University of Sussex

Mean-Field Models in Network Inference and Epidemic Control

Francesco Di Lauro

Submitted for the degree of Doctor of Philosophy

University of Sussex

April 2021

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

Francesco Di Lauro

UNIVERSITY OF SUSSEX

Francesco Di Lauro

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN MATHEMATICS

Mean-Field Models in Network Inference and Epidemic Control

Summary

Systems that are comprised of agents and pairwise interactions between agents can be studied through the lenses of Network theory. As a general framework, Network Theory has applications in various disciplines, including Statistical Physics, Economics, and Biology. The interplay between the contact structure of a population and epidemic spreading is one of the most studied research areas in Epidemiology, where network based research has offered many breakthroughs in recent years. Since an individual based description is computationally intractable, as state spaces scale exponentially with the number of agents modelled, many mathematical approximations have been developed to describe the system in terms of low dimensional aggregate statistics, such as the average number of infected people. This thesis is focused on the application of such approximation techniques, in particular the well known mean-field models, to two key problems in Epidemiology: inference and epidemic control.

In the first part of this work, the theme is the inference of network properties from the observation of outbreaks at a population-level. Typically, readily available information during an outbreak is (daily) case counts. With this in mind, a new mean-field like model is introduced to approximate epidemics on networks via Birth-and-Death processes, whose rates are random variables which depend implicitly on the structure of the underlying network and disease dynamics. By using Bayesian model selection, it is possible to recover the most likely underlying network class from datasets that consist only of discrete-time observations from one single epidemic. Further, having a description in terms of Birth-and-Death processes allows to study the large N limit of the process as a one-dimensional Fokker-Planck equation, that implies an even greater reduction in dimensionality.

In the second part of this thesis more standard mean-field models are adopted to perform epidemic control. The aim is to reduce the burden of an outbreak on a target population. Intervention policies may consist of one time interventions either to minimise the epidemic peak or the final size, or to maximise the average time to infection. Homogeneous mixing models are a nice tool to showcase how interventions that achieve such goals can be optimised. A network perspective is introduced to study the so-called disease-induced herd immunity: in principle, epidemics act like targeted vaccinations, preferentially immunising higher-risk individuals. This means that the herd-immunity threshold might be reached at lower levels compared to that derived from homogeneous mixing models, and this might be relevant for epidemic control. However, it is shown that the magnitude of this effect depends heavily on how both the topology of the contact network and the way non-pharmaceutical interventions are modelled. Finally, epidemic response can be thought of as a feedback process, that is, social distancing policies might be deployed depending on the observed epidemic curve, rather than being pre-determined from theoretical arguments. In this case, the goal is to maintain the epidemic at manageable levels throughout its course, by tailoring interventions that aim to be as less disruptive as possible. This possibility is investigated on a high dimensional network model, by deriving a feedback-loop control action that at its core is based on a mean-field approximation

Acknowledgements

Firstly, I would like to thank my academic supervisors Professor István Z. Kiss, Professor Luc Berthouze and Doctor Masoumeh Dashti, for their guidance throughout my time as a PhD researcher. I would like to thank the Leverhulme Trust for funding my PhD studentship and I would also like to give thanks to the researchers I collaborated with on the work contained within this thesis. There are numerous people I have had the pleasure of knowing or meeting during my time as a PhD researcher in the School of Mathematical and Physical Sciences and I wish to acknowledge their contributions in supporting me to do this work. From friends, colleagues and university facilities staff to professional services staff and students, this piece of work would not be the same without all of these people and I am immensely grateful to all those unnamed individuals for their presence and inspiration. Finally, and most importantly, I would like to thank my family, who have continuously supported and encouraged me throughout this process, and my fiancé who helped me to survive abroad.

List of publications

- **Network inference from population-level observation of epidemics**
  Francesco Di Lauro, Jean-Charles Croix, Masoumeh Dashti, Luc Berthouze and István Z. Kiss. Published in *Nature Scientific Reports* 10.1 (2020), available at https://doi.org/10.1093/comnet/cnaa043

- **PDE limits of stochastic SIS epidemics on networks**
  Francesco Di Lauro, Jean-Charles Croix, Luc Berthouze and István Z. Kiss. Published in *Journal of Complex Networks* 8.4 (2020), available at https://doi.org/10.1038/s41598-020-75558-9.

- **Optimal timing of one-shot interventions for epidemic control**
  Francesco Di Lauro, István Z. Kiss, Joel C. Miller. Published in *PLOS Computational Biology* 2021, available at https://doi.org/10.1371/journal.pcbi.1008763.

- **The impact of network properties and mixing on control measures and disease-induced herd immunity in epidemic models: a mean-field model perspective** Francesco Di Lauro, Luc Berthouze, Matthew D. Dorey, Joel C. Miller, István Z. Kiss. (2020) Arxiv preprint available at arXiv:2007.06975. Currently under revision at Bulletin of Maths Biology. Parts of it published as a Chapter in **Beyond COVID-19: network science and sustainable exit strategies**, (2021) *Journal of Physics: Complexity*, Volume 2, Issue 2.

- **Covid-19 and Flattening the Curve: A Feedback Control Perspective**
  Francesco Di Lauro, István Z. Kiss, Daniela Rus, Cosimo Della Santina. (2020)

Published in *IEEE Control Systems Letters* 5 (4), 1435-1440, available at https://ieeexplore.ieee.org/abstract/document/9264204.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and thesis overview

Hundreds of years of Epidemiological Research have resulted in a well-developed theory that embraces Biology, Sociology, Statistics and Network Science. Although contributions to the theory of epidemics go as far back as 1760, when Bernoulli developed a model for the growth of a population affected by smallpox Bernoulli [1766], it was only after the seminal work of Kermack and McKendrick Kermack and McKendrick [1927] that Epidemiology as a mathematical theory flourished. In it, the famous Susceptible-Infected-Recovered (SIR) model for infectious diseases was first introduced and studied. To quote from the introduction of the very same paper:

> "*The problem may be summarised as follows: One (or more) infected person is introduced into a community of individuals, more or less susceptible to the disease in question. The disease spreads from the affected to the unaffected by contact infection. Each infected person runs through the course of his sickness, and finally is removed from the number of those who are sick, by recovery or by death. [...] As the epidemic spreads, the number of unaffected members of the community becomes reduced. Since the course of an epidemic is short compared with the life of an individual, the population may be considered as remaining constant, except in as far as it is modified by deaths due to the epidemic disease itself. In the course of time the epidemic may come to an end.* "

In its essence, the original SIR model is a system of ordinary differential equations for three variables S, I, R representing the number of people susceptible to a disease, infected and infectious, or removed, respectively. Models in which the population is partitioned into compartments with labels, such as S, I, R, are known as *compartmental models* [Kiss et al., 2017]. Compartmental models are perhaps the most widely used tools in mathematical modelling of infectious diseases. As the epidemic progresses, people flow between different compartments, and the name of the model usually summarises the flow patterns; for example, SIR indicates that the typical individual is initially susceptible, may become infected/infectious after being in contact with an infected, and eventually recovers.

The key idea in the classical SIR model is that the epidemic is driven by a non-linear term proportional to both the number of susceptible and infected people, that approximates the number of contacts at risk between infected and susceptible individuals. This term gives rise to epidemic curves such as the ones depicted in figure 1.1. It is perhaps surprising how a simple-yet-ingenious model is still subject to research, and many insights are found by approaching the same model from novel perspectives [Britton et al., 2020; Casella, 2021].

The equations that describe the dynamics are the following:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N} SI, \\ \frac{dI}{dt} &= \frac{\beta}{N} SI - \gamma I, \\ \frac{dR}{dt} &= \gamma I, \end{aligned} \tag{1.1}$$

where $\beta, \gamma > 0$, are the per-contact rate of infection and rate of recovery, respectively. The initial condition is usually set to $(S, I, T) = (N - 1, 1, 0)$. Note that, at any given time, we have $S(t) + I(t) + R(t) = N$.

This dynamical system is quite simple to describe, although an exact closed-form solution does not exist [Barlow and Weinstein, 2020]. First of all, note that a condition for the epidemic to happen is $I > 0$ and $\frac{dI}{dt} > 0$, which immediately gives the following:

$$\frac{\beta}{N} SI - \gamma I > 0 \rightarrow \frac{S}{N} > \frac{\gamma}{\beta} \equiv \frac{1}{\mathcal{R}_0}, \tag{1.2}$$

where $\mathcal{R}_0$ is a quantity called *basic reproduction number*, and is interpreted as the number of people that will get infected when one typical infected individual is intro-

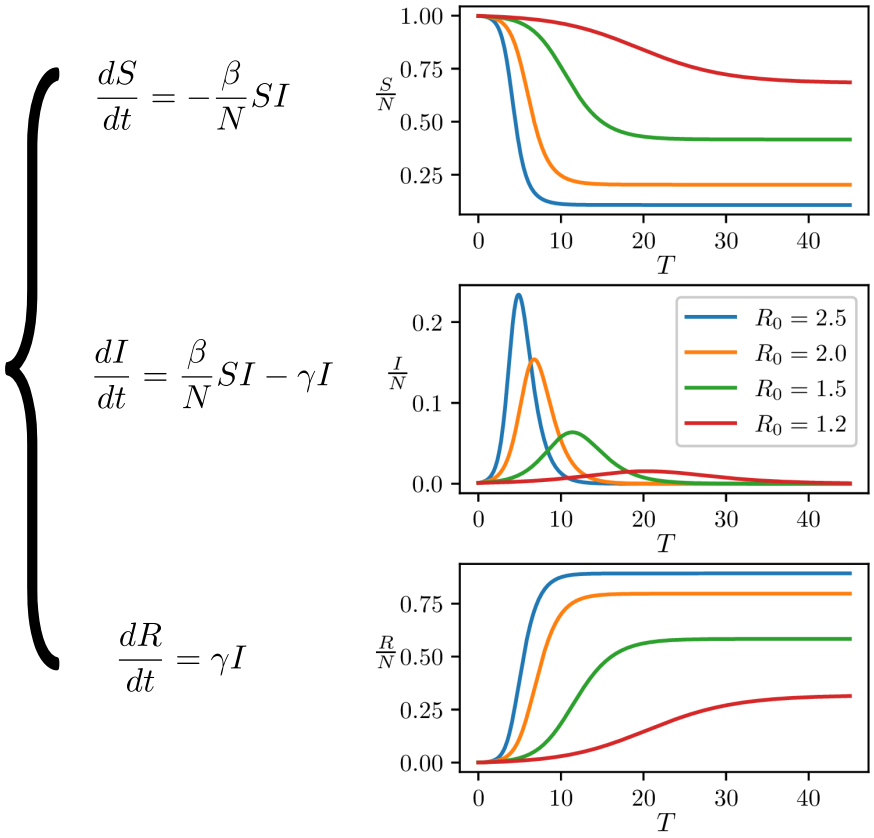

Figure 1.1: The SIR model as introduced by Kermack and Mckendrick Kermack and McKendrick [1927] is a system of differential equations that depends on two parameters, β and γ, named infection and recovery rate, respectively. The combination of these two gives rise to $\mathcal{R}_0 = \frac{\beta}{\gamma}$, a quantity of great importance, that determines the peak and the final size of the epidemic. On the right, various solutions for different values of $\mathcal{R}_0$, when $\gamma = 1$.

