds will go up. At the same time the recovery reaction will decrease the number of infecteds and turn them into immune individuals. The outbreak in this node eventually dies out. Therefore the SIR wave front has this characteristic shape with an exponential rise and a slower decaying tail.

In the next figure, 5.3 on page 35, the additional diffusion model is added to the dynamics and the differences in comparison to the plain models can be seen. The only striking differences are in

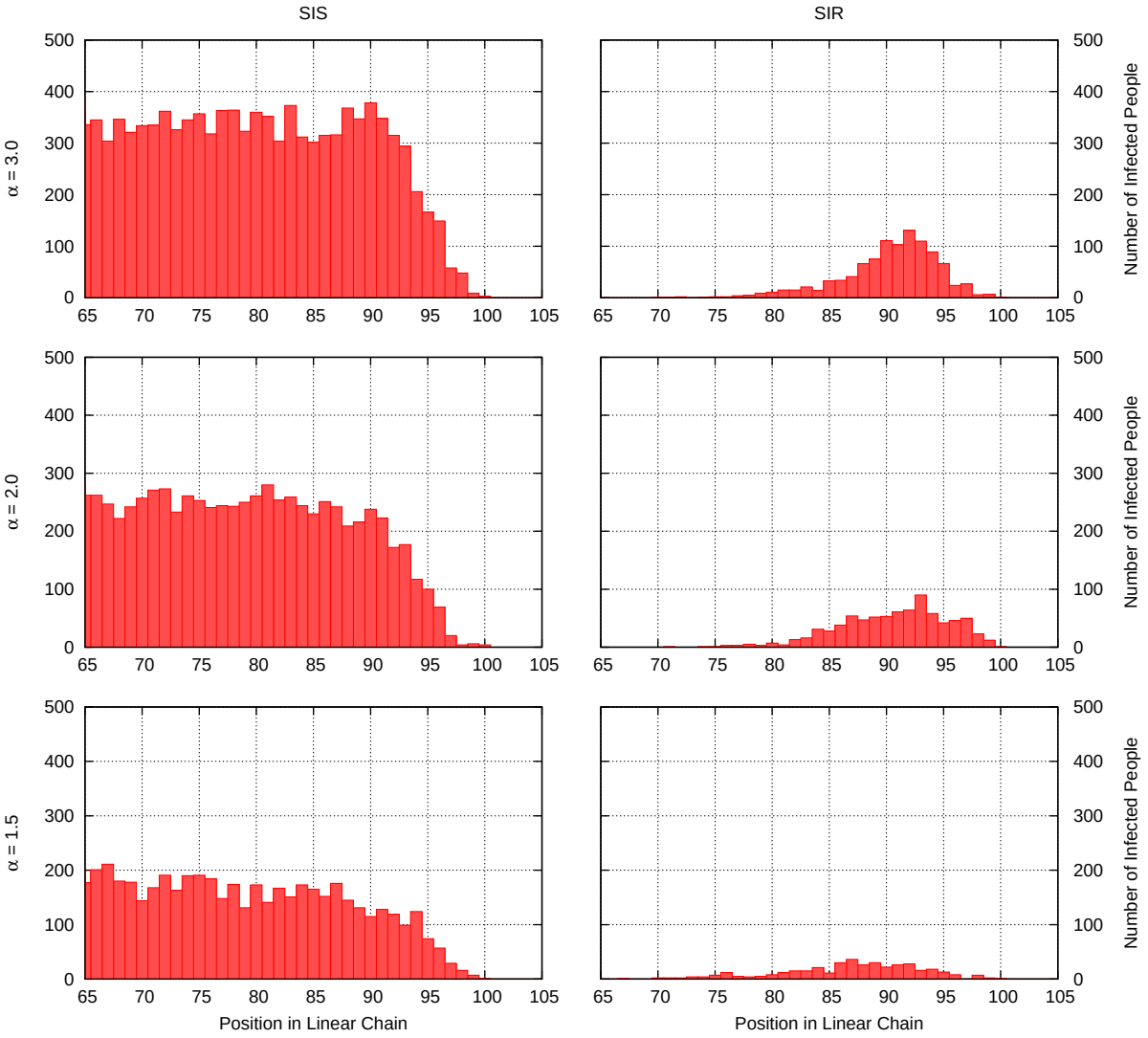

Figure 5.1: Snapshots of the wave front at the moment it hits the 100th node in the linear chain. On the left we have the endemic SIS model and on the right the SIR model outbreak travelling through the system. In the SIS system the level of infected people per node gets lower with decreasing α values and the SIR wave front gets smaller and weaker. Since each plot is a snapshot from one simulation trajectory the stochasticity can be clearly seen in the SIS wave front tail.

