

Pairwise models for non-Markovian epidemics on networks

Ph.D. thesis

ZSOLT VIZI

Supervisors:

János Karsai, Ph.D.

Gergely Röst, Ph.D.

Doctoral School of Mathematics
and Computer Science
University of Szeged, Bolyai Institute

Szeged

2016.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors János Karsai and Gergely Röst for the continuous support of my Ph.D. study and related research, for their patience, motivation and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis.

I am also thankful to István Kiss, whom we have been working for three years with. He motivated me greatly and taught me a lot about disease spread on networks. My sincere thanks also goes to Péter Simon, who suggested the research topic of non-Markovian network epidemics.

I would like to thank my high school teacher, Judit Szilágyi for bringing me back to the marvellous world of mathematics. Her personality and teaching attitude inspired me to fulfill my dream of being an applied mathematician.

I am also grateful for the support of Bolyai Institute and Doctoral School in Mathematics and Computer Science. My sincere thanks also goes to the members of EPIDELAY research group, the Ph.D. students of Bolyai Institute, who encouraged me to go a long way.

Last but not the least, I would like to thank my wife, Kitti Szatmári, my family and my friends for supporting me spiritually throughout writing this thesis and my life in general.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Mathematical framework | 6 |
| 2.1 | <i>SIR</i> epidemic with homogeneous mixing | 6 |
| 2.1.1 | Stochastic approach | 6 |
| 2.1.2 | Deterministic approach | 7 |
| 2.2 | Heterogeneity in interactions: networks | 8 |
| 2.2.1 | Node-level and link-level approximations | 9 |
| 2.2.2 | Stochastic simulation | 13 |
| 2.3 | Non-Markovian stochastic simulations | 15 |
| 2.4 | Functional differential equations | 16 |
| 3 | Reproduction numbers and final size relations | 19 |
| 3.1 | Results for classical <i>SIR</i> model (2.1) | 19 |
| 3.2 | Concept of reproduction numbers for networks | 20 |
| 3.3 | Special cases | 21 |
| 3.4 | Final size relations | 22 |
| 4 | Impact of distribution on disease spread | 24 |
| 4.1 | Gamma distribution | 24 |
| 4.2 | Uniform distribution | 27 |
| 4.3 | Lognormal distribution | 29 |
| 4.4 | General case | 30 |
| 4.5 | Conclusions | 33 |
| 5 | Models with fixed recovery time | 34 |
| 5.1 | Model development | 34 |
| 5.2 | Analytical results | 37 |
| 5.2.1 | First integral | 37 |

| | | |
|----------|---|------------|
| 5.2.2 | Positivity | 38 |
| 5.2.3 | Epidemic outbreak and reproduction numbers | 39 |
| 5.2.4 | Final size relation | 41 |
| 5.3 | Numerical simulations | 45 |
| 5.4 | Conclusions | 47 |
| 6 | General recovery time | 49 |
| 6.1 | The model | 49 |
| 6.2 | Analytical results | 54 |
| 6.2.1 | Positivity | 54 |
| 6.2.2 | Epidemic outbreak and reproduction numbers | 55 |
| 6.2.3 | Final size relation | 57 |
| 6.3 | Numerical methods | 61 |
| 6.4 | Special cases | 68 |
| 6.4.1 | Markovian recovery time: exponential distribution with parameter γ | 68 |
| 6.4.2 | Fixed recovery time σ | 68 |
| 6.4.3 | Gamma distribution with shape $K \in \mathbb{Z}^+$ and rate $K\gamma$ | 69 |
| 6.4.4 | Uniform distribution on interval $[A, B]$ | 71 |
| 6.5 | Conclusions | 72 |
| 7 | Summary | 74 |
| 8 | Összefoglalás | 78 |
| | Appendix | A-1 |

List of Figures

| | | |
|-----|---|----|
| 2.1 | Flowchart of SIR dynamics | 6 |
| 2.2 | Disease spread on a small network | 8 |
| 2.3 | Several outputs of individual-based stochastic simulation for Markovian epidemic on network and the average epidemic curve | 13 |
| 2.4 | Several outputs of individual-based stochastic simulation for non-Markovian epidemic on network and the average epidemic curve | 15 |
| 4.1 | Epidemic curves as explicit stochastic simulations for non-Markovian epidemics with gamma distributed infectious period | 25 |
| 4.2 | Contour lines of \mathcal{R}_0^p as a two-variable function of the parameters of the gamma distribution | 26 |
| 4.3 | Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics with uniformly distributed infectious period | 27 |
| 4.4 | Contour lines of \mathcal{R}_0^p as a two-variable function of the parameters of the uniform distribution | 28 |
| 4.5 | Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics with lognormally distributed infectious period | 29 |
| 4.6 | Contour lines of \mathcal{R}_0^p as a two-variable function of the parameters of the lognormal distribution | 30 |
| 4.7 | Comparison of three epidemic curves after averaging explicit stochastic simulations with three different distributions of recovery times | 33 |
| 5.1 | Simulations of non-Markovian epidemics on network with fixed infectious period | 46 |
| 6.1 | Stochastic and numerical experiments for non-Markovian epidemic with various recovery time distribution | 67 |

1

Introduction

This thesis aims to give a generalisation of deterministic models for network epidemics with Markovian infectious process and non-Markovian recovery process. We develop the models for the fixed and the general case, simulate explicitly the stochastic process for comparison with the solutions of the resulting systems and study the most important features, such as reproduction number and implicitly given final size relation. These results are summarised in three scientific papers of István Kiss, Gergely Röst and the author of this thesis, see [49], [70] and [71]. We consider *SIR* dynamics, but the introduced framework may be applied to more exotic dynamics as well.

Mathematical epidemiology has a long history, going back to the smallpox model of Daniel Bernoulli in 1760. The first fundamental results were developed in the first half of the 20th century and this field is in continuous progression nowadays. Mathematical epidemiology differs from most sciences in availability of experimental validation of models. Experiments are usually impossible and would probably be unethical. This gives great importance to mathematical models as a possible tool for the comparison of strategies to plan for an anticipated epidemic or pandemic, and real-time control of disease outbreak.

The predominant majority of disease models are based on a compartmentalization of individuals or hosts according to their disease status ([48], [5], [7], [33], [69]). The basic models describe the number of individuals (or proportion of the population) that are susceptible to, infected with and recovered from a particular disease (the *SIR models*). Many of the details of the progression of infection are therefore neglected, as are differences in response between individuals. The *SIR* model is appropriate for infectious diseases that confer lifelong immunity, such as measles or whooping cough ([48], [5], [33], [69]). The *SIS* model, where the infected individuals become susceptible again after recovery from the disease, is primarily used for sexually transmitted diseases, such as chlamydia or gonorrhoea, where repeated infections are common ([38], [28]). Many

modifications have been made to this basic framework, usually involving the inclusion of more heterogeneities by further subdividing the S , I and R classification to reflect either more complex pathogen biology ([4], [34]) or greater structure within the host population ([38], [29], [42]).

It has long been acknowledged that the connectivity pattern between individuals in a population is an important factor in determining the properties of a disease spreading process ([10], [62], [65], [20]). Using networks to model disease transmission, where individuals are represented as nodes in a network and the connectivity between individuals is represented by links between the nodes, allowed us to capture a high level of detail of many realistic processes and led to more accurate models, especially when compared to classical compartmental models which operate on the assumption of homogeneous random mixing.

There is a significant amount of research investigating network epidemic models with the aim of understanding how network properties influence the disease spread ([46]). Many different modelling approaches have been introduced, which fall into three broad classes: exact Markovian or state-based models ([77], [80]), individual-based stochastic simulation or micro models ([46]) and deterministic ODE-based macro models ([72], [66], [43], [81], [80]). This classification refers to the scale (e.g. individual level or population level) at which the modelling is being carried out. The links between state-based, micro and macro models are explored in detail by [32].

While networks provide a clear departure from classic compartmental models, the role of simplest macro-level models, i.e. mean-field models remains crucial. These offer us a reliable way to obtain analytical results, such as epidemic threshold ([64], [53]) and final epidemic size [43], which in turn can be used to uncover the interplay between network properties and dynamic processes on networks. Probably the most well-known mean-field model for network epidemics is the degree-based or heterogeneous mean-field model ([64], [65]).

Pairwise approximation - a link-level macro-modelling approach - aims to model the spread of infection on generic networks where higher-order structure has been ignored. Rather than modelling a network of interactions in its entirety, pairwise models, as the name suggests, examine the various types of connected pairs found within a population ([12], [27], [35], [37], [42], [43], [67]). Pairwise models have been used to examine a number of epidemiological issues: fade-out and critical community size for childhood infections ([42]); evolution of pathogen virulence ([12]); spread and control of sexually transmitted diseases in heterogeneous populations ([23], [27]) and modelling epidemics on adaptive networks ([35], [79]). This modelling scheme can be applied in study of social interactions

([21]) and ecological systems ([37]).

While networks offer an accurate representation of contact patterns, many network epidemic models consider only Markovian epidemics with both the infectious and recovery processes being memoryless. However, empirical observations show that the Markovian framework is not satisfactory in describing temporal statistics, such as time intervals between discrete, consecutive events. Examples include inter-trade durations in financial markets ([73]), socio-networks ([55]), or contacts between individuals being dynamic ([59]). In the context of epidemiology, the period of infectiousness has a key role ([44], [54]). The empirical distribution of infectious periods of various diseases is often approximated by log-normal and gamma (smallpox [24], [63]), fixed-length (measles [6]) or Weibull distributions (ebola [19]).

Nowadays, there is a renewed interest in non-Markovian processes, such as epidemics on networks ([11], [17], [18], [40], [47], [58], [82]), random walks ([39]) and temporal networks ([59]). Several different approaches exist to model non-Markovian epidemics on networks. These are largely guided by the choice of the model and variables to be tracked. Notable examples include the message passing approach, often referred to as the cavity model ([41], [85], [86]), and the percolation based approach ([47], [56], [57], [61]). While the latter only offers information about the final state of the epidemic, the former describes the temporal evolution of the epidemic. Generalisations of the pairwise model to gamma-distributed infectious periods have also been proposed and this has been developed for both homogeneous and heterogeneous networks ([75], [76]), non-Markovian *SIS* model is considered in [17, 82].

This thesis is motivated by this renewed interest in non-Markovian processes and aims to extend the pairwise model from Markovian to non-Markovian epidemic dynamics where the infection process remains Markovian but the infectious period is taken from an arbitrary distribution. Unfortunately, the reliable tools and mathematical machinery of Markovian theory do not translate directly to modelling and analysis of non-Markovian systems, and this leads to many significant challenges.

In Chapter 2, we present the fundamental models of mathematical epidemiology and network epidemics to give a complete picture about the preliminary results of this relevant topic. We start with the classical *SIR* model and introduce the stochastic and deterministic approaches considering Markovian infection and recovery processes and homogeneous mixing in the population. As a possible improvement, we recall the results of Sharkey et. al [74] about exact master equations for stochastic process on networks. By introducing the node-level and link-level approximations, we write down the mean-field and pairwise models for the simplest *SIR* type epidemic process. These equations

are the starting points of this research. In the last part of this chapter, we introduce the most commonly used simulation algorithms for stochastic processes, such as synchronously and asynchronously updating paradigms. In the last section, we summarise briefly two possible simulation approaches for non-Markovian systems: the one we use through this thesis and a generalised version of Gillespie algorithm from Boguna et al [11].

The most important quantities derived for *SIR* dynamics are the reproduction number and the implicitly given epidemic final size. We recall the classical results for *SIR* model with mass action and introduce the general concept of reproduction numbers in Chapter 3. We obtain analytical formulae for the basic reproduction number of mean-field type systems and the newly introduced concept of pairwise reproduction number for link-level models. These definitions were published in our first paper about non-Markovian epidemics in [49]. As an application, we calculate these measurements for exponentially distributed and fixed recovery times. In the last section, we summarise the general functional forms of final size relations for both type of models, highlighting the relation between these equations.

As it is expected, the distribution of the recovery time has a great impact on the disease spread and knowing as much as possible about it is a key factor in understanding the dynamics. In Chapter 4 we illustrate the behaviour of the epidemic process for the most widely observed distributions, such as gamma, uniform and lognormal distributions and study the interesting phenomenon, that the pairwise reproduction number is decreasing if the variance of the recovery time is increasing. We can analytically prove this statement for the case of gamma and uniform distribution and find numerical evidence for lognormal distribution. We prove a theorem for the general case with a condition for the integral functions of cumulative distribution functions and show an example for its application. As a conclusion, we see that this result holds within one family of distributions, but different types of distributions with the same mean and variance may produce significantly different pairwise reproduction numbers.

The first generalisations of mean-field and pairwise models are presented in Chapter 5, where we consider fixed recovery time and develop the models from first principles. We investigate the basic properties, such as positivity of solutions of the resulting delayed differential equations and reduce the dimensionality by finding a first integral to the pairwise system. We explore the relation between the occurrence of an epidemic and reproduction numbers by investigating linearisation of models and the associated characteristic roots. The main results are the derivation of implicit equations for final epidemic size. In the last part of this chapter, we test the validity of our models with nu-

merical simulations and present a possible extension of the model for more heterogeneous networks.

After all these studies, the question arises naturally, how to develop a model for arbitrarily distributed recovery time. First we consider a hyperbolic system of partial differential equations, where the population of infective nodes and links are structured by the time elapsed since the beginning of the infectious period, also called as age since infection. By solving the partial differential equations, the model is transformed into a system of integro-differential equations, which is analysed both from a mathematical and numerical point of view. The associated lengthy calculations are included in the first section of Chapter 6. The further sections follow the scheme of Chapter 5: after the positivity results and performing calculations for showing the relation between epidemic outbreak and reproduction numbers, the final size relations are obtained for mean-field and pairwise models through several pages of calculations. Since there is no general solver for integro-differential equations, we implement an algorithm adapting collocation schemes of [15]. In the last part, we investigate the special choices for distributions and from the general model we get back the Markovian models (for exponential distribution), systems for fixed case in Chapter 5 and system of ordinary differential equations published in [75] (for gamma distribution).

2

Mathematical framework

2.1 *SIR* epidemic with homogeneous mixing

During the Great Plague in London in 1665-1666, the disease appeared suddenly, grew in intensity and then disappeared, leaving the 85 % of the population untouched. One of the early triumphs of mathematical epidemiology was the formulation of the model by Kermack and McKendrick [48], whose predictions are very similar to this behaviour, observed in countless epidemics. The model is based on relatively simple assumptions on the rates of flow between different classes of members of the population.

We will work with compartmental models, i.e. we divide the population studied into several disjoint classes. The most used classes are *S* (susceptible), *I* (infected), *R* (recovered), *E* (exposed) and *A* (asymptomatic) compartments. In the case of *SIR* models, the susceptible members of the population are affected by the disease due to a contact with an infected individual and recover at the end of the infectious period (this duration may be called *recovery time* in this simplified model), see Fig. 2.1.

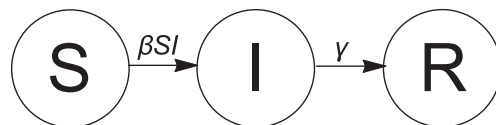


Figure 2.1: Flowchart of SIR dynamics

2.1.1 Stochastic approach

In a stochastic model we consider, that each individual may have only one state *S*, *I* or *R*. The state of the system is a triple (x, y, z) , where the elements are the number of individuals in state *S*, *I* and *R*, respectively. The probability of transition from (x, y, z)

to $(x-1, y+1, z)$ is depending on x and y and changing to $(x, y-1, z+1)$ is proportional to y . Clearly, for a given population and considering precisely given rules for transitions (see [3]), an epidemic may be considered as an recorded output of the projection of the stochastic model and thus may differ for repeated projections.

Throughout this thesis, we neglect the effects of births and deaths, thus population size N is assumed to be constant, because the duration of an epidemic is measured on a shorter temporal scale, thus the demographic changes can be neglected. Initially, we have a fully susceptible population, and a small number of infected individuals is added at time $t = 0$.

One of the most important assumptions, that the population is homogeneously mixed, i.e. every individual interacts identically with every other. This is called the *law of mass action*. Events occur at discrete time steps, furthermore the infection and recovery occur with constant probability at each time step. This latter assumption is called *Markovian property*.

2.1.2 Deterministic approach

If a population is large, however, we may consider probabilities of changing between states as rates, and derive differential equations. These describe the expected proportion of the (large) population in each state at each time, as the continuum limit of the stochastic processes. The usefulness of a differential equation approach is that it is deterministic; it tells us something about a typical epidemic, whereas any one projection output of the stochastic model does not give insight into how the next epidemic might arise.

We can introduce variables $S(t), I(t), R(t)$ for the number/proportion of susceptible, infected and recovered individuals, respectively at time t . Therefore, we have the following system:

$$\begin{aligned}\dot{S}(t) &= -\beta S(t)I(t) \\ \dot{I}(t) &= \beta S(t)I(t) - \gamma I(t) \\ \dot{R}(t) &= \gamma I(t),\end{aligned}\tag{2.1}$$

where overdot represents the derivative with respect to time. We can interpret the equations as follows. First, the susceptible individuals may only be depleted, because there is no mechanism allowing infected or recovered to re-enter the S state. Next, the law of mass action implies that this depletion happens at a constant rate β proportional to $S(t)I(t)$ and this decrease in $S(t)$ causes replenishment in $I(t)$. Finally, the depletion from $I(t)$ happens by the recovery at the constant rate γ proportional to $I(t)$ and recovered population fills at this rate. Naturally, the initial conditions satisfy $S(0), I(0), R(0) \geq 0$

and $S(0) + I(0) + R(0) = N$. Adding up equations in (2.1), we have $\dot{S} + \dot{I} + \dot{R} = \dot{N} = 0$, thus the population has constant size over time.

This model appears as a very special case in [48], which was the basis for research in mathematical epidemiology during the last 90 years. In Chapter 3, we define the key parameters, e.g. basic reproduction number and study the most important properties of model (2.1).

2.2 Heterogeneity in interactions: networks

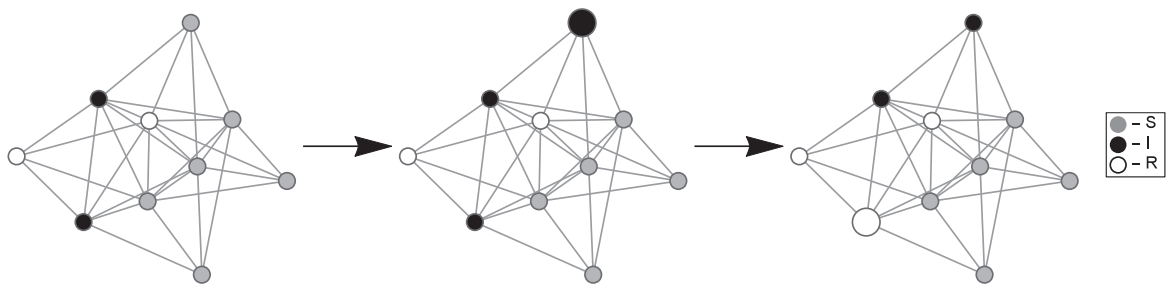


Figure 2.2: Disease-spread on a small network. Grey, black and white nodes are susceptible, infected and recovered, respectively. The contact network is generated in Wolfram Mathematica 10 with Barabasi-Albert graph distribution. In the second and third figure, a new infected and recovered node (highlighted by enlarging) appears, respectively.

In general, a disease outbreak starts with a small number of infective individuals and the transmission of infection is a stochastic event depending on the pattern of contacts between members of the population. For a more realistic modelling approach we should take this pattern into account.

Determining a complete mixing network requires knowledge of every individual in a population and every relationship between individuals. For all but the smallest populations, this is an impractically time-consuming task. The networks commonly used are generated by computer simulation to conform to several observed social characteristics. Several forms of computer-generated networks have been studied in the context of disease transmission. Each of these idealized networks can be defined in terms of how individuals are distributed in space (which may be geographical or social) and how connections are formed, thereby simplifying and making explicit the many and complex processes involved in network formation within real populations.

In random networks, the spatial position of individuals is irrelevant, and connections are formed at random. In the most analytically tractable version of the random network,

each individual has a fixed number of contacts (called degree in graph theory) through which infection can spread. In special case, when all degrees are equal, we get a *homogeneous random network*. An alternative formulation of the random network is to connect any two nodes with probability p . This leads to the *Erdős-Rényi network model* ([25]). Beside random networks, where the connections are unstructured, lattices ([68]), small-world- ([84]), spatial- ([45]) and scale-free networks ([8]) are the most popular network types in applications.

Any given contact network, whether generated using a theoretical algorithm or constructed from real-world observations, can be described by a simple adjacency matrix which is then used in the construction of exact Kolmogorov/ master equations, which describe the probability of the population being in each and every possible configuration. In the following, let us now recall the notations and models from [74].

We consider the general stochastic system Γ whose state is denoted by $\Gamma^\alpha, \alpha \in \{1, 2, \dots, 3^N\}$. The evolution in the state space can be described by a continuous time Markov-process. The time-dependent probabilities

$$P_{(\alpha)}(t) = \mathbb{P}(\text{system } \Gamma \text{ is in state } \Gamma^\alpha \text{ at time } t)$$

can be calculated from the master equations:

$$\dot{P}_{(\alpha)}(t) = \sum_{\beta=1}^{3^N} \left[R^{\beta,\alpha} P_{(\beta)}(t) - R^{\alpha,\beta} P_{(\alpha)}(t) \right], \quad (2.2)$$

where $R^{\beta,\alpha}$ denotes the Poissonian transition rate from state Γ^β to state Γ^α . The solution of these master equations leads to an exact description of the dynamics of the full system. Since the system size of the master equations scales exponentially with population size, even with modern advances in computing power we are unable to numerically integrate these equations for realistic populations.

2.2.1 Node-level and link-level approximations

We suppose that within the system Γ , there exist well-defined smaller systems which we refer to as subsystems. One way to reduce the dimensionality of system (2.2) to split Γ into a set of Z coupled subsystems $\psi_i, i = 1, 2, \dots, Z$.

For each subsystem ψ_i , we can write master equations to describe the state probabilities $P_{(i,a)}(t) = \mathbb{P}(i\text{th subsystem } \psi_i \text{ is in state } \psi_i^a \text{ at time } t)$:

$$\dot{P}_{(i,a)}(t) = \sum_{\beta=1}^{m_i} \left[R_i^{b,a} P_{(i,a)}(t) - R_i^{a,b} P_{(i,b)}(t) \right], \quad (2.3)$$

where the indices a and b denote two of the m_i possible states of the i th subsystem and R_i denotes the matrix of transition rates between states for the i th subsystem and is, in general, dependent on the states of the other subsystems. The number of equations $m = \sum_{i=1}^Z m_i$ can be far smaller than the 3^N master equations for the complete system, but at the cost that these equations are not closed.

An obvious subsystem is formed by the nodes themselves. For probabilities

$$P_i^X(t) = \mathbb{P}(\text{node } i \text{ is in state } X \text{ at time } t),$$

where $X \in \{S, I, R\}$, Eq. (2.3) becomes

$$\begin{aligned} \dot{P}_i^S(t) &= -R_i^{S,I} P_i^S(t) \\ \dot{P}_i^I(t) &= R_i^{S,I} P_i^S(t) - R_i^{I,R} P_i^I(t). \\ \dot{P}_i^R(t) &= R_i^{I,R} P_i^I(t). \end{aligned} \quad (2.4)$$

Similarly, we can introduce the following notations:

$$\begin{aligned} P_{i,j}^{X,Y}(t) &= \mathbb{P}(\text{node } i \text{ is in state } X, \text{ node } j \text{ is in state } Y \text{ at time } t), \\ P_{i,j,k}^{X,Y,Z}(t) &= \mathbb{P}(\text{node } i \text{ is in state } X, \text{ node } j \text{ is in state } Y, \text{ node } k \text{ is in state } Z \text{ at time } t), \end{aligned}$$

where $X, Y, Z \in \{S, I, R\}$. It can be seen, if the contact network is represented by an undirected graph $\mathbb{G} = (V, E)$ with vertex/node set V with N vertices, edge/link set E and the adjacency matrix, which is an $N \times N$ matrix $G = (G_{j,k})$ defined as:

$$G_{j,k} = \begin{cases} 1, & \text{if vertices } j \text{ and } k \text{ are linked,} \\ 0, & \text{otherwise.;} \end{cases}$$

we have

$$\begin{aligned} R_i^{S,I} &= \sum_{j=1}^N \tau G_{j,i} \frac{P_{j,i}^{I,S}}{P_i^S} \\ R_i^{I,S} &= \gamma. \end{aligned}$$

where τ denotes the transition rate, γ denotes the recovery rate. It is not always necessary to determine all time-dependent probabilities, since the expected values of the number/proportion of susceptible, infected or recovered nodes are equally valuable. These expected values at time t are denoted by $[S](t)$, $[I](t)$ and $[R](t)$ respectively and can be expressed as follows,

$$[X](t) = \sum_{i=1}^N P_i^X(t), \quad X \in \{S, I, R\}.$$

Similar expressions can be obtained for the expected number of $X - Y$ links and $X - Y - Z$ triplets,

$$\begin{aligned} [XY](t) &= \sum_{i=1}^N \sum_{j=1}^N G_{i,j} P_{i,j}^{X,Y}(t), \\ [XYZ](t) &= \sum_{i,j,k,i \neq k} G_{i,j} G_{j,k} P_{i,j,k}^{X,Y,Z}(t), \end{aligned} \quad (2.5)$$

where $X, Y, Z \in \{S, I, R\}$. It follows from (2.4) that:

$$\begin{aligned} [\dot{S}](t) &= -\tau[SI](t) \\ [\dot{I}](t) &= \tau[SI](t) - \gamma[I](t) \\ [\dot{R}](t) &= \gamma[I](t). \end{aligned} \quad (2.6)$$

The interpretation is clear, since $[S](t)$ is depleted at constant rate $\tau[SI](t)$ (i.e. proportional to the expected number of $S - I$ links), which is a rate of increase of infected nodes. Infection is, as in the previous section, depleted at rate $\gamma[I](t)$, which is the rate of increase in $[R](t)$. We observe, that the system (2.6) is not closed, we have variable $[SI](t)$ without any governing equation. A possible way to solve this problem to close the system (2.6) at the level of nodes. For elaborating this, we need a formula for $[SI](t)$, which depends only on the node-level variables $[S](t)$ and $[I](t)$. Notice, that the expected number of $X - Y$ links in a homogeneous network with uniform degree distribution $\langle k \rangle = n$ (which is an n -regular graph) is

$$[XY] \approx n[X] \frac{[Y]}{N}. \quad (2.7)$$

Using (2.7), we obtain the node-level system, called *mean-field model*:

$$\begin{aligned} [\dot{S}](t) &= -\tau \frac{n}{N} [S][I](t) \\ [\dot{I}](t) &= \tau \frac{n}{N} [S][I](t) - \gamma[I](t) \\ [\dot{R}](t) &= \gamma[I](t), \end{aligned} \quad (2.8)$$

which has the same form as (2.1).

On the other hand, we can introduce the subsystem of links as an extension of node-level view. In this case, the complete reduced master equations for the link dynamics

are:

$$\begin{aligned}
\dot{P}_{j,i}^{S,S}(t) &= - \sum_{k,k \neq i} \tau G_{k,j} P_{k,j,i}^{I,S,S}(t) - \sum_{k,k \neq j} \tau G_{k,i} P_{j,i,k}^{S,S,I}(t), \\
\dot{P}_{j,i}^{I,S}(t) &= \sum_{k,k \neq i} \tau G_{k,j} P_{k,j,i}^{I,S,S}(t) - \sum_{k,k \neq j} \tau G_{k,i} P_{j,i,k}^{I,S,I}(t) - \tau G_{j,i} P_{j,i}^{I,S}(t) - \gamma P_{j,i}^{I,S}(t), \\
\dot{P}_{j,i}^{R,S}(t) &= - \sum_{k,k \neq j} \tau G_{k,i} P_{j,i,k}^{R,S,I}(t) + \gamma P_{j,i}^{I,S}(t), \\
\dot{P}_{j,i}^{I,I}(t) &= \sum_{k,k \neq i} \tau G_{k,i} P_{k,j,i}^{I,S,I}(t) + \sum_{k,k \neq j} \tau G_{j,i} P_{j,i,k}^{I,S,I}(t) - \tau G_{i,j} P_{j,i}^{S,I}(t) - 2\gamma P_{j,i}^{I,I}(t), \\
\dot{P}_{j,i}^{R,I}(t) &= \sum_{k,k \neq j} \tau G_{k,i} P_{j,i,k}^{R,S,I}(t) + \gamma P_{j,i}^{I,I}(t) - \gamma P_{j,i}^{R,I}(t), \\
\dot{P}_{j,i}^{R,R}(t) &= \gamma P_{j,i}^{R,I}(t) + \gamma P_{j,i}^{I,R}(t)
\end{aligned} \tag{2.9}$$

Due to the symmetry of the graph, we have $G_{i,j} = G_{j,i}$, so $[XY] = [YX]$. Furthermore, applying formulae in (2.5) for $P_{j,i}^{S,S}$, $P_{j,i}^{I,S}$ and $P_{j,i}^{I,I}$ expressions in (2.9), we obtain

$$\begin{aligned}
\dot{[S]}(t) &= -2\tau[SSI](t), \\
\dot{[SI]}(t) &= \tau[SSI](t) - \tau[ISI](t) - \tau[SI](t) - \gamma[SI](t), \\
\dot{[II]}(t) &= 2\tau[ISI](t) + 2\tau[SI](t) - 2\gamma[II](t).
\end{aligned} \tag{2.10}$$

We observe that equations in (2.10) depend on the variables of triplets. We will approximate the variable $[XYZ]$ by a function of $[XY]$, $[YZ]$ and $[Y]$. The detailed explanation of the following moment-closure approximation formula can be found in [18] and [43]:

$$[XYZ] = \frac{n-1}{n} \frac{[XY][YZ]}{[Y]}. \tag{2.11}$$

Applying (2.11), we can obtain the following self-consistent system by including the equations for variables of nodes and links:

$$\begin{aligned}
\dot{[S]}(t) &= -\tau[SI](t) \\
\dot{[I]}(t) &= \tau[SI](t) - \gamma[I](t) \\
\dot{[SS]}(t) &= -2\tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)}, \\
\dot{[SI]}(t) &= \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)[SI](t)}{[S](t)} - \tau[SI](t) - \gamma[SI](t).
\end{aligned} \tag{2.12}$$

System (2.12) is the simplest link-level approximation, called *pairwise SIR model*. The number of links in an n -homogeneous network with N nodes is $Nn/2$, thus the sum of variables $[XY]$, $X, Y \in \{S, I, R\}$ equals to Nn . This gives a condition for initial values

$[XY]_0$. In our simulations in this thesis, we consider a small number of infecteds in a fully susceptible population, thus the initial values $[SS]_0 = \frac{n}{N}[S]_0^2$, $[SI]_0 = \frac{n}{N}[S]_0[I]_0$ are reasonable for (2.12).