duced into a fully susceptible population. When the proportion of the susceptible becomes lower than $\frac{1}{\mathcal{R}_0}$, the number of infected people decreases in time. This phenomenon is referred to as *herd-immunity threshold*, and, in an uncontrolled epidemic, it is reached at the peak of the I curve. Figure 1.1 shows different solutions to the SIR model for different values of $\mathcal{R}_0$. Note that when the peak of the epidemic is reached, as the number of infected people decreases towards 0, some people shall get infected, a phenomenon known as *overshoot*. If instead enough people are recovered at the beginning of the epidemic, so that fewer than $\frac{N}{\mathcal{R}_0}$ susceptible are available, then no outbreak is possible. This is particularly useful when planning a vaccination campaign ahead of an expected outbreak, such as with the yearly influenza strains - in the hypothesis that the vaccine confers total immunity from the disease. Another important analytical result that can be derived from equation (1.1) is the final size $R(\infty)$, that is the cumulative number of infected people expected at the end of the epidemic, which satisfies the implicit equation

$$N - R(\infty) = S(0) \exp\left(-\frac{\beta}{N\gamma} R(\infty)\right).$$

The ones reported here are just a few examples of the numerous insights that are hidden in such a simple model. The interest in epidemic models has been renewed recently due to the on-going pandemic of Covid-19, with tens of thousands of papers published during the first months of 2020 [Teixeira da Silva et al., 2021]. It is remarkable that whilst the mathematical theory of epidemics developed in a wide variety of directions, it maintained the two key concepts expressed in the paper, i.e. that the population can be split into compartments with respect to the different stages of the disease, e.g. susceptible (S), infected (I) and recovered/removed (R), and that the evolution of the prevalence depends strongly on the number of interactions between those who are susceptible and those who are able to transmit the disease upon contact. For instance, many deterministic epidemic models akin to the SIR can be derived from this simple framework, see figure 1.2.

Epidemic models such as the SIR take into account the impact of the population's contact structure through the term $S \times I$, i.e. by explicitly assuming that everyone is in contact with everyone. To compensate for this assumption, the parameter β needs to account both for the probability of transmitting the disease upon contact and the rate at which contacts happen [Roberts and Heesterbeek, 1993].

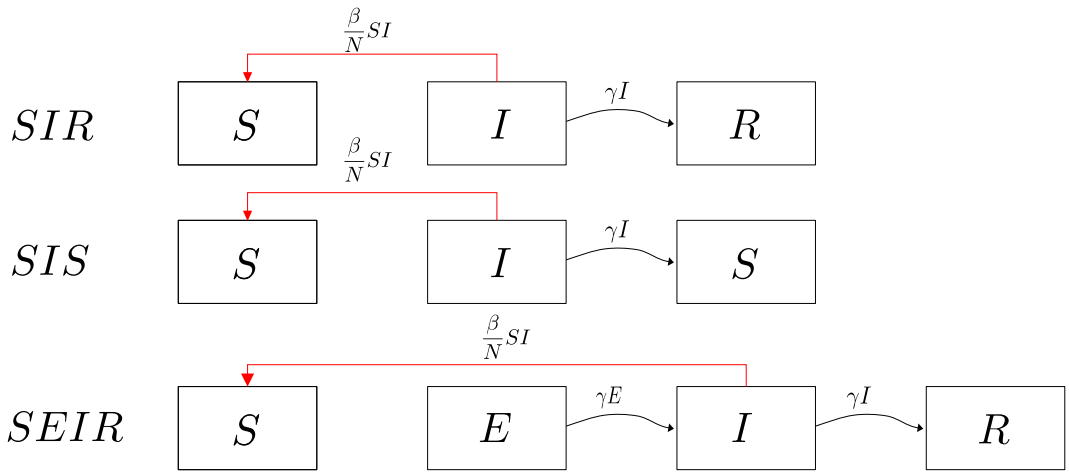

Figure 1.2: Flow diagrams for three different epidemic models, namely SIR, SIS and SEIR. The boxes on each line represent the number of states an individual can be in. The forward arrows between boxes represent transitions between different compartments. The red backward arrows represent ways in which the susceptible people are depleted, transitioning to the first non-susceptible compartment. In these cases, epidemics are due to contacts between people in compartments S and I, with rates indicated over the arrows.

This dual interpretation of β overloads it with meaning, making its interpretation quite difficult. For this reason, mathematical models able to consider these two characteristics of epidemics separately potentially offer more insights on this dynamics. Indeed, the knowledge of the network of contacts is so important that a whole branch of modern Epidemiology is dedicated to describe and model it. Among others, the role of network science and epidemic models on networks has been prominent since the late 20th century [Pastor-Satorras et al., 2015]. Networks represent perhaps the most natural way to describe agents (as nodes) and pairwise interactions between agents (as edges or links between nodes). Depending on the rules that assign edges to nodes, an incredibly vast number of networks or contact patterns may be created, see figure 1.3 for a few examples. As A.L. Barabási, an eminent researcher in this field asserts in his book *Network Science*[Barabási, 2013]: "Networks are everywhere, from the Internet, to social networks, and the genetic networks that determine our biological existence". Indeed, Network Science has revolutionised not only Epidemiology, but the whole discipline of Complex Systems. Within this framework, epidemics can be modelled as random processes taking place on top of a network, in which nodes that carry the disease infect their neighbours through links, which represent contacts at risks. Figure 1.4 provides a pictorial representation of this kind of models for a SIR. The delicate interplay between network topology and spreading processes on top of it is now a well-researched topic [Pastor-Satorras et al., 2015; Porter and Gleeson, 2016; Kiss et al., 2017]. The network-based representation of epidemics continuously evolved to better capture the complex nature of interactions that most systems exhibit. For instance, most real-world networks are temporal [Holme and Saramäki, 2012], meaning that contacts between nodes are a dynamic quantity. Similarly, it proves useful to distinguish between different types of contacts, and this has led to the introduction of multi-layer networks in Epidemiology [Kivelä et al., 2014]. It is worth mentioning that often, for real life applications (such as policy informing) these achievements are made possible thanks to the availability of large quantity of data, the so-called *Big Data* [Bansal et al., 2016] revolution. This has drastically modified the way much research in Complex Systems is carried out - a sudden change of paradigm with its own challenges [Cai and Zhu, 2015]. Adopting the point of view of Network Science allows to differen-