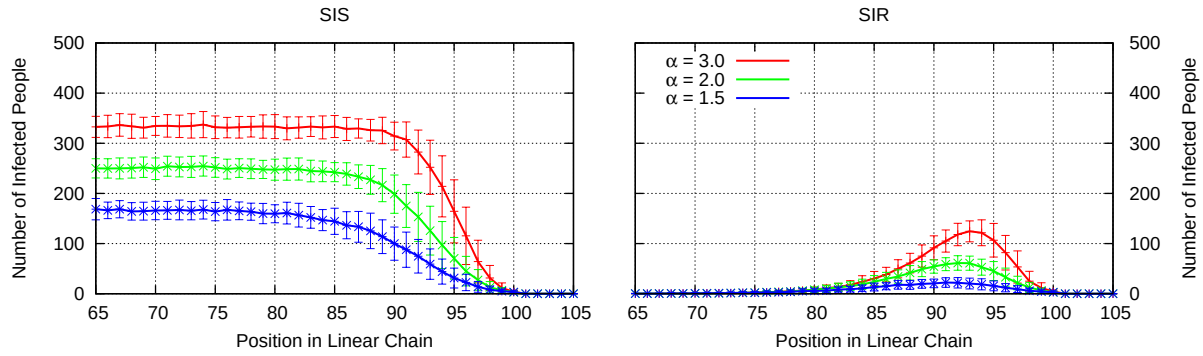

Figure 5.2: Averaged wave front form for the same simulation parameters as in Figure 5.1. The effect of different α values can be seen clearly. For SIS, the theoretical stable equilibrium point is at $I_{eq}^{\alpha} = N(1 - \beta/\alpha)$. For simplicity, β has been always set to 1 and N_0 was 500 at the beginning of the simulation. Therefore we get $I_{eq}^{3.0} \approx 333$, $I_{eq}^{2.0} \approx 250$ and $I_{eq}^{1.5} \approx 167$. The simulation results reproduce the theoretical values in a more than pleasing way.
The averaging smoothed out the stochastic effects of the diffusion reaction which can be seen in the single trajectory data. This helps to analyse the change in the SIR wave front: the total amount of infected individuals, which make up the wave front, decrease with lower α values. For both SIR and SIS, a higher α results in a faster rise of the wave front and therefore a faster spreading through the linear chain. For the following simulations α is always fixed at 3.0.

the models where everybody is affected by the additional diffusion. The wave front has a stronger rise in both, SIS and SIR, models and in the SIR model the tail is longer and the wavefront consists of more infecteds in total. Since the reaction dynamics time scales do not change, the long tail, strong rise and the many more infecteds can only mean that the disease spreads faster through the system. The same observation can be made in the directed flight model data (Figure 5.4 on the facing page). In those, the wave front shape for the variant where only susceptibles run away is also very interesting. We have an even stronger rise but the tail stays normal. The Figures 5.6 and 5.7 on page 37 show the measured speeds of the wave fronts. For all model combinations we can measure an increased speed. Interestingly, all variants where everybody is affected have a higher wave front speed than their counterpart where only the susceptibles run away. Since the wave front is only able to spread into the next node if infected individuals move to this node, it is obvious, that if all different classes have an increased rate to travel to their neighbour nodes, the spreading speed increases. In both models, additional diffusion and directed flight, the rate from nodes with infecteds to neighbour nodes without infecteds is higher than normal, resulting in the increased speed. In normal SIS and SIR dynamics the spreading speed is dominated by the diffusion of the infected individuals. One could simply shutdown the diffusion of susceptibles completely without slowing the wavefront down. This makes the results of the other variant, where only the susceptibles are affected even more interesting. There is an increase in the speed and we only changed the behaviour of the susceptibles. The difference between simply increasing the diffusion of the susceptibles and the additional diffusion model is that in our additional diffusion model, the diffusion for susceptibles is increased only locally, depending on the density of infected individuals. So if we imagine the linear chain, we have areas with higher diffusion rates and areas with lower diffusion rates. If all rates were equal this would be the normal situation. The diffusion would make sure that after some time the number of individuals in each node is roughly the same. But now if there are different rates in different areas this will also result in

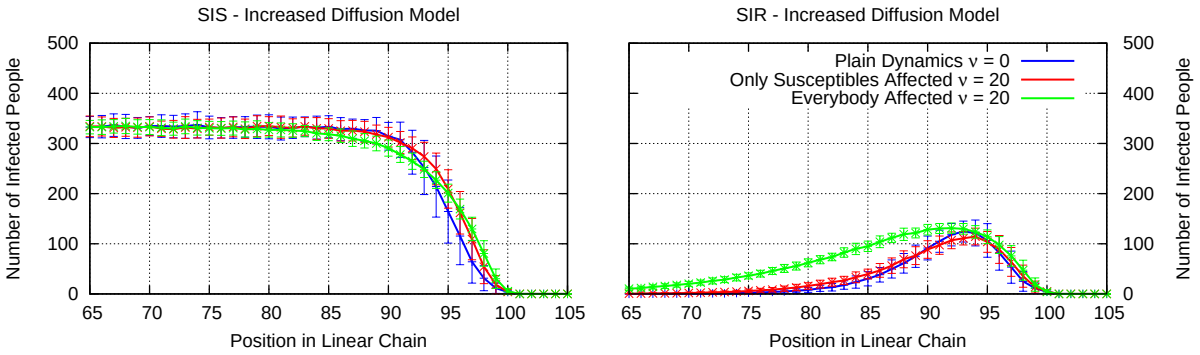

Figure 5.3: Averaged wave front shape for the additional diffusion model. The data is based on one hundred simulation runs for each parameter set. We have the plain SIS and SIR dynamics and data for the two variants of the additional diffusion model with $\nu = 20$. In one case only the susceptibles are affected by the additional diffusion and in the second case all species are influenced.
The red line, which is the average where only susceptibles have the additional diffusion, is nearly the same as the plain dynamics. The additional diffusion in this variant seems to make no real difference in the wave front shape.
In contrast, the green line, depicting the case where all individual classes are affected, has a stronger uprise in the SIS model and a longer tail in the SIR model. The long tail could be an indicator for a faster moving of the wave front. As we will see, the speed measurements agree with this interpretation. The long tail also means that we have more infected individuals in the spreading wave in comparison to the plain dynamics.
The additional diffusion model where every class of individuals reacts to the local density of infected people helps the disease spreading faster.