The models (2.8) and (2.12) are describing epidemics with *Markovian* recovery, because the recovery process is assumed to be Markovian. The main goal of this dissertation is giving a generalisation for these fundamental models by relaxing the Markovian assumption for the recovery process.

2.2.2 Stochastic simulation

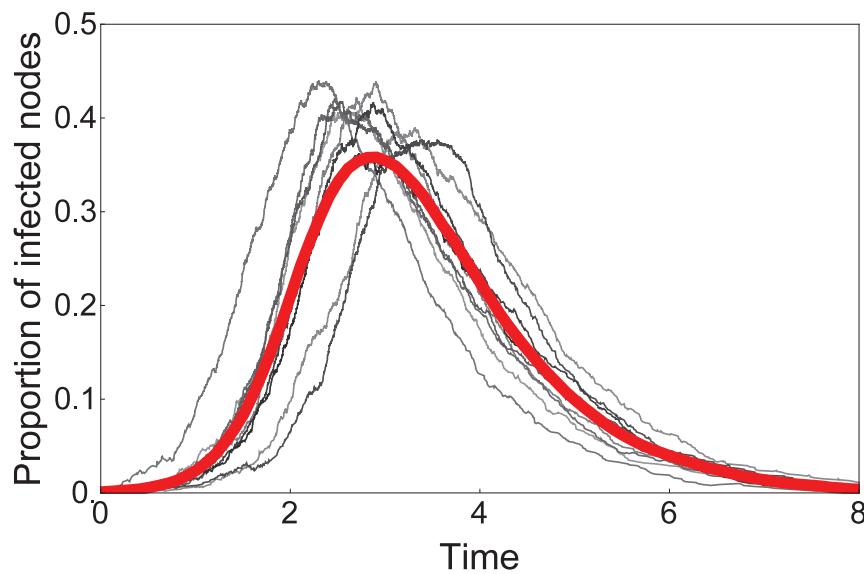


Figure 2.3: Several outputs of individual-based stochastic simulation for Markovian epidemic on network and the average epidemic curve. We consider homogeneous networks with $N = 1000$ nodes and degree $n = 15$. The initial number of susceptibles is $[S]_0 = 999$, the transmission rate is $\tau = 0.3$ and recovery rate is $\gamma = 1$. The thick curve shows the mean of 100 simulations.

As an alternative for solving master equations, individual-based stochastic simulations can be performed on the full network. Taking the average of a large number of such simulations remains the most viable way to explore dynamics on large, complex networks (see Fig. 2.3). Unfortunately, all information about rare events is lost in the averaging process, whereas if the master equation could be solved, the probability of any possible event could be calculated.

A simple technique for simulating stochastic systems is the synchronously updating implementation. Here, discrete-time simulations can be performed by fixing a time step Δt and, at every step, allowing every possible event to occur with a probability calculated

from the various disease parameters, the current state of the network and the magnitude of Δt , and the state of the system is updated at each time step. An example from this type of simulations is the *tau-leaping method*.

Synchronously updating simulations can be very efficient to implement in terms of coding complexity. However, allowing multiple events to occur simultaneously is not in keeping with the Markovian nature of stochastic models and hence Δt needs to be small enough to avoid this happening. Moreover, in general synchronous updating approach may generate non-realistic behaviour and strange patterns ([26]).

An alternative simulation approach is the asynchronously updating paradigm. In general, this technique allows at most only one event at each time step, which simulates better the continuous-time phenomena. For asynchronously updating algorithms with time step Δt_{async} , where inter-event time is fixed and synchronous version with time step Δt_{sync} allowing n events, we have $\Delta t_{sync} \approx n\Delta t_{async}$.

A special type of asynchronously updating algorithms with dynamically changing inter-event times is known as Gillespie algorithm [30], whereby the time to next event, T , is an exponentially distributed random variable chosen from an exponential distribution parametrised by R_{total} , where R_{total} is the rate of all possible transitions given the current status of all individuals. It follows that working out R_{total} amounts to summing all infection and recovery rates across the whole network. As the inter-event time is directly related to the total rate, large rates result in small inter-event times. Once the time to next event is determined, a single event out of all possible is chosen at random but proportionally to its rate.

In detail, assuming Markovian infection and recovery process, the Gillespie algorithm for simulating network epidemics has the following steps:

1. *Initialization.* Set initial time, states of the nodes and the associated rates based on the state of the node and the neighbouring nodes.
2. *Sample an inter-event time.* Generate an exponentially distributed inter-event time, where the parameter is the reciprocal of the sum of all rates.
3. *Sample a node.* Choose a node to be updated by generating a uniformly distributed random number and using the cumulative sums of rates.
4. *Update.* Update time, the state of the node and the rates calculated to the nodes.

It is clear, that the Gillespie algorithm can be applied to system of discrete Markovian stochastic processes. The algorithm takes advantage of the theory of superposition of a fixed number of renewal processes.

2.3 Non-Markovian stochastic simulations

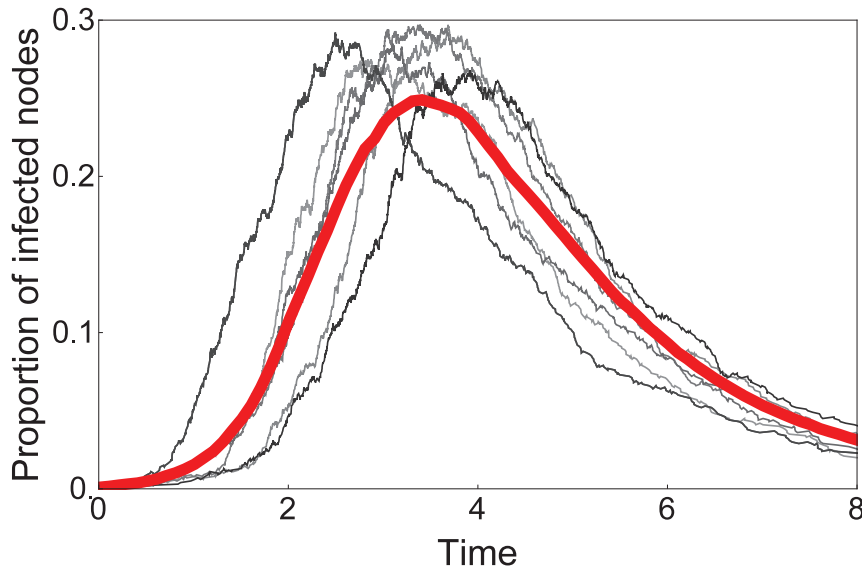


Figure 2.4: Several outputs of individual-based stochastic simulation for non-Markovian epidemic on network and the average epidemic curve. We consider Weibull distributed recovery time with scale $\lambda = 1$ and shape $k = 1$ on homogeneous networks with $N = 1000$ nodes. The initial number of susceptibles is $[S]_0 = 999$ and infection rate is $\tau = 0.3$. The thick curve shows the mean of 100 simulations.

Since the Gillespie algorithm builds on the memorylessness property of the inter-event times, the generalisation of this algorithm is a challenging task. In the following, we describe the algorithm we used in the simulations, which was implemented in *Wolfram Mathematica* 10. The code highly exploits the new developments of the previous version 9, namely, the built-in functions for generating arbitrary random networks. The program is very flexible for further extension of the stochastic model, such as dynamically changing, more realistic networks.

The main program executes the event-based simulation, where waiting times for all possible events are generated from appropriate distributions. During an update the event with the smallest waiting time is executed followed by the necessary update of the waiting times of events affected by the most recent change. The two types of update are implemented in two modules, which update the states and waiting times according to the type of the event. Several outputs of a sample run and the average epidemic curve can be seen in Fig. 2.4. For more details, see Appendix.

The simulation framework for non-Markovian processes has also been improved by Boguñá et al. [11], by proposing an equivalent of the efficient Gillespie algorithm for Poisson processes. In this paper, they describe an algorithm simulating statistically

independent discrete stochastic processes with arbitrary inter-event distribution. This approach stores the time elapsed since the last occurrence of each process up to a given point in time and uses these values and the associated survival and density functions to calculate the time until the next event. The nMGA ('non-Markovian Gillespie algorithm') is an approximation of the exact simulations assuming that $N \gg 1$, but uses the list of elapsed times and survival functions.

2.4 Functional differential equations

In this thesis, we introduce deterministic models for non-Markovian *SIR* network epidemics, which fall in a more general class of differential equations, namely delay and integro-differential equations. Delay differential equations are widely used in infectious disease modelling, applied for malaria ([60]), influenza ([50]) or in immuno-epidemiology ([9]). In population dynamics, Volterra-type integral and integro-differential equations are applied for modelling e.g. predator-prey systems ([83]).

A general form of a delay differential equation (DDE) is

$$\dot{x}(t) = f(t, x_t),$$

where $x_t(\theta) = x(t + \theta)$, $-r \leq \theta \leq 0$ and $f : \mathbb{R} \times C([-r, 0], \mathbb{R}^n) \rightarrow \mathbb{R}^n$ is a given continuous function. One of the most important difference compared to ordinary differential equations (ODE), that for a DDE we have to prescribe an initial function $\phi(\theta) \in C([-r, 0], \mathbb{R}^n)$ to define an initial-value problem.

If $f(t, x_t) = f(t, x(t), x(t - r))$, where $r > 0$ constant, we have a differential equation with *discrete delay*. A term like

$$\int_{t-r}^t k(t-s)x(s)ds = \int_0^r k(s)x(t-s)ds,$$

where $0 \leq r < \infty$ is referred to as a *distributed delay*. On the one hand, if we want to define a dynamical system for an autonomous DDE, the state space will be clearly $C([-r, 0], \mathbb{R}^n)$, which is an infinite dimensional Banach space, compared to an ODE, where the state space is finite dimensional. On the other hand, considering the simplest DDE

$$\dot{x}(t) = -x(t - 1),$$

and deriving the associated characteristic equation

$$\lambda + e^{-\lambda} = 0,$$

we have infinitely many characteristic roots and linearly independent solutions of the form $e^{\lambda t}$. An excellent introduction about DDEs with applications in mathematical biology can be found in [36].

Clearly, the stability definitions can be extended to DDEs. We consider the system of delay differential equations

$$\begin{aligned}\dot{x}(t) &= f(t, x_t), \\ x_\sigma &= \phi\end{aligned}\tag{2.13}$$

where $f(t, 0) = 0, t \in \mathbb{R}$, thus $x(t) \equiv 0$ is a solution. The solution $x = 0$ is *stable*, if for any $\sigma \in \mathbb{R}$ and $\epsilon > 0$, there exists $\delta = \delta(\sigma, \epsilon) > 0$ such that $\phi \in C([-r, 0], \mathbb{R}^n)$ and $\|\phi\| < \delta$ implies that $x_t(\sigma, \phi) < \epsilon$, where $\|\phi\| = \sup\{|\phi(\theta)| : -r \leq \theta \leq 0\}$ and $x(t, \sigma, \phi)$ is the solution of (2.13). It is *asymptotically stable*, if it is stable and if there exists $b(\sigma) > 0$ such that, whenever $\phi \in C([-r, 0], \mathbb{R}^n)$ and $\|\phi\| < b(\sigma)$, then $x(t, \sigma, \phi) \rightarrow 0, t \rightarrow \infty$. Finally, $x = 0$ is *unstable*, if it is not stable.

An *integral equation* is an equation in which the unknown function $x(t)$ appears under an integral sign, e.g.,

$$f(t) = h(t, x(t)) + \int_a^{b(t)} k(t, s, x(s))ds, \quad t \geq a,$$

where $h(t), f(t), b(t)$ and $k(t, s, x)$ are given functions and we wish to determine $x(t)$. The function $k(t, s, x)$ is called the kernel of the integral equation. A first-order *integro-differential equation* involves both integral and derivative of the unknown function, e.g.

$$\dot{x}(t) = h(t, x(t)) + \int_a^{b(t)} k(t, s, x(s))ds, \quad t \geq a.$$

An equation is *Volterra-type*, if $b(t) = t$ and *Fredholm-type*, if $b(t) \equiv b$, where b is constant. An integral/integro-differential equation is *of the first kind*, if $h \equiv 0$ and *of the second kind* otherwise. In this thesis, we will work with Volterra integro-differential equations of the second kind

$$\dot{x}(t) = h(t, x(t)) + \int_0^t k(t, s, x(s))ds, \quad t \geq 0,$$

and investigate a *convolution-type linear Volterra integral equation of the first kind* has the form

$$x(t) = \int_0^t x(t-s)K(s)ds, \quad t \geq 0.$$

Despite the fact, that for the integral/integro-differential equations above only an initial value at the starting time point is enough to determine an initial-value problem, the simplest linear equations have infinitely many solutions in the form of $e^{\lambda t}$. However, it is possible to specify the unknown function on an initial interval, which may influence greatly

the solutions of the equation. More details about Volterra integral/integro-differential equations can be found in [16]. The stability definitions for initial-value problems are similar to definitions in ODE theory (prescribed initial value) and DDE theory (for initial functions).

3

Reproduction numbers and final size relations

An epidemic, which acts on a short temporal scale, may be described as a sudden outbreak of a disease that infects a substantial portion of the population in a region before it disappears. Epidemics usually leave many members untouched. The number of untouched individuals appears in the final size relation, which gives a relationship between the size of the epidemic (number of members of the population who are infected over the course of the epidemic) and the associated reproduction number.

Reproduction numbers play a crucial role in mathematical epidemiology and are defined as the expected number of secondary infections caused by a ‘typical’ infected individual during its infectious period when placed in a fully susceptible population, which is a definition understood at the level of individuals ([22]).

3.1 Results for classical *SIR* model (2.1)

The most important results for *SIR* models are the explicit formula of basic reproduction number and an implicit equation for the final epidemic size containing R_0 . For the classical *SIR* model (2.1), it is easy to see, that $I_\infty = \lim_{t \rightarrow \infty} I(t) = 0$. Epidemic may occur, if $I'(0) > 0$, which is equivalent to condition $R_0 := \frac{\tau S_0}{\gamma} > 1$. Here, R_0 is the basic reproduction number associated to model (2.1) and can be interpreted as the number of newly infected individuals per unit time (τS_0) over the average infectious period ($\frac{1}{\gamma}$). Furthermore, we can reduce dI/dS and integrate it to obtain

$$\ln\left(\frac{S_\infty}{S_0}\right) = R_0\left(\frac{S_\infty}{S_0} - 1\right).$$

We will use the notation $s_\infty = \frac{S_\infty}{S_0}$. Clearly, attack rate is $1 - s_\infty$. (Technically, the attack rate should be called an attack *ratio*, since it is dimensionless and is not a rate.)

Using these formulae, we have

$$\ln s_\infty = R_0 (s_\infty - 1). \quad (3.1)$$

This equation is called *final size relation* and gives an implicit equation for the proportion of remaining individuals after the disease outbreak. Clearly, larger the R_0 , smaller the s_∞ (thus larger the attack rate) we have, which is in harmony with the meaning of basic reproduction number.

3.2 Concept of reproduction numbers for networks

In the following, we introduce a general concept for the reproduction number associated to the mean-field and pairwise models. We introduced these definitions in [49] and use these concepts and results in our further papers in this topic ([70], [71]). The pairwise model is written at the level of links and describes the dynamics of susceptible ($S - S$) and infected ($S - I$) links. This leads to the definition of a new type of reproduction numbers, which we call *pairwise reproduction number*. More precisely, we distinguish the following two useful quantities:

- (a) the *basic* reproduction number is the expected lifetime of an I **node** multiplied by the number of newly infected **nodes** per unit time (denoted by \mathcal{R}_0);
- (b) the *pairwise* reproduction number is the expected lifetime of an $S - I$ **link** multiplied by the number of newly generated $S - I$ **links** per unit time (denoted by \mathcal{R}_0^p).

The expected life time of an infectious node is the expected value of a random variable \mathcal{I} corresponding to the distribution of the length of infectious periods. In contrast, an $S - I$ link can be removed either due to the recovery of the I node or the infection of the S node. Therefore, the expected lifetime of the $S - I$ link is the expected value of the minimum of two random variables. If we assume that the transmission process of infection along such a link has density function f_i with survival function ξ_i , and the process of recovery has density function $f_{\mathcal{I}}$ with survival function $\xi_{\mathcal{I}}$, then, denoting by Z the random variable defined by the lifetime of an $S - I$ link, we have

$$\mathbb{E}(Z) = \int_0^\infty t (f_i(t)\xi_{\mathcal{I}}(t) + f_{\mathcal{I}}(t)\xi_i(t)) dt. \quad (3.2)$$

From the assumption that the infection time along $S - I$ links is exponentially distributed (i.e. $f_i(t) = \tau e^{-\tau t}$, $\xi_i(t) = e^{-\tau t}$), the number of newly infected nodes per unit

time in the mean-field and pairwise model are $\frac{n}{N}\tau[S]_0$ and $\tau\frac{n-1}{n}\frac{[SS]_0}{[S]_0} = \tau\frac{n-1}{N}[S]_0$, respectively, where we used the approximation

$$[SS]_0 = \frac{n}{N}[S]_0^2, \quad (3.3)$$

which is consistent with the assumption of introducing small amount of infected nodes at $t = 0$. Moreover, we will see in Ch. 6, that Eq. (3.3) comes automatically from the general model.

To calculate the expected lifetime of an $S - I$ link, if the infection is Markovian and the recovery is arbitrary with density function $f_{\mathcal{I}}(t)$ and survival function $\xi_{\mathcal{I}}(t)$, we integrate by parts and we obtain

$$\begin{aligned} \mathbb{E}(Z) &= \int_0^\infty t (f_i(t)\xi_{\mathcal{I}}(t) + f_{\mathcal{I}}(t)\xi_i(t)) dt = \int_0^\infty t (\tau e^{-\tau t}\xi_{\mathcal{I}}(t) + e^{-\tau t}f_{\mathcal{I}}(t)) dt \\ &= \int_0^\infty t \tau e^{-\tau t}\xi_{\mathcal{I}}(t) dt + \int_0^\infty t e^{-\tau t}f_{\mathcal{I}}(t) dt \\ &= \left[\left(-te^{-\tau t} - \frac{e^{-\tau t}}{\tau} \right) \xi_{\mathcal{I}}(t) \right]_0^\infty - \int_0^\infty \left(te^{-\tau t} + \frac{e^{-\tau t}}{\tau} \right) f_{\mathcal{I}}(t) dt + \int_0^\infty t e^{-\tau t}f_{\mathcal{I}}(t) dt \\ &= \frac{1}{\tau} - \frac{1}{\tau} \int_0^\infty e^{-\tau t}f_{\mathcal{I}}(t) dt = \frac{1 - \mathcal{L}[f_{\mathcal{I}}](\tau)}{\tau}, \end{aligned}$$

where $\mathcal{L}[f_{\mathcal{I}}](\tau)$ denotes the Laplace transform of $f_{\mathcal{I}}$ at τ . Multiplying this formula with the expected number of newly generated $S - I$ links $\tau\frac{n-1}{N}[S]_0$, we have the following general formula for pairwise reproduction number:

$$\mathcal{R}_0^p = \frac{n-1}{N}[S]_0(1 - \mathcal{L}[f](\tau)). \quad (3.4)$$

Clearly, the basic reproduction number \mathcal{R}_0 for arbitrary recovery time is

$$\mathcal{R}_0 = \frac{n}{N}\tau\mathbb{E}(\mathcal{I})[S]_0. \quad (3.5)$$

3.3 Special cases

For Markovian recovery (where recovery time is exponentially distributed with parameter γ , thus $f_{\mathcal{I}}(t) = \gamma e^{-\gamma t}$), it is straightforward to obtain, that pairwise reproduction number is

$$\mathcal{R}_{0,\text{Exp}(\gamma)}^p = \frac{n-1}{N}[S]_0(1 - \mathcal{L}[f_{\mathcal{I}}](\tau)) = \frac{n-1}{N}[S]_0 \left(1 - \frac{\gamma}{\tau + \gamma} \right) = \frac{n-1}{N}[S]_0 \frac{\tau}{\tau + \gamma}.$$

where $\mathcal{R}_{0,\text{Exp}(\gamma)}^p$ denotes the pairwise reproduction number for exponentially distributed recovery time with mean length $\frac{1}{\gamma}$. Since the Markovian mean-field model (2.8) has the

same functional form as the classical *SIR* model (2.1), the basic reproduction number is clearly

$$\mathcal{R}_{0,\text{Exp}(\gamma)} = \tau \frac{n}{N} [S]_0 \mathbb{E}(\mathcal{I}) = \frac{n}{N} \tau [S]_0 \frac{1}{\gamma}.$$

For fixed recovery time σ , the survival function is $\xi_{\mathcal{I}}(t) = 1$ if $0 \leq t < \sigma$ and $\xi_{\mathcal{I}}(t) = 0$ if $t \geq \sigma$, and the density function $f_{\mathcal{I}}(t)$ is the Dirac-delta $\delta(t - \sigma)$. Using fundamental properties of the delta function, we have

$$\mathcal{R}_{0,\text{Fixed}(\sigma)}^p = \frac{n-1}{N} [S]_0 (1 - \mathcal{L}[f_{\mathcal{I}}](\tau)) = \frac{n-1}{N} [S]_0 (1 - e^{-\tau\sigma}).$$

where $\mathcal{R}_{0,\text{Fixed}(\sigma)}^p$ denotes the pairwise reproduction number for fixed recovery time σ . Obviously, the basic reproduction number \mathcal{R}_0 for fixed recovery time σ is

$$\mathcal{R}_{0,\text{Fixed}(\sigma)} = \frac{n}{N} \tau [S]_0 \mathbb{E}(\mathcal{I}) = \frac{n}{N} \tau [S]_0 \sigma.$$

In Ch. 5, we develop the mean-field and pairwise models for fixed recovery time and investigate the essential properties of these systems. The general case is covered in Ch. 6 and the study of general model gives an insight to the role of reproduction numbers defined above. The most important cases for \mathcal{R}_0 and \mathcal{R}_0^p are summarised in Table 3.1.

| | \mathcal{R}_0 | \mathcal{R}_0^p |
|-----------|--|--|
| Markovian | $\frac{n}{N} \frac{\tau}{\gamma} [S]_0$ | $\frac{n-1}{N} \frac{\tau}{\tau+\gamma} [S]_0$ |
| Fixed | $\frac{n}{N} \tau \sigma [S]_0$ | $\frac{n-1}{N} (1 - e^{-\tau\sigma}) [S]_0$ |
| General | $\frac{n}{N} \tau \mathbb{E}(\mathcal{I}) [S]_0$ | $\frac{n-1}{N} (1 - \mathcal{L}[f_{\mathcal{I}}](\tau)) [S]_0$ |

Table 3.1: Basic and pairwise reproduction numbers for different recovery distributions. $\mathcal{L}[f_{\mathcal{I}}](\tau)$ denotes the Laplace transform of $f_{\mathcal{I}}$, the density of the recovery process, at τ .

3.4 Final size relations

It is easy to see for mean-field model (2.8), that the final size relation is

$$\ln s_{\infty} = \mathcal{R}_0 (s_{\infty} - 1), \quad (3.6)$$

where $\mathcal{R}_0 = \mathcal{R}_{0,\text{Exp}(\gamma)}$. It is clear, that the derivation of the final size relation is more complicated for the pairwise model. A possible derivation may have the following major steps:

- (a) find an invariant to reduce the dimensionality of the system,

- (b) integrate the equation for $[SI](t)$,
- (c) integrate the equation for $[S](t)$ on $[0, \infty)$,
- (d) employ algebraic manipulations to obtain the final size relation.

For (2.12) as we will show for general case in Ch. 6, this procedure yields

$$\frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n-1}} = \mathcal{R}_0^p \left(s_\infty^{\frac{n-1}{n}} - 1 \right), \quad (3.7)$$

where $\mathcal{R}_0^p = \mathcal{R}_{0, \text{Exp}(\gamma)}^p$. Observe, that taking the limit of $n \rightarrow \infty$ in (3.7) gives rise to $\ln(s_\infty) = \mathcal{R}_0^p(s_\infty - 1)$, which is equivalent to the ‘standard’ form of (3.6). This shows that, in a sense, by a suitable scaling of parameters one can obtain the mean-field model as a limit of the pairwise model, since in fully connected graphs the pairwise approximation tells us that the number of $[AB]$ pairs can be expressed as $[AB] = [A][B]$.

4

Impact of distribution on disease spread

In this chapter, we study how the distribution of infectious periods influences the dynamics of epidemics on networks. These results are summarised in our second paper in this topic [70]. Note, while \mathcal{R}_0 depends on the expected value only, (see Table 3.1, case 'General'), the pairwise reproduction number \mathcal{R}_0^p , uses the complete density function, thus the average length of infectious period does not determine exactly the reproduction number. It implies, that for an epidemic we have to know as precisely as possible the shape of the distribution.

Besides exponential, the most commonly used distributions in epidemiology - because of their simplicity and their good fit to empirical observations - are gamma, uniform and lognormal distributions. We will show for these typical families of distributions that higher variance in the recovery times generates lower reproduction numbers and different epidemic curves within each distribution family. We also show that knowing the expected value and the variance of the recovery times is not sufficient to determine the key characteristics of the epidemics such as initial growth rate, peak size, peak time and final epidemic size. For accurate predictions, more detailed information on the distribution of the infectious period is required, thus carefully estimating this distribution in the case of a real epidemics has paramount public health importance.

4.1 Gamma distribution

The gamma distribution is one of the most commonly used distributions in the epidemiology literature to approximate empirically observed latent periods and infectious periods. For example, it has been fitted to the incubation period and infectious period of smallpox [24], bluetongue disease [31] and so on. It is applied in a wide spectrum of models,

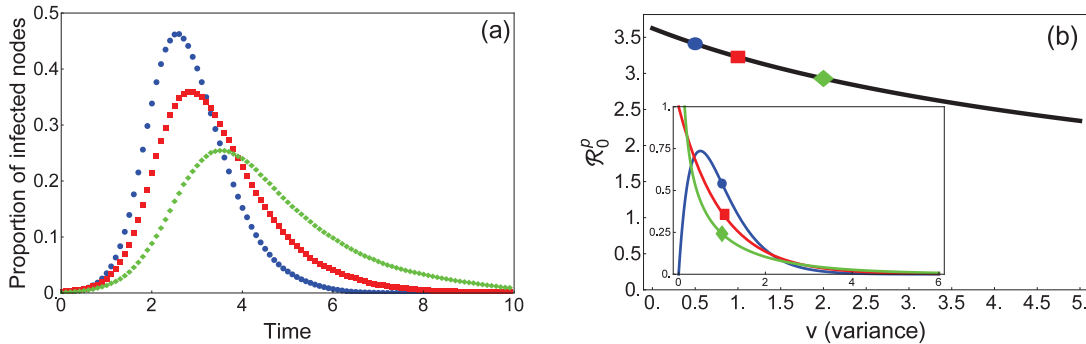


Figure 4.1: **(a)** Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics, where the transmission rate is $\tau = 0.3$ and the initial number of susceptibles is $[S]_0 = 999$ on a homogeneous network with $N = 1000$ nodes and degree $n = 15$. The circles/squares/diamonds correspond to simulations for gamma distributed recovery time with parameters $(a, b) = (2, 0.5)/(1, 1)/(0.5, 2)$, respectively. **(b)** The solid curve shows the reproduction number \mathcal{R}_0^p as a function of variance v for fixed $m = 1$, and the circle/square/diamond represent the cases simulated in Fig. (a). In the inset figure, the shapes of the three corresponding probability density functions are presented.

because of its flexibility and the possibility of incorporating it into ordinary differential equation models by the method of stages (also called linear chain) [54].

The usual notation of gamma distribution is $\text{Gamma}(a, b)$, where a is called the shape parameter and b is called the scale parameter. The probability density function and its Laplace transform are

$$f_{\mathcal{I}}(x) = \frac{x^{a-1} e^{-\frac{x}{b}}}{\Gamma(a) b^a}, \quad \mathcal{L}[f_{\mathcal{I}}](s) = \left(\frac{1}{1 + sb} \right)^a,$$

where $\Gamma(a)$ is the gamma function evaluated at a . If the infectious period \mathcal{I} is gamma distributed with shape parameter a and scale parameter b , that is $\mathcal{I} \sim \text{Gamma}(a, b)$, then the expected value is $m := \mathbb{E}(\mathcal{I}) = ab$, and the variance is $v := \text{Var}(\mathcal{I}) = ab^2$, and for simplicity later we shall use the notation m and v to denote the mean (expected value) and the variance of distributions.