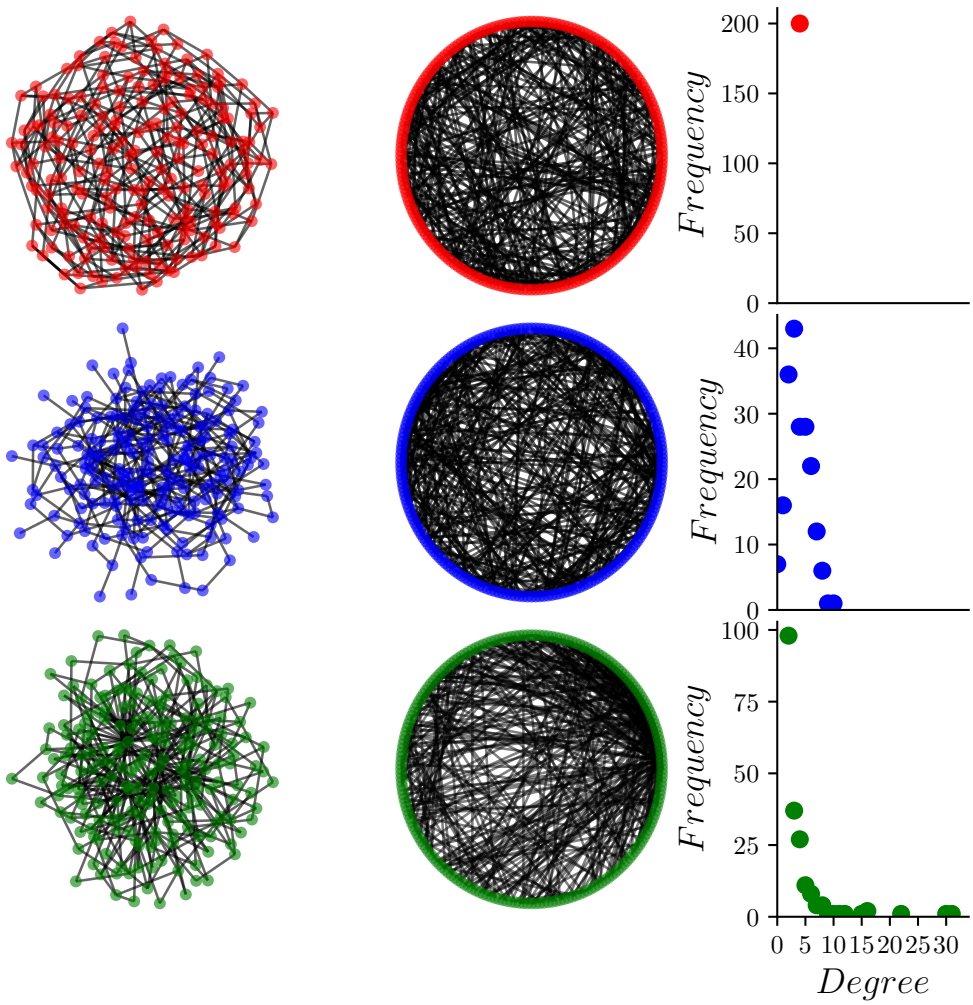

Figure 1.3: Three different networks with the same size and the same average degree 4 but different degree distribution. (top) Regular network, (middle) Erdős-Rényi network, (bottom) Barabási-Albert network. On the first two columns, two different representations of such networks that highlight degree heterogeneity. The first one, known as Kamada-Kaway, displays nodes in such a way that the higher the number of links, the more central the place of the node. It is helpful to recognise the difference between regular networks and Erdős-Rényi. In the second column, nodes are placed on a ring and links are straight lines between nodes. In this one, it is clear that the Barabási-Albert network has hubs, displayed on the right of the image, while the other two are more homogeneous. This shows how drawing networks in different ways may help visualising their properties. Finally, the right column shows the non-normalized degree distributions, that is, the distribution of the number of neighbours of nodes.

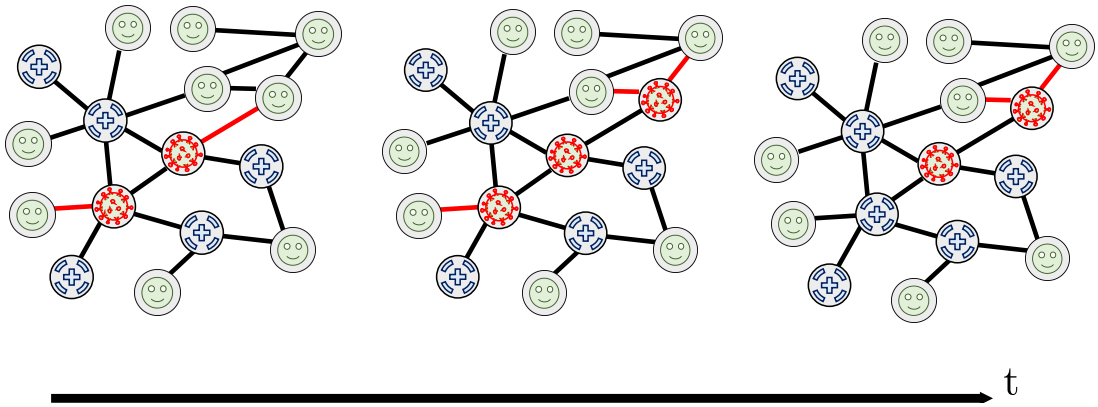

Figure 1.4: Pictorial representation of a SIR epidemic spreading on a network. There are three states associated with a node: susceptible (green smiling face), infected (red coronavirus symbol) and recovered (blue shielded plus). An infected node spreads the disease to its susceptible neighbours (red links). As time goes by, it might happen that one susceptible turns into an infected (second panel) and starts infecting, or that an infected recovers (third panel) and does not contribute to the spreading anymore. Note how some nodes become shielded from infection, when all their neighbours are recovered.

tiate between the biological characteristics of a pathogen and the topology of the contact network, in contrast to the so-called mass action models, where all the individuals are connected to each other (such as the Kermack–McKendrick model). The extent to which an epidemic invades and spreads in a population depends strongly on the underlying contact structure, which can be studied and acted upon by policy makers when needed. In times of Covid-19, questions such as "What is the role of households and communities in sustaining epidemics?", "Which groups of people are likely to be infected early on?", "Who should get a vaccine first?" are almost everyday on newspapers. To a good extent, these questions may be answered through the lens of Network Science. Although nowadays explicit networks may not enter directly in the most cutting-edge models that govern pandemic response, which are known as agent-based-models (such as the ones described in [Ferguson et al., 2020; Siettos and Russo, 2013]), they are often utilised to inform such models, and remain a useful conceptual tool for the understanding of the key aspects of contagion processes. For instance, the rate at which new infection events happen is directly proportional to the number of contacts between susceptible and infected individuals, that is the number of links that connect infected nodes to susceptible ones. This