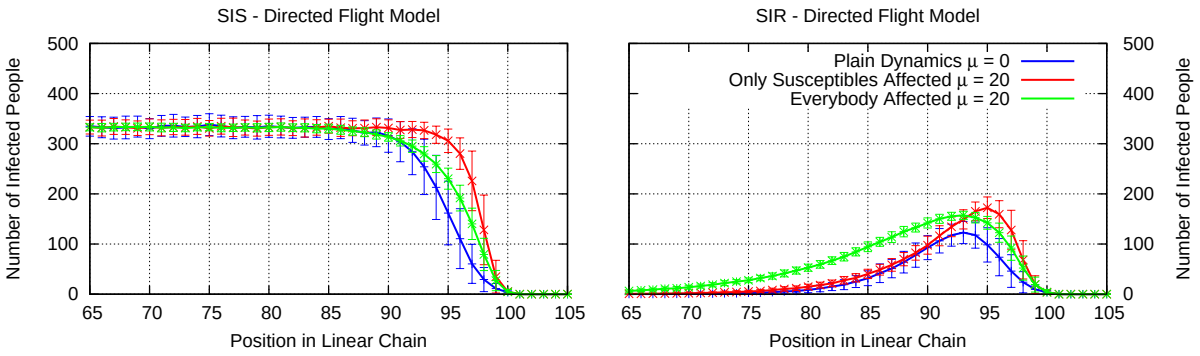

Figure 5.4: Averaged wave front shape for the directed flight model. Again data is the average over one hundred trajectories for each parameter set. The force strength μ for the two model variants is set to 20 and 0 for the plain dynamics.
Now, when the reaction on the density of infecteds is directed and not random, we can see that the variant where only susceptibles are affected differs from the plain dynamics and rises very fast. The tail after the maximum in the SIR model behaves similar to the plain dynamics tail.
In the case where all individual classes do the directed flight the wave front rises stronger than in the plain dynamics but not as fast as in the other variant. The tail in the SIR model variant of this case is again longer and we can also assume a faster propagation speed of the wave front.