The monotonicity of the reproduction number in the variance is depicted in Fig. 4.1(b). For a fixed mean but different variances of the gamma distribution, we can observe different epidemic curves in Fig. 4.1(a), and correspondingly different reproduction numbers (see Fig. 4.1(b)). The dependence of $R_{0,\mathcal{I}}^p$ on the distribution parameters is detailed in Fig. 4.2. In the following proposition we summarise our observations.

Proposition 4.1.1. *Consider two random variables $\mathcal{I}_1 \sim \text{Gamma}(a_1, b_1)$ and $\mathcal{I}_2 \sim \text{Gamma}(a_2, b_2)$ such that $\mathbb{E}(\mathcal{I}_1) = \mathbb{E}(\mathcal{I}_2)$ and $\text{Var}(\mathcal{I}_1) \leq \text{Var}(\mathcal{I}_2)$. If \mathcal{I}_1 and \mathcal{I}_2 represent the recovery time distribution, then for the corresponding reproduction numbers the*

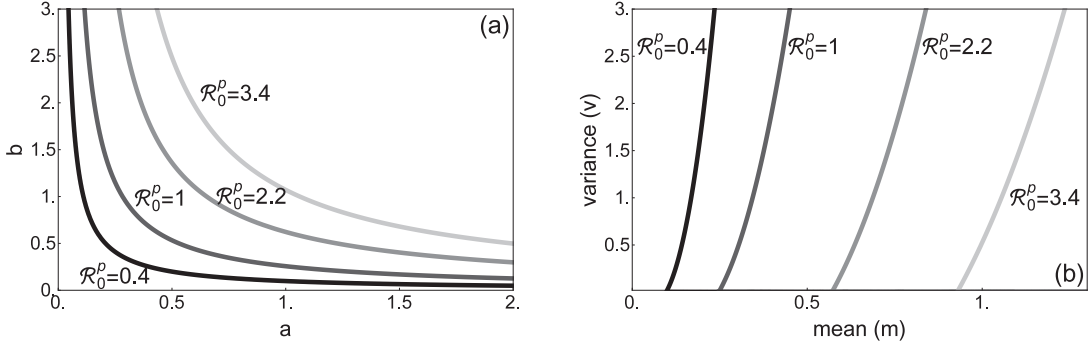


Figure 4.2: Contour lines of \mathcal{R}_0^p as a two-variable function of the parameters of the gamma distribution. The transmission rate is $\tau = 0.3$, the network has $N = 1000$ nodes and degree $n = 15$ and initial number of susceptibles is $[S]_0 = 999$. In **(a)**, the contour lines are given as the function of the shape and scale parameter, while in **(b)** they are depicted as the function of the mean and the variance of the gamma distribution.

relation $\mathcal{R}_{0,\mathcal{I}_1}^p \geq \mathcal{R}_{0,\mathcal{I}_2}^p$ holds (i.e. for gamma distributions with a given mean, the pairwise reproduction number is monotonically decreasing with respect to the variance).

Let us remark, that the exponential distribution is also an element of this family (Gamma($1, \frac{1}{\lambda}$) = Exp(λ)) and the sum n of independent exponentially distributed random variables with the same parameter λ is Gamma distributed with parameters shape n and scale $\frac{1}{\lambda}$ (i.e. $\sum_{i=1}^n \text{Exp}(\lambda) = \text{Gamma}(n, \frac{1}{\lambda})$).

Proof. Fix m as the common mean of \mathcal{I}_1 and \mathcal{I}_2 , then the scale parameter can be expressed as $b = \frac{m}{a}$. Using that $v = \text{Var}(\mathcal{I}) = ab^2$ and

$$\mathcal{L}[f_{\mathcal{I}}](\tau) = \left(\frac{1}{1 + \tau b} \right)^a,$$

we can express the parameters in terms of the mean m and variance v , and thus the Laplace transform can be written as

$$\lambda_m(\tau; v) := \mathcal{L}[f_{\mathcal{I}}](\tau) = \left(\frac{1}{1 + \frac{\tau v}{m}} \right)^{\frac{m^2}{v}},$$

where the notation $\lambda_m(\tau; v)$ for the Laplace transform is meant to emphasize that the Laplace transform evaluated at τ for a fixed m is a function of v . For arbitrary $0 < x < y < 1$ and $0 < a < b$, the inequalities

$$x^{\frac{m^2}{a}} < x^{\frac{m^2}{b}}, \quad x^{\frac{m^2}{b}} < y^{\frac{m^2}{b}}$$

hold. For $v_1 \leq v_2$ and for $\tau > 0$ we have $\left(1 + \frac{\tau}{m} v_2\right)^{-1} \leq \left(1 + \frac{\tau}{m} v_1\right)^{-1}$, and using the relations above we obtain

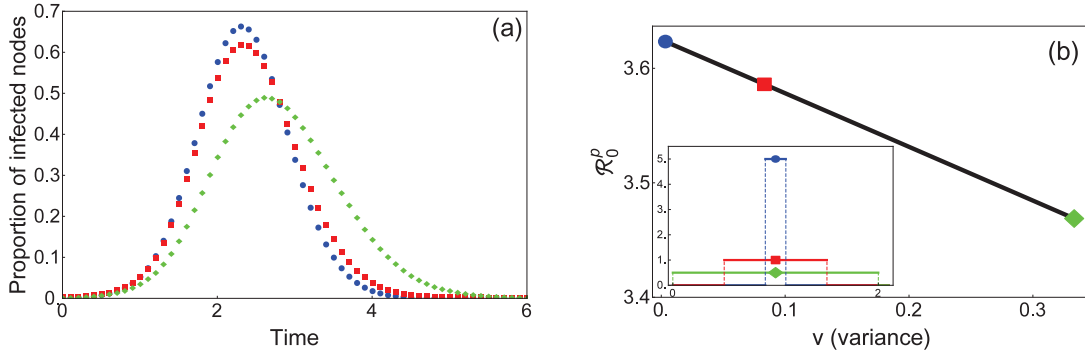


Figure 4.3: **(a)** Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics, where transmission rate is $\tau = 0.3$ and initial number of susceptibles is $[S]_0 = 999$, on homogeneous network with $N = 1000$ nodes and degree $n = 15$. The circles/squares/diamonds correspond to simulations for uniformly distributed recovery time with parameters $(a, b) = (0.9, 1.1)/(0.5, 1.5)/(0, 2)$, respectively. **(b)** The solid curve shows the reproduction number \mathcal{R}_0^p as the function of variance v for fixed $m = 1$, and the circle/square/diamond represent the cases simulated in Fig. (a). In the inset figure, the three uniform density functions are depicted.

$$\left(\frac{1}{1 + \frac{\tau}{m} v_1} \right)^{\frac{m^2}{v_1}} \leq \left(\frac{1}{1 + \frac{\tau}{m} v_2} \right)^{\frac{m^2}{v_2}} \leq \left(\frac{1}{1 + \frac{\tau}{m} v_2} \right)^{\frac{m^2}{v_2}},$$

which means that $\lambda_m(\tau; v)$ is monotone increasing in v . Therefore, the pairwise reproduction number

$$\mathcal{R}_{0,\mathcal{I}}^p = \frac{n-1}{N} [S]_0 (1 - \lambda_m(\tau; v))$$

is monotone decreasing in v . □

4.2 Uniform distribution

Since its simplicity allows us to make explicit calculations, in this section we outline how the reproduction number and the disease dynamics behave when the recovery time follows uniform distribution. Uniformly distributed incubation and infectious periods have been used for example, in the modelling of avian influenza [87].

Let $\text{Uniform}(a, b)$ denote a uniform distribution corresponding to the interval $[a, b]$, where $a \geq 0, b > a$. If $\mathcal{I} \sim \text{Uniform}(a, b)$, then the expected value is $m = \mathbb{E}(\mathcal{I}) = \frac{a+b}{2}$, and the variance is $v = \text{Var}(\mathcal{I}) = \frac{(b-a)^2}{12}$. The probability density function and its Laplace transform are given as

$$f_{\mathcal{I}}(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}, \quad \mathcal{L}[f_{\mathcal{I}}](s) = \frac{e^{-sa} - e^{-sb}}{s(b-a)}.$$

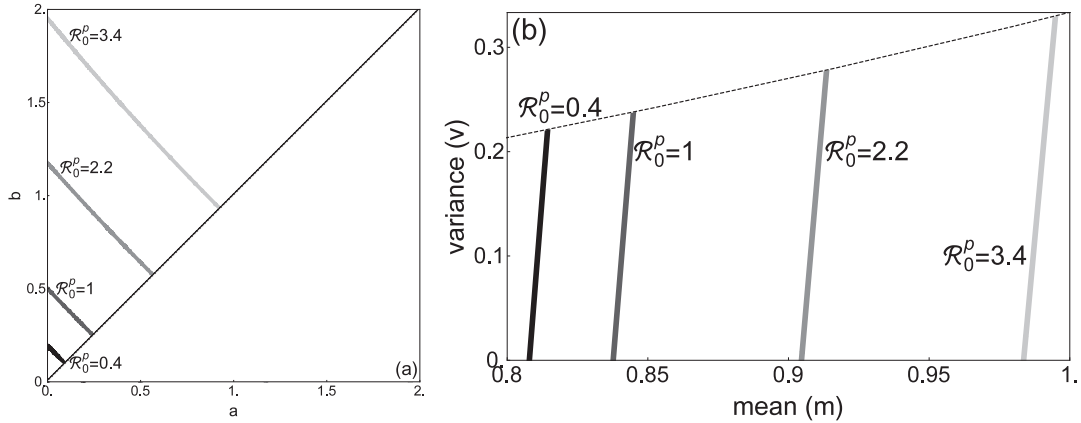


Figure 4.4: **(a)** Contour lines of \mathcal{R}_0^p as a two-variable function of the parameters of the uniform distribution. **(b)** Contour lines of \mathcal{R}_0^p as a two-variable function of m and v for uniform distribution. For both (a) and (b), the transmission rate is $\tau = 0.3$, the network has $N = 1000$ nodes and degree $n = 15$ with an initial number of susceptibles $[S]_0 = 999$.

The monotonicity of the reproduction number in the variance is depicted in Fig. 4.3(b). Similarly to the gamma distribution, for a fixed mean but different variances of the uniform distribution we can observe different epidemic curves in Fig. 4.3(a), and correspondingly different reproduction numbers (see Fig. 4.3(b)). The dependence of $\mathcal{R}_{0,\mathcal{I}}^p$ on the distribution parameters is detailed in Fig. 4.4.

Proposition 4.2.1. *Consider two random variables $\mathcal{I}_1 \sim \text{Uniform}(a_1, b_1)$ and $\mathcal{I}_2 \sim \text{Uniform}(a_2, b_2)$ such that $\mathbb{E}(\mathcal{I}_1) = \mathbb{E}(\mathcal{I}_2)$ and $\text{Var}(\mathcal{I}_1) \leq \text{Var}(\mathcal{I}_2)$. If \mathcal{I}_1 and \mathcal{I}_2 represent the recovery time distribution, then for the corresponding reproduction numbers the relation $\mathcal{R}_{0,\mathcal{I}_1}^p \geq \mathcal{R}_{0,\mathcal{I}_2}^p$ holds (i.e. for uniform distributions with a given mean, the pairwise reproduction number is monotonically decreasing with respect to the variance).*

Proof. Fixing the mean m , the right endpoint of the interval is $b = 2m - a$. Using that $v = \text{Var}(\mathcal{I}) = \frac{(b-a)^2}{12}$ and $\mathcal{L}[f_{\mathcal{I}}](\tau) = \frac{e^{-\tau a} - e^{-\tau b}}{\tau(b-a)}$, by simple algebra we can express the parameters by m and the variance v , and consequently the Laplace transform can be written as the function of v :

$$\lambda_m(\tau; v) = \frac{e^{-\tau m} e^{\tau\sqrt{3v}} - e^{-\tau\sqrt{3v}}}{2m \tau\sqrt{3v}},$$

where $v \in \left(0, \frac{m^2}{3}\right]$ from our assumptions on a and b . Expanding the exponentials in $\lambda_m(\tau; v)$ into Taylor series, we can notice that the negative terms cancel out, and we

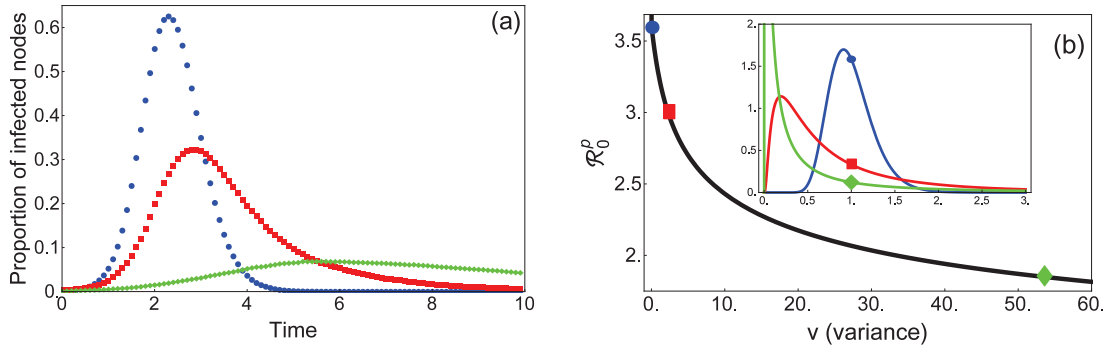


Figure 4.5: **(a)** Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics, where transmission rate is $\tau = 0.3$ and initial number of susceptibles is $[S]_0 = 999$, on homogeneous network with $N = 1000$ nodes and degree $n = 15$. The circles/squares/diamonds correspond to simulations for lognormally distributed recovery time with parameters $(\mu, \sigma) = (-0.03125, 0.25)/(-\frac{\ln(3)}{2}, \sqrt{\ln 3})/(0, 2)$, respectively. **(b)** The solid curve shows the reproduction number \mathcal{R}_0^p as the function of variance v for fixed $m = 1$, and the circle/square/diamond represent the cases simulated in Fig. (a). Inset figure shows the shape of these three distributions.

obtain

$$\begin{aligned}
 \lambda_m(\tau; v) &= \frac{e^{-\tau m} (1 + (\tau\sqrt{3v}) + \dots) - (1 - (\tau\sqrt{3v}) + \dots)}{2m \tau\sqrt{3v}} \\
 &= \frac{e^{-\tau m} 2\tau\sqrt{3v} + 2\frac{(\tau\sqrt{3v})^3}{3!} + \dots}{2m \tau\sqrt{3v}} \\
 &= \frac{e^{-\tau m}}{m} \sum_{n=0}^{\infty} \frac{(\tau\sqrt{3v})^{2n}}{(2n+1)!},
 \end{aligned} \tag{4.1}$$

which is monotone increasing in v for $m > 0$ and $\tau > 0$. Therefore, the pairwise reproduction number $R_{0,\mathcal{I}}^p = \frac{n-1}{N} [S]_0 (1 - \lambda_m(\tau; v))$ is monotone decreasing in v . \square

4.3 Lognormal distribution

The lognormal distribution is also widely used in epidemiology. They have been fitted, among others, to the incubation and infectious periods of smallpox [63]. Let $\ln\mathcal{N}(\mu, \sigma^2)$ denote a lognormal distribution, i.e. its logarithm is a normal distribution with expected value μ and variance σ . Then for the lognormal distribution $m = \mathbb{E}(\mathcal{I}) = e^{\mu + \frac{\sigma^2}{2}}$, $v = \text{Var}(\mathcal{I}) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$, and the probability density function is, for $x > 0$,

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{(-\mu + \ln(x))^2}{2\sigma^2}}.$$

Unfortunately a closed form does not exist for its Laplace transform, thus we cannot perform a full analysis as in the previous two sections. We can still investigate numerically

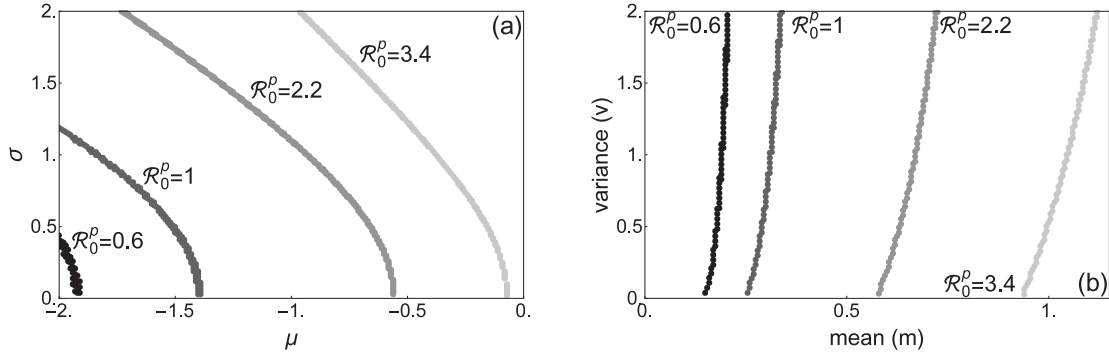


Figure 4.6: **(a)** Contour lines of \mathcal{R}_0^p as a two-variable function of the parameters of the lognormal distribution. **(b)** Contour lines of \mathcal{R}_0^p as a two-variable function of the mean and variance for lognormal distribution. For both (a) and (b), the transmission rate is $\tau = 0.3$, the network has $N = 1000$ nodes and degree $n = 15$ with an initial number of susceptibles is $[S]_0 = 999$.

the impact of m and v on the reproduction number and the time course of the epidemic. The density function can again be expressed in terms of m and v by the formula

$$f_m(x; v) = \frac{1}{x\sqrt{2\pi}\sqrt{\ln\left(\frac{v}{m^2} + 1\right)}} e^{-\frac{\left(\ln(x) - \ln(m) + \frac{1}{2}\ln\left(\frac{v}{m^2} + 1\right)\right)^2}{2\ln\left(\frac{v}{m^2} + 1\right)}}, \text{ for } x > 0. \quad (4.2)$$

By straightforward calculation, we can find $\mu = \ln(m) - \frac{1}{2}\ln\left(1 + \frac{v}{m^2}\right)$ and $\sigma^2 = \ln\left(1 + \frac{v}{m^2}\right)$, and then the formula above can be derived. Using this formula for the density, we can numerically determine its Laplace transform and plot the pairwise reproduction number as a function of the variance for any given m , see Fig. 4.5(b). The epidemic curves corresponding to these distributions can be seen in Fig. 4.5(a), and the dependence of \mathcal{R}_0^p on the distribution parameters is detailed in Fig. 4.6.

4.4 General case

In the last section we give some simple conditions for the general case, which may guarantee, that smaller variance induces higher pairwise reproduction number. We consider a random variable \mathcal{I} corresponding to recovery times with probability density functions $f_{\mathcal{I}}(t)$, cumulative distribution function $F_{\mathcal{I}}(t) = \int_0^t f_{\mathcal{I}}(s)ds$ and integral function of CDF $\mathcal{F}_{\mathcal{I}}(t) := \int_0^t F_{\mathcal{I}}(s)ds$. Clearly, $\frac{d^2}{dt^2}\mathcal{F}_{\mathcal{I}}(t) = \frac{d}{dt}F_{\mathcal{I}}(t) = f_{\mathcal{I}}(t)$. Moreover, $F_{\mathcal{I}}(0) = \mathcal{F}_{\mathcal{I}}(0) = 0$.

Theorem 4.4.1. *Consider two random variables \mathcal{I}_1 and \mathcal{I}_2 such that*

$$\mathbb{E}(\mathcal{I}_1) = \mathbb{E}(\mathcal{I}_2) < \infty, \quad (4.3)$$

and

$$\text{Var}(\mathcal{I}_1) < \text{Var}(\mathcal{I}_2) < \infty. \quad (4.4)$$

Let us assume, that

$$\lim_{t \rightarrow \infty} t^3 f_{\mathcal{I}}(t) = 0 \quad (4.5)$$

and for all $t > 0$,

$$\mathcal{F}_{\mathcal{I}_1}(t) \neq \mathcal{F}_{\mathcal{I}_2}(t). \quad (4.6)$$

holds. If \mathcal{I}_1 and \mathcal{I}_2 represent the recovery time distribution, then for the corresponding reproduction numbers the relation $\mathcal{R}_{0,\mathcal{I}_1}^p > \mathcal{R}_{0,\mathcal{I}_2}^p$ holds.

Proof. Using assumption (4.3), we deduce

$$\begin{aligned} \int_0^\infty t (f_{\mathcal{I}_1}(t) - f_{\mathcal{I}_2}(t)) &= [t(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t))]_0^\infty - \int_0^\infty (F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) dt \\ &= \lim_{t \rightarrow \infty} t(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) - [\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t)]_0^\infty \\ &\stackrel{[*]}{=} - \lim_{t \rightarrow \infty} (\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t)) = 0 \end{aligned}$$

thus

$$\lim_{t \rightarrow \infty} (\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t)) = 0. \quad (4.7)$$

To see $[*]$, i.e. $\lim_{t \rightarrow \infty} t(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) = 0$, we need some algebraic manipulations:

$$\begin{aligned} \lim_{t \rightarrow \infty} t(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) &= \lim_{t \rightarrow \infty} \frac{F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)}{\frac{1}{t}} \stackrel{\text{L'H}}{=} \lim_{t \rightarrow \infty} \frac{f_{\mathcal{I}_1}(t) - f_{\mathcal{I}_2}(t)}{-\frac{1}{t^2}} \\ &= - \lim_{t \rightarrow \infty} t^2 (f_{\mathcal{I}_1}(t) - f_{\mathcal{I}_2}(t)) \stackrel{(4.4)}{=} 0, \end{aligned}$$

where L'H refers to the L'Hospital rule. From assumption (4.4), we have

$$\begin{aligned} \text{Var}(\mathcal{I}_1) &= \mathbb{E}(\mathcal{I}_1^2) - (\mathbb{E}(\mathcal{I}_1))^2 < \mathbb{E}(\mathcal{I}_2^2) - (\mathbb{E}(\mathcal{I}_2))^2 = \text{Var}(\mathcal{I}_2) \\ &\stackrel{(4.3)}{\Rightarrow} \mathbb{E}(\mathcal{I}_1^2) < \mathbb{E}(\mathcal{I}_2^2). \end{aligned}$$

or equivalently $\int_0^\infty t^2 (f_{\mathcal{I}_1} - f_{\mathcal{I}_2}) dt < 0$. We can carry out some calculation on the left-hand side of this inequality:

$$\begin{aligned} \int_0^\infty t^2 (f_{\mathcal{I}_1} - f_{\mathcal{I}_2}) dt &= [t^2 (F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t))]_0^\infty - 2 \int_0^\infty t (F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) dt \\ &= \lim_{t \rightarrow \infty} t^2 (F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) - 2 [t(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t))]_0^\infty \\ &+ 2 \int_0^\infty \mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t) dt \\ &\stackrel{[**]}{=} -2 \lim_{t \rightarrow \infty} t(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t)) + 2 \int_0^\infty \mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t) dt \\ &\stackrel{[**]}{=} 2 \int_0^\infty \mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t) dt, \end{aligned}$$

consequently

$$\int_0^\infty \mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t) dt < 0 \quad (4.8)$$

To prove $[**]$, i.e. $\lim_{t \rightarrow \infty} t^2(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) = \lim_{t \rightarrow \infty} t(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t)) = 0$, we have

$$\begin{aligned} \lim_{t \rightarrow \infty} t(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t)) &= \lim_{t \rightarrow \infty} \frac{\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t)}{\frac{1}{t}} \stackrel{\text{L'H}}{=} \lim_{t \rightarrow \infty} \frac{F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)}{-\frac{1}{t^2}} \\ &= -\lim_{t \rightarrow \infty} t^2(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t)) \\ &\stackrel{\text{L'H}}{=} \lim_{t \rightarrow \infty} \frac{f_{\mathcal{I}_1}(t) - f_{\mathcal{I}_2}(t)}{\frac{2}{t^3}} = \frac{1}{2} \lim_{t \rightarrow \infty} t^3(f_{\mathcal{I}_1}(t) - f_{\mathcal{I}_2}(t)) \\ &\stackrel{(4.5)}{=} 0. \end{aligned}$$

Since $F_{\mathcal{I}}(t) \geq 0, t \geq 0$ and monotone increasing, the integral function of CDF $\mathcal{F}_{\mathcal{I}}(t)$ is monotone increasing and convex. Using (4.6) and (4.8), we obtain

$$\mathcal{F}_{\mathcal{I}_1}(t) < \mathcal{F}_{\mathcal{I}_2}(t), \quad (4.9)$$

for all $t > 0$. Clearly, for $\mathcal{R}_{0,\mathcal{I}_1}^p > \mathcal{R}_{0,\mathcal{I}_2}^p$, it is enough to prove, that $\mathcal{L}[f_{\mathcal{I}_1}](\tau) < \mathcal{L}[f_{\mathcal{I}_2}](\tau)$, i.e. $\int_0^\infty e^{-\tau t}(f_{\mathcal{I}_1}(t) - f_{\mathcal{I}_2}(t))dt < 0$. First, we perform some algebraic manipulation on the left-hand side:

$$\begin{aligned} \int_0^\infty e^{-\tau t}(f_{\mathcal{I}_1}(t) - f_{\mathcal{I}_2}(t))dt &= [e^{-\tau t}(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t))]_0^\infty \\ &\quad + \tau \int_0^\infty e^{-\tau t}(F_{\mathcal{I}_1}(t) - F_{\mathcal{I}_2}(t))dt \\ &= \tau[e^{-\tau t}(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t))]_0^\infty \\ &\quad + \tau^2 \int_0^\infty e^{-\tau t}(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t))dt \\ &\stackrel{(4.7)}{=} \tau^2 \int_0^\infty e^{-\tau t}(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t))dt. \end{aligned}$$

In conclusion, we have

$$\tau^2 \int_0^\infty e^{-\tau t}(\mathcal{F}_{\mathcal{I}_1}(t) - \mathcal{F}_{\mathcal{I}_2}(t))dt \stackrel{(4.9)}{<} 0,$$

therefore $\mathcal{L}[f_{\mathcal{I}_1}](\tau) < \mathcal{L}[f_{\mathcal{I}_2}](\tau)$, which gives $\mathcal{R}_{0,\mathcal{I}_1}^p > \mathcal{R}_{0,\mathcal{I}_2}^p$. \square

As an example, we consider $\mathcal{I}_1 \sim \text{Exp}(\gamma)$ and $\mathcal{I}_2 \sim \text{Fixed}\left(\frac{1}{\gamma}\right)$, i.e. $f_{\mathcal{I}_1}(t) = \gamma e^{-\gamma t}, t \geq 0$ and $f_{\mathcal{I}_2}(t) = \delta\left(t - \frac{1}{\gamma}\right)$, where $\delta(t)$ denotes the Dirac delta function. Clearly, we obtain $\mathcal{F}_{\mathcal{I}_1}(t) = t + \frac{1}{\gamma}e^{-\gamma t} - \frac{1}{\gamma}$ and $\mathcal{F}_{\mathcal{I}_2}(t) = t - \frac{1}{\gamma}$, thus there is no $t_0 > 0$, such that $\mathcal{F}_{\mathcal{I}_1}(t_0) = \mathcal{F}_{\mathcal{I}_2}(t_0)$. Since $\mathbb{E}(\mathcal{I}_1) = \mathbb{E}(\mathcal{I}_2) = \frac{1}{\gamma}, \frac{1}{\gamma^2} = \text{Var}(\mathcal{I}_1) > \text{Var}(\mathcal{I}_2) = 0$ and the other conditions of Theorem 4.4.1 are satisfied, we find $\mathcal{R}_{0,\mathcal{I}_1}^p < \mathcal{R}_{0,\mathcal{I}_2}^p$.

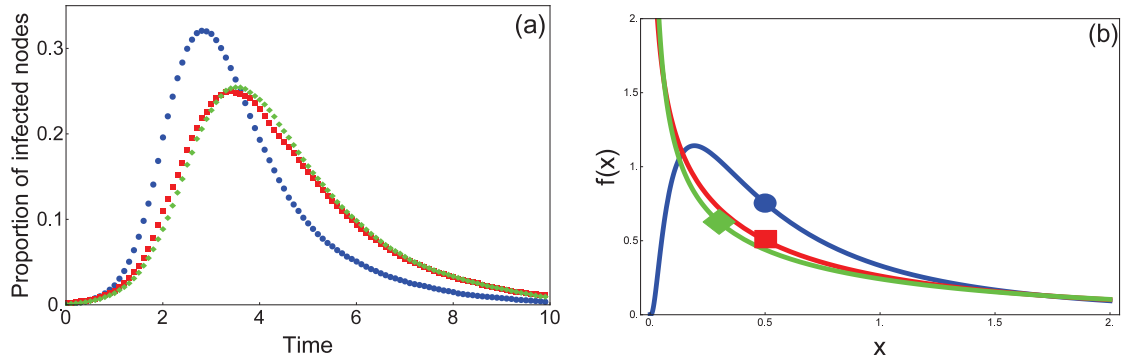


Figure 4.7: **(a)** Comparison of three epidemic curves after averaging explicit stochastic simulations with three different distributions of recovery times. The diamond/circle/square corresponds to $\text{Gamma}(0.5, 2)$, $\ln\mathcal{N}\left(\frac{\ln(3)}{2}, \sqrt{\ln(3)}\right)$ and $\text{Weibull}(0.72, 0.81)$ distributions, respectively. All three distributions have mean $m = 1$ and variance $v = 2$. **(b)** Probability density functions corresponding to the three distributions.