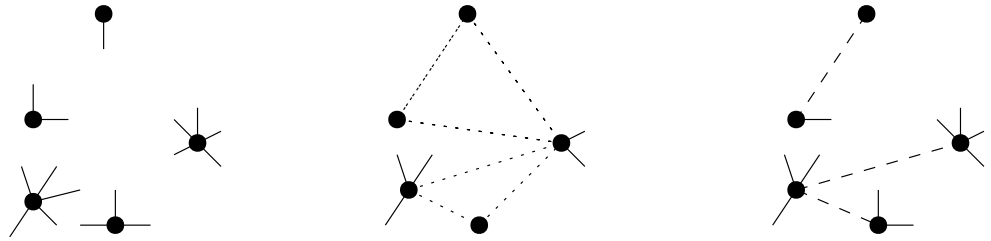

Figure 1.5: When approaching the study of networks and epidemics, one often starts from the degree distribution (left panel), followed by clustering, the tendency of nodes to form triangles - dotted lines (mid) and assortativity, the tendency of nodes to be neighbours of nodes with similar degree - dashed lines (right).

term in turn depends strongly on the contact structure of the population itself, and particularly on the number of links in the network and their distribution, called *degree distribution*, where the degree of a node is the number of neighbours of that node.

Further, years of research led to the realisation that certain properties of the contact structure, such as degree heterogeneity (how much the number of contacts varies between nodes), clustering (the propensity of nodes that are neighbours of a node to be neighbours of each other), and assortativity (who mixes with whom) have a profound impact on how epidemics spread in a population, see figure 1.5. For instance, it is well-known that highly connected nodes tend to be infected first during an epidemic, and have a prominent role in spreading the disease early on [Cohen et al., 2003; Pastor-Satorras et al., 2003], that clustering reduces the final size of epidemics [Ritchie et al., 2014, 2015], and that assortativity accelerates the speed at which epidemics grow [Badham and Stocker, 2010]. All of these breakthroughs are a direct consequence of the adoption of a network-based point of view.

In general, a stochastic epidemic model with m possible disease states on a network with N nodes can be modelled as a continuous-time Markov chain with m^N states, that is, all the unique combinations of m states among N nodes, and as many linear differential equations. Although this is in principle easy to encode and formalise, the numerical complexity of such models implies that exact mathematical formulations to describe an epidemic process unfolding on a network are intractable except in a few, rather exceptional, circumstances. This results in a trade-off between *tractability* and *complexity* that is common to many other areas of Complex

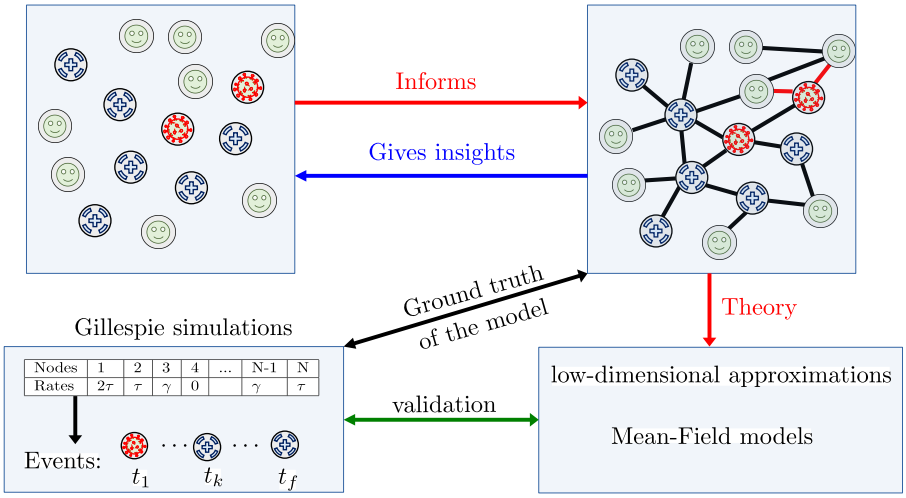

Figure 1.6: Networks approximate the contact structure of a population that might experience an outbreak. Network epidemic models are at the basis of both numerical simulations and theoretical considerations that aim to offer insights on epidemic dynamics.

Systems. These two terms are borrowed from the language of Computer Science, but their meaning is enriched to comprise the typical situation modellers find themselves in, that is, the more features the model has, the less mathematically tractable it is. Figure 1.6 shows the typical approach to the theoretical study of epidemics on network is shown: a real-world epidemic is studied by formalising the problem as a network model with an epidemic process spreading on it. Theory developed on such models aims at reducing the dimensionality of the problem. The validation of such theoretical models is usually done by means of simulations of epidemic processes on networks, which constitute the "ground truth" of epidemic models on networks.

In many cases, the only way to acquire a theoretical understanding of the problem is to introduce approximations that reduce the dimensionality of the system. In the case of epidemics on networks, the paradigm is to switch the focus onto relevant aggregate statistics, such as the number of nodes in each possible state at a given time, rather than studying their dynamics of individual nodes or of the full system. These are indeed the quantities that policy makers are interested in and it is therefore of great importance in the community. Further, readily available estimates are often given in terms of anonymised, aggregate data, such as the number of new positive tests each day in a particular country.

There is a number of different techniques used to derive approximate quantit-

ative results, such as edge-based models [Miller et al., 2012; Wang et al., 2018], percolation-theory approach [Allard et al., 2009], and finally mean-field, pairwise and higher-order approximations [Kiss et al., 2017]. Mean-field modeling techniques are borrowed from Statistical Physics, in particular Ising Models [Dorogovtsev et al., 2002]. The simplest of this class of models is also the most intuitive one, and relies on the following consideration: to approximate the number of links between susceptible and infected nodes, at a first order one considers only the average number of links per node in the network. This quantity reflects only the most basic property of the topology, i.e. the average degree. In this setting, each susceptible node has approximately the same number of neighbours, of which roughly a proportion $\frac{I}{N}$ is susceptible, resulting in a system of differential equations similar in form to (1.1), that is rather naive but surprisingly accurate for a wide variety of cases; nonetheless it can be improved, for instance, by considering higher moments of the degree distribution, or by looking at other topological features of the network - such as clustering [Barnard et al., 2019]. Figure 1.7 shows how the average of many epidemic simulations is well captured by mean-field models.