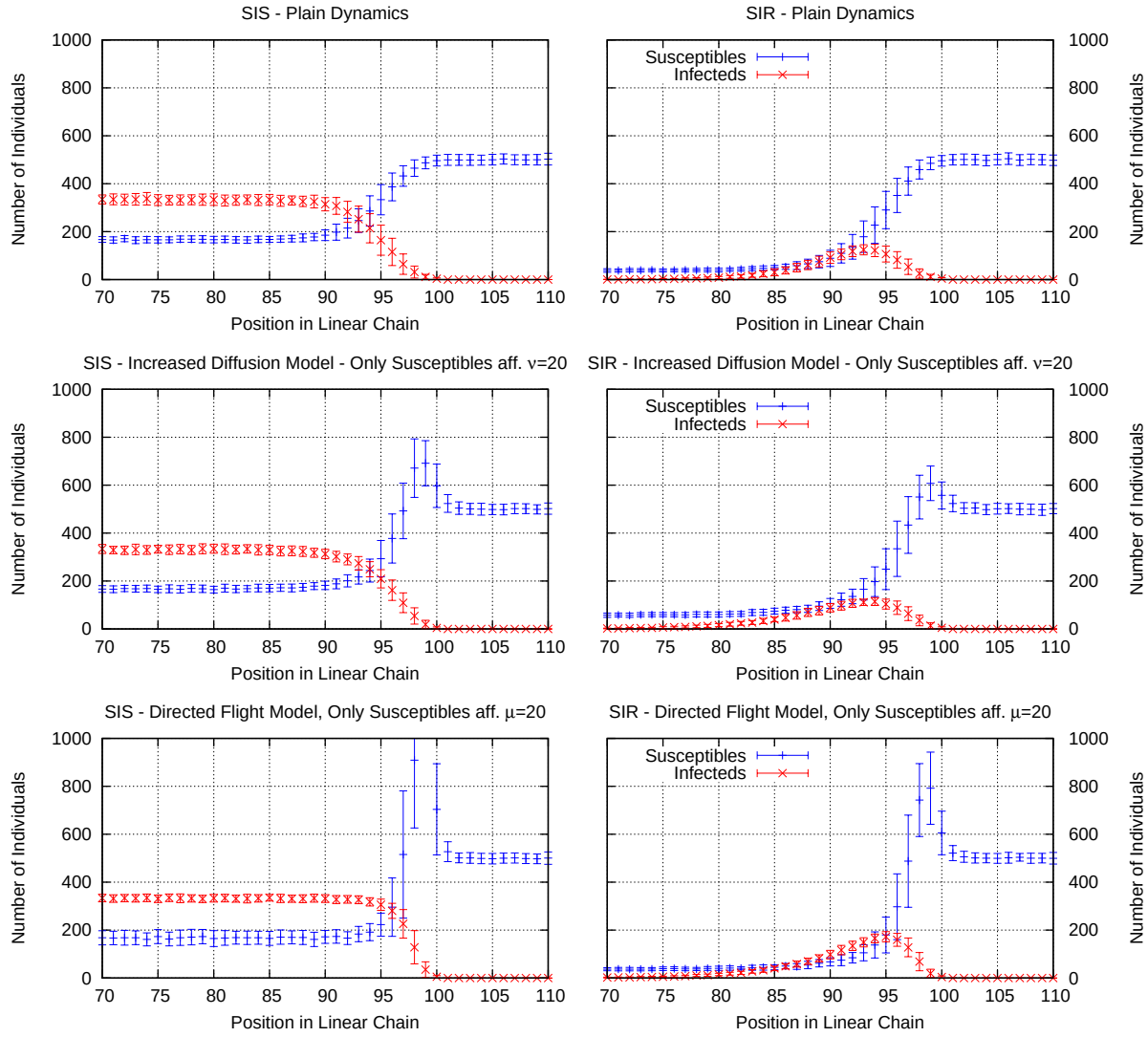

Figure 5.5: Averaged wave front data: The red points correspond to the total number of infected individuals in each node. The blue points indicate the number of susceptibles. The different models change the number of susceptibles in front of the spreading wave drastically. This higher density allows the disease to spread more easily into these nodes. Therefore the wave front speed inreases: The disease can spread faster.

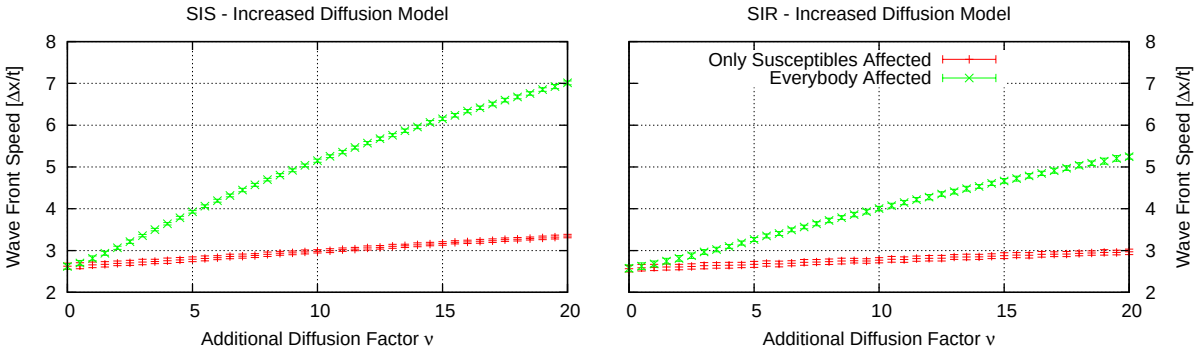

Figure 5.6: Averaged wave front speed for the four increased diffusion model variants. The speed has been measured by the number of nodes the wave front travelled per simulation time unit. The speed is on the y-axis whereas the additional diffusion factor ν has been varied and is used as the x-axis. The plain SIS and SIR dynamis correspond to $\nu = 0$ and therefore they have a spreading speed of $^{\Delta x}/_{t} \approx 2.6$ units.
One can see that for both SIS and SIR and only susceptibles running away the front wave speed increases slowly. For $\nu = 20$ the speed is about 15% to 30% percent higher than the plain dynamics.
If all individuals are affected by the increased diffusion the speed increases even more and is about 170%, for the SIS model, and about 115% faster for SIR. The dependency between ν and the speed seems to be not linear, as the rise is stronger for smaller ν.