4.5 Conclusions

For two-parameter distribution families, it is possible to regard \mathcal{R}_0^p as a function depending on two variables, e.g. the mean m and variance v , see Figs. 4.2, 4.4, 4.6. Since the general final size relation (3.7) is monotone in \mathcal{R}_0^p , we conclude that smaller variance generates more infections. In Fig. 4.7, we compared three distributions from different families, each having $m = 1$ and $v = 2$. Besides the gamma and the lognormal distributions, for the sake of comparison we selected a third type of continuous distribution, namely Weibull distribution, which has been fitted to the infectious period for the recent ebola outbreak [19]. Fig. 4.7 illustrates that the mean and the variance of the recovery times alone are not able to determine the key characteristics of the epidemic curves, and a large variety of outbreaks can be generated from having the same mean and variance. This is especially the case in Fig. 4.7, where the gamma distributed infectious period leads to a very different epidemic, compared to that corresponding to the lognormally distributed infectious period, despite the mean and the variance are being identical. Therefore, in a real life situation, it is crucial to estimate the empirical distribution of the infectious period as accurately as possible, since the mean and the variance alone do not provide enough information for accurate predictions.

5

Models with fixed recovery time

In this chapter, we consider a non-Markovian epidemic process with fixed recovery time. We generalise the concept of mean-field and pairwise models by developing their analysis for this case. First, a fixed infectious period, denoted by σ , is considered, and the derivation of the approximating deterministic models from first principles is illustrated. We show that the non-Markovian dynamics can be described by a system of delay differential equations with constant and distributed delays. The results of this chapter are summarised in our first paper on non-Markovian network epidemics [49].

5.1 Model development

The infection process is assumed to be Markovian, thus the equation for $[S](t)$ is the same as before (see Eq. (2.12)), i.e.

$$[\dot{S}](t) = -\tau[SI](t).$$

The number of infected nodes at time t is replenished by $\tau[SI](t)$ and we assume through this chapter, that all initial infected nodes are newborn at $t = 0$, thus there is no recovery for $0 \leq t < \sigma$ and we have

$$[\dot{I}](t) = \tau[SI](t),$$

for $t < \sigma$ and $[I](t)$ is depleted by $\tau[SI](t - \sigma)$ for $t > \sigma$, that yields

$$[\dot{I}](t) = \tau[SI](t) - \tau[SI](t - \sigma).$$

The equation for the number of $S - S$ links is the same as in Eq. (2.12), because the infection process is Markovian thus

$$[\dot{SS}](t) = -2\tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)}.$$

In a similar manner, the number of $S - I$ links is replenished by

$$\tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)},$$

which is the half of the rate of depletion of $S - S$ links (due to symmetry property $[SI] = [IS]$). Furthermore, depletion of $S - I$ links occurs due to the infection within $S - I$ pairs, $\tau[SI](t)$, and due to the infection of the S node from outside the pair,

$$\tau \frac{n-1}{n} \frac{[SI](t)[SI](t)}{[S](t)}.$$

It is clear, that for $0 \leq t < \sigma$ no recovery happens, thus the governing equation for $S - I$ links is

$$[\dot{SI}](t) = \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau[SI](t) - \tau \frac{n-1}{n} \frac{[SI](t)[SI](t)}{[S](t)}.$$

The derivation above gives the following model for $0 \leq t < \sigma$:

$$[\dot{S}](t) = -\tau[SI](t), \tag{5.1a}$$

$$[\dot{I}](t) = \tau[SI](t), \tag{5.1b}$$

$$[\dot{SS}](t) = -2\tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)}, \tag{5.1c}$$

$$[\dot{SI}](t) = \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau[SI](t) - \tau \frac{n-1}{n} \frac{[SI](t)[SI](t)}{[S](t)} \tag{5.1d}$$

The Eq. (5.1) is a system of ordinary differential equations, given initial values $[S]_0, [I]_0, [SS]_0$ and $[SI]_0$ at $t = 0$ are sufficient to guarantee a unique solution. Let us denote the solution of (5.1) on the time interval $[0, \sigma]$ by

$$X^*(t) = ([S]^*(t), [I]^*(t), [SS]^*(t), [SI]^*(t)).$$

At time $t = \sigma$, the initial infected nodes recover 'instantly', thus a discontinuity appears and obviously, the solution for $t > \sigma$ starts from

$$\widetilde{X} = ([S]^*(\sigma), [I]^*(\sigma) - [I]_0, [SS]^*(\sigma), [SI]^*(\sigma) - [SI]_0).$$

On the other hand, for $t > \sigma$, there are $S - I$ links, which survive for time σ , but will be removed due to the recovery of the I node. Next, we need to take into account for the removal of $S - I$ links which were created precisely σ times ago. By the first intuition, one would believe that this term is simply proportional to

$$\tau \frac{n-1}{n} \frac{[SS](t-\sigma)[SI](t-\sigma)}{[S](t-\sigma)}.$$

However, one must take into consideration that, in the time interval $(t - \sigma, t)$, an $S - I$ link could have been destroyed due to either within pair infection or by infection of the S node from outside. Hence, a discount factor needs to be determined to capture this effect. To calculate this factor, $S - I$ links, that are created at the same time, are considered as a cohort denoted by x , and we model infection within and from outside by writing down the following evolution equation,

$$\dot{x}(t) = -\tau(n-1)\frac{[SI](t)}{n[S](t)}x(t) - \tau x(t),$$

where, the first term denotes the ‘outer’ infection of the S node, while the second term stands for ‘inner’ infection of the S node. We note that the outside infection is simply proportional to the probability that an S node with an already engaged link has a further infected neighbour, $\tau(n-1)\frac{[SI](t)}{n[S](t)}$. The solution of the latter evolution equation in $[t-\sigma, t]$ is

$$x(t) = x(t-\sigma)e^{-\int_{t-\sigma}^t (\tau\frac{n-1}{n}\frac{[SI](u)}{[S](u)} + \tau) du}$$

and this provides the depletion or discount rate of $S - I$ links. In this case,

$$x(t-\sigma) = \tau\frac{n-1}{n}\frac{[SS](t-\sigma)[SI](t-\sigma)}{[S](t-\sigma)},$$

which is the replenishment of $S - I$ links. Therefore, summarising all the above, the pairwise approximation for $t > \sigma$ is

$$[\dot{S}](t) = -\tau[SI](t), \tag{5.2a}$$

$$[\dot{I}](t) = \tau[SI](t) - \tau[SI](t-\sigma) \tag{5.2b}$$

$$[\dot{S}S](t) = -2\tau\frac{n-1}{n}\frac{[SS](t)[SI](t)}{[S](t)}, \tag{5.2c}$$

$$[\dot{S}I](t) = \tau\frac{n-1}{n}\frac{[SS](t)[SI](t)}{[S](t)} - \tau[SI](t) - \tau\frac{n-1}{n}\frac{[SI](t)[SI](t)}{[S](t)} - \tau\frac{n-1}{n}\frac{[SS](t-\sigma)[SI](t-\sigma)}{[S](t-\sigma)}e^{-\int_{t-\sigma}^t \tau\frac{n-1}{n}\frac{[SI](u)}{[S](u)} + \tau du}. \tag{5.2d}$$

Let us mention, that Eq. (5.2) is a system of delay differential equations with discrete and distributed delays. Despite the fact, that the initial function

$$X(t) = \begin{cases} X^*(t), & \text{if } t < \sigma, \\ \widetilde{X}, & \text{if } t = \sigma, \end{cases}$$

is not continuous, the solution of Eq. (5.2) is continuous for $t > \sigma$. Similarly to Markovian

case, the non-Markovian mean-field model for fixed infectious period is

$$[\dot{S}](t) = -\tau \frac{n}{N} [S](t)[I](t), \quad (5.3a)$$

$$[\dot{I}](t) = \tau \frac{n}{N} [S](t)[I](t), \quad (5.3b)$$

for $0 \leq t < \sigma$ and

$$[\dot{S}](t) = -\tau \frac{n}{N} [S](t)[I](t), \quad (5.4a)$$

$$[\dot{I}](t) = \tau \frac{n}{N} [S](t)[I](t) - \tau \frac{n}{N} [S](t - \sigma)[I](t - \sigma), \quad (5.4b)$$

for $t > \sigma$. Here, if we denote the solution of (5.3) for initial values $[S]_0, [I]_0$ and for time interval $t \in [0, \sigma]$ by $X_m^*(t) = ([S]^*(t), [I]^*(t))$, the initial function associated to (5.4) is $X_m^*(t)$ for $0 \leq t < \sigma$ and $([S]^*(\sigma), [I]^*(\sigma) - [I]_0)$ at $t = \sigma$. These systems are now the main subjects of our investigation from analytical and numerical point of view.

5.2 Analytical results

In this section, we explore the most important features of systems (5.1)-(5.2) and (5.3)-(5.4). First, we find a first integral of the pairwise model (5.1)-(5.2), which allows us to reduce the dimensionality. We show that the solutions of the models are biologically meaningful, i.e. solutions with non-negative data remain non-negative for $t \geq 0$. The paramount results of this part are the theorems for occurrence of an outbreak and the implicit relations between the reproduction number and the final epidemic size.

5.2.1 First integral

We use (5.2a) and (5.2c) to find an invariant of the system.

Proposition 5.2.1. *The function $U(t) = \frac{[SS](t)}{[S]^{2\frac{n-1}{n}}(t)}$ is a first integral of system (5.1)-(5.2).*

Proof. To see this, let us divide Eq.(5.2c) by Eq.(5.2a), which gives

$$\frac{d[SS]}{d[S]} = \frac{-2\tau \frac{n-1}{n} \frac{[SS][SI]}{[S]}}{-\tau [SI]} = 2 \frac{n-1}{n} \frac{[SS]}{[S]}.$$

Solving this equation, we find $[SS] = K[S]^{2\frac{n-1}{n}}$, where K is a constant, thus $U(t) = \frac{[SS](t)}{[S]^{2\frac{n-1}{n}}(t)}$ is an invariant quantity in the system and its value is

$$U(0) = K = \frac{[SS](0)}{[S]^{2\frac{n-1}{n}}(0)} = \frac{[SS]_0}{[S]_0^{2\frac{n-1}{n}}} = \frac{n[S]_0 \frac{[S]_0}{N}}{[S]_0^{2\frac{n-1}{n}}} = \frac{n}{N} [S]_0^{\frac{2}{n}}.$$

□

Consequently, using this first integral, we obtain

$$[SS](t) = \frac{n}{N} [S]_0^{\frac{2}{n}} [S]^{2\frac{n-1}{n}}(t). \quad (5.5)$$

Applying Eq.(5.5), we can reduce our pairwise model to a two-dimensional system:

$$\begin{aligned} [\dot{S}](t) &= -\tau[SI](t), \\ [\dot{SI}](t) &= \tau\kappa[S]^{\frac{n-2}{n}}(t)[SI](t) - \tau[SI](t) - \tau\frac{n-1}{n}\frac{[SI](t)}{[S](t)}[SI](t) \\ &\quad - \tau\kappa[S]^{\frac{n-2}{n}}(t-\sigma)[SI](t-\sigma)e^{-\int_{t-\sigma}^t \tau\frac{n-1}{n}\frac{[SI](u)}{[S](u)} + \tau du}, \end{aligned} \quad (5.6)$$

where

$$\kappa = \frac{n-1}{N} [S]_0^{\frac{2}{n}}. \quad (5.7)$$

On the other hand, it is clear, that $[SS]_0 = \frac{n}{N} [S]_0^{\frac{2}{n}} [S]_0^{2\frac{n-1}{n}} = \frac{n}{N} [S]_0^2$ holds, which is exactly the same as Eq. (3.3).

5.2.2 Positivity

We are interested only in nonnegative solutions of system (5.1)-(5.2). The following proposition shows, that the solutions remain nonnegative provided that the initial conditions are nonnegative.

Proposition 5.2.2. *If initial conditions $[S]_0, [SS]_0, [I]_0$ and $[SI]_0$ for (5.1) and (5.3) are nonnegative, then $[S](t) \geq 0, [SS](t) \geq 0, [I](t) \geq 0$ and $[SI](t) \geq 0$ hold for $t \geq 0$ in both mean-field model (5.3)-(5.4) and pairwise model (5.1)-(5.2).*

Proof. It is clear, that $[SS](t)$ remains nonnegative, if the initial condition $[SS](0)$ is nonnegative, because $[SS](t)$ can be expressed from Eq.(5.2c) in the form

$$[SS](t) = [SS]_0 e^{-2\tau\frac{n-1}{n} \int_0^t \frac{[SI](s)}{[S](s)} ds}. \quad (5.8)$$

Moreover, if $[SS]_0$ is positive, then $[SS](t) > 0$ for all $t \geq 0$. From Eq. (5.5) we obtain that $[S](t)$ cannot be zero, if $[SS](t)$ is positive for all $t \geq 0$, which implies (from continuity of solutions) $[S](t) > 0$ for $t \geq 0$. For $[I](t)$ and $[SI](t)$, we have the following formulae:

$$[I](t) = \begin{cases} [I]_0 + \int_0^t \tau[SI](a) da, & \text{if } 0 \leq t < \sigma; \\ \int_{t-\sigma}^t \tau[SI](a) da, & \text{if } t \geq \sigma. \end{cases} \quad (5.9)$$

for $0 \leq t < \sigma$,

$$[SI](t) = [SI]_0 + \int_0^t \tau \frac{n-1}{n} \frac{[SS](a)[SI](a)}{[S](a)} e^{-\int_a^t \tau \frac{n-1}{n} \frac{[SI](u)}{[S](u)} + \tau du} da, \quad (5.10)$$

and for $t > \sigma$,

$$[SI](t) = \int_{t-\sigma}^t \tau \frac{n-1}{n} \frac{[SS](a)[SI](a)}{[S](a)} e^{-\int_a^t \tau \frac{n-1}{n} \frac{[SI](u)}{[S](u)} + \tau du} da. \quad (5.11)$$

It can be seen that $[I](t)$ remains nonnegative if $[SI](t)$ is nonnegative for $t \geq 0$. On the other hand, $[SI](t_0)$ cannot be zero for some $t_0 \geq 0$, because the formulae (5.10)-(5.11) depends on the $[S](t; t \in [t-\sigma, t])$ and $[SS](t; t \in [t-\sigma, t])$, which are positive, hence $[SI](t) > 0$. \square

In the case of the mean-field model (5.3)-(5.4), the positivity of $[S](t)$ is clear. To see the positivity of $[I](t)$, we substitute (2.7) into (5.9), which gives

$$[I](t) = \begin{cases} [I]_0 + \int_0^t \tau \frac{n}{N} [S](a)[I](a) da, & \text{if } 0 \leq t < \sigma; \\ \int_{t-\sigma}^t \tau \frac{n}{N} [S](a)[I](a) da, & \text{if } t \geq \sigma. \end{cases} \quad (5.12)$$

Notice that $[I](t)$ remains nonnegative if $[S](t)$ is nonnegative for $t \geq 0$.

5.2.3 Epidemic outbreak and reproduction numbers

A disease with a very long exposed period in a population into which a small number of infectives is introduced, the number of infectives could decrease initially before starting to grow as exposed individuals become infective. The following definition was proposed in [14].

Definition 1. *In a disease transmission model with no demographic effects, there is no epidemic if the equilibrium with all members of the population susceptible is (locally) asymptotically stable, and there is an epidemic if this equilibrium is unstable, in each case considering only perturbations of the equilibrium with positive infected initial states.*

Using this concept, we state the following theorems for the relation between epidemic outbreak and reproduction numbers.

Theorem 5.2.1. *There is an epidemic for the model (5.3)-(5.4) if and only if $\mathcal{R}_0 > 1$, where the basic reproduction number is $\mathcal{R}_0 = \tau \frac{n}{N} [S]_0 \sigma$.*

Proof. Clearly, $(\bar{S}, \bar{I}) = ([S]_0, 0), 0 \leq [S]_0 \leq N$ are equilibria of the mean-field model (5.3)-(5.4), thus we have infinitely many equilibria. On the other hand, $[I](t)$ is monotone increasing on time interval $[0, \sigma]$, but for long-term behaviour we will study (5.4). This system can be rewritten in the form

$$\begin{aligned} [\dot{S}](t) &= -\tau \frac{n}{N} [S](t)[I](t), \\ [I](t) &= \int_{t-\sigma}^t \tau \frac{n}{N} [S](a)[I](a) da = \tau \frac{n}{N} \int_0^\sigma [S](t-a)[I](t-a) da. \end{aligned}$$

Formal linearisation at $(\bar{S}, \bar{I}) = ([S]_0, 0)$ gives

$$\begin{aligned} \dot{s}(t) &= -\tau \frac{n}{N} [S]_0 i(t), \\ i(t) &= \tau \frac{n}{N} [S]_0 \int_0^\sigma i(t-a) da, \end{aligned}$$

thus the characteristic equation is

$$\det \begin{pmatrix} -\lambda & -\tau \frac{n}{N} [S]_0 \\ 0 & \tau \frac{n}{N} [S]_0 \int_0^\sigma e^{-\lambda a} da - 1 \end{pmatrix} = 0.$$

There are two real roots of the characteristic equation, namely $\lambda = 0$ and the real solution of equation

$$G(\lambda) := \tau \frac{n}{N} [S]_0 \int_0^\sigma e^{-\lambda a} da = 1.$$

First, the eigenspace associated to characteristic root $\lambda = 0$ contains only the equilibrium points. Indeed, searching for solution in form $\begin{pmatrix} [S]^* \\ [I]^* \end{pmatrix} e^{\lambda t}$, we find

$$\begin{aligned} \lambda [S]^* &= -\tau \frac{n}{N} [S]_0 [I]^*, \\ [I]^* &= \tau \frac{n}{N} [S]_0 \int_0^\sigma [I]^* e^{-\lambda a} da, \end{aligned}$$

thus for $\lambda = 0$, we obtain $[I]^* = 0$.

Second, $G(\lambda)$ is monotone decreasing with respect to λ and the real solution of the equation $G(\lambda) = 1$ is positive, if and only if $G(0) > 1$ or equivalently, $G(0) = \tau \frac{n}{N} [S]_0 \int_0^\sigma da = \tau \frac{n}{N} [S]_0 \sigma = \mathcal{R}_0 > 1$.

Third, let us assume there is a non-real characteristic root μ with positive real part for $\mathcal{R}_0 < 1$. Then we have

$$1 = \tau \frac{n}{N} [S]_0 \left| \int_0^\sigma e^{-\mu a} da \right| \leq \tau \frac{n}{N} [S]_0 \int_0^\sigma |e^{-\mu a}| da < \frac{n}{N} [S]_0 \int_0^\sigma da = \tau \frac{n}{N} [S]_0 \sigma = \mathcal{R}_0,$$

which is a contradiction. In conclusion, we have an epidemic if and only if $\mathcal{R}_0 > 1$. \square

Similar procedure can be done for the pairwise model (5.1)-(5.2). In that case, we study the system in form

$$\begin{aligned} \dot{[S]}(t) &= -\tau[SI](t), \\ [SI](t) &= \int_{t-\sigma}^t \tau \frac{n-1}{N} [S]_0^{\frac{2}{n}} [S]^{\frac{n-2}{n}}(a) [SI](a) e^{-\int_a^t \tau \frac{n-1}{N} \frac{[SI](u)}{[S](u)} + \tau du} da \\ &= \int_0^\sigma \tau \frac{n-1}{N} [S]_0^{\frac{2}{n}} [S]^{\frac{n-2}{n}}(t-a) [SI](t-a) e^{-\int_{t-a}^t \tau \frac{n-1}{N} \frac{[SI](u)}{[S](u)} + \tau du} da, \end{aligned}$$

and the associated formally linearised system

$$\begin{aligned} \dot{si}(t) &= -\tau si(t), \\ si(t) &= \tau \frac{n-1}{N} [S]_0 \int_0^\sigma si(t-a) e^{-\tau a} da. \end{aligned}$$

The non-trivial characteristic roots come from the equation

$$G_p(\lambda) := \tau \frac{n-1}{N} [S]_0 \int_0^\sigma e^{-\tau a} e^{-\lambda a} da = 1,$$

and obviously,

$$G_p(0) = \tau \frac{n-1}{N} [S]_0 \int_0^\sigma e^{-\tau a} da = \tau \frac{n-1}{N} [S]_0 \left[\frac{e^{-\tau a}}{-\tau} \right]_0^\sigma = \frac{n-1}{N} [S]_0 (1 - e^{-\tau\sigma}) = \mathcal{R}_0^p.$$

The remarks above can be summarised in the following theorem:

Theorem 5.2.2. *There is an epidemic for the model (5.1)-(5.2) if and only if $\mathcal{R}_0^p > 1$, where the pairwise reproduction number is $\mathcal{R}_0 = \frac{n-1}{N} [S]_0 (1 - e^{-\tau\sigma})$.*

5.2.4 Final size relation

In this part, we derive final size relations that allow us to calculate the total number of infected nodes during an epidemic outbreak on the network. We use the notation $s_\infty = \frac{[S]_\infty}{[S]_0}$, where $[S]_\infty = \lim_{t \rightarrow \infty} [S](t)$ (this limit exists, since $[S](t)$ is positive for $t > 0$).

Theorem 5.2.3. *The final size relation associated to the mean-field model (5.3)-(5.4) is*

$$\ln(s_\infty) = \mathcal{R}_0 (s_\infty - 1), \quad (5.13)$$

where the basic reproduction number is $\mathcal{R}_0 = \tau \frac{n}{N} [S]_0 \sigma$, see details in Section 3.2.

Notice, that Eq. (5.13) has the same functional form as the classical final size relation.

Proof. From Eq. (5.4a), we have

$$[S](t) = [S]_\sigma e^{-\tau \frac{n}{N} \int_\sigma^t [I](u) du}$$

for $t > \sigma$ and

$$[S]_\infty - [S]_\sigma = -\tau \frac{n}{N} \int_\sigma^\infty [S](u)[I](u) du.$$

On the other hand, for $t > \sigma$

$$[I](t) = \int_0^\sigma \tau \frac{n}{N} [S](t-w)[I](t-w) dw,$$

where $\tau \frac{n}{N} [S](t-w)[I](t-w)$ is the new infections at $t-w$. Hence

$$\begin{aligned} \ln \left(\frac{[S]_\infty}{[S]_\sigma} \right) &= -\tau \frac{n}{N} \int_\sigma^\infty [I](u) du = -\tau \frac{n}{N} \int_\sigma^\infty \int_0^\sigma \tau \frac{n}{N} [S](u-w)[I](u-w) dw du \\ &= -\left(\tau \frac{n}{N} \right)^2 \int_0^\sigma \int_\sigma^\infty [S](u-w)[I](u-w) du dw \\ &= -\left(\tau \frac{n}{N} \right)^2 \int_0^\sigma \left(\int_\sigma^\infty [S](q)[I](q) dq + \int_{\sigma-w}^\sigma [S](q)[I](q) dq \right) dw \\ &= -\left(\tau \frac{n}{N} \right)^2 \int_0^\sigma \left(\int_\sigma^\infty [S](q)[I](q) dq + \int_{\sigma-w}^\sigma [S](q)[I]_0 e^{\tau \frac{n}{N} \int_0^q S(a) da} dq \right) dw, \end{aligned}$$

where we used the fact, that $[I](t) = [I]_0 e^{\tau \frac{n}{N} \int_0^t S(u) du}$ from Eq. (5.3b). By neglecting the small amount of initial infected nodes, on the one hand, we have the approximation

$$-\ln \left(\frac{[S]_\infty}{[S]_\sigma} \right) = \left(\tau \frac{n}{N} \right)^2 \sigma \int_\sigma^\infty [S](q)[I](q) dq = -\tau \frac{n}{N} \sigma ([S]_\infty - [S]_\sigma).$$

On the other hand, from Eq. (5.3a) we have $[S]_\sigma = [S]_0 e^{-\tau \frac{n}{N} \int_0^\sigma I(u) du}$, thus (from small fraction of initial infected nodes) we obtain the approximation $[S]_\sigma \approx [S]_0$. Therefore,

$$\ln \left(\frac{[S]_\infty}{[S]_0} \right) = \tau \frac{n}{N} \sigma [S]_0 \left(\frac{[S]_\infty}{[S]_0} - 1 \right),$$

which is equivalent to (5.13). □

In the following, we derive the final-size relation for the pairwise system (5.1)-(5.2).

Theorem 5.2.4. *The final size relation associated to the pairwise model (5.1)-(5.2) is*

$$\frac{\frac{1}{s_\infty^n} - 1}{\frac{1}{n-1}} = \mathcal{R}_0^p \left(\frac{n-1}{s_\infty^n} - 1 \right), \quad (5.14)$$

where the pairwise reproduction number $\mathcal{R}_0^p = \frac{n-1}{N} [S]_0 (1 - e^{-\tau\sigma})$, see details in Section 3.2.

Proof. As we have seen earlier, if we have an equation in the following form

$$w'(t) = \text{in}(t) - \text{out}(t)w(t) - \text{in}(t - \sigma)e^{-\int_{t-\sigma}^t \text{out}(u)du},$$

its solution can be expressed as

$$w(t) = \int_{t-\sigma}^t \text{in}(u)e^{-\int_u^t \text{out}(s)ds} du,$$

for $t > \sigma$. From Eq. (5.6), our setting is

$$\begin{aligned} \text{in}(t) &= \tau\kappa[S]^{\frac{n-2}{n}}(t)[SI](t), \\ \text{out}(t) &= \tau + \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)}, \end{aligned}$$

the equation for $[SI](t)$ is

$$[SI](t) = \int_{t-\sigma}^t \tau\kappa[S]^{\frac{n-2}{n}}(u)[SI](u)e^{-\int_u^t \tau + \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} du.$$

Applying $[\dot{S}](t) = -\tau[SI](t)$, we obtain

$$\begin{aligned} [SI](t) &= \int_{t-\sigma}^t \tau\kappa[S]^{\frac{n-2}{n}}(u)[SI](u)e^{-\int_u^t \tau + \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} du \\ &= -\int_{t-\sigma}^t \kappa[\dot{S}](u)[S]^{\frac{n-2}{n}}(u)e^{-\tau(t-u)}e^{\int_u^t \frac{n-1}{n} \frac{[\dot{S}](s)}{[S](s)} ds} du \\ &= -\int_{t-\sigma}^t \kappa[\dot{S}](u)[S]^{\frac{n-2}{n}}(u)e^{-\tau(t-u)}e^{\ln([S]^{\frac{n-1}{n}}(t)) - \ln([S]^{\frac{n-1}{n}}(u))} du \\ &= -\kappa[S]^{\frac{n-1}{n}}(t) \int_{t-\sigma}^t [S]^{-\frac{1}{n}}(u)[\dot{S}](u)e^{-\tau(t-u)} du. \end{aligned}$$

Substituting back to Eq. (5.2a), we get

$$[\dot{S}](t) = \tau\kappa[S]^{\frac{n-1}{n}}(t) \int_{t-\sigma}^t [S]^{-\frac{1}{n}}(u)[\dot{S}](u)e^{-\tau(t-u)} du.$$

Formally solving this scalar equation as a separable first-order differential equation, we derive

$$[S]^{1-\frac{n-1}{n}}(s) = [S]_0^{1-\frac{n-1}{n}} + \tau\kappa \left(1 - \frac{n-1}{n}\right) \int_0^s \int_{t-\sigma}^t [S]^{-\frac{1}{n}}(u)[\dot{S}](u)e^{-\tau(t-u)} dudt.$$

For the final size relation, we consider the equation at $s = \infty$ and obtain

$$[S]_\infty^{\frac{1}{n}} = [S]_0^{\frac{1}{n}} + \kappa \frac{\tau}{n} \int_0^\infty e^{-\tau t} \int_{t-\sigma}^t [S]^{-\frac{1}{n}}(u)[\dot{S}](u)e^{\tau u} dudt. \quad (5.15)$$

First, we compute the double integral:

$$\begin{aligned}
& \int_0^\infty e^{-\tau t} \int_{t-\sigma}^t [S]^{-\frac{1}{n}}(u) [\dot{S}](u) e^{\tau u} du dt \\
&= \int_0^\infty [S]^{-\frac{1}{n}}(u) [\dot{S}](u) e^{\tau u} \int_u^{u+\sigma} e^{-\tau t} dt du \\
&= \int_0^\infty [S]^{-\frac{1}{n}}(u) [\dot{S}](u) e^{\tau u} \left[\frac{e^{-\tau t}}{-\tau} \right]_u^{u+\sigma} du \\
&= -\frac{1}{\tau} \int_0^\infty [S]^{-\frac{1}{n}}(u) [\dot{S}](u) e^{\tau u} \left(e^{-\tau(u+\sigma)} - e^{-\tau u} \right) du \\
&= \frac{1}{\tau} (1 - e^{-\tau\sigma}) \int_0^\infty [S]^{-\frac{1}{n}}(u) [\dot{S}](u) du \\
&= \frac{1}{\tau} (1 - e^{-\tau\sigma}) \left[\frac{[S]^{\frac{n-1}{n}}(u)}{\frac{n-1}{n}} \right]_0^\infty \\
&= \frac{1}{\tau} (1 - e^{-\tau\sigma}) \frac{n}{n-1} \left([S]_\infty^{\frac{n-1}{n}} - [S]_0^{\frac{n-1}{n}} \right).
\end{aligned}$$

Plugging into (5.15) we obtain

$$[S]_\infty^{\frac{1}{n}} = [S]_0^{\frac{1}{n}} + \frac{\kappa}{n} \frac{n}{n-1} (1 - e^{-\tau\sigma}) \left([S]_\infty^{\frac{n-1}{n}} - [S]_0^{\frac{n-1}{n}} \right).$$

Thus,

$$\frac{[S]_\infty^{\frac{1}{n}} - [S]_0^{\frac{1}{n}}}{\frac{1}{n}} = \kappa \frac{n}{n-1} (1 - e^{-\tau\sigma}) \left([S]_\infty^{\frac{n-1}{n}} - [S]_0^{\frac{n-1}{n}} \right).$$

Using (5.7), we have:

$$\frac{[S]_\infty^{\frac{1}{n}} - [S]_0^{\frac{1}{n}}}{\frac{1}{n}} = \frac{n}{N} [S]_0^{\frac{2}{n}} (1 - e^{-\tau\sigma}) \left([S]_\infty^{\frac{n-1}{n}} - [S]_0^{\frac{n-1}{n}} \right).$$

Therefore, the relation

$$\frac{[S]_\infty^{\frac{1}{n}} - [S]_0^{\frac{1}{n}}}{\frac{1}{n}} = \frac{n}{N} (1 - e^{-\tau\sigma}) [S]_0^{\frac{n+1}{n}} \left(\frac{[S]_\infty^{\frac{n-1}{n}}}{[S]_0^{\frac{n-1}{n}}} - 1 \right),$$

holds, which can be written as

$$\frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n-1}} = \frac{n-1}{N} (1 - e^{-\tau\sigma}) [S]_0 \left(s_\infty^{\frac{n-1}{n}} - 1 \right).$$

□

Note, that taking the limit of $n \rightarrow \infty$ in (5.14) gives rise to $\ln(s_\infty) = \mathcal{R}_0^p(s_\infty - 1)$, which is equivalent to the ‘standard’ form of final size relations. Indeed, using L’Hospital’s rule, it is easy to see, that the limit of the left-hand side as $n \rightarrow \infty$ is

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n-1}} &= \lim_{n \rightarrow \infty} \frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n}} \frac{n-1}{n} = \lim_{n \rightarrow \infty} \frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n}} \\ &= \lim_{m \rightarrow 0} \frac{s_\infty^m - 1}{m} = \lim_{m \rightarrow 0} \frac{s_\infty^m \ln s_\infty}{1} \\ &= \ln s_\infty. \end{aligned}$$

Moreover, the implicit relation between final size and \mathcal{R}_0^p is conserved between the Markovian and non-Markovian DDE model (see Eq. (3.7)).