Using mean-field models to approximate epidemics on networks is a well-established research area of Network Science and mathematical Epidemiology, and a vast literature is available on this topic (see for example [Pastor-Satorras et al., 2015; Kiss et al., 2017] for a summary of the state of the art). Many mean-field models (some exact in the large network size limit [Janson et al., 2014]) are able to capture various network properties. For example, some only capture the number of edges/links in the network, others capture the degree heterogeneity but not who mixes with whom (assortativity), others account for clustering. In this sense all mean-field models reflect a different level of abstraction from real contact network structures and make different assumptions. Typically, the goal is to derive (and validate by means of simulations on explicit networks) relatively low-dimensional systems of ordinary differential equations that can be solved numerically and that might elicit some theoretical results on the underlying system. Theoretical approximations to the dynamics of epidemics on networks can be used for a vast variety of purposes. Most notably, the study of the steady-state or final size [Nåsell, 2011], prediction of the epidemic curve [Szabó-Solticzky et al., 2015], herd-immunity threshold [Britton

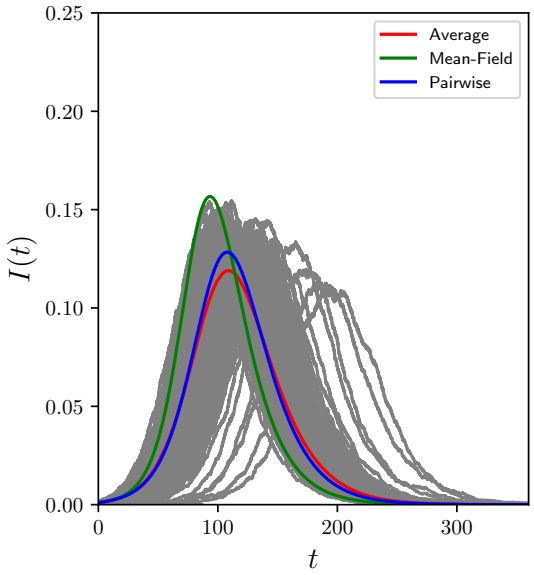

Figure 1.7: Several realisations of SIR epidemics on a network of size 1000 are plotted in gray. Then, the average is considered and compared against mean-field model (green curve) and pairwise approximation. The degree distribution of the network is approximately normal with average 10 and standard deviation 1, the epidemic parameters are $\tau = 0.016$, $\gamma = 0.07$.

et al., 2020] and network inference [Britton and O'Neill, 2002].

1.1.1 Network Inference

The first part of this thesis (Chapters 2,3) focuses on network inference from observation of epidemics, in particular for Susceptible-Infected-Susceptible (SIS) models.

While theoretical advancements in Epidemiology, and in particular in epidemic control through vaccination [Holme and Litvak, 2017], describe in great detail how to act on the contact structure of a population to alter significantly the course of an epidemic with minimum disruptions, many countries relied on global lockdowns during the Covid-19 pandemic. The reason for this can be interpreted in the language of network science as a lack of detailed knowledge of the network of contacts, which in turn prevents targeted measures to be put in place. It is therefore of paramount importance to build up tools for network inference during an on-going outbreak. This can be done via observation of the epidemic as it unfolds on the network of contacts. This task is usually [O'Neill and Roberts, 1999; Britton and O'Neill, 2002;

Gomez Rodriguez et al., 2014] performed when explicit information about nodes is available. For instance, in [Gomez Rodriguez et al., 2014], the full network of contacts is reconstructed, provided that one can observe the status of all the nodes through multiple realisations of the epidemic. When only one outbreak is observed, but still information about nodes is available, it is possible to infer some generative parameters of the topology [Britton and O'Neill, 2002].

When explicit information on nodes is not available, only a few models attempt to assess network properties based only on population-level observations. However, this is likely the most accessible quantity available during an on-going pandemic and it is of interest to have a framework that is able to capture information about the network structure when only daily observations at the level of the population is available. To achieve this, in Chapter 2 we conjecture that epidemics on networks may be well approximated by a class of Birth-Death processes, in which rates encode information on both the network and the disease dynamics. This implies a reduction from 2^N equation to $N + 1$. This conjecture is validated numerically, and it is at the basis of a novel scheme for network inference: rather than inferring the contact structure, we infer the rates of this surrogate model, and use Bayesian model selection to recover information about the network structure. Elaborating further on this approximation, in Chapter 3 we explore the behaviour of epidemics on networks in the thermodynamic limit ($N \to \infty$). The idea is to test whether the rates of the model satisfy the density dependence relation, meaning that we can express the dynamics of a SIS epidemic on a network as a solution of a one-dimentional Fokker-Planck equation. Apart from a significant reduction of dimensionality, this PDE can be used to perform network inference on high dimensional networks, as it is shown with a fully worked out example. Chapter 2 is based on my first paper, and the original abstract reads

> "*Using the continuous-time susceptible-infected-susceptible (SIS) model on networks, we investigate the problem of inferring the class of the underlying network when epidemic data is only available at population-level (i.e. the number of infected individuals at a finite set of discrete times of a single realisation of the epidemic), the only information likely to be available in real world settings. To tackle this, epidemics on networks*

are approximated by a Birth-and-Death process which keeps track of the number of infected nodes at population level. The rates of this surrogate model encode both the structure of the underlying network and disease dynamics. We use extensive simulations over Regular, Erdős-Rényi and Barabási-Albert networks to build network class-specific priors for these rates. We then use Bayesian model selection to recover the most likely underlying network class, based only on a single realisation of the epidemic. We show that the proposed methodology yields good results on both synthetic and real-world networks."