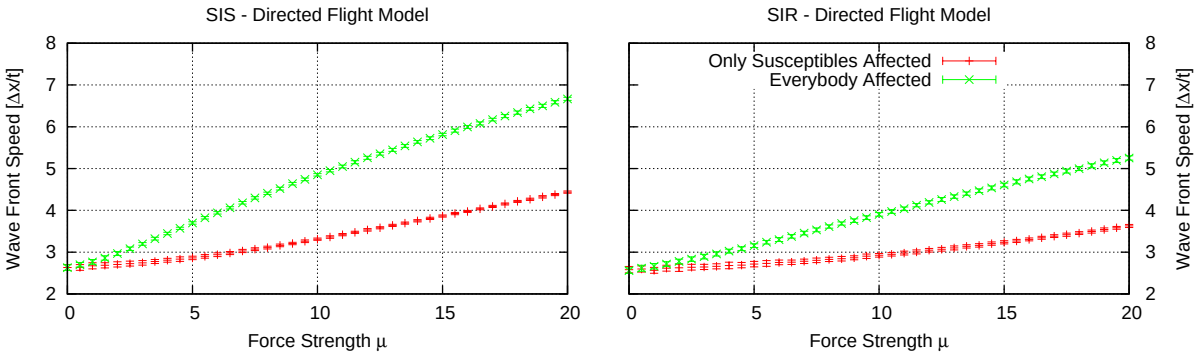

Figure 5.7: Averaged wave front speed for the directed flight models. The x axis denotes the force strength parameter μ and gets varied to see the influence on the wave front speed.
When only susceptibles do the directed flight due to the local infection density the wave front speed first increases slowly and later seems to accelerate for higher μ. For $\mu = 20$ the speed in the SIS model increased by 70% and for the SIR model by 50%. The wave front speed for the model where all individuals are allowed to take flight increases nearly to the same speed as in the additional diffusion model. We end up with speeds 160%, for SIS, and 100%, for SIR, higher than the plain SIR/SIS dynamics.

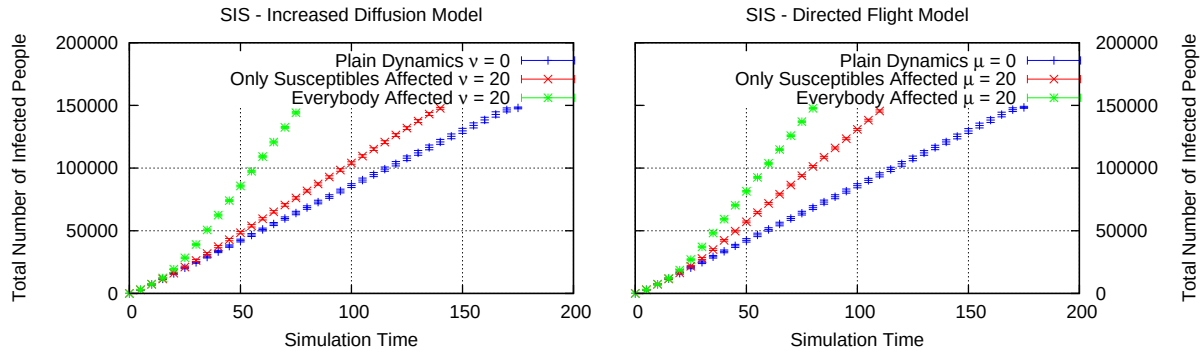

Figure 5.8: The averaged time evolution of the total number of infected people for the SIS model. On the left the data for the increased diffusion model with the three variants is shown and on the right the same for the directed flight model. Since the simulation increases the ν slowly up to the desired value between simulation time 10 and 20 there is an arc in the data points until the wave front found a stable form. From there on the points describe a straight line until the wave front reaches the end of the simulation system. The errorbars and the linear behaviour suggests that the wave front is pretty stable and moves with a constant speed.
Since the gradient of the data describes the amount of new infections per time unit it is obvious, that in both models the variant where all individuals try to run away has the effect that the disease spreads much faster in the population and infects more than two times of persons in the same time than the plain dynamics.
For the other variant, where only the susceptibles are allowed the extra mobility, the effect differs between increased diffusion and directed flight. With increased diffusion the gradient is not as high as for the directed flight model.

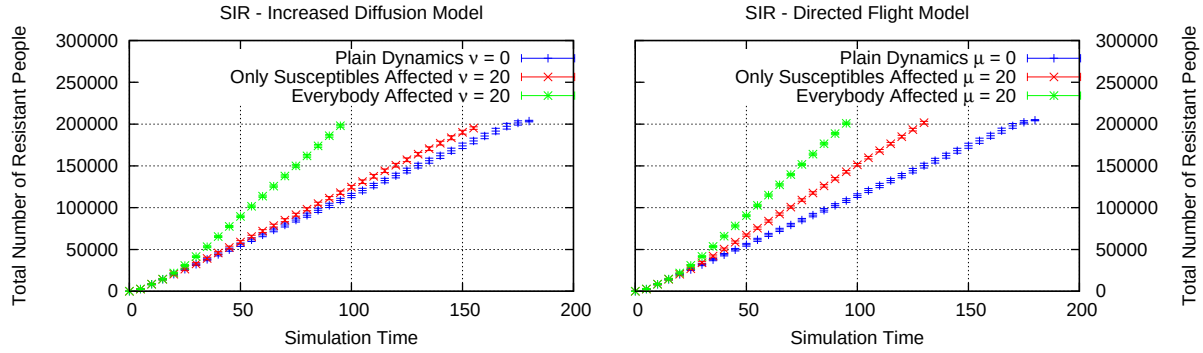

Figure 5.9: Since the total number of infected people in the SIR model would be a meausure of the size of the front wave and the gradient would be no meausure for the number of infection reactions per time unit the data plotted here are the total number of resistant people. Since all infected people will eventually end up as resistant people one can take the gradient of the total number to get the number of infections per time unit.
Qualitatively the result is the same as for the SIS model. When everybody is affected the rate of infection is doubled and when only susceptibles are affected the rate only increases marginally for the increased diffusion and is higher for the directed flight model.