5.3 Numerical simulations

In Fig. 5.1(a,b) homogeneous (or regular random) and Erdős-Rényi random networks are considered, respectively. Here, the mean of 100 simulations is compared to the solution of system (5.2). The agreement is excellent for homogeneous networks, even for low degrees. Despite the pairwise model not explicitly accounting for degree heterogeneity, the agreement is surprisingly good for relatively dense Erdős-Rényi networks. The figure also shows that the fixed infectious period significantly accelerates the growth and turnover of the epidemic compared to the purely Markovian case.

In Fig. 5.1(c), the differences between simulations, mean-field and pairwise models for the non-Markovian case are compared. For denser networks, $\langle k \rangle = 15$, both models perform well with the pairwise yielding a better agreement. However, the difference is striking for sparser networks, $\langle k \rangle = 5$, where the mean-field approximation performs poorly, while the pairwise DDE model leads to good agreement with simulation, even in this case.

In Fig. 5.1(d), analytic final size relations are tested against simulation results for a range of different infectious period distributions, all sharing the same mean. Surprisingly, the final epidemic size can vary by as much as 15%, see $\tau \sim 0.083$, simply due to the recovery time distributions. The inset in Fig. 1(d) shows that the same value of \mathcal{R}_0^p produces the same attack rate, regardless of the distribution from where it originates from, in accordance with our formula (5.14).

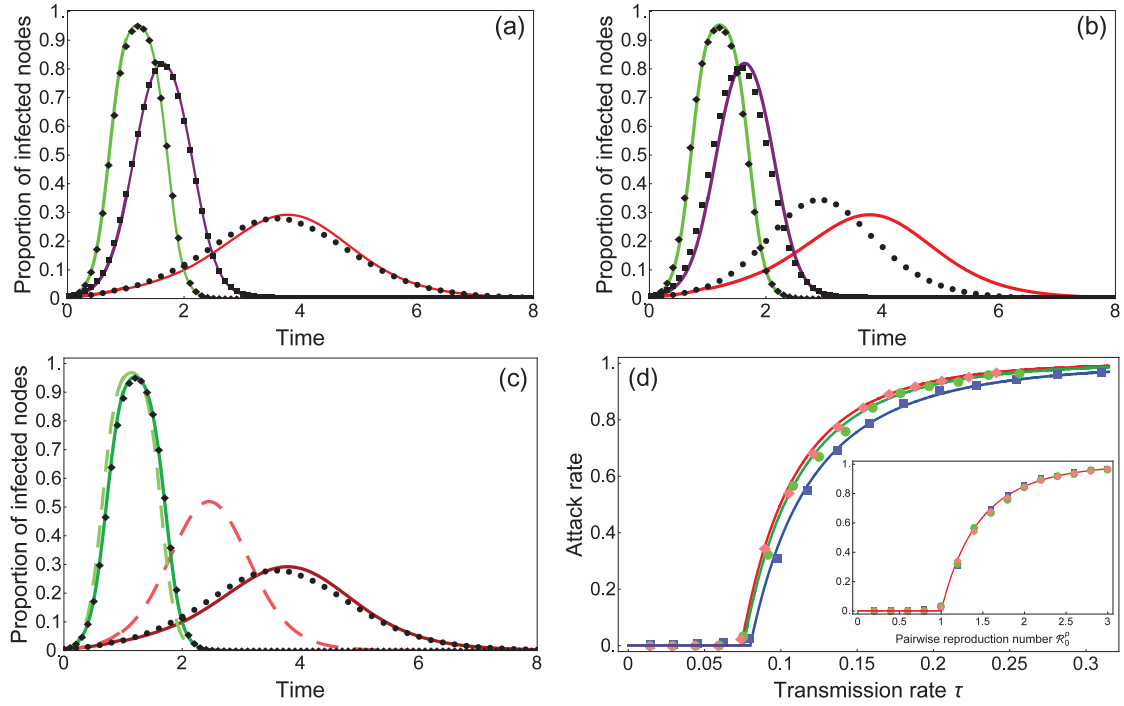


Figure 5.1: Simulations of non-Markovian epidemics on networks with $N = 1000$ nodes: (a) solid lines show the solution of (5.1)-(5.2) and the circles/squares/diamonds correspond to simulations for homogeneous (random regular) graphs with $\langle k \rangle = 5/10/15$, respectively; dotted, $\langle k \rangle = 5$, and dashed, $\langle k \rangle = 15$, lines correspond to purely Markovian epidemics given by (2.12); (b) the same as before but for Erdős-Rényi random graphs with $\langle k \rangle = 5/10/15$; (c) the solid and dashed lines show the solution of pairwise (5.1)-(5.2) and mean-field (5.4)-(5.4) models, respectively and, for regular random graphs with $\langle k \rangle = 5$ and $\langle k \rangle = 15$. For (a), (b) and (c) the transmission rate is $\tau = 0.55$ and the infectious period is fixed, $\sigma = 1$. Finally, (d) the diamonds/circles/squares correspond to simulations using regular random graphs with $\langle k \rangle = 15$ and using fixed and two different but gamma distributed infectious periods (\circ - shape $\alpha = 2$, scale $\beta = \frac{1}{2}$, \square - shape $\alpha = \frac{1}{2}$, scale $\beta = 2$), respectively. The solid lines correspond to the analytical final size for fixed (5.1)-(5.2) and general (5.14) infectious periods, with the dashed line denoting the purely Markovian case. The inset shows the analytical and the simulated final epidemic sizes plotted against the reproduction number.

Based on Table 3.1, the analytical expressions for \mathcal{R}_0^p are

$$\begin{aligned}\mathcal{R}_{0,\Gamma(\frac{1}{2},2)}^p &= \frac{(n-1)[S]_0}{N} \left(1 - \frac{1}{\sqrt{1+2\tau}}\right), \mathcal{R}_{0,\text{Exp}(1)}^p = \frac{(n-1)[S]_0}{N} \left(\frac{\tau}{\tau+1}\right), \\ \mathcal{R}_{0,\Gamma(2,\frac{1}{2})}^p &= \frac{(n-1)[S]_0}{N} \left(1 - \frac{4}{(2+\tau)^2}\right), \mathcal{R}_{0,\text{Fixed}(1)}^p = \frac{(n-1)[S]_0}{N} (1 - e^{-\tau}),\end{aligned}$$

where $\Gamma(a, b)$ denotes Gamma-distribution with parameters a and b , and satisfy the following inequality

$$\mathcal{R}_{0,\Gamma(\frac{1}{2},2)}^p \leq \mathcal{R}_{0,\text{Exp}(1)}^p \leq \mathcal{R}_{0,\Gamma(2,\frac{1}{2})}^p \leq \mathcal{R}_{0,\text{Fixed}(1)}^p. \quad (5.16)$$

We note that (a) all recovery time distributions have the same mean 1 and (b) the variances satisfy the converse inequality, with higher variance in recovery time (i.e. 2, 1, 1/2 and 0) giving a smaller \mathcal{R}_0^p value, despite τ being fixed. We have seen more details about this phenomenon in Ch. 4.

5.4 Conclusions

The proposed model provides a viable framework for a more systematic analysis of non-Markovian processes on networks with several future research directions. Similarly to the evolution of the original pairwise model for Markovian dynamics, the proposed model and new closure can be extended to networks with heterogeneous degree distribution [23], clustering or to directed and weighted networks. For example, we show how the current pairwise equations extend naturally to heterogeneous networks. In this case variables, such as

1. $[S_i](t)$ - expected number of susceptible nodes of degree i ,
2. $[I_i](t)$ - expected number of infected nodes of degree i ,
3. $[S_i S_j](t)$ - expected number of $S - S$ links, where S and S have degrees i and j , respectively,
4. $[S_i I_j](t)$ - expected number of $S - I$ links, where S and I have degrees i and j , respectively,

need to be considered, where $i, j \in \{k_{\min}, k_{\min}+1, \dots, k_{\max}\}$ represent the various degrees in the networks.

The slightly more technical part is replicating derivation of Eq. (5.2d) to degree dependent $[SI]$ pairs. This can be done as follows. Let $x(t)$ denote the factor by which $[S_i I_j]$ links needs to be discounted by. The equation for $x(t)$ is given by

$$\dot{x}(t) = -\tau \frac{i-1}{i} \frac{\sum_k [S_i I_k]}{[S_i]} x(t) - \tau x(t), \quad (5.17)$$

where in fact the factor x only depends on the degree of the susceptible node so it could be denoted by x_i . It is worth noting that

$$\frac{\sum_k [S_i I_k]}{i[S_i]}$$

gives the probability that a stub emanating from a susceptible node with i links will connect to an infected node, and $(i-1)$ stands for the remaining stubs emanating from an S_i node which is already connected to another node, in this case an infected node.

This can be integrated as before and the non-Markovian pairwise system for heterogeneous network yields

$$\begin{aligned} [\dot{S}_i](t) &= -\tau \sum_j [S_i I_j](t), \\ [\dot{I}_i](t) &= \tau \sum_j [S_i I_j](t) - \tau \sum_j [S_i I_j](t - \sigma), \\ [S_i \dot{S}_j](t) &= -\tau \frac{j-1}{j} \frac{[S_i S_j](t)}{[S_j](t)} \sum_k [S_j I_k](t) - \tau \frac{i-1}{i} \frac{[S_i S_j](t)}{[S_i](t)} \sum_k [S_i I_k](t), \\ [S_i \dot{I}_j](t) &= \tau \frac{j-1}{j} \frac{[S_i S_j](t)}{[S_j](t)} \sum_k [S_j I_k](t) - \tau \frac{i-1}{i} \frac{[S_i I_j](t)}{[S_i](t)} \sum_k [I_k S_i](t) - \tau [S_i I_j](t) \\ &\quad - \tau \sum_k \frac{j-1}{j} \frac{[S_i S_j](t - \sigma) [S_j I_k](t - \sigma)}{[S_j](t - \sigma)} e^{-\int_{t-\sigma}^t \left(\tau \frac{i-1}{i} \frac{\sum_k [S_i I_k](u)}{[S_i](u)} + \tau \right) du}. \end{aligned} \quad (5.18)$$

where $i, j, k \in \{k_{min}, k_{min} + 1, \dots, k_{max}\}$.

Additionally, this framework can be employed to model different dynamics, such as SIS epidemics or more complex systems, such as adaptive networks.

6

General recovery time

In this chapter, we present the generalised mean-field and pairwise models for non-Markovian epidemics on networks with any kind of recovery time distributions. A lengthy, but very instructive derivation gives a system of integro-differential equations, which is analysed both from a mathematical and numerical point of view. After the study of asymptotic behaviour, as an illustration of the applicability of the general model we recover known results.

6.1 The model

We want to build mean-field and pairwise models for the *SIR* type epidemic process with exponentially distributed transmission and general recovery time distribution. First, let $i(t, a)$ represent the density of infected nodes with respect to the age of infection a at the current time t , then $[I](t) = \int_0^\infty i(t, a) da$. Similarly, $Si(t, a)$ and $ISi(t, a)$ describe the density of $S - i$ links and $I - S - i$ triplets, respectively, where the infected node i has age a at time t and $[SI](t) = \int_0^\infty Si(t, a) da$, $[ISI](t) = \int_0^\infty ISi(t, a) da$. We assume that the infection process along $S - I$ links is Markovian with transmission rate $\tau > 0$. The recovery part is considered to be non-Markovian, with a cumulative distribution function $F_{\mathcal{I}}(a)$ and probability density function $f_{\mathcal{I}}(a)$. We use the associated survival function $\xi_{\mathcal{I}}(a) = 1 - F_{\mathcal{I}}(a)$ and hazard function $h_{\mathcal{I}}(a) = -\frac{\xi'_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a)} = \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a)}$.

Using the notations above, we arrive at the following model

$$[\dot{S}](t) = -\tau[SI](t), \quad (6.1a)$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) i(t, a) = -h_{\mathcal{I}}(a)i(t, a), \quad (6.1b)$$

$$[\dot{SS}](t) = -2\tau[SSI](t), \quad (6.1c)$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) Si(t, a) = -\tau ISi(t, a) - (\tau + h_{\mathcal{I}}(a))Si(t, a), \quad (6.1d)$$

subject to the boundary conditions

$$i(t, 0) = \tau[SI](t), \quad (6.2a)$$

$$Si(t, 0) = \tau[SSI](t), \quad (6.2b)$$

and initial conditions

$$[S](0) = [S]_0, [SS](0) = [SS]_0, i(0, a) = \varphi(a), \quad (6.3a)$$

$$Si(0, a) = \chi(a) \approx \frac{n}{N}[S]_0 i(0, a) = \frac{n}{N}[S]_0 \varphi(a). \quad (6.3b)$$

We shall use the biologically feasible assumption $\lim_{a \rightarrow \infty} \varphi(a) = 0$. To break the dependence on higher order moments, we apply the closure approximation formula (2.11) for $ISi(t, a)$ in the form

$$ISi(t, a) = \frac{n-1}{n} \frac{[SI](t)Si(t, a)}{[S](t)}. \quad (6.4)$$

To obtain a self-consistent system for classical network variables $[S]$, $[SS]$, $[I]$ and $[SI]$, further calculations are needed. Repeating Eq. (6.1a) and applying the moment-closure formula (2.11) to Eq. (6.1c), we have

$$\begin{aligned} [\dot{S}](t) &= -\tau[SI](t), \\ [\dot{SS}](t) &= -2\tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)}. \end{aligned}$$

Using $[I](t) = \int_0^\infty i(t, a)da$, from Eq. (6.1b) we obtain

$$\begin{aligned} [\dot{I}](t) &= \int_0^\infty \frac{\partial}{\partial t} i(t, b)db = \int_0^\infty \left(-h_{\mathcal{I}}(b)i(t, b) - \frac{\partial}{\partial b} i(t, b) \right) db \\ &= -\int_0^\infty h_{\mathcal{I}}(b)i(t, b)db - (i(t, \infty) - i(t, 0)) \\ &= -\int_0^\infty h_{\mathcal{I}}(b)i(t, b)db - i(t, \infty) + i(t, 0). \end{aligned} \quad (6.5)$$

Solving the first-order linear PDE (6.1b) along characteristic lines, we obtain

$$i(t, a) = \begin{cases} i(t - a, 0)e^{-\int_0^a h_{\mathcal{I}}(b)db}, & \text{if } t > a; \\ i(0, a - t)e^{-\int_{a-t}^a h_{\mathcal{I}}(b)db}, & \text{if } t \leq a. \end{cases}$$

Plugging (6.2a) and (6.3a) into the solution above, we have

$$i(t, a) = \begin{cases} \tau[SI](t - a)e^{-\int_0^a h_{\mathcal{I}}(b)db}, & \text{if } t > a; \\ \varphi(a - t)e^{-\int_{a-t}^a h_{\mathcal{I}}(b)db}, & \text{if } t \leq a. \end{cases} \quad (6.6)$$

Applying this formula for $[I](t) = \int_0^\infty i(t, a)da$, we find

$$[I](t) = \int_0^t \tau[SI](t - a)e^{-\int_0^a h_{\mathcal{I}}(b)db} da + \int_t^\infty \varphi(a - t)e^{-\int_{a-t}^a h_{\mathcal{I}}(b)db} da. \quad (6.7)$$

Finally, using that along the characteristic lines, $i(t, \infty) = i(0, \infty) = \varphi(\infty) = 0$ from the assumption, substituting (6.6) and the boundary condition (6.2a) into (6.5), we get

$$\begin{aligned} [\dot{I}](t) &= \tau[SI](t) - \int_0^t \tau[SI](t - a)h_{\mathcal{I}}(a)e^{-\int_0^a h_{\mathcal{I}}(b)db} da \\ &\quad - \int_t^\infty \varphi(a - t)h_{\mathcal{I}}(a)e^{-\int_{a-t}^a h_{\mathcal{I}}(b)db} da. \end{aligned} \quad (6.8)$$

Using the definition and properties of hazard function, we can deduce the following formulae:

$$e^{-\int_0^a h_{\mathcal{I}}(b)db} = \frac{\xi_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(0)} = \xi_{\mathcal{I}}(a), \quad (6.9a)$$

$$e^{-\int_{a-t}^a h_{\mathcal{I}}(b)db} = \frac{\xi_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a - t)}, \quad (6.9b)$$

$$h_{\mathcal{I}}(a)e^{-\int_0^a h_{\mathcal{I}}(b)db} = f_{\mathcal{I}}(a), \quad (6.9c)$$

$$h_{\mathcal{I}}(a)e^{-\int_{a-t}^a h_{\mathcal{I}}(b)db} = \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a - t)}. \quad (6.9d)$$

Applying these formulae to Eq.(6.7) and (6.8), we have

$$[I](t) = \int_0^t \tau[SI](t - a)\xi_{\mathcal{I}}(a)da + \int_t^\infty \varphi(a - t)\frac{\xi_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a - t)}da, \quad (6.10)$$

and

$$[\dot{I}](t) = \tau[SI](t) - \int_0^t \tau[SI](t - a)f_{\mathcal{I}}(a)da - \int_t^\infty \varphi(a - t)\frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a - t)}da.$$

To compute the equation for $[SI](t)$, we follow the calculation process above. First, applying (6.4) to Eq. (6.1d), we get

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) Si(t, a) = -\frac{\tau(n-1)}{n} \frac{[SI](t)}{[S](t)} Si(t, a) - (\tau + h_{\mathcal{I}}(a))Si(t, a). \quad (6.11)$$

Using $[SI](t) = \int_0^\infty Si(t, a)da$, from Eq. (6.11) we find

$$\begin{aligned}
[\dot{SI}](t) &= \int_0^\infty \frac{\partial}{\partial t} Si(t, a) da \\
&= \int_0^\infty \left(-\frac{\tau(n-1)}{n} \frac{[SI](t)}{[S](t)} Si(t, a) \right) da - \int_0^\infty (\tau + h_{\mathcal{I}}(a)) Si(t, a) da \\
&\quad - \int_0^\infty \frac{\partial}{\partial a} Si(t, a) da \\
&= -\tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau [SI](t) - \int_0^\infty h_{\mathcal{I}}(a) Si(t, a) da \\
&\quad - Si(t, \infty) + Si(t, 0). \tag{6.12}
\end{aligned}$$

We want to express the variable $Si(t, a)$ as a function of classical network variables. To achieve this, let us consider the following first-order PDE:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} \right) x(t, a) = -f_{\mathcal{I}}(t)x(t, a) - g(a)x(t, a)$$

with boundary conditions

$$x(t, 0) = \phi(t), \quad x(0, a) = \psi(a).$$

Solving along the characteristic lines $t - a = c$, we find that

$$x(t, a) = \begin{cases} \phi(t-a) e^{-\int_{t-a}^t f_{\mathcal{I}}(s) ds} e^{-\int_0^a g(b) db}, & \text{if } t > a; \\ \psi(a-t) e^{-\int_0^t f_{\mathcal{I}}(s) ds} e^{-\int_{a-t}^a g(b) db}, & \text{if } t \leq a. \end{cases} \tag{6.13}$$

In our case, $x(t, a) = Si(t, a)$, $f_{\mathcal{I}}(t) = \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)}$, $g(a) = \tau + h_{\mathcal{I}}(a)$, $\phi(t) = \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)}$, (from closure approximation (6.4)) and $\psi(a) = \frac{n}{N} [S]_0 \varphi(a)$, hence from Eq. (6.13) we get

$$Si(t, a) = \begin{cases} \tau \frac{n-1}{n} \frac{[SS](t-a)[SI](t-a)}{[S](t-a)} e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} e^{-\int_0^a \tau + h_{\mathcal{I}}(b) db}, & \text{if } t > a; \\ \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} e^{-\int_{a-t}^a \tau + h_{\mathcal{I}}(b) db}, & \text{if } t \leq a. \end{cases} \tag{6.14}$$

Again, along the characteristic lines we have $Si(t, \infty) = Si(0, \infty) = \chi(\infty) = 0$. Putting (6.14) into $[SI](t) = \int_0^\infty Si(t, a) da$, we obtain

$$\begin{aligned}
[SI](t) &= \int_0^t \tau \frac{n-1}{n} \frac{[SS](t-a)[SI](t-a)}{[S](t-a)} e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} e^{-\int_0^a \tau + h_{\mathcal{I}}(b) db} da \\
&\quad + \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} e^{-\int_{a-t}^a \tau + h_{\mathcal{I}}(b) db} da. \tag{6.15}
\end{aligned}$$

If we substitute (6.14) back to Eq.(6.12), we derive

$$\begin{aligned}
[\dot{S}I](t) &= \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \frac{\tau(n-1)}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau[SI](t) \\
&\quad - \int_0^t \tau \frac{n-1}{n} \frac{[SS](t-a)[SI](t-a)}{[S](t-a)} e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} \\
&\quad \times e^{-\int_0^a \tau + h_{\mathcal{I}}(b) db} h_{\mathcal{I}}(a) da \\
&\quad - \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} e^{-\int_{a-t}^a \tau + h_{\mathcal{I}}(b) db} h_{\mathcal{I}}(a) da.
\end{aligned} \tag{6.16}$$

Applying the formulae (6.9a)-(6.9d) for Eq. (6.15) and (6.16), we have

$$\begin{aligned}
[SI](t) &= \int_0^t \tau \frac{n-1}{n} \frac{[SS](t-a)[SI](t-a)}{[S](t-a)} e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \xi_{\mathcal{I}}(a) da \\
&\quad + \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \frac{\xi_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da,
\end{aligned} \tag{6.17}$$

and

$$\begin{aligned}
[\dot{S}I](t) &= \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau[SI](t) \\
&\quad - \int_0^t \tau \frac{n-1}{n} \frac{[SS](t-a)[SI](t-a)}{[S](t-a)} e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} f_{\mathcal{I}}(a) da \\
&\quad - \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da.
\end{aligned}$$

Putting together the results above, the pairwise system is the following integro-differential equation:

$$[\dot{S}](t) = -\tau[SI](t) \tag{6.18a}$$

$$[\dot{S}S](t) = -2\tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} \tag{6.18b}$$

$$[\dot{I}](t) = \tau[SI](t) - \int_0^t \tau[SI](t-a) f_{\mathcal{I}}(a) da - \int_t^\infty \varphi(a-t) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da \tag{6.18c}$$

$$\begin{aligned}
[\dot{S}I](t) &= \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau[SI](t) \\
&\quad - \int_0^t \tau \frac{n-1}{n} \frac{[SS](t-a)[SI](t-a)}{[S](t-a)} e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} f_{\mathcal{I}}(a) da \\
&\quad - \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da.
\end{aligned} \tag{6.18d}$$

From Eq.(6.18), the associated mean-field model can be easily deduced by using the closure approximation formula (2.7), the node-level system becomes

$$[\dot{S}](t) = -\tau \frac{n}{N} [S](t)[I](t) \quad (6.19a)$$

$$[\dot{I}](t) = \tau \frac{n}{N} [S](t)[I](t) - \int_0^t \tau \frac{n}{N} [S](t-a)[I](t-a) f_{\mathcal{I}}(a) da \\ - \int_t^\infty \varphi(a-t) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da. \quad (6.19b)$$

In the following, we investigate these systems from mathematical and numerical point of view, focussing on the epidemiologically meaningful properties of the models.

6.2 Analytical results

Next, similarly to Section 5.2, we summarize the biologically meaningful properties of systems (6.18) and (6.19). Note, that Prop. 5.2.1 holds for the pairwise system, thus we can reduce the model (6.18) to the following two-dimensional system:

$$[\dot{S}](t) = -\tau [SI](t), \\ [\dot{SI}](t) = \tau \kappa [S]^{\frac{n-2}{n}}(t) [SI](t) - \tau [SI](t) - \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) \\ - \int_0^t \tau \kappa [S]^{\frac{n-2}{n}}(t-a) [SI](t-a) e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} f_{\mathcal{I}}(a) da \\ - \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da, \quad (6.20)$$

where

$$\kappa = \frac{n-1}{N} [S]_0^{\frac{2}{n}}.$$

6.2.1 Positivity

The first proposition of this section states, that the solutions remain nonnegative provided that the initial conditions are nonnegative.

Proposition 6.2.1. *If initial conditions $[S]_0, [SS]_0$ are nonnegative and $\varphi(a) \geq 0$ for $a \geq 0$, then $[S](t) \geq 0, [SS](t) \geq 0, [I](t) \geq 0$ and $[SI](t) \geq 0$ hold for $t \geq 0$.*

Proof. It is clear, that $[SS](t)$ remains nonnegative, if the initial condition $[SS](0)$ is nonnegative, because $[SS](t)$ can be expressed in form (5.8) Moreover, if $[SS]_0$ is positive, then $[SS](t) > 0$ for all $t \geq 0$. From Eq. (5.5) we obtain that $[S](t)$ cannot be zero, if $[SS](t)$ is positive for all $t \geq 0$, which implies (from continuity of solutions) $[S](t) > 0$ for

$t \geq 0$. It can be seen from (6.10) that $[I](t)$ remains nonnegative if $[SI](t)$ is nonnegative for $t \geq 0$. On the other hand, $[SI](t_0)$ cannot be zero for some $t_0 \geq 0$, because the right-hand side of (6.17) depends on the $[S](t; t < t_0)$, $[SS](t; t < t_0)$ and $[S](t; t < t_0)$, which are positive, hence $[SI](t) > 0$. \square

6.2.2 Epidemic outbreak and reproduction numbers

As we have seen in Chapter 5, there is a relation between reproduction numbers and occurrence of epidemic outbreak. We investigated a transformed version of the mean-field and pairwise models by formal linearisation and proved, that the disease-free equilibrium is unstable if and only if the associated reproduction number is above one, thus epidemic occurs. This analysis could be performed, because we could apply the methods of stability theory of dynamical systems. Unfortunately, the general mean-field (6.19) and pairwise (6.18) models do not induce a dynamical system directly, thus stability analysis does not work. However, the relation is preserved for the general case, which will be shown in the following with some formal and intuitive calculations. First we can rewrite the mean-field model (6.19) to the form

$$\begin{aligned} [\dot{S}](t) &= -\tau \frac{n}{N} [S](t)[I](t), \\ [I](t) &= \int_0^t \tau \frac{n}{N} [S](t-a)[I](t-a) \xi_{\mathcal{I}}(a) da + \int_t^\infty \varphi(a-t) \frac{\xi_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da, \end{aligned} \quad (6.21)$$

and pairwise system (6.20) to the form

$$\begin{aligned} [\dot{S}](t) &= -\tau [SI](t), \\ [SI](t) &= \int_0^t \tau \frac{n-1}{N} [S]_0^{\frac{2}{n}} [S]^{\frac{n-2}{n}}(t-a) [SI](t-a) e^{-\int_{t-a}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \xi_{\mathcal{I}}(a) da \\ &\quad + \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \frac{\xi_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da. \end{aligned} \quad (6.22)$$

Following the procedure in [14], we may study the associated limit equations to investigate the asymptotic behaviour, which is

$$\begin{aligned} [\dot{S}](t) &= -\tau \frac{n}{N} [S](t)[I](t), \\ [I](t) &= \int_0^\infty \tau \frac{n}{N} [S](t-a)[I](t-a) \xi_{\mathcal{I}}(a) da, \end{aligned} \quad (6.23)$$

for (6.21) and

$$\begin{aligned} [\dot{S}](t) &= -\tau [SI](t), \\ [SI](t) &= \int_0^\infty \tau \frac{n-1}{N} [S]_0^{\frac{2}{n}} [S]^{\frac{n-2}{n}}(t-a) [SI](t-a) e^{-\tau a} \xi_{\mathcal{I}}(a) da, \end{aligned} \quad (6.24)$$

for (6.22). Then formal linearisation gives

$$\begin{aligned}\dot{s}(t) &= -\tau \frac{n}{N} [S]_0 i(t), \\ i(t) &= \int_0^\infty \tau \frac{n}{N} [S]_0 i(t-a) \xi_{\mathcal{I}}(a) da,\end{aligned}\tag{6.25}$$

for (6.23) and

$$\begin{aligned}\dot{s}(t) &= -\tau s i(t), \\ s i(t) &= \tau \frac{n-1}{N} [S]_0 \int_0^\infty s i(t-a) e^{-\tau a} \xi_{\mathcal{I}}(a) da\end{aligned}\tag{6.26}$$

for (6.24). Calculating the characteristic equations, the non-trivial roots are given by the equation

$$\mathcal{G}_m(\lambda) := \tau \frac{n}{N} [S]_0 \int_0^\infty e^{-\lambda a} \xi_{\mathcal{I}}(a) da = 1$$

for (6.25) and

$$\mathcal{G}_p(\lambda) := \tau \frac{n-1}{N} [S]_0 \int_0^\infty e^{-\tau a} e^{-\lambda a} \xi_{\mathcal{I}}(a) da = 1,$$

for (6.26). Applying the same argument as in the proof of Theorem 5.2.1, the positive real root appears, when $\mathcal{G}_m(0) > 1$ and $\mathcal{G}_p(0) > 1$, or equivalently

$$\begin{aligned}\mathcal{G}_m(0) &= \tau \frac{n}{N} [S]_0 \int_0^\infty \xi_{\mathcal{I}}(a) da = \tau \frac{n}{N} [S]_0 \int_0^\infty \int_a^\infty f_{\mathcal{I}}(v) dv da \\ &= \tau \frac{n}{N} [S]_0 \int_0^\infty \int_0^v f_{\mathcal{I}}(v) da dv \\ &= \tau \frac{n}{N} [S]_0 \int_0^\infty v f_{\mathcal{I}}(v) dv = \tau \frac{n}{N} [S]_0 \mathbb{E}(\mathcal{I}) = \mathcal{R}_0 > 1,\end{aligned}$$

for the mean-field approach and

$$\begin{aligned}\mathcal{G}_p(0) &= \tau \frac{n-1}{N} [S]_0 \int_0^\infty e^{-\tau a} \xi_{\mathcal{I}}(a) da = \tau \frac{n-1}{N} [S]_0 \int_0^\infty e^{-\tau a} \int_a^\infty f_{\mathcal{I}}(v) dv da \\ &= \tau \frac{n-1}{N} [S]_0 \int_0^\infty f_{\mathcal{I}}(v) \left(\int_0^v e^{-\tau a} da \right) dv \\ &= \tau \frac{n-1}{N} [S]_0 \int_0^\infty f_{\mathcal{I}}(v) \left[\frac{e^{-\tau a}}{-\tau} \right]_0^v dv \\ &= \tau \frac{n-1}{N} [S]_0 \int_0^\infty f_{\mathcal{I}}(v) \left(-\frac{1}{\tau} e^{-\tau v} + \frac{1}{\tau} \right) dv \\ &= \frac{n-1}{N} [S]_0 \left(\int_0^\infty f_{\mathcal{I}}(v) dv - \int_0^\infty f_{\mathcal{I}}(v) e^{-\tau v} dv \right) \\ &= \frac{n-1}{N} [S]_0 (1 - \mathcal{L}(f_{\mathcal{I}})[\tau]) = \mathcal{R}_0^p > 1.\end{aligned}$$

for the pairwise approximation.