Chapter 3 is based on a follow-up paper that has been published in 2020, and the original abstract reads

"*Stochastic epidemic models on networks are inherently high-dimensional and the resulting exact models are intractable numerically even for modest network sizes. Mean-field models provide an alternative but can only capture average quantities, thus offering little or no information about variability in the outcome of the exact process. In this paper we conjecture and numerically demonstrate that it is possible to construct PDE-limits of the exact stochastic SIS epidemics on Regular, Erdős-Rényi, Barabási-Albert networks and lattices. To do this we first approximate the exact stochastic process at population level by a Birth-and-Death process (BD) (with a state space of $O(N)$ rather than $O(2^N)$) whose coefficients are determined numerically from Gillespie simulations of the exact epidemic on explicit networks. We numerically demonstrate that the coefficients of the resulting BD process are density-dependent, a crucial condition for the existence of a PDE limit. Extensive numerical tests for Regular, Erdős-Rényi, Barabási-Albert networks and lattices show excellent agreement between the outcome of simulations and the numerical solution of the Fokker-Planck equations. Apart from a significant reduction in dimensionality, the PDE also provides the means to derive the epidemic outbreak threshold linking network and disease dynamics parameters, albeit in an implicit way. Perhaps more importantly, it enables the formulation and numerical evaluation of likelihoods for epidemic and network*

inference as illustrated in a fully worked out example."

1.1.2 Epidemic Control

The second part of this thesis instead focuses on the problem of epidemic control, which has received a lot of attention because of the ongoing pandemic of Covid-19. The pandemic played a major role in driving my second and third year of research, with many projects suspended due to the particular opportunity of researching on Covid-19 related topics. Epidemic control is perhaps the most debated topic of these times, with many contributions from researchers from various fields [Teixeira da Silva et al., 2021]. The main question is which policies should be employed to make sure (at least) that the healthcare system of a country is not overwhelmed as the epidemic spreads, possibly without forcing the whole population to undergo and remain into "full lockdown" more than the necessary. My research in years 2 and 3 focused on using mean-field approximations to describe how to optimally perform control on Covid-19 inspired epidemic models.

In Chapter 4, we consider SIR models to explore how the timing of different short-term non-repeated interventions impacts different quantities related to the severity of the epidemic, in terms of final size, peak prevalence, and duration. We find that, depending on which strategy policy makers adopt, the timing of optimal intervention changes drastically. This investigation is extended to multiple sub-populations to assess how to adapt interventions in each sub-populations to achieve optimality. The work in this chapter is based on my most influential paper to date, and its original abstract reads

> "*The apparent early success in China's large-scale intervention to control the COVID-19 epidemic has led to interest in whether other countries can replicate it as well as concerns about disease resurgence as China relaxes the interventions. In this paper we look at the impact of a single short-term intervention on an epidemic and consider the impact of the intervention's timing. To minimize the total number infected, the intervention should start close to the peak so that there is no rebound once the intervention is stopped. To minimise the peak prevalence, it should start earlier, allowing two peaks of comparable size rather than one very large*

peak. To delay infections as much as possible (as might be appropriate if we expect improved interventions or treatments to be developed), earlier interventions have clear benefit, and waiting until the optimal time gives only small improvements. In populations with distinct subgroups, synchronized interventions are less effective than targeting the interventions in each sub-population separately. We do not attempt to determine what makes an intervention sustainable or not. We believe that is a policy question, and the answer will depend on the disease severity. If an intervention is sustainable, it should implemented early and kept in place. Our intent is to offer insight into how best to time an intervention whose impact on society is too great to maintain beyond a specified duration. "

In Chapter 5, the topic of disease-induced herd-immunity, as introduced in [Britton et al., 2020], is extended to include models able to capture more realistic scenarios, such as heterogeneity in contacts and clustering. The observation at the basis of this line of research is that the disease acts on the population like a targeted vaccine, that preferentially immunises (through infection) high-risk individuals first. This means that, once interventions are lifted, the residual susceptible population may not be able to sustain further outbreaks. Although this is true in the context of mean-field models, when heterogeneity and clustering are in play, the effect may be more subtle. Contact network properties strongly impact the disease-induced herd immunity levels, and therefore a proper assessment of such properties proves necessary for policy-making. The original abstract of this work currently under revision in Bulletin of Mathematics Biology reads

"*The contact structure of a population plays an important role in transmission of infection. Many "structured models" capture aspects of the contact structure through an underlying network or a mixing matrix. An important observation in unstructured models of a disease that confers immunity, is that once a fraction* $1 - 1/\mathcal{R}_0$ *has been infected, the residual susceptible population can no longer sustain an epidemic. A recent observation of some structured models is that this threshold can be crossed with a smaller fraction of infected individuals, because the disease acts like a targeted vaccine, preferentially immunizing higher-risk*

individuals who play a greater role in transmission. Therefore, a limited "first wave" may leave behind a residual population that cannot support a second wave once interventions are lifted. In this paper, we systematically analyse a number of well-known mean-field models for networks and other structured populations to shed further light on some important questions relevant to the current COVID-19 pandemic. In particular, we consider the question of herd-immunity under several scenarios. When modelling interventions as changes in transmission rates, we confirm that in networks with significant degree heterogeneity, the first wave of the epidemic confers herd-immunity with significantly fewer infections than equivalent models with less or no degree heterogeneity. However, if modelling the intervention as a change in the contact network, then this effect might become much more subtle. Indeed, modifying the structure disproportionately can shield highly connected nodes from becoming infected during the first wave and therefore make the second wave more substantial. We strengthen this finding by using an age-structured compartmental model parameterised with real data and comparing lockdown periods implemented either as a global scaling of the mixing matrix or age-specific structural changes. Overall, we find that results regarding (disease-induced) herd immunity levels are strongly dependent on the model, the duration of the lockdown and how the lockdown is implemented in the model."

Finally, in Chapter 6, the topic of translating control measures derived from mean-field-like models onto more realistic network-based epidemic stochastic dynamics. In particular, we focus not on the problem of eradication, but rather on the topic of flattening the curve (also known as the hammer and the dance [Pueyo, 2020]), that is bringing the epidemic curve down to acceptable levels via enforcement of social distancing. This allows to gradually relax social distancing measures, meaning that the population does not have to undergo an indefinite lockdown, but with the option of reintroducing them if the epidemic becomes too large. The tools to devise this strategy are borrowed from feedback control theory, a natural candidate to perform the task of keeping a system close to a pre-determined set point. The

equations of the optimally controlled systems are derived for the simple SIR model. Then, through the use of mean-field models, a map from the simple SIR to a more realistic model on networks is defined. This map allows to translate the control as defined on the original system on a more realistic, stochastic one. This framework was showed to be robust even when the real system employed faced strong delays and perturbations, and may prove to be useful to build up an optimal pandemic response. This last chapter is based on my most-recent paper, and its original abstract reads

> "*Many of the policies that were put into place during the Covid-19 pandemic had a common goal: to flatten the curve of the number of infected people so that its peak remains under a critical threshold. This letter considers the challenge of engineering a strategy that enforces such a goal using control theory. We introduce a simple formulation of the optimal flattening problem, and provide a closed form solution. This is augmented through nonlinear closed loop tracking of the nominal solution, with the aim of ensuring close-to-optimal performance under uncertain conditions. A key contribution of this paper is to provide validation of the method with extensive and realistic simulations in a Covid-19 scenario, with particular focus on the case of Codogno - a small city in Northern Italy that has been among the most harshly hit by the pandemic.*"

1.2 Technical Introduction

In this section, I provide an overview of the tools that will be used extensively through all the chapters of the thesis. To be able to carry out my research, I had to use a number of tools from Network Science, Dynamical Systems, Epidemiology, and Statistical Analysis. In figure 1.8, a simple Venn diagram highlights the most important techniques used in my research.

1.2.1 Networks

A network -or graph- is a pair $\mathcal{G} = (V, \mathcal{E})$, in which V is a set of N nodes, and $\mathcal{E}$ is a set of tuples $\{u, v\}$, such that $u, v \in V$, called links or edges. In the language

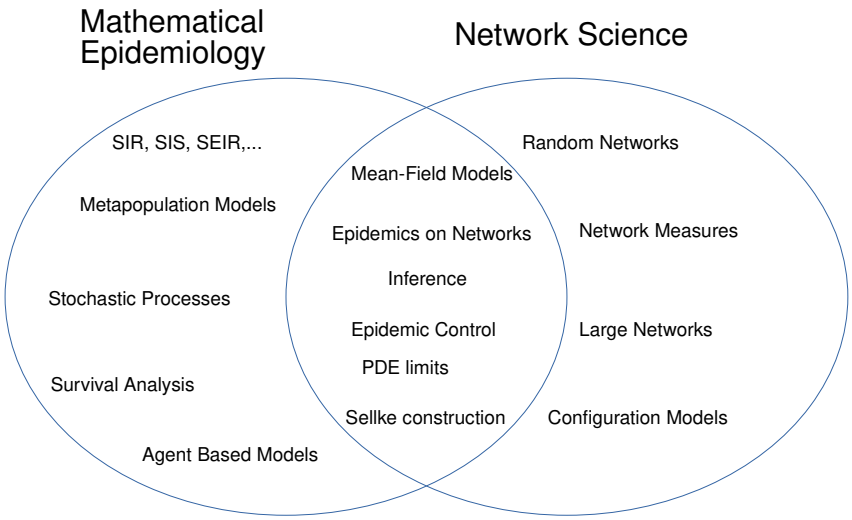

Figure 1.8: The Venn diagram shows the principal research techniques used in my thesis. Mathematical models in Epidemiology include SIR/SIS, Stochastic SIS/SIR and Agent Based Models, whereas tools borrowed from Network Science range from statistical properties of random networks to algorithms for network generation. Network Epidemiology emerges at the intersection between the two disciplines. All the techniques presented and developed in this thesis are part of this multidisciplinary field.

of Network Science, a population contact structure can be modelled by a network in which nodes represent people, and links represent contacts between people. In a realistic scenario, individuals change their contacts in time, but, for simplicity, we will consider static networks, and interpret the links as persistent contacts among individuals. Further, we consider undirected networks, such that $\{u, v\} \in \mathcal{E} \iff \{v, u\} \in \mathcal{E}$, with no multiple edges and no self-links. Given a network $\mathcal{G}$, we define a matrix $G = (g_{ij})$ of size $N \times N$ such that $g_{ij} = 1$ if node i shares a link with node j, 0 otherwise, namely *Adjacency matrix*. Furthermore, $g_{ii} = 0$ and $g_{ij} = g_{ji}$.

Although a network is completely described by its adjacency matrix, often full information about it is completely out of reach. Further, knowing exactly the neighborhood of each node is not always necessary to derive good quantitative results in Epidemiology. There is a number of mathematical tools apt to summarise the most important properties of a network. The single most important quantity of interest is the density of links in the network, that can be summarised by the average number of links per node, or average degree, that is the expected value of the discrete probability distribution $p(k)$ of the degrees of the nodes, where the degree of node i is defined as $k_i = \sum_j g_{ij}$. Different degree distributions can lead to very different epidemics, even if the average degree is constant, as shown in figure 1.9.

The degree distribution allows to describe the neighbourhood of individual nodes. However, although it may not have a significant effect in all cases [Ritchie et al., 2014], it is important to study also the properties of such neighbourhoods. Notably, the most influential quantities in this sense are the *clustering coefficient* and *assortativity*. The global clustering coefficient is defined as the number of closed triplets in a network over the number of all triplets (open and closed), or, in terms of the adjacency matrix

$$C = \frac{\sum_{(i,j,k)} g_{ij} g_{gk} g_{ki}}{\sum_i k_i (k_i - 1)}. \tag{1.3}$$

This quantity reflects the tendency of nodes that have a common neighbour to be neighbours of each other. In terms of epidemics, having a high degree of clustering tends to reduce the spread of the disease. For instance, the value of the rate of transmission needed to generate an epidemic is larger for networks which are clustered when compared to equivalent networks with no clustering. The other quantity of interest is assortativity. This is formally defined as the Pearson correlation coefficient