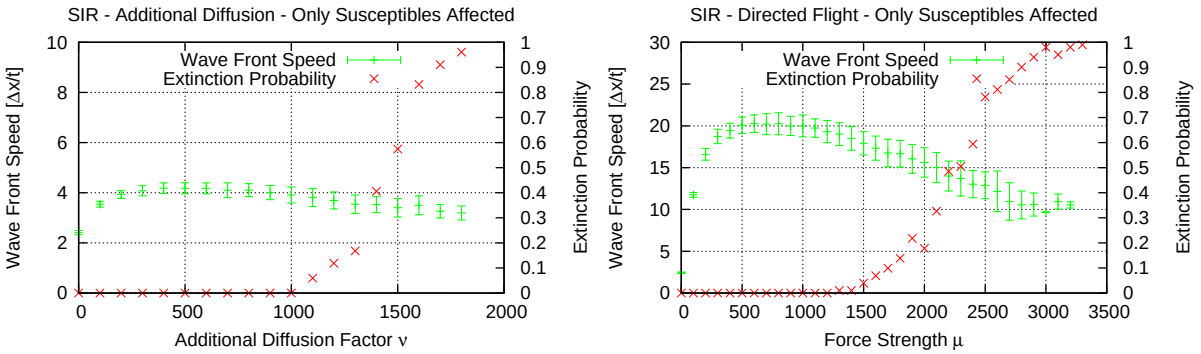

Figure 5.10: Since the simulations are stochastic there is a chance that the disease can die out. The chance that a SIS outbreak, which has pandemic characteristics, dies out due to stochastic effects is quasi zero where as in the SIR model there is always a small but realistic propability. These graphs show how a further increase of μ and ν lead to regions where the disease does not survive the simulation.
Interestingly it can be observed that the wave front speed first increases to a maximum and then starts to slowly decrease again. At the same time as the speed decreases again the standard deviation also increases. This means that there is a high variability in the wave front speeds of different simulations. The wave front speed for the directed flight model reaches far higher speeds than the additional diffusion model. The maximum is around 6.7 times the plain dynamics speed. In the additional diffusion the maxium is around 1.6 times the normal speed.
The extinction probability in this data is the percentage of simulations where the number of infected people reaches zero before the wave front arrives at the end of the linear chain system. The probability in the additional diffusion model rises earlier and reaches a probability of nearly one faster. This means that for high enough parameters ν and μ no disease spreading can occur: the outbreak is locally confined.

different numbers of individuals proportional to the differences in the diffusion rates. If only the susceptibles have this difference in the rates we will end up with an heterogenous distribution of susceptibles. Now we need to remember that the process where one infected meets a susceptible individual is strongly depending on the local density of susceptibles. This means that if one infected comes into a node with a higher than average density of susceptibles, the number of infected individuals will, on average, reach a given level of infecteds faster as in a node with a normal density of susceptibles. Higher numbers of infected individuals in a node increase the probability to infect the neighbouring nodes. All this, first the higher densities of susceptibles around areas with infected individuals and second the higher infection speed in exactly these nodes, eventually leads to an increased wave front speed. In the directed flight model the effect is even stronger: The susceptibles do not move randomly anymore, but directed. This supports the process of cumulating susceptibles in front of the wave and therefore helps the disease spread more easily forward. Figure 5.5 shows the distribution of susceptibles and infected individuals at the wave front. The increased density of susceptibles in front of the wave can be clearly seen and lacks in the plain SIS/SIR model.

It looks like both types of reactions we modeled here have a bad effect on the population; the diseases spreads faster. But maybe there are some aspects that are beneficial for the population. For example it could be possible that the total number of individuals that are and were infected is lower if they try to evade the wave. The Figures 5.8 and 5.9 on page 38 show the time evolution of the integral over the infected or immune individuals. The derivation of these plots give the number of new infecteds per simulation time unit. It can be seen as the stress that the disease puts on the population's health system. Unfortunately, the simulation results are not very healthy for the population. All model variants increase the number of infected individuals faster than the plain dynamics. Again the models where all classes are affected are the fastest at new infections and the variants with the susceptibles are in between those and the plain dynamics.

If the parameters ν and μ are increased even further an interesting behaviour can be observed. The corresponding figure is 5.10 on the previous page which shows the additionl diffusion and directed flight for the SIR model. On the y-axis we have both the speed of the wave front and the probability that the wave reached node number 400 and successfully finished. The x-axis is, as before, the value for ν and μ respectively. In both cases the wave front speed reaches a maximum and then slowly decreases but does not reach the speed of the plain dynamics again. For those higher values of ν and μ the extinction probability leaves the area around zero and quickly increases to one: The total number of infected individuals in these simulations reached zero before the wave front reached the 400th node. So there is a possibility that too many susceptibles leave an infected node such that the disease can not spread further. Since in the SIR model all infecteds will eventually end up as immune individuals that can not infect others anymore the disease dies out. The parameters ν and μ are so high that the rates for the node switching events of the susceptibles are increased in such a way that all areas of the linear chain where infecteds exist get completely abandoned by them. This effect also hinders the spread of the disease for parameter values where the disease does not die out. When the rate for the susceptibles is much higher than the rate for the infecteds, stochastic effects begin to dominate the spreading. The time the disease has to infect susceptibles in a node decreases due to the high rate of susceptibles leaving the node. Therefore the number of infecteds does not go as high as in the plain dynamics. Less infecteds also mean a smaller chance to infect the neighbour nodes and in that way decreases the wave front speed. The extreme case would be that in the very moment one infected would reach a node, all susceptibles would leave it instantly. The stochasticity of this effect explains the more and more increasing standard deviation of the measured speeds in Figure 5.10 on the preceding page. Christian Thiemann's mean-field simulations of the same two models show no disease extinction of this kind. The mean-field approximation does not catch the stochastic fluctuations that eventually lead there.