6.2.3 Final size relation

In this section, we prove, that the functional forms (5.13) and (5.14) hold for arbitrary recovery time distribution.

Theorem 6.2.1. *The final size relation associated to the mean-field model (6.19) is*

$$\ln(s_\infty) = \mathcal{R}_0(s_\infty - 1), \quad (6.27)$$

where the basic reproduction number $\mathcal{R}_0 = \frac{n}{N}\tau[S]_0\mathbb{E}(\mathcal{I})$, see details in Section 3.2.

Proof. Similarly to proof for fixed recovery time, from (6.19a), we obtain

$$[S]_\infty - [S]_0 = -\tau \frac{n}{N} \int_0^\infty [S](u)[I](u)du \quad (6.28)$$

and

$$\ln\left(\frac{[S]_\infty}{[S]_0}\right) = -\tau \frac{n}{N} \int_0^\infty [I](u)du. \quad (6.29)$$

Substituting (6.19b) into (6.29), we get

$$\begin{aligned} \ln\left(\frac{[S]_\infty}{[S]_0}\right) &= -\tau \frac{n}{N} \int_0^\infty \int_0^u \tau \frac{n}{N} [S](u-a)[I](u-a)\xi_{\mathcal{I}}(a)dadu \\ &\quad -\tau \frac{n}{N} \int_0^\infty \int_u^\infty \varphi(a-u) \frac{\xi_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-u)} dadu. \end{aligned}$$

Neglecting the small initial amount of infected nodes ($\varphi(a) \approx 0$), we obtain

$$\ln\left(\frac{[S]_\infty}{[S]_0}\right) = -\left(\tau \frac{n}{N}\right)^2 \int_0^\infty \int_0^u [S](u-a)[I](u-a)\xi_{\mathcal{I}}(a)dadu. \quad (6.30)$$

After some algebraic manipulation, we obtain

$$\begin{aligned} \ln\left(\frac{[S](\infty)}{[S](0)}\right) &= -\left(\tau \frac{n}{N}\right)^2 \int_0^\infty \int_0^u [S](u-a)[I](u-a)\xi_{\mathcal{I}}(a)dadu \\ &= -\left(\tau \frac{n}{N}\right)^2 \int_0^\infty \int_0^u [S](v)[I](v)\xi_{\mathcal{I}}(u-v)dvdu \\ &= -\left(\tau \frac{n}{N}\right)^2 \int_0^\infty \int_v^\infty [S](v)[I](v)\xi_{\mathcal{I}}(u-v)dudv \\ &= -\left(\tau \frac{n}{N}\right)^2 \int_0^\infty [S](v)[I](v) \left[\int_v^\infty \xi_{\mathcal{I}}(u-v)du \right] dv \\ &= -\left(\tau \frac{n}{N}\right)^2 \int_0^\infty [S](v)[I](v) \left[\int_0^\infty \xi_{\mathcal{I}}(p)dp \right] dv \end{aligned}$$

$$\begin{aligned}
&= \tau \frac{n}{N} \left[\int_0^\infty \xi_{\mathcal{I}}(p) dp \right] ([S](\infty) - [S](0)) \\
&= \tau \frac{n}{N} \left[\int_0^\infty \int_p^\infty f_{\mathcal{I}}(q) dq dp \right] ([S](\infty) - [S](0)) \\
&= \tau \frac{n}{N} \left[\int_0^\infty \int_0^q f_{\mathcal{I}}(q) dp dq \right] ([S](\infty) - [S](0)) \\
&= \tau \frac{n}{N} \left[\int_0^\infty q f_{\mathcal{I}}(q) dq \right] ([S](\infty) - [S](0)) \\
&= \tau \frac{n}{N} \mathbb{E}(\mathcal{I}) ([S](\infty) - [S](0)).
\end{aligned}$$

where \mathcal{I} denotes the infectious period of an infected node. Therefore, we found

$$\ln(s_\infty) = \mathcal{R}_0 (s_\infty - 1),$$

where $\mathcal{R}_0 = \tau \frac{n}{N} \mathbb{E}(\mathcal{I}) [S]_0$. □

In the following, we derive the final-size relation for the pairwise system (6.18).

Theorem 6.2.2. *The final size relation associated to the pairwise model (6.18) is*

$$\frac{s_\infty^{\frac{1}{n-1}} - 1}{\frac{1}{n-1}} = \mathcal{R}_0^p \left(s_\infty^{\frac{n-1}{n}} - 1 \right),$$

where the pairwise reproduction number is $\mathcal{R}_0^p = \frac{n-1}{N} (1 - \mathcal{L}[f_{\mathcal{I}}](\tau)) [S]_0$, see details Section 3.2.

Proof. The second equation of the two-dimensional system (6.20) has the general form

$$x'(t) = a(t) - b(t)x(t),$$

where

$$\begin{aligned}
a(c) &= \tau \kappa [S]^{\frac{n-2}{n}}(c) [SI](c) \\
&\quad - \int_0^c \tau \kappa [S]^{\frac{n-2}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) e^{-\int_{c-a}^c \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} da \\
&\quad - \int_c^\infty \frac{n}{N} [S]_0 \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} e^{-\int_0^c \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} da, \\
b(w) &= \tau + \tau \frac{n-1}{n} \frac{[SI](w)}{[S](w)}, \\
x(t) &= [SI](t),
\end{aligned}$$

and has the solution

$$x(u) = e^{-\int_0^u b(w) dw} x(0) + \int_0^u e^{-\int_c^u b(w) dw} a(c) dc. \quad (6.31)$$

Using (6.1a), simple calculations give the relations

$$\begin{aligned} e^{-\int_0^u \tau + \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} &= e^{-\tau u} \frac{[S]^{\frac{n-1}{n}}(u)}{[S]^{\frac{n-1}{n}}(0)}, \\ e^{-\int_c^u \tau + \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} ds} &= e^{-\tau(u-c)} \frac{[S]^{\frac{n-1}{n}}(u)}{[S]^{\frac{n-1}{n}}(c)}. \end{aligned}$$

Using these relations, from (6.31) we get

$$\begin{aligned} [SI](u) &= \frac{[SI](0)}{[S]^{\frac{n-1}{n}}(0)} e^{-\tau u} [S]^{\frac{n-1}{n}}(u) \\ &+ \int_0^u \tau \kappa [S]^{-\frac{1}{n}}(c) [SI](c) e^{\tau c} e^{-\tau u} [S]^{\frac{n-1}{n}}(u) dc \\ &- \int_0^u \int_0^c \tau \kappa [S]^{-\frac{1}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) \times e^{-\tau a} e^{\tau c} e^{-\tau u} [S]^{\frac{n-1}{n}}(u) da dc \\ &- \int_0^u \int_c^\infty \frac{n}{N} [S]_0^{\frac{1}{n}} \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} e^{-\tau u} [S]^{\frac{n-1}{n}}(u) da dc. \end{aligned}$$

Then, substituting this formula into the first equation of (6.20), we find an equation in general form

$$[S]'(t) = -\tau A(t) [S]^{\frac{n-1}{n}}(t),$$

where

$$\begin{aligned} A(u) &= \frac{[SI](0)}{[S]^{\frac{n-1}{n}}(0)} e^{-\tau u} + \int_0^u \tau \kappa [S]^{-\frac{1}{n}}(c) [SI](c) e^{\tau c} e^{-\tau u} dc \\ &- \int_0^u \int_0^c \tau \kappa [S]^{-\frac{1}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) e^{-\tau a} e^{\tau c} e^{-\tau u} da dc \\ &- \int_0^u \int_c^\infty \frac{n}{N} [S]_0^{\frac{1}{n}} \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} e^{-\tau u} da dc. \end{aligned}$$

Solving this scalar equation, we have

$$[S]^{\frac{1}{n}}(t) = [S]_0^{\frac{1}{n}} - \frac{\tau}{n} \int_0^t A(u) du.$$

For the final size relation, we consider the following equation

$$[S]_\infty^{\frac{1}{n}} = [S]_0^{\frac{1}{n}} - \frac{\tau}{n} \int_0^\infty A(u) du.$$

Using the linearity of integration, we have to calculate the four integrals on the right-hand side. For the first integral, we have

$$I_1 = \int_0^\infty \frac{[SI](0)}{[S]^{\frac{n-1}{n}}(0)} e^{-\tau u} du = \frac{[SI](0)}{[S]^{\frac{n-1}{n}}(0)} \left[\frac{e^{-\tau u}}{-\tau} \right]_0^\infty = \frac{[SI](0)}{[S]^{\frac{n-1}{n}}(0)} \frac{1}{\tau}.$$

After some algebraic manipulation, we obtain the following expression for the second integral I_2 :

$$\begin{aligned}
I_2 &= \int_0^\infty \int_0^u \tau \kappa [S]^{-\frac{1}{n}}(c) [SI](c) e^{\tau c} e^{-\tau u} dc du \\
&= \int_0^\infty \int_c^\infty \tau \kappa [S]^{-\frac{1}{n}}(c) [SI](c) e^{\tau c} e^{-\tau u} du ddc \\
&= \int_0^\infty \tau \kappa [S]^{-\frac{1}{n}}(c) [SI](c) e^{\tau c} \left[\frac{e^{-\tau u}}{-\tau} \right]_c^\infty dc \\
&= \int_0^\infty \tau \kappa [S]^{-\frac{1}{n}}(c) [SI](c) e^{\tau c} \frac{e^{-\tau c}}{\tau} dc \\
&= -\frac{1}{\tau} \kappa \int_0^\infty [S]^{-\frac{1}{n}}(c) \dot{[S]}(c) dc = -\frac{1}{\tau} \kappa \frac{n}{n-1} \left([S]_\infty^{\frac{n-1}{n}} - [S]_0^{\frac{n-1}{n}} \right).
\end{aligned}$$

The most challenging one is the third integral I_3 :

$$\begin{aligned}
I_3 &= \int_0^\infty \int_0^u \int_0^c \tau \kappa [S]^{-\frac{1}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) e^{-\tau a} e^{\tau c} e^{-\tau u} da dc du \\
&= \int_0^\infty \int_c^\infty \int_0^c \tau \kappa [S]^{-\frac{1}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) e^{-\tau a} e^{\tau c} e^{-\tau u} da du dc \\
&= \int_0^\infty \int_c^\infty \int_0^c \tau \kappa [S]^{-\frac{1}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) e^{-\tau a} e^{\tau c} e^{-\tau u} du da dc \\
&= \int_0^\infty \int_0^c \tau \kappa [S]^{-\frac{1}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) e^{-\tau a} e^{\tau c} \left[\frac{e^{-\tau u}}{-\tau} \right]_c^\infty da dc \\
&= \frac{1}{\tau} \int_0^\infty \int_0^c \tau \kappa [S]^{-\frac{1}{n}}(c-a) [SI](c-a) f_{\mathcal{I}}(a) e^{-\tau a} da dc \\
&= -\frac{1}{\tau} \kappa \int_0^\infty f_{\mathcal{I}}(a) e^{-\tau a} \left[\frac{[S]^{-\frac{n-1}{n}}(c-a)}{\frac{n-1}{n}} \right]_a^\infty da \\
&= -\frac{1}{\tau} \kappa \frac{n}{n-1} \left([S]_\infty^{\frac{n-1}{n}} - [S]_0^{\frac{n-1}{n}} \right) \int_0^\infty f_{\mathcal{I}}(a) e^{-\tau a} da.
\end{aligned}$$

For the fourth integral, we compute

$$\begin{aligned}
I_4 &= \int_0^\infty \int_0^u \int_c^\infty \frac{n}{N} [S]_0^{\frac{1}{n}} \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} e^{-\tau u} da dc du \\
&= \int_0^\infty \int_c^\infty \int_c^\infty \frac{n}{N} [S]_0^{\frac{1}{n}} \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} e^{-\tau u} da du dc \\
&= \int_0^\infty \int_c^\infty \int_c^\infty \frac{n}{N} [S]_0^{\frac{1}{n}} \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} e^{-\tau u} du da dc \\
&= \int_0^\infty \int_c^\infty \frac{n}{N} [S]_0^{\frac{1}{n}} \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} \left[\frac{e^{-\tau u}}{-\tau} \right]_c^\infty da dc \\
&= \frac{1}{\tau} \int_0^\infty \int_c^\infty \frac{n}{N} [S]_0^{\frac{1}{n}} \varphi(a-c) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-c)} e^{-\tau c} da dc.
\end{aligned}$$

Having a small amount of initial infecteds (i.e. $[I](0) = \int_0^\infty \varphi(a)da \ll 1$), the integrals I_1 and I_4 are approximately zero. We arrived to the relation

$$[S]_\infty^{\frac{1}{n}} = [S]_0^{\frac{1}{n}} + \kappa \frac{1}{n-1} \left(1 - \int_0^\infty f_{\mathcal{I}}(a)e^{-\tau a} da \right) \left([S]_\infty^{\frac{n-1}{n}} - [S]_0^{\frac{n-1}{n}} \right).$$

After some algebraic manipulation and substituting back the formula of κ , we have

$$\frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n-1}} = \frac{n-1}{N} (1 - \mathcal{L}[f](\tau)) [S]_0 \left(s_\infty^{\frac{n-1}{n}} - 1 \right),$$

where $\mathcal{L}[f_{\mathcal{I}}](\tau)$ denotes the Laplace transform of $f_{\mathcal{I}}$, the PDF of recovery time at τ . \square

6.3 Numerical methods

For the numerical solution of integro-differential equations (6.18) and (6.19), we developed a numerical scheme based on collocation method. The numerical methods in [15] were adapted to the mean-field model and the reduced, but highly nonlinear pairwise system. The development of this scheme for more general integral equations is an active research topic nowadays ([1], [52], [88]).

Since we want to compare the result of stochastic simulation and the numerical solutions of deterministic models, we assume, that the initial infecteds are 'newborn', i.e. the initial distribution of infected nodes $\varphi(a) = [I]_0 \delta(a)$, where $\delta(a)$ is the Dirac delta function. Then,

$$\int_t^\infty \varphi(a-t) \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da = [I]_0 f_{\mathcal{I}}(t) \quad (6.32)$$

and

$$\begin{aligned} & \int_t^\infty \frac{n}{N} [S]_0 \varphi(a-t) e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \frac{f_{\mathcal{I}}(a)}{\xi_{\mathcal{I}}(a-t)} da \\ &= \frac{n}{N} [S]_0 [I]_0 e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} f_{\mathcal{I}}(t). \end{aligned} \quad (6.33)$$

A collocation solution u_h to a functional equation on an interval I is an element from some finite-dimensional function space (the collocation space) which satisfies the equation on an appropriate finite subset of points in I (the set of collocation points), whose cardinality essentially matches the dimension of the collocation space. For integro-differential equations the collocation equations are not yet in a form amenable to numerical computation, due to the presence of the memory term given by the integral operator, thus another discretisation step, based on appropriate quadrature approximations, is necessary to obtain the fully discretised collocation scheme. In the following, we derive the numerical algorithm for solving the general systems. First, we introduce the following notations:

- $I = [0, T]$, $\Delta = \{t_k \in I : 0 = t_1 < t_2 < t_3 < \dots < t_N = T\}$;
- $h_k = t_{k+1} - t_k$, $I_k = [t_k, t_{k+1}]$, ($k = 1, \dots, N - 1$);
- $0 \leq c_1 < c_2 \leq 1$ are the collocation parameters;
- P_2 : space of real polynomials of degree not exceeding 2.

The systems (6.19) and (6.20) are two-dimensional, thus the coordinates of $u_h(t)$ are elements of $S_1^{(0)}(I)$, where

$$S_2^{(0)}(I) = \left\{ p \in C(I, \mathbb{R}) : p|_{I_k} \in P_2, 0 \leq k \leq N - 1 \right\}.$$

Let $u_h(t) = (x_h(t), y_h(t))$. The quintessence of the collocation method is the local Lagrange representation for $x_h(t)$ and $y_h(t)$:

$$y'_h(t_k + vh_k) = L_1(v)Y_{k,1} + L_2(v)Y_{k,2} \quad (6.34)$$

where $0 \leq v \leq 1$ and

$$\begin{aligned} Y_{k,1} &= y'_h(t_k + c_1 h_k), \\ Y_{k,2} &= y'_h(t_k + c_2 h_k), \\ L_1(v) &= \frac{c_2 - v}{c_2 - c_1}, \\ L_2(v) &= \frac{v - c_1}{c_2 - c_1} \end{aligned} \quad (6.35)$$

thus we obtain

$$y_h(t_k + vh_k) = y_h(t_k) + h_k Y_{k,1} \beta_1(v) + h_k Y_{k,2} \beta_2(v), \quad (6.36)$$

where

$$\begin{aligned} \beta_1(v) &= \int_0^v L_1(v) dv = \frac{v(2c_2 - v)}{2(c_2 - c_1)}, \\ \beta_2(v) &= \int_0^v L_2(v) dv = \frac{v(v - 2c_1)}{2(c_2 - c_1)}. \end{aligned} \quad (6.37)$$

We consider (6.19) in the general form

$$\begin{aligned} x'(t) &= M_1(x(t), y(t)), \\ y'(t) &= M_2(t, x(t), y(t)) - \int_0^t \mathcal{M}(t - a, x(a), y(a)) da, \end{aligned} \quad (6.38)$$

where

$$\begin{aligned}
M_1(x, y) &= -\tau \frac{n}{N} xy, \\
M_2(x, y) &= \tau \frac{n}{N} xy - [I]_0 f_{\mathcal{I}}(t), \\
\mathcal{M}(t, x, y) &= \tau \frac{n}{N} f_{\mathcal{I}}(t) xy.
\end{aligned} \tag{6.39}$$

For pairwise model (6.20), we consider

$$\begin{aligned}
x'(t) &= P_1(x(t), y(t)), \\
y'(t) &= P_2(x(t), y(t)) - \int_0^t \mathcal{P}_1(t-a, x(a), y(a)) e^{-\int_a^t \mathcal{P}_2(x(s), y(s)) ds} da \\
&\quad - \mathcal{P}_3\left(t, \int_0^t \mathcal{P}_2(x(s), y(s)) ds\right),
\end{aligned} \tag{6.40}$$

where

$$\begin{aligned}
P_1(x, y) &= -\tau y, \\
P_2(x, y) &= \tau \frac{n-1}{N} [S]_0^{\frac{2}{n}} x^{\frac{n-2}{n}} y - \tau y - \tau \frac{n-1}{n} \frac{y^2}{x}, \\
\mathcal{P}_1(t, x, y) &= f_{\mathcal{I}}(t) \tau \frac{n-1}{N} x_0^{\frac{2}{n}} x^{\frac{n-2}{n}} y, \\
\mathcal{P}_2(x, y) &= \tau \frac{n-1}{n} \frac{y}{x} + \tau, \\
\mathcal{P}_3(t, z) &= \frac{n}{N} [S]_0 [I]_0 e^{-z} f_{\mathcal{I}}(t).
\end{aligned} \tag{6.41}$$

Considering the mean-field case, formally we have the following equations:

$$\begin{aligned}
x'_h(t_k + c_1 h_k) &= M_1(x_h(t_k + c_1 h_k), y_h(t_k + c_1 h_k)), \\
x'_h(t_k + c_2 h_k) &= M_1(x_h(t_k + c_2 h_k), y_h(t_k + c_2 h_k)), \\
y'_h(t_k + c_1 h_k) &= M_2(t_k + c_1 h_k, x_h(t_k + c_1 h_k), y_h(t_k + c_1 h_k)) \\
&\quad - \int_0^{t_k + c_1 h_k} \mathcal{M}(t_k + c_1 h_k - a, x_h(a), y_h(a)) da \\
y'_h(t_k + c_2 h_k) &= M_2(t_k + c_2 h_k, x_h(t_k + c_2 h_k), y_h(t_k + c_2 h_k)) \\
&\quad - \int_0^{t_k + c_2 h_k} \mathcal{M}(t_k + c_2 h_k - a, x_h(a), y_h(a)) da.
\end{aligned}$$

Using the Lagrange representation, we obtain

$$\begin{aligned}
X_{k,1} &= M_1(x_h(t_k) + h_k X_{k,1} \beta_1(c_1) + h_k X_{k,2} \beta_2(c_1), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_1) + h_k Y_{k,2} \beta_2(c_1)), \\
X_{k,2} &= M_1(x_h(t_k) + h_k X_{k,1} \beta_1(c_2) + h_k X_{k,2} \beta_2(c_2), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_2) + h_k Y_{k,2} \beta_2(c_2)), \\
Y_{k,1} &= M_2(t_k + c_1 h_k, x_h(t_k) + h_k X_{k,1} \beta_1(c_1) + h_k X_{k,2} \beta_2(c_1), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_1) + h_k Y_{k,2} \beta_2(c_1)) \\
&\quad - \int_0^{t_k + c_1 h_k} \mathcal{M}(t_k + c_1 h_k - a, x_h(a), y_h(a)) da, \\
Y_{k,2} &= M_2(t_k + c_2 h_k, x_h(t_k) + h_k X_{k,1} \beta_1(c_2) + h_k X_{k,2} \beta_2(c_2), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_2) + h_k Y_{k,2} \beta_2(c_2)) \\
&\quad - \int_0^{t_k + c_2 h_k} \mathcal{M}(t_k + c_2 h_k - a, x_h(a), y_h(a)) da.
\end{aligned}$$

Let us focus on the integral term; first, we can exploit the linearity of integral, thus we get for $\alpha \in \{1, 2\}$

$$\begin{aligned}
&\sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \mathcal{M}(t_k + c_\alpha h_k - a, x_h(a), y_h(a)) da \\
&\quad + \int_{t_k}^{t_k + c_\alpha h_k} \mathcal{M}(t_k + c_\alpha h_k - a, x_h(a), y_h(a)) da.
\end{aligned}$$

Applying substitutions $a = t_i + h_i b$ and $a = t_k + h_k b c_\alpha$, we have

$$\begin{aligned}
&\sum_{i=0}^{k-1} h_i \int_0^1 \mathcal{M}(t_k + c_\alpha h_k - (t_i + h_i b), x_h(t_i + h_i b), y_h(t_i + h_i b)) db \\
&\quad + c_\alpha h_k \int_0^1 \mathcal{M}(t_k + c_\alpha h_k - (t_k + b c_\alpha h_k), x_h(t_k + b c_\alpha h_k), y_h(t_k + b c_\alpha h_k)) db.
\end{aligned}$$

For integrals we apply the interpolatory quadrature formula using the abscissas based on the collocation parameters, i.e.

$$\int_0^1 \mathcal{M}(t_k + v h_k) dv \approx \mathcal{M}(t_k + c_1 h_k) \beta_1(1) + \mathcal{M}(t_k + c_2 h_k) \beta_2(1) \quad (6.42)$$

Notice, that

$$\int_0^1 \mathcal{M}(t_k + v c_\alpha h_k) dv \approx \mathcal{M}(t_k + c_1 c_\alpha h_k) \beta_1(1) + \mathcal{M}(t_k + c_2 c_\alpha h_k) \beta_2(1) dv. \quad (6.43)$$

Considering the pairwise case, we have the following equations:

$$\begin{aligned}
x'_h(t_k + c_1 h_k) &= P_1(x_h(t_k + c_1 h_k), y_h(t_k + c_1 h_k)), \\
x'_h(t_k + c_2 h_k) &= P_1(x_h(t_k + c_2 h_k), y_h(t_k + c_2 h_k)), \\
y'_h(t_k + c_1 h_k) &= P_2(x_h(t_k + c_1 h_k), y_h(t_k + c_1 h_k)) \\
&\quad - \int_0^{t_k + c_1 h_k} \mathcal{P}_1(t_k + c_1 h_k - a, x_h(a), y_h(a)) e^{-\int_a^{t_k + c_1 h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} da \\
&\quad - \mathcal{P}_3 \left(t_k + c_1 h_k, \int_0^{t_k + c_1 h_k} \mathcal{P}_2(x(s), y(s)) ds \right), \\
y'_h(t_k + c_2 h_k) &= P_2(x_h(t_k + c_2 h_k), y_h(t_k + c_2 h_k)) \\
&\quad - \int_0^{t_k + c_2 h_k} \mathcal{P}_1(t_k + c_2 h_k - a, x_h(a), y_h(a)) e^{-\int_a^{t_k + c_2 h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} da \\
&\quad - \mathcal{P}_3 \left(t_k + c_2 h_k, \int_0^{t_k + c_2 h_k} \mathcal{P}_2(x(s), y(s)) ds \right).
\end{aligned}$$

Using the Lagrange representation, we obtain

$$\begin{aligned}
X_{k,1} &= P_1(x_h(t_k) + h_k X_{k,1} \beta_1(c_1) + h_k X_{k,2} \beta_2(c_1), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_1) + h_k Y_{k,2} \beta_2(c_1)),
\end{aligned}$$

$$\begin{aligned}
X_{k,2} &= P_1(x_h(t_k) + h_k X_{k,1} \beta_1(c_2) + h_k X_{k,2} \beta_2(c_2), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_2) + h_k Y_{k,2} \beta_2(c_2)),
\end{aligned}$$

and

$$\begin{aligned}
Y_{k,1} &= P_2(t_k + c_1 h_k, x_h(t_k) + h_k X_{k,1} \beta_1(c_1) + h_k X_{k,2} \beta_2(c_1), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_1) + h_k Y_{k,2} \beta_2(c_1)) \\
&\quad - \int_0^{t_k + c_1 h_k} \mathcal{P}_1(t_k + c_1 h_k - a, x_h(a), y_h(a)) e^{-\int_a^{t_k + c_1 h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} da \\
&\quad - \mathcal{P}_3 \left(t_k + c_1 h_k, \int_0^{t_k + c_1 h_k} \mathcal{P}_2(x(s), y(s)) ds \right),
\end{aligned}$$

$$\begin{aligned}
Y_{k,2} &= P_2(t_k + c_2 h_k, x_h(t_k) + h_k X_{k,1} \beta_1(c_2) + h_k X_{k,2} \beta_2(c_2), \\
&\quad y_h(t_k) + h_k Y_{k,1} \beta_1(c_2) + h_k Y_{k,2} \beta_2(c_2)) \\
&\quad - \int_0^{t_k + c_2 h_k} \mathcal{P}_1(t_k + c_2 h_k - a, x_h(a), y_h(a)) e^{-\int_a^{t_k + c_2 h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} da \\
&\quad - \mathcal{P}_3 \left(t_k + c_2 h_k, \int_0^{t_k + c_2 h_k} \mathcal{P}_2(x(s), y(s)) ds \right).
\end{aligned}$$

Again, for the integral term we obtain

$$\begin{aligned} & \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \mathcal{P}_1(t_k + c_\alpha h_k - a, x_h(a), y_h(a)) e^{-\int_a^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} da \\ & + \int_{t_k}^{t_k+c_\alpha h_k} \mathcal{P}_1(t_k + c_\alpha h_k - a, x_h(a), y_h(a)) e^{-\int_a^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} da. \end{aligned}$$

Applying substitutions $a = t_i + h_i b$ and $a = t_k + h_k b c_\alpha$, we have

$$\begin{aligned} & \sum_{i=0}^{k-1} h_i \int_0^1 \mathcal{P}_1(t_k + c_\alpha h_k - (t_i + h_i b), x_h(t_i + h_i b), y_h(t_i + h_i b)) \\ & \quad \times e^{-\int_{t_i+h_i b}^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} db \\ & + c_\alpha h_k \int_0^1 \mathcal{P}_1(t_k + c_\alpha h_k - (t_k + b c_\alpha h_k), x_h(t_k + b c_\alpha h_k), y_h(t_k + b c_\alpha h_k)) \\ & \quad \times e^{-\int_{t_k+b c_\alpha h_k}^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds} db. \end{aligned}$$

By using the substitution $s = t_k + w h_k$, the exponent of the last term gets

$$\int_{t_k+b c_\alpha h_k}^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds = h_k \int_{c_\alpha b}^{c_\alpha} \mathcal{P}_2(x_h(t_k + w h_k), y_h(t_k + w h_k)) dw.$$

For $k \geq 3$ and $i < k - 1$, we can expand the exponents of the summands as follows:

$$\begin{aligned} \int_{t_i+h_i b}^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds & = \int_{t_i+h_i b}^{t_{i+1}} \mathcal{P}_2(x_h(s), y_h(s)) ds \\ & + \sum_{j=i+1}^{k-1} \int_{t_j}^{t_{j+1}} \mathcal{P}_2(x_h(s), y_h(s)) ds \\ & + \int_{t_k}^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds. \end{aligned}$$

Applying the recently used substitutions, we find

$$\begin{aligned} \int_{t_i+h_i b}^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds & = h_i \int_b^1 \mathcal{P}_2(x_h(t_i + h_i w), y_h(t_i + h_i w)) dw \\ & + \sum_{j=i+1}^{k-1} h_j \int_0^1 \mathcal{P}_2(x_h(t_j + h_j w), y_h(t_j + h_j w)) dw \\ & + c_\alpha h_k \int_0^1 \mathcal{P}_2(x_h(t_k + c_\alpha h_k w), y_h(t_k + c_\alpha h_k w)) dw. \end{aligned}$$

On the other hand, we can obtain

$$\begin{aligned} & \mathcal{P}_3 \left(t_k + c_\alpha h_k, \int_0^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds \right) \\ & = \mathcal{P}_3 \left(t_k + c_\alpha h_k, \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \mathcal{P}_2(x_h(s), y_h(s)) ds + \int_{t_k}^{t_k+c_\alpha h_k} \mathcal{P}_2(x_h(s), y_h(s)) ds \right) \\ & = \mathcal{P}_3 \left(t_k + c_\alpha h_k, \sum_{i=1}^{k-1} h_i \int_0^1 \mathcal{P}_2(x^{(i)}(w), y^{(i)}(w)) dw + c_\alpha h_k \int_0^1 \mathcal{P}_2(x^{(k)}(c_\alpha w), y^{(k)}(c_\alpha w)) ds \right), \end{aligned}$$

where

$$\begin{aligned} x^{(i)}(w) &= x_h(t_i + h_i w) = x_h(t_i) + h_i X_{i,1} \beta_1(w) + h_i X_{i,2} \beta_2(w) \\ y^{(i)}(w) &= y_h(t_i + h_i w) = y_h(t_i) + h_i Y_{i,1} \beta_1(w) + h_i Y_{i,2} \beta_2(w). \end{aligned}$$

Similarly to the interpolatory formulae (6.42)-(6.43), we can derive

$$\int_0^1 \mathcal{P}_1(t_k + v h_k) dv \approx \mathcal{P}_1(t_k + c_1 h_k) \beta_1(1) + \mathcal{P}_1(t_k + c_2 h_k) \beta_2(1)$$

and

$$\int_0^1 \mathcal{P}_1(t_k + v c_\alpha h_k) dv \approx \mathcal{P}_1(t_k + c_1 c_\alpha h_k) \beta_1(1) + \mathcal{P}_1(t_k + c_2 c_\alpha h_k) \beta_2(1) dv.$$

We implemented this recursive algorithm and solved the Eqs. (6.19) and (6.18) with it. In Fig. 6.1, homogeneous (or regular random) networks were considered and the average of 100 simulations is compared to the numerical solutions of mean-field (6.19) and pairwise (6.18) models. Several observations can be made: (a) the agreement of the simulation results with the numerical solution of pairwise model is excellent, and (b) the mean-field model, which largely ignores the network structure, performs poorly. This gives us great confidence that the generalised pairwise model can and will be used in different contexts as dictated by empirical or other theoretical studies.

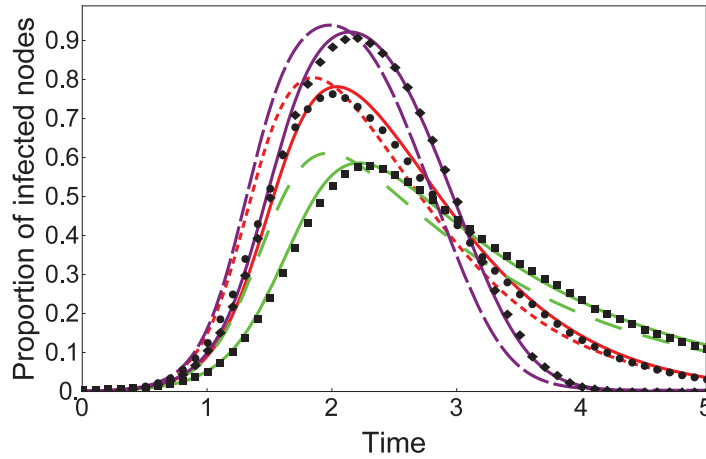


Figure 6.1: Stochastic and numerical experiments for non-Markovian epidemic with various recovery time distributions on homogeneous networks with $N = 1000$ nodes and infection rate $\tau = 0.35$. Squares, circles, diamonds show the mean of 100 simulations on random regular graphs with average degree $\langle k \rangle = 15$ for exponential distribution with parameter $\lambda = \frac{2}{3}$ (mean = $\frac{3}{2}$, variance = $\frac{9}{4}$), gamma distribution with shape $\alpha = 3$ and rate $\beta = 2$ (mean = $\frac{3}{2}$, variance = $\frac{3}{4}$), uniform distribution on interval $[a, b] = [1, 2]$ (mean = $\frac{3}{2}$, variance = $\frac{1}{12}$), respectively. Dashed and solid lines correspond to the numerical solution of the mean-field (6.19) and pairwise (6.18) models, respectively.

6.4 Special cases

In this section, we investigate some common choices for the recovery time. As we expect, if $\mathcal{I} \sim \text{Exp}(\gamma)$ (i.e. the infectious period \mathcal{I} is exponentially distributed), we get back the classical Markovian models (2.8) and (2.12). In the case of fixed recovery time, the models reduce to the systems (5.3)-(5.4) and (5.1)-(5.2). We can also recover the multi-stage infection model of [75] with gamma distributed recovery time. Finally, we consider $\mathcal{I} \sim \text{Uniform}(A, B)$ and write down the associated equations in a compact form. In this section we assume, that the initial infecteds are 'newborn', thus we use formulae (6.32)-(6.33).

6.4.1 Markovian recovery time: exponential distribution with parameter γ

The most widely used distribution in disease modelling is the exponential distribution. Both the stochastic and deterministic approaches exploit the memorylessness property to build tractable models. The resulting deterministic systems are ordinary differential equations with the advantage of relatively simple structure and numerical solvability. In the exponential case, $\xi_{\mathcal{I}}(t) = e^{-\gamma t}$ and $f_{\mathcal{I}}(t) = \gamma e^{-\gamma t}$. Using the assumption $\varphi(a) = [I]_0 \delta(a)$, (6.10) and $f_{\mathcal{I}}(t) = \gamma \xi_{\mathcal{I}}(t)$, from (6.18c) we obtain

$$[\dot{I}](t) = \tau[SI](t) - \gamma[I](t),$$

which gives the classical Markovian type pairwise equation for $[I](t)$. With similar arguments, from (6.18d) we obtain

$$[\dot{SI}](t) = \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau[SI](t) - \gamma[SI](t).$$

For the mean-field model (6.19), the same calculation gives the classical Markovian mean-field equation for $\dot{I}(t)$:

$$[\dot{I}](t) = \tau[S](t)[I](t) - \gamma[I](t).$$

6.4.2 Fixed recovery time σ

In several models, it is a reasonable assumption for the infectious period to have a fixed, constant duration, e.g. for measles [6]. In the case of fixed recovery time σ , we have

$$\xi_{\mathcal{I}}(t) = \begin{cases} 1 & \text{if } 0 \leq t < \sigma, \\ 0 & \text{if } t \geq \sigma, \end{cases}$$

and

$$f_I(t) = \delta(t - \sigma),$$

where $\delta(t)$ denotes the Dirac-delta function. Applying the fundamental property of $\delta(t)$, from (6.18c) and $\varphi(a) = [I]_0\delta(a)$ we have

$$[\dot{I}](t) = \tau[SI](t) - \begin{cases} 0 & \text{if } 0 \leq t < \sigma, \\ \tau[SI](t - \sigma) & \text{if } t \geq \sigma, \end{cases}$$

and from (6.18d), if $0 \leq t < \sigma$, we obtain

$$[\dot{S}I](t) = \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau[SI](t).$$

If $t > \sigma$, we get

$$\begin{aligned} [\dot{S}I](t) &= \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau[SI](t) \\ &\quad - \tau \frac{n-1}{n} \frac{[SS](t-\sigma)[SI](t-\sigma)}{[S](t-\sigma)} e^{-\int_{t-\sigma}^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds}. \end{aligned}$$

which is exactly the same system, that was studied in detail in Ch. 5. The mean-field model (5.3)-(5.4) can also be derived from (6.19b) using the same arguments.

6.4.3 Gamma distribution with shape $K \in \mathbb{Z}^+$ and rate $K\gamma$

The case of gamma distributed recovery time was studied in [75]. Using pairwise approximation with a standard closure, the authors have been able to analytically derive a number of important characteristics of disease dynamics. These included the final size of an epidemic and the epidemic threshold. Their results have shown that a higher number of disease stages, but with the same average duration of the infectious period, results in faster epidemic outbreaks with a higher peak prevalence and a larger final size of the epidemic. The pairwise model in [75] has the following equations for nodes:

$$\begin{aligned} [\dot{S}] &= -\tau \sum_{i=1}^K [SI_i], \\ [\dot{I}_1] &= \tau \sum_{i=1}^K [SI_i] - K\gamma[I_1] \\ [\dot{I}_j] &= K\gamma[I_{j-1}] - K\gamma[I_j], \quad j = 2, 3, \dots, K, \end{aligned} \tag{6.44}$$

where I_i , $i = 1, 2, \dots, K$ are the infectious stages, where nodes spend an exponentially distributed time with parameter $K\gamma$. The distribution of the total infectious period is

the sum of K exponential distributions with parameter $K\gamma$, which gives the gamma distribution with shape K and rate $K\gamma$ (thus the expected infectious period is $K \times 1/K\gamma = 1/\gamma$). Clearly, $[I](t) = \sum_{j=1}^K [I_j](t)$ and $[SI](t) = \sum_{i=1}^K [SI_i](t)$ and the sum of equations for infectious stages gives

$$[\dot{I}](t) = \tau[SI](t) - K\gamma[I_K](t).$$

On the other hand, using (6.32), the PDF and survival function of Gamma distribution

$$\begin{aligned} f_{\mathcal{I}}(a) &= \frac{(K\gamma)^K}{(K-1)!} a^{K-1} e^{-K\gamma a}, \\ \xi_{\mathcal{I}}(a) &= e^{-K\gamma a} \sum_{k=0}^{K-1} \frac{(K\gamma)^k}{k!} a^k, \end{aligned}$$

and inserting into (6.18c) and (6.10), we have

$$\begin{aligned} [\dot{I}](t) &= \tau[SI](t) - K\gamma \int_0^t \tau[SI](t-a) \frac{(K\gamma)^{K-1}}{(K-1)!} a^{K-1} e^{-K\gamma a} da \\ &\quad - K\gamma [I]_0 \frac{(K\gamma)^{K-1}}{(K-1)!} t^{K-1} e^{-K\gamma t} \end{aligned} \quad (6.45)$$

and

$$[I](t) = \sum_{k=0}^{K-1} \left(\int_0^t \tau[SI](t-a) \frac{(K\gamma)^k}{k!} a^k e^{-K\gamma a} da - [I]_0 \frac{(K\gamma)^k}{k!} t^k e^{-K\gamma t} \right). \quad (6.46)$$

These equations suggest the relations

$$\begin{aligned} [I_j](t) &= \int_0^t \tau[SI](t-a) \frac{(K\gamma)^{j-1}}{(j-1)!} a^{j-1} e^{-K\gamma a} da + [I]_0 \frac{(K\gamma)^{j-1}}{(j-1)!} t^{j-1} e^{-K\gamma t}, \\ &\quad j = 1, 2, \dots, K. \end{aligned} \quad (6.47)$$

To show this, we consider the equations for infectious stages in (6.44) as a first-order, linear differential equations with variation of constants formulae

$$[I_1](t) = [I_1](0) e^{-K\gamma t} + \int_0^t e^{-K\gamma(t-s)} \tau[SI](s) ds \quad (6.48)$$

and

$$\begin{aligned} [I_j](t) &= [I_j](0) e^{-K\gamma t} + \int_0^t e^{-K\gamma(t-s)} K\gamma [I_{j-1}](s) ds, \\ &\quad j = 2, 3, \dots, K. \end{aligned} \quad (6.49)$$

If all infecteds are newborn, we have $[I_1](0) = [I]_0$ and $[I_2](0) = [I_3](0) = \dots = [I_K](0) = 0$. Proceeding by induction yields that (6.47) satisfies (6.48) for $j = 1$ and (6.49) for $j = 2, 3, \dots, K$. Indeed, letting $j = 1$ in (6.47), we have

$$[I_1](t) = \int_0^t \tau[SI](t-a)e^{-K\gamma a} da + [I]_0 e^{-K\gamma t},$$

which comes directly from (6.48). Assuming that (6.47) holds for $1 < j$, we prove that it holds for $j + 1$. Indeed, we can do the following elaboration:

$$\begin{aligned} [I_{j+1}](t) &= [I_{j+1}](0)e^{-K\gamma t} + \int_0^t e^{-K\gamma(t-s)} K\gamma [I_j](s) ds \\ &= \int_0^t e^{-K\gamma(t-s)} K\gamma \int_0^s \tau[SI](s-a) \frac{(K\gamma)^{j-1}}{(j-1)!} a^{j-1} e^{-K\gamma a} da ds \\ &\quad + \int_0^t e^{-K\gamma(t-s)} K\gamma [I]_0 \frac{(K\gamma)^{j-1}}{(j-1)!} s^{j-1} e^{-K\gamma s} ds \\ &= \int_0^t \frac{(K\gamma)^j}{(j-1)!} e^{-K\gamma(t-s)} \left(\int_0^s \tau[SI](s-a) a^{j-1} e^{-K\gamma a} da \right) ds \\ &\quad + [I]_0 \frac{(K\gamma)^j}{(j-1)!} e^{-K\gamma t} \int_0^t s^{j-1} ds \\ &= \int_0^t \frac{(K\gamma)^j}{(j-1)!} e^{-K\gamma(t-s)} \left(\int_0^s \tau[SI](u) (s-u)^{j-1} e^{-K\gamma(s-u)} du \right) ds \\ &\quad + [I]_0 \frac{(K\gamma)^j}{j!} t^j e^{-K\gamma t} \\ &= \int_0^t \frac{(K\gamma)^j}{(j-1)!} e^{-K\gamma(t-u)} \tau[SI](u) \left(\int_u^t (s-u)^{j-1} ds \right) du \\ &\quad + [I]_0 \frac{(K\gamma)^j}{j!} t^j e^{-K\gamma t} \\ &= \int_0^t \tau[SI](u) \frac{(K\gamma)^j}{j!} (t-u)^j e^{-K\gamma(t-u)} du + [I]_0 \frac{(K\gamma)^j}{j!} t^j e^{-K\gamma t} \\ &= \int_0^t \tau[SI](t-a) \frac{(K\gamma)^j}{j!} a^j e^{-K\gamma a} da + [I]_0 \frac{(K\gamma)^j}{j!} t^j e^{-K\gamma t}. \end{aligned}$$

It is analogous to derive the equations for $[SI_j](t)$.

6.4.4 Uniform distribution on interval $[A, B]$

The uniform distribution is one of the most natural probability distributions and preferred in agent-based modeling [51], and was applied also for avian influenza [87]. Let the recovery time be distributed uniformly on interval $[A, B]$ (we assume $0 < A < B$), i.e.,

$$f_{\mathcal{I}}(t) = \begin{cases} \frac{1}{B-A} & \text{if } t \in (A, B), \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\xi_{\mathcal{I}}(t) = \begin{cases} 1 & \text{if } t \leq A, \\ \frac{B-t}{B-A} & \text{if } t \in (A, B), \\ 0 & \text{if } t \geq B. \end{cases}$$

We have to study the three cases $t < A$, $A < t < B$ and $t > B$. Writing the equation for $[\dot{I}](t)$, we have (after changing the variable):

$$[\dot{I}](t) = \tau[SI](t) - \begin{cases} 0 & \text{if } t < A, \\ \int_0^{t-A} \frac{\tau[SI](u)}{B-A} du + \frac{[I]_0}{B-A} & \text{if } t \in [A, B], \\ \int_{t-B}^{t-A} \frac{\tau[SI](u)}{B-A} du & \text{if } t > B. \end{cases}$$

With a more compact notation,

$$[\dot{I}](t) = \tau[SI](t) - \int_{\max(0, t-B)}^{\max(0, t-A)} \frac{\tau[SI](u)}{B-A} du - \frac{[I]_0}{B-A} \iota_{[A, B]}(t),$$

where $\iota_{[A, B]}(t)$ is the indicator function of interval $[A, B]$. The same argument gives

$$\begin{aligned} [\dot{S}I](t) &= \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)}{[S](t)} [SI](t) - \tau[SI](t) \\ &\quad - \int_{\max(0, t-B)}^{\max(0, t-A)} \frac{\tau}{B-A} \frac{n-1}{n} \frac{[SS](u)[SI](u)}{[S](u)} e^{-\int_u^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} du \\ &\quad - \frac{n}{N} [S]_0 e^{-\int_0^t \tau \frac{n-1}{n} \frac{[SI](s)}{[S](s)} + \tau ds} \frac{[I]_0}{B-A} \iota_{[A, B]}(t). \end{aligned}$$

For $t > B$ the model becomes a system of differential equations with distributed delays.

6.5 Conclusions

The generalised pairwise model provides a description of a possible deterministic approximation of non-Markovian epidemic processes on networks. The integro-differential system, which describes the dynamics at the level of nodes and links, is a powerful tool for investigating the classical quantities of an *SIR*-type epidemic, such as the reproduction number and final epidemic size. The generalised model is more challenging to analyse due to its complexity but it largely relies on tools from the theory of integro-differential equations. Further extensions of the model could focus on relaxing the assumption of homogeneous networks and extend the model to networks with heterogeneous degree distribution, see for example [76, 78], or to consider modelling the situation where both the infectious and recovery processes are non-Markovian. With the model proposed we wanted to emphasise opportunities to frame problems and models of network epidemics in more rigorous mathematical terms and use existing mathematical theory to enhance

our understanding of stochastic processes on networks and their average behaviour as captured by mean-field models.

7

Summary

Networks (or graphs) offer a flexible framework to explicitly incorporate various heterogeneities in how individuals within a population interact. This framework has led to a number of models where the strong assumptions of random mixing of the classical compartmental models can be relaxed. Because of the flexibility of the network approach, nodes can represent not only single individuals but also groups of individuals or locations. Similarly, links can represent contacts between individuals along which diseases can spread, or interactions between groups such as flight routes between different locations.

Most *SIR* (susceptible-infected-recovered) models on networks assume that both the disease transmission and recovery process are Markovian. The assumption of Markovianity is a strong simplifying assumption, as especially in the context of epidemiology, the period of infectiousness has paramount importance, and often this is approximated from the empirical distribution of observed infectious periods of various diseases by non-exponential distributions. Recently, however there is renewed interest in modelling non-Markovian processes, such as epidemics on networks. A possible modelling approach involves mean-field approximations, which are based on the classical compartmental principles and pairwise models, which have been very successful in capturing the average behaviour of a stochastic epidemics on networks.

This thesis aims to extend the pairwise model from Markovian to non-Markovian epidemic dynamics where the infection process remains Markovian but the infectious period is taken from an arbitrary distribution. In addition, we want to perform the full mathematical analysis of the resulting systems, with focus on the positivity of solutions, associated reproduction numbers and the implicit relation concerning the final epidemic size and implement explicit stochastic simulations and numerical solvers to test the validity of these models.

In Chapter 3, we introduce the new concept of reproduction numbers for mean-field

and pairwise models. In this systematic approach, we give analytical formulae for the basic reproduction number \mathcal{R}_0 of mean-field type systems and the pairwise reproduction number \mathcal{R}_0^p of link-level models. As an illustration, we calculate these key parameters for Markovian and fixed recovery times. In the last part, we summarise the general forms of implicitly given final size equations for both type of models.

It can be easily seen, while in general \mathcal{R}_0 depends on the expected value only, the pairwise reproduction number \mathcal{R}_0^p uses the complete density function, thus the average length of infectious period does not determine exactly the reproduction number. We study how the distribution of infectious periods influences the dynamics of epidemics on networks in Chapter 4. From studying typical families of distributions, we obtain the following results:

- For gamma and uniformly distributed recovery times it is shown, that higher variance in the recovery times generates lower reproduction numbers and different epidemic curves within each distribution family.
- The same phenomenon is numerically evidenced for lognormal distribution.
- We prove, that lower variance produces higher reproduction number in the general case.
- By comparing epidemics generated by different types of the recovery time distributions, we illustrate that estimating the expected value and the variance of the recovery time is not sufficient to give a realistic description of the epidemics.

The first generalisation of mean-field and pairwise models is presented in Chapter 5. For the case of infectious periods of fixed length, the resulting pairwise model is a system of delay differential equations, which shows excellent agreement with results based on stochastic simulations. The proposed framework includes the following results:

- The derivation of the pairwise model from first principles is illustrated. The term associated to the recovery is determined by an evolution equation and results in a differential equation with discrete and distributed delays.
- Investigating the basic properties of the systems, an invariant can be found and positivity of the solutions can be shown.
- An important relation between reproduction numbers and epidemic outbreaks can be explored by linearisation and study of the characteristic roots. The condition for an epidemic is exactly the same as in the Markovian case.

- The final size relations can be derived for both mean-field and pairwise models via lengthy calculations. The results are consistent with the classical functional form of final size relation.
- Stochastic simulations for the epidemic process are performed and compared with the numerical solution of the deterministic models and very good agreement is found.
- A possible extension for more heterogeneous networks is proposed.

The generalised mean-field and pairwise models for non-Markovian epidemics on networks with arbitrary recovery time distributions are presented in Chapter 6. The main section of this dissertation involves a wide spectrum of analytical and numerical results proposing a systematic framework to investigate non-Markovian *SIR* models on networks. We have the following results:

- Considering a first-order partial differential equation, where the population of infective nodes and links are structured by age since infection, the model is transformed into a system of integro-differential equations.
- Positivity of the solutions and condition for a disease outbreak is obtained. An intuitive picture is given about the role of the reproduction number in stability analysis of equilibria of linearised equations.
- Complete mathematical analysis is performed to obtain the general functional form of final size relations. The relation between final size and reproduction number is exactly described.
- A numerical scheme based on collocation method of Volterra integro-differential equations is developed and implemented to solve the mean-field and pairwise systems.
- As an illustration of the applicability of the general model we recover known results for the fixed, exponentially and gamma distributed recovery times and obtain new pairwise models with uniformly distributed infectious period.

The general framework that we proposed shows a more complete picture of the impact of non-Markovian recovery on network epidemics. With the model proposed we wanted to emphasise opportunities to frame problems and models of network epidemics in more rigorous mathematical terms and use existing and new mathematical theories to enhance our understanding of stochastic processes on networks.

This dissertation is based on three papers with co-authors István Kiss and Gergely Röst. These publications are the following:

- Kiss, I.Z., Röst, G. and Vizi, Z., 2015. Generalization of pairwise models to non-Markovian epidemics on networks. *Physical review letters*, 115(7), p.078701. <http://dx.doi.org/10.1103/PhysRevLett.115.078701>
- Röst, G., Vizi, Z. and Kiss, I.Z., 2015. Impact of non-Markovian recovery on network epidemics. In *Biomat 2015: Proceedings of the International Symposium on Mathematical and Computational Biology*
- Röst, G., Vizi, Z. and Kiss, I.Z., 2016. Pairwise approximation for SIR type network epidemics with non-Markovian recovery. *arXiv preprint arXiv:1605.02933*.

Further publications of the author are the following:

- Röst, G. and Vizi, Z., 2014. Backward bifurcation for pulse vaccination. *Nonlinear Analysis: Hybrid Systems*, 14, pp.99-113.
- Barbarossa, M.V., Dénes, A., Kiss, G., Nakata, Y., Röst, G. and Vizi, Z., 2015. Transmission dynamics and final epidemic size of Ebola virus disease outbreaks with varying interventions. *PloS one*, 10(7), p.e0131398.

8

Összefoglalás

A hálózatok (vagy gráfok) olyan struktúrák, amelyek lehetőséget nyújtanak a populáció tagjai közötti kapcsolatok heterogenitásának feltárására, modellekbe történő beépítésére. Számos modell, amely korábban feltételezte a klasszikus rekeszrendszeres megközelítéshez tartozó egyenletes (homogén) keveredést, a hálózatok alkalmazásával fejlődött tovább. A struktúra rugalmasságát fémjelzi, hogy a csúcsok az egyedek mellett csoportokat és területeket is reprezentálhatnak. Hasonlóan, az összeköttetések (élek) az egyedek közötti kapcsolatokon túl, csoportok közötti interakciót és földrajzi helyek közötti útvonalakat is jelenthetnek.

A legtöbb *SIR* (fogékony (susceptible)-fertőzött (infected)-felgyógyult (recovered) típusú hálózatos járványterjedési modell a fertőzési és felgyógyulási folyamatot egyaránt Markovinak (memória nélkülinek) feltételezi. Ez a feltétel túlságosan erősnek bizonyul, például az epidemiológia esetében, ahol kimagasló jelentőséggel bír a fertőzési periódus eloszlása és számtalan esetben a mérési adatokból nyert empirikus eloszlásokat nem-exponenciális eloszlással közelítik. Napjainkban újra előtérbe került a nem-Markovi folyamatok vizsgálata, különösen járványterjedés hálózaton történő modellezésénél. Egy lehetséges megközelítés az átlag-tér (mean-field) és páronkénti (pairwise) modellek alkalmazása: míg előbbi a rekeszrendszerek elvén alapul, utóbbival eredményesen közelíthetők a hálózatos sztochasztikus járványterjedési folyamatok.

A disszertáció célkitűzése a páronkénti modellek kidolgozása nem-Markovi járványterjedési dinamikára, ahol a fertőzés Markovi marad, de a felgyógyulási folyamatban a fertőzéstől a felgyógyulásig eltelt idő tetszőleges eloszlású lehet. A kapott rendszerek teljes analízise esetén a megoldások pozitívítására, a modellekhez tartozó reprodukciós számokra és a végállapot-egyenletekre összpontosítunk. A modellek érvényességét egyrészt explicit sztochasztikus szimulációk, másrészt numerikus megoldó algoritmusok implementálásával vizsgáljuk.

A 3. fejezetben bevezetjük az átlag-tér és páronkénti modellekhez tartozó reprodukciós számok fogalmát. A szisztematikus felépítés lehetőséget ad analitikus formulák levezetésére az átlag-tér típusú modellek \mathcal{R}_0 alap reprodukciós száma és a kapcsolat szintű modellek \mathcal{R}_0^p páronkénti reprodukciós száma esetén. Példaként levezetjük az exponenciális eloszlású és konstans felgyógyulási időtartamhoz tartozó formulákat. A fejezet utolsó részében összegezzük az implicit módon felírt végállapot egyenletek funkcionális alakjára vonatkozó eredményeket mindkét modell esetében.

Ezekből a kifejezésekből könnyen látható, hogy míg \mathcal{R}_0 csak a fertőzési periódus hosszának várható értékétől függ, addig az \mathcal{R}_0^p páronkénti reprodukciós szám kiszámításához a teljes sűrűségfüggvényre szükség van, így az átlagos periódusidő nem határozza meg egyértelműen \mathcal{R}_0^p értékét. A 4. fejezetben azt vizsgáljuk, hogy a fertőzési időtartam hosszának eloszlása hogyan befolyásolja a páronkénti reprodukciós számot. A jellegzetes eloszláscsaládokat vizsgálva a következő eredményeket kapjuk:

- Gamma és egyenletes eloszlású felgyógyulási időtartamokra belátjuk, hogy a nagyobb varianciájú periódushoz alacsonyabb reprodukciós szám és más járványgörbe tartozik.
- Ezt a jelenséget numerikusan vizsgáljuk a lognormális eloszlás esetére.
- Bebizonyítjuk az általános esetben is, hogy az azonos várható értékkel rendelkező eloszlások esetén kisebb variancia nagyobb reprodukciós számot generál.
- Az azonos várható értékkel és varianciával rendelkező eloszlásokra generált járványgörbék összehasonlításával illusztráljuk, hogy a várható érték és a variancia ismerete önmagában még nem elégséges a járvány teljeskörű leírására.