6 Discussion

The additional diffusion model has been created with a population in mind that reacts panically to a spreading disease. The people try to escape and simply want to change their location if there are infecteds that could possibly infect them. They do not think about their escape strategy and just choose a direction by random. If this model is applicable to our modern society with television, mobilphones and internet is questionable, but if the information about infections is only locally known this model would be more realistic. The directed flight model uses one more level of information. The people's decision is based on the infection levels of their node and their neighbours. This increases the chance that more people travel to nodes with lower infection rates, but they do not care about the global situation. So this strategy seems better than the panic like behaviour of the additional diffusion model. Of course the directed flight is also not the optimal strategy because it only reacts to events in a small area and not on the whole system's dynamics. Interestingly, in both models the people's change in behaviour at first only supports the disease outbreak. The disease spreads faster and, what is maybe more critical, harder: more people are getting infected per time unit and more people are infected at every time, putting a greater workload on the health system. Surprisingly, the directed flight model, where the people react more intelligent, has a much higher maximum wave front speed than the additional diffusion model. It seems that flight reactions do more harm than good. Since we do not know what realistic values for ν and μ would be, the parameter regions where the stochastic effects begin to slow the wave front down and increase the extinction probability are as important as the results for small values. This means that we can not draw a uniform conclusion for what will happen in a more realistic system or in the real world. More importantly the models presented here should be really seen as toy models only. Made as simple as possible to examine the effects.

Many scientists try to go a different way: They build very complex systems and include as much information as possibe into the simulations. They model travelling behaviour, society structures, geographic differences and many more things into one gigantic model and start the simulations with many million agents. At the end they have a lot of data which is based on so many assumptions that nobody can really use it. So a much better way to understand how diseases spread is to concentrate on simpler models that only add one new effect at a time, building up the epidemics puzzle piece for piece. Since networks, which are not linear chains or regular lattices, dominate the human society, the next step will be to exchange the linear chain with a human traffic network[1] and study what effect the topology has on the different measures. Other interesting areas are in the field of spatially extended predator-prey models in biology, where e.g. the predator could change its hunting behaviour when the prey density changes. I think there are many different applications in ecology. In his thesis, Christian Thiemann, introduced a third model which could not be used in my current stochastic simulation framework yet. The model allows the people of a node a much further look and therefore they can react and begin to travel away before the front of the disease is near them. This seems to be a really effective strategy since it can stop the wave front in a mean-field simulation and it would be really interesting to see the stochastic results if the model can be implemented properly. Another idea is to include an active response of the society in the model e.g. vaccination or quarantine of nodes.

[1] For example the data from http://www.wheresgeorge.com is a good candidate for an approximation of the human travel network in the US.

A Appendix

A.1 Direct Method Implementation

A complete implementation of the Direct Method from Gillespie is very straightforward and simple. Here as an example a simple system with four different particle types and three reaction channels. The possible reactions are:

$$V + X \xrightarrow{k_0} 2X$$

$$X + Y \xrightarrow{k_1} 2Y$$

$$Y \xrightarrow{k_2} W$$

The source code should be more or less self explanatory. At least for persons with a minimum of programming knowledge.

```
1  #include <stdio.h>
2  #include <stdlib.h>
3  #include <math.h>
4
5  int V=180, W=0, X=10, Y=10;
6  double k[]={1.0, 1.0, 1.0}, a[3];
7  double a0=0.0, t=0.0, r, tau, as;
8  int mu;
9
10 int main() {
11    while(1) {
12       printf("%lf: V=%d X=%d Y=%d W=%d\n", t, V, X, Y, W);
13
14       // calculate the propensities for all reactions
15       a[0] = k[0] * V * X;
16       a[1] = k[1] * X * Y;
17       a[2] = k[2] * Y;
18       a0 = a[0] + a[1] + a[2];
19       if(a0 <= 0.0) break;
20
21       // draw random number and choose one reaction accordingly
22       as = 0;
23       r = a0 * drand48();
24       for(mu=0; mu<3; mu++) {
25          as += a[mu];
26          if(as >= r) break;
27       }
28
29       // draw random number to choose an appropiate time
30       r=drand48();
31       tau=-log(r) / a0;
32
33       // execute the chosen reaction and increase simulation time
34       switch(mu) {
35          case 0:
36             V--;
37             X++;
38             break;
39          case 1:
40             X--;
41             Y++;
42             break;
43          case 2:
44             Y--;
45             W++;
46             break;
47       }
48       t+=tau;
49    }
50    return 0;
51 }
```