Az átlag-tér és páronkénti modellek első általánosítását a 5. fejezetben tárgyaljuk. A rögzített fertőzési periódus esetén a modellek késleltetett differenciálegyenletekből álló rendszereket eredményeznek, amelyek megoldásai kiváló egyezést mutatnak a sztochasztikus szimulációkból kapott eredményekkel. A fejezet a következő eredményeket tartalmazza:

- Levezetjük a páronkénti modellt, amelyben a felgyógyulást leíró tagot egy populáció egyedszámának változását leíró differenciálegyenletből kapjuk. A felírt modell konstans és eloszlásbeli (distributed) késleltetéseket tartalmaz.
- A modell alapvető tulajdonságait vizsgálva felírható egy első integrál és megmutatható a megoldások pozitivitása.

- A járvány kitörése és a reprodukciós számok közötti fontos kapcsolatot egy tételben mondjuk ki, amelyben a rendszerek linearizáltjának karakterisztikus gyökeit vizsgáljuk. A felírt feltétel megegyezik a Markovi esetben levezethetővel.
- A végállapot-egyenletet mind az átlag-tér, mind a páronkénti modellre levezetjük. Az eredmények konzisztensek a végállapot egyenletek klasszikus funkcionális alakjára vonatkozó eredményekkel.
- A járványterjedési folyamatra lefuttatott szimulációkat összehasonlítjuk a determinisztikus modellek numerikus megoldásával és kiváló egyezést tapasztalunk.
- Végül felírjuk a modell erősen heterogén hálózatok egy családjára vonatkozó általánosítását.

Az átlag-tér és páronkénti modellek tetszőleges felgyógyulási időtartamra történő általánosítását a 6. fejezetben tárgyaljuk. A disszertáció legfontosabb fejezete az analitikus és numerikus eredmények széles spektrumát tartalmazza, szisztematikusan vizsgáljuk a nem-Markovi *SIR* modellekkel kapcsolatos következő, fontos eredményeket:

- A kiindulásként tekintett parciális differenciálegyenletet, ahol a fertőzött csúcsok és összeköttetések a fertőződéstől eltelt idő szerint is struktúrálva vannak, egy integro-differenciálegyenletté transzformáljuk.
- Bizonyítjuk a megoldások pozitivitását és feltételt adunk a járvány kitörésének bekövetkezésére. Egy intuitív levezetésű kapcsolatot teremtünk a reprodukciós számok és a linearizált egyenletek egyensúlyi helyzetének stabilitása között.
- A végállapot-egyenletek levezetése lehetőséget ad a reprodukciós számok és a végállapot közötti kapcsolat felírására.
- A Volterra típusú integro-differenciálegyenletekhez kidolgozott kollokációs módszeren (collocation method) alapuló numerikus sémát fejlesztünk az átlag-tér és a páronkénti modellek numerikus megoldására.
- Néhány speciális esetet vizsgálva az általános rendszerekből levezetjük a már létező modelleket (Markovi, exponenciális és gamma), illetve egy új modellt írunk fel egyenletes eloszlású felgyógyulási időtartam esetén.

A leírt módszerek teljesebb képet adnak a nem-Markovi típusú felgyógyulás hálózaton történő járványterjedésre gyakorolt hatásáról. A modellek a hálózati járványterjedéssel

kapcsolatos problémák matematikailag is precízebb tárgyalására adnak lehetőséget és segítik a hálózaton vett sztochasztikus folyamatok megértését.

A disszertáció az alábbi három, Röst Gergellyel és Kiss Istvánval közös publikációra épül:

- Kiss, I.Z., Röst, G. and Vizi, Z., 2015. Generalization of pairwise models to non-Markovian epidemics on networks. *Physical review letters*, 115(7), p.078701. <http://dx.doi.org/10.1103/PhysRevLett.115.078701>
- Röst, G., Vizi, Z. and Kiss, I.Z., 2015. Impact of non-Markovian recovery on network epidemics. In *Biomat 2015: Proceedings of the International Symposium on Mathematical and Computational Biology*
- Röst, G., Vizi, Z. and Kiss, I.Z., 2016. Pairwise approximation for SIR type network epidemics with non-Markovian recovery. *arXiv preprint arXiv:1605.02933*.

A szerző további publikációi:

- Röst, G. and Vizi, Z., 2014. Backward bifurcation for pulse vaccination. *Nonlinear Analysis: Hybrid Systems*, 14, pp.99-113.
- Barbarossa, M.V., Dénes, A., Kiss, G., Nakata, Y., Röst, G. and Vizi, Z., 2015. Transmission dynamics and final epidemic size of Ebola virus disease outbreaks with varying interventions. *PloS one*, 10(7), p.e0131398.

Bibliography

- [1] Allaei, S.S., Yang, Z.W. and Brunner, H., 2016. Collocation methods for third-kind VIEs. *IMA Journal of Numerical Analysis*, p.drw033.
- [2] Allen, L.J., Brauer, F., Van den Driessche, P. and Wu, J., 2008. *Mathematical epidemiology* (pp. 81-130). Berlin: Springer.
- [3] Allen, L.J., 2008. An introduction to stochastic epidemic models. In *Mathematical epidemiology* (pp. 81-130). Springer Berlin Heidelberg.
- [4] Anderson, R.M. and May, R.M., 1988. Epidemiological parameters of HIV. *Nature*, 333.
- [5] Anderson, R.M., May, R.M. and Anderson, B., 1992. *Infectious diseases of humans: dynamics and control* (Vol. 28). Oxford: Oxford university press.
- [6] Bailey, N.T., 1956. On estimating the latent and infectious periods of measles: I. Families with two susceptibles only. *Biometrika*, pp.15-22.
- [7] Bailey, N.T., 1957. Stochastic theory continuous infection. *The mathematical theory of Epidemics*, 108.
- [8] Barabási, A.L. and Albert, R., 1999. Emergence of scaling in random networks. *science*, 286(5439), pp.509-512.
- [9] Barbarossa, M.V. and Röst, G., 2015. Immuno-epidemiology of a population structured by immune status: a mathematical study of waning immunity and immune system boosting. *Journal of mathematical biology*, 71(6-7), pp.1737-1770.
- [10] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.U., 2006. Complex networks: Structure and dynamics. *Physics reports*, 424(4), pp.175-308.
- [11] Boguñá, M., Lafuerza, L.F., Toral, R. and Serrano, M.Á., 2014. Simulating non-Markovian stochastic processes. *Physical Review E*, 90(4), p.042108.

- [12] Boots, M. and Sasaki, A., 1999. ‘Small worlds’ and the evolution of virulence: infection occurs locally and at a distance. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1432), pp.1933-1938.
- [13] Brauer, F. and Castillo-Chavez, C., 2013. *Mathematical Models in Population Biology and Epidemiology* (Vol. 40). Springer Science and Business Media.
- [14] Brauer, F. and Chowell, G. 2013. On epidemic growth rates and the estimation of the basic reproduction number, *Notes on modeling and numerical methods. Computational modeling of biological systems*. Morales Vazquez, M.A., Botello Rionda, S (eds.). Publisher: CIMAT, Guanajuato, Mexico
- [15] Brunner, H., 2004. Collocation methods for Volterra integral and related functional differential equations (Vol. 15). *Cambridge University Press*.
- [16] Burton, T.A., 2005. *Volterra integral and differential equations* (Vol. 202). Elsevier.
- [17] Cator, E., Van de Bovenkamp, R. and Van Mieghem, P., 2013. Susceptible-infected-susceptible epidemics on networks with general infection and cure times. *Physical Review E*, 87(6), p.062816.
- [18] Cooper, F., 2013. Non-Markovian network epidemics (*MA Thesis*, 2013. www.dtc.ox.ac.uk/people/13/cooperf/files/MA469ThesisFergusCooper.pdf).
- [19] Chowell, G. and Nishiura, H., 2014. Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC medicine*, 12(1), p.1.
- [20] Danon, L., Ford, A.P., House, T., Jewell, C.P., Keeling, M.J., Roberts, G.O., Ross, J.V. and Vernon, M.C., 2011. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011.
- [21] Demirel, G., Vazquez, F., Böhme, G.A. and Gross, T., 2014. Moment-closure approximations for discrete adaptive networks. *Physica D: Nonlinear Phenomena*, 267, pp.68-80.
- [22] Diekmann, O., Heesterbeek, J.A.P. and Metz, J.A., 1990. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4), pp.365-382.
- [23] Eames, K.T. and Keeling, M.J., 2002. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proceedings of the National Academy of Sciences*, 99(20), pp.13330-13335.

- [24] Eichner, M. and Dietz, K., 2003. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *American Journal of Epidemiology*, 158(2), pp.110-117.
- [25] Erdős, P. and Rényi, A., 1959. On random graphs, I. *Publicationes Mathematicae* (Debrecen), 6, pp.290-297.
- [26] Fates, N., 2013, September. A guided tour of asynchronous cellular automata. In *International Workshop on Cellular Automata and Discrete Complex Systems* (pp. 15-30). Springer Berlin Heidelberg.
- [27] Ferguson, N.M. and Garnett, G.P., 2000. More realistic models of sexually transmitted disease transmission dynamics: sexual partnership networks, pair models, and moment closure. *Sexually transmitted diseases*, 27(10), pp.600-609.
- [28] Garnett, G.P. and Anderson, R.M., 1996. Sexually transmitted diseases and sexual behavior: insights from mathematical models. *Journal of Infectious Diseases*, 174(Supplement 2), pp.S150-S161.
- [29] Ghani, A.C., Swinton, J. and Garnett, G.P., 1997. The role of sexual partnership networks in the epidemiology of gonorrhoea. *Sexually transmitted diseases*, 24(1), pp.45-56.
- [30] Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25), pp.2340-2361.
- [31] Gubbins, S., Carpenter, S., Baylis, M., Wood, J.L. and Mellor, P.S., 2008. Assessing the risk of bluetongue to UK livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *Journal of the Royal Society Interface*, 5(20), pp.363-371.
- [32] Gustafsson, L. and Sternad, M., 2010. Consistent micro, macro and state-based population modelling. *Mathematical biosciences*, 225(2), pp.94-107.
- [33] Grenfell, B.T., 1992. Chance and chaos in measles dynamics. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.383-398.
- [34] Grenfell, B.T., Bjørnstad, O.N. and Kappey, J., 2001. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865), pp.716-723.
- [35] Gross, T., D'Lima, C.J.D. and Blasius, B., 2006. Epidemic dynamics on an adaptive network. *Physical review letters*, 96(20), p.208701.

- [36] Smith, H., 2010. *An introduction to delay differential equations with applications to the life sciences* (Vol. 57). Springer Science & Business Media.
- [37] Hébert-Dufresne, L., Patterson-Lomba, O., Goerg, G.M. and Althouse, B.M., 2013. Pathogen mutation modeled by competition between site and bond percolation. *Physical review letters*, 110(10), p.108103.
- [38] Hethcote, H.W. and Yorke, J.A., 1984. *Lecture notes in biomathematics*.
- [39] Hoffmann, T., Porter, M.A. and Lambiotte, R., 2012. Generalized master equations for non-Poisson dynamics on networks. *Physical Review E*, 86(4), p.046102.
- [40] Jo, H.H., Perotti, J.I., Kaski, K. and Kertész, J., 2014. Analytically solvable model of spreading dynamics with non-Poissonian processes. *Physical Review X*, 4(1), p.011041.
- [41] Karrer, B. and Newman, M.E., 2010. Message passing approach for general epidemic models. *Physical Review E*, 82(1), p.016101.
- [42] Keeling, M.J., Rand, D.A. and Morris, A.J., 1997. Correlation models for childhood epidemics. *Proceedings of the Royal Society of London B: Biological Sciences*, 264(1385), pp.1149-1156.
- [43] Keeling, M.J., 1999. Correlation equations for endemic diseases: externally imposed and internally generated heterogeneity. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1422), pp.953-960.
- [44] Keeling, M.J. and Grenfell, B.T., 2002. Understanding the persistence of measles: reconciling theory, simulation and observation. *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1489), pp.335-343.
- [45] Keeling, M., 2005. The implications of network structure for epidemic dynamics. *Theoretical population biology*, 67(1), pp.1-8.
- [46] Keeling, M.J. and Eames, K.T., 2005. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4), pp.295-307.
- [47] Kenah, E. and Robins, J.M., 2007. Second look at the spread of epidemics on networks. *Physical Review E*, 76(3), p.036113.

- [48] Kermack, W.O. and McKendrick, A.G., 1927, August. A contribution to the mathematical theory of epidemics. *In Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences* (Vol. 115, No. 772, pp. 700-721). The Royal Society.
- [49] Kiss, I.Z., Röst, G. and Vizi, Z., 2015. Generalization of pairwise models to non-Markovian epidemics on networks. *Physical review letters*, 115(7), p.078701.
- [50] Knipf, D.H., Röst, G. and Wu, J., 2013. Epidemic spread and variation of peak times in connected regions due to travel-related infections—dynamics of an antigravity-type delay differential model. *SIAM Journal on Applied Dynamical Systems*, 12(4), pp.1722-1762.
- [51] Laskowski, M., Mostaçõ-Guidolin, L.C., Greer, A.L., Wu, J. and Moghadas, S.M., 2011. The impact of demographic variables on disease spread: influenza in remote communities. *Scientific Reports*, 1.
- [52] Liang, H. and Brunner, H., 2016. On the convergence of collocation solutions in continuous piecewise polynomial spaces for Volterra integral equations. *BIT Numerical Mathematics*, pp.1-29.
- [53] Liu, M., Röst, G. and Vas, G., 2013. SIS model on homogeneous networks with threshold type delayed contact reduction. *Computers and Mathematics with Applications*, 66(9), pp.1534-1546.
- [54] Lloyd, A.L., 2001. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theoretical population biology*, 60(1), pp.59-71.
- [55] Malmgren, R.D., Stouffer, D.B., Motter, A.E. and Amaral, L.A., 2008. A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences*, 105(47), pp.18153-18158.
- [56] Miller, J.C., 2007. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E*, 76(1), p.010101.
- [57] Miller, J.C., 2008. Bounding the size and probability of epidemics on networks. *Journal of Applied Probability*, pp.498-512.
- [58] Min, B., Goh, K.I. and Kim, I.M., 2013. Suppression of epidemic outbreaks with heavy-tailed contact dynamics. *EPL (Europhysics Letters)*, 103(5), p.50002.

- [59] Moinet, A., Starnini, M. and Pastor-Satorras, R., 2015. Burstiness and aging in social temporal networks. *Physical review letters*, 114(10), p.108701.
- [60] Nah, K., Nakata, Y. and Röst, G., 2014. Malaria dynamics with long incubation period in hosts. *Computers & Mathematics with Applications*, 68(9), pp.915-930.
- [61] Newman, M.E., 2002. Spread of epidemic disease on networks. *Physical review E*, 66(1), p.016128.
- [62] Newman, M.E., 2003. The structure and function of complex networks. *SIAM review*, 45(2), pp.167-256.
- [63] Nishiura, H. and Eichner, M., 2007. Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiology and infection*, 135(07), pp.1145-1150.
- [64] Pastor-Satorras, R. and Vespignani, A., 2001. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14), p.3200.
- [65] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. and Vespignani, A., 2015. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3), p.925.
- [66] Rand, D.A., 1999. Correlation equations and pair approximations for spatial ecologies. *Advanced ecological theory: principles and applications*, 100.
- [67] Rand, D.A., 2009. *Advanced Ecological Theory: Principles and Applications* (John Wiley and Sons, New York), Chap. 4, pp. 100–142.
- [68] Rhodes, C.J. and Anderson, R.M., 1997. Epidemic thresholds and vaccination in a lattice model of disease spread. *Theoretical Population Biology*, 52(2), pp.101-118.
- [69] Rohani, P., Earn, D.J. and Grenfell, B.T., 2000. Impact of immunisation on pertussis transmission in England and Wales. *The Lancet*, 355(9200), pp.285-286.
- [70] Röst, G., Vizi, Z. and Kiss, I.Z., 2015. Impact of non-Markovian recovery on network epidemics. In *Biomat 2015*, World Scientific, pp. 40-53.
- [71] Röst, G., Vizi, Z. and Kiss, I.Z., 2016. Pairwise approximation for SIR type network epidemics with non-Markovian recovery. *arXiv preprint arXiv:1605.02933*.
- [72] Sato, K., Matsuda, H. and Sasaki, A., 1994. Pathogen invasion and host extinction in lattice structured populations. *Journal of mathematical biology*, 32(3), pp.251-268.

- [73] Scalas, E., Kaizoji, T., Kirchler, M., Huber, J. and Tedeschi, A., 2006. Waiting times between orders and trades in double-auction markets. *Physica A: Statistical Mechanics and its Applications*, 366, pp.463-471.
- [74] Sharkey, K.J., 2008. Deterministic epidemiological models at the individual level. *Journal of Mathematical Biology*, 57(3), pp.311-331.
- [75] Sherborne, N., Blyuss, K.B. and Kiss, I.Z., 2015. Dynamics of multi-stage infections on networks. *Bulletin of mathematical biology*, 77(10), pp.1909-1933.
- [76] Sherborne, N., Blyuss, K.B. and Kiss, I.Z., 2016. Compact pairwise models for epidemics with multiple infectious stages on degree heterogeneous and clustered networks. arXiv preprint arXiv:1602.05227.
- [77] Simon, P.L., Taylor, M. and Kiss, I.Z., 2011. Exact epidemic models on graphs using graph-automorphism driven lumping. *Journal of mathematical biology*, 62(4), pp.479-508.
- [78] Simon, P.L. and Kiss, I.Z., 2015. Super compact pairwise model for SIS epidemic on heterogeneous networks. *Journal of Complex Networks*, p.cnv018.
- [79] Szabó-Solticzky, A., Berthouze, L., Kiss, I.Z. and Simon, P.L., 2016. Oscillating epidemics in a dynamic network model: stochastic and mean-field analysis. *Journal of mathematical biology*, 72(5), pp.1153-1176.
- [80] M. Taylor, P. L. Simon, D. M. Green, T. House, and I. Z. Kiss. From Markovian to pairwise epidemic models and the performance of moment closure approximations. *J. Math. Biol.*, 64:1021-1042, 2012.
- [81] Van Baalen, M., 2000. Pair approximations for different spatial geometries. *The geometry of ecological interactions: simplifying spatial complexity*, 742, pp.359-387.
- [82] Van Mieghem, P. and Van de Bovenkamp, R., 2013. Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. *Physical review letters*, 110(10), p.108701.
- [83] Volterra, V., 2005. *Theory of functionals and of integral and integro-differential equations*. Courier Corporation.
- [84] Watts, D.J. and Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684), pp.440-442.

- [85] Wilkinson, R.R. and Sharkey, K.J., 2014. Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks. *Physical Review E*, 89(2), p.022808.
- [86] Wilkinson, R.R., Ball, F.G. and Sharkey, K.J., 2016. The relationships between message passing, pairwise, Kermack-McKendrick and stochastic SIR epidemic models. *arXiv preprint arXiv:1605.03555*.
- [87] Yang, Y., Halloran, M.E., Sugimoto, J.D. and Longini Jr, I.M., 2007. Detecting human-to-human transmission of avian influenza A (H5N1). *Emerg Infect Dis*, 13(9), pp.1348-1353.
- [88] Zhang, R., Liang, H. and Brunner, H., 2016. Analysis of Collocation Methods for Generalized Auto-Convolution Volterra Integral Equations. *SIAM Journal on Numerical Analysis*, 54(2), pp.899-920.

Appendix

Sample run

□ Description

A code of sample run to simulate the non-Markovian stochastic process on networks. We consider Weibull distributed recovery time with parameters a and b , homogeneous random network with uniform degree distribution $\langle k \rangle = 15$ and $N = 1000$ nodes. The loop in simulation stores the actual state of the system in a list to export the output at the end of the epidemic.

□ Initialisation

```
Nn = 1000;
network =
  RandomGraph[DegreeGraphDistribution[
    Table[15, {i, 1, Nn}]]];
tau = 0.3;
a = 1.0;
b = 1.0;
t = 0;
timelist = {t};
counter = 1;
S = 999;
II = Nn - S;
R = 0;
config = {{S, II, R}};
tstop = 1000;
(*edge list of the network*)
links =
  N[Table[{EdgeRules[network][[k, 1]],
    EdgeRules[network][[k, 2]],
    {k, 1, EdgeCount[network]}]];
(*labelling the nodes by a state: S = -1;
I = positive number/time to recovery; R = ∞ *)
label = Join[Table[-1., {i, 1, S}],
  Table[0., {i, S + 1, S + II}];
(*list of S-I links*)
inlinks =
  Select[links,
    ((label[[#[[1]]]] == -1. && label[[#[[2]]]] ≥ 0) ||
    (label[[#[[2]]]] == -1. && label[[#[[1]]]] ≥ 0)) &];
```

```

SIlinks =
  Table[If[label[[inflinks[[n, 1]]]] == -1.,
    {inflinks[[n, 1]], inflinks[[n, 2]]},
    {inflinks[[n, 2]], inflinks[[n, 1]]}],
    {n, 1, Length[inflinks]};
(*sampling exponentially distributed waiting time
for all S-I links*)
timetoinf = Table[
  RandomVariate[ExponentialDistribution[tau]],
  {k, 1, Length[SIlinks]};
(*list of infected nodes*)
infectednodes = Table[i, {i, S + 1, S + II}];
(*sampling time to recovery for infected nodes*)
timetorec = Table[
  RandomVariate[WeibullDistribution[a, b]],
  {k, 1, Length[infectednodes]};
(*lists for export*)
labels = {label};
infectednodeslist = {infectednodes};
timetoinflist = {timetoinf};
SIlinkslist = {SIlinks};
timetoreclist = {timetorec};
(*variables for minimum of list 'timetorec' and '
timetoinf'*)
minforinf = 0;
minforrec = 0;

```

□ *Simulating the epidemics*

```

While[(t < tstop && minforrec < ∞ && minforinf < ∞ && II > 0),
  (*time to the next infection and recovery,
  respectively*)
  minforrec = Min[timetorec];
  minforinf = Min[timetoinf];
  (*finding the infecting link and recovering*)
  eventSIlink = If[minforinf < ∞,
    SIlinks[[Flatten[Position[timetoinf, minforinf]]]],
    {}];
  eventInode = If[minforrec < ∞,
    infectednodes[[
      Flatten[Position[timetorec, minforrec]]]], {}];
  (*executing infection or recovery*)
  simulation = process[tau, a, b, t, label, links, S,
    II, R, minforrec, eventInode, infectednodes,
    timetorec, minforinf, eventSIlink, SIlinks, timetoinf];
  (*updating lists and variables for next step and

```



```

    export*)
    t = simulation[[1, 1]];
    label = simulation[[2]];
    currentconfig = simulation[[3]];
    S = currentconfig[[1]];
    II = currentconfig[[2]];
    R = currentconfig[[3]];
    SIlinks = simulation[[4]];
    timetoinf = simulation[[5]];
    infectednodes = simulation[[6]];
    timetorec = simulation[[7]];
    labels = Join[labels, {label}];
    timelist = Join[timelist, {t}];
    SIlinkslist = Join[SIlinkslist, {SIlinks}];
    config = Join[config, {currentconfig}];
    timetoinflist = Join[timetoinflist, {timetoinf}];
    counter++;
];

```

Module “inf”: the infecting process

□ *Description*

This module executes the infection subprocess. First, it updates the labels of nodes, which are infected in this step. Next, we modify the lists associated to S-I links. Finally, we change the lists associated to infected nodes and calculate the returning values of the module.

□ *Parameters:*

tau_: transmission parameter

a_: shape parameter of Gamma distribution

b_: scale parameter of Gamma distribution

t_: time of system

label_List: list of nodes' labels

links_List: links of the network

S_: number of susceptible nodes at time t

II_: number of infected nodes at time t

R_: number of recovered nodes at time t

minforrec_: time to the next recovery

eventInode_List: list of recovering nodes

infectednodes_List: list of infected nodes

timetorec_List: list of time to recovery for all infected nodes

minforinf_: time to the next infection

eventSIlink_List: list of S-I links, where infection goes along

SIlinks_List: list of S-I links

timetoinf_List: list of time to infection for all S-I links

```
(*Module executing the infecting process*)
inf[tau_, a_, b_, t_, label_List, links_List, S_, II_,
  R_, minforrec_, eventInode_List, infectednodes_List,
  timetorec_List, minforinf_, eventSIlink_List,
  SIlinks_List, timetoinf_List] :=
Module[{noofchanges, outputSIlinks, remSI, newSI,
  updatedlabels, suscnod, updremSIw, newSIw,
  outputSIlinkswaiting, remSIw, updIw, outputInodes,
  outputInodewaiting, remrp},
(*updating labels of all nodes*)
noofchanges = Length[eventSIlink];
updatedlabels =
  Table[If[(MemberQ[eventSIlink[[All, 1]], N[n]]),
    N[0], If[(label[[n]] < ∞ && label[[n]] ≥ 0),
      label[[n]] + minforinf, label[[n]]]],
    {n, 1, Length[label]}];
(*updating list of S-I links*)
suscnod = N[Flatten[Position[updatedlabels, -1.]]];
remSI = Select[SIlinks,
  ! MemberQ[eventSIlink[[All, 1]], #[[1]]] &];
newSI =
  Select[
  N[
  Flatten[
  Table[
  If[
  (MemberQ[links, {eventSIlink[[1, 1]],
    suscnod[[k]]}] ||
  MemberQ[links, {suscnod[[k]],
    eventSIlink[[1, 1]]})],
  {suscnod[[k]], eventSIlink[[1, 1]]}, {0, 0}},
  {1, 1, noofchanges}, {k, 1, Length[suscnod]}, 1]],
  #[[1]] ≠ 0 &];
outputSIlinks = Join[remSI, newSI];
(*updating waiting times of S-I links*)
remrp =
  Flatten[
  Table[If[MemberQ[eventSIlink[[All, 1]],
    SIlinks[[i, 1]]], {}, {i}],
  {i, 1, Length[SIlinks]}]];
remSIw = timetoinf[[remrp]];
```

```

updremSIw = remSIw - minforinf;
newSIw =
  Table[RandomVariate[ExponentialDistribution[tau]],
    {i, 1, Length[newSI]}];
outputSIlinkswaiting = Join[updremSIw, newSIw];
(*updating list of infected nodes*)
outputInodes = Join[infectednodes,
  eventSIlink[[All, 1]]];
(*updating waiting times of infected nodes*)
updIw = Table[timetorec[[i]] - minforinf,
  {i, 1, Length[timetorec]}];
outputInodewaiting =
  Join[updIw,
    Table[RandomVariate[WeibullDistribution[a, b]],
      {i, 1, noofchanges}]];
(*returning values*)
{{t + minforinf, updatedlabels,
  {S - noofchanges, II + noofchanges, R}, outputSIlinks,
  outputSIlinkswaiting, outputInodes, outputInodewaiting}
];

```

Module 'rec': the infecting process

□ Description

This module executes the recovery subprocess. First, it updates the labels of nodes, which are recovered in this step. Next, we modify the lists associated to S-I links. Finally, we change the lists associated to infected nodes and calculate the returning values of the module.

□ Parameters

See Section 'Parameters' of module 'inf'.

```

(*Module executing the recovery process*)
rec[tau_, a_, b_, t_, label_List, links_List, S_, II_,
  R_, minforrec_, eventInode_List, infectednodes_List,
  timetorec_List, minforinf_, eventSIlink_List,
  SIlinks_List, timetoinf_List] :=
Module[{outputSIlinks, outputSIlinkswaiting, updremSIw,
  newSIw, remSI, newSI, updatedlabels, noofchanges,
  remSIw, outputInodes, remI, remIw, outputInodewaiting,
  remrp},
  (*updating labels*)
  noofchanges = Length[eventInode];
  updatedlabels =

```

```

Table[If[MemberQ[eventInode, N[n]], ∞,
  If[(label[[n]] ≥ 0 && label[[n]] < ∞),
    label[[n]] + minforrec, label[[n]]],
{n, 1, Length[label]};
(*updating list of S-I links*)
remrp =
Flatten[
  Table[If[MemberQ[eventInode // N, N[SILinks[[i, 2]]],
    {}, {i}], {i, 1, Length[SILinks]}];
outputSILinks = SILinks[[remrp]];
(*updating waiting times of S-I links*)
remSIw = timetoinf[[remrp]];
outputSILinkswaiting =
  Table[remSIw[[i]] - minforrec, {i, 1, Length[remSIw]}];
(*updating list of infected nodes*)
remI =
Flatten[
  Table[If[MemberQ[eventInode, infectednodes[[i]]],
    {}, {i}], {i, 1, Length[infectednodes]}];
outputInodes = infectednodes[[remI]];
(*updating waiting times of infected nodes*)
remIw = timetorec[[remI]];
outputInodewaiting =
  Table[remIw[[i]] - minforrec, {i, 1, Length[remIw]}];
{{t + minforrec}, updatedlabels,
  {S, II - noofchanges, R + noofchanges}, outputSILinks,
  outputSILinkswaiting, outputInodes, outputInodewaiting}
];

```

Module 'process': executing the next subprocess

□ Description

This module makes the decision, which subprocess is executed in this step.

□ Parameters:

See Section 'Parameters' of module 'inf'.

```

(*Main function for choosing and executing the event
according to the minimum of 'minforinf' and 'minforrec'*)
process[tau_, a_, b_, t_, label_List, links_List, S_,
  II_, R_, minforrec_, eventInode_List, infectednodes_List,
  timetorec_List, minforinf_, eventSILink_List,
  SILinks_List, timetoinf_List] :=
If[minforrec < minforinf,

```

```
rec[tau, a, b, t, label, links, S, II, R, minforrec,  
    eventInode, infectednodes, timetorec, minforinf,  
    eventSIlink, SIlinks, timetoinf],  
inf[tau, a, b, t, label, links, S, II, R, minforrec,  
    eventInode, infectednodes, timetorec, minforinf,  
    eventSIlink, SIlinks, timetoinf]  
];
```