A.2 Next Reaction Method Implementation

The complete source code of my Next Reaction Method Implementation would be too much for the scope of this thesis I therefore only present the fundamental function which implements the basic idea of this algorithm as commented source code. The function below first accesses the binary tree to get the reaction with the earliest time, then executes this reaction and updates all reaction channels that depend on the reaction. Finally the binary tree update function is called to move the updated reactions to their new positions according to their reaction times.

```
1  int global_make_step(struct btree** btree, double* t) {
2     struct dlist* temp;
3     struct reaction* reaction;
4     double aold;
5
6     // get the earliest reaction from the sorted btree == first entry
7     reaction=(*btree)->reaction;
8
9     // if its time is infinity no reaction will occur anymore -> nothing to do
10    if(isinf(reaction->t)) return 1;
11
12    // save the time of the reaction and execute the reaction
13    *t=reaction->t;
14    reaction_step(reaction);
15
16    // get the list of dependend reactions which now
17    // need recalculation of their propensities and time
18    temp=reaction->dlist;
19    while(temp) {
20       aold=temp->reaction->a;
21       // calculate the new propensity for a dependend reaction
22       reaction_calc_prop(temp->reaction);
23       if(temp->reaction!=reaction) {
24          if(isinf(temp->reaction->t)) {
25             // if needed generate a new time from random number
26             temp->reaction->t=reaction_calc_time(temp->reaction, *t);
27          } else {
28             // else just recalculate the time from old and new propensity
29             temp->reaction->t=(aold/temp->reaction->a)*(temp->reaction->t - *t) + *t;
30          }
31       }
32       // also generate a new time for the reaction we executed
33       if(temp->reaction==reaction) {
34          temp->reaction->t=reaction_calc_time(temp->reaction, *t);
35       }
36       if(isnan(temp->reaction->t)) {
37          temp->reaction->t=HUGE_VAL;
38       }
39
40       // update the btree entry for the reaction we just modified
41       // this assures that the btree is always sorted right
42       btree_update(temp->reaction->leaf);
43       temp=temp->next;
44    }
45    return 0;
46 }
```

Bibliography

[And79] R. M. Anderson and R. M. May: *Population Biology of Infectious Diseases: Part I.* Nature **280**:361–367 (1979). 5

[Bro06] Dirk Brockmann, Lars Hufnagel, and Theo Geisel: *The scaling laws of human travel.* Nature **439**:462–465 (2006). 5

[Cor01] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein: *Introduction to Algorithms.* The MIT Press, second edition (2001). 25

[Fer05] Neil M. Ferguson, Derek A.T. Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S. Burke: *Strategies for containing an emerging influenza pandemic in Southeast Asia.* Nature **437**(7056):209–214 (2005). 5

[Gar96] C. W. Gardiner: *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences (Springer Series in Synergetics).* Springer (1996). 5, 10

[Gib00] M. A. Gibson and J. Bruck: *Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels.* J. Phys. Chem. A **104**(9):1876–1889 (2000). 5, 19, 24

[Gil76] D. T. Gillespie: *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions.* Journal of Computational Physics **2**:403–434 (1976). 5, 19

[Gil77] D. T. Gillespie: *Exact Stochastic Simulation of Coupled Chemical Reactions.* The Journal of Physical Chemistry **81**(25):2340–2361 (1977). 5, 19

[Huf04] Lars Hufnagel, Dirk Brockmann, and Theo Geisel: *Forecast and Control of epidemics in a globalized world.* PNAS **101**:15124–15129 (2004). 5

[Ker27] W. O. Kermack and A. G. McKendrick: *A Contribution to the Mathematical Theory of Epidemics.* Proc. Roy. Soc. Lond. **A 115**:700–721 (1927). 5

[Mur02] J. D. Murray: *Mathematical Biology I.* Springer (2002). 5

[Pop82] K. Popper: *Postscript to the Logic of Scientific Discovery, Vol. 3*, chapter Quantum Theory and the Schism in Physics. Hutchinson (1982). 8

[Pre07] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery: *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press (2007). 21

[Thi08] Christian Thiemann: *Panic Reactions and Global Disease Dynamics.* Master's thesis, Georg August University Göttingen (2008). 10

Acknowledgements

First of all I have to thank Theo Geisel for having created such a wonderful institute and for allowing me to participate in it. Thanks also go to:

Dirk Brockmann for supervising me and to spark my interest in the field of epidemological research.

Bernd Blasius from the ICBM Oldenburg for having the initial idea for this thesis and the fruitful discussions about it.

Christian Thiemann. Without him this thesis would have not been possible. Thank you for the many discussions, jokes, tips, corrections, help and all the fish.

Denny Fliegner for managing and caring for the computing cluster all the time.

All the epidemic research people from the institute: Vincent David, Vitaly Belik, Alejandro Morales Gallardo, Benjamin Schwenker and Jan Nagler for constant feedback and input.

All my friends that helped[1] me completing the thesis: Christian Thiemann, Sarah Müller, Christin Kalkert, Katharina Wilhelm and many more.

My parents Barbara and Detlef Brune and all my brothers and sisters, Daniel, Leonora, Magdalena, Gabriel and Ludmila for always supporting me and making all this possible.

All the other people from the institute and especially the MPI Think Tank Band.

The Fast Food Industry for trying to increase my BMI but totally failed to do so.

[1] Help means: Corrections, cookies, cooking, support, the occasional hustling, DVD watching, telling me about the nice weather when I have to work and many other services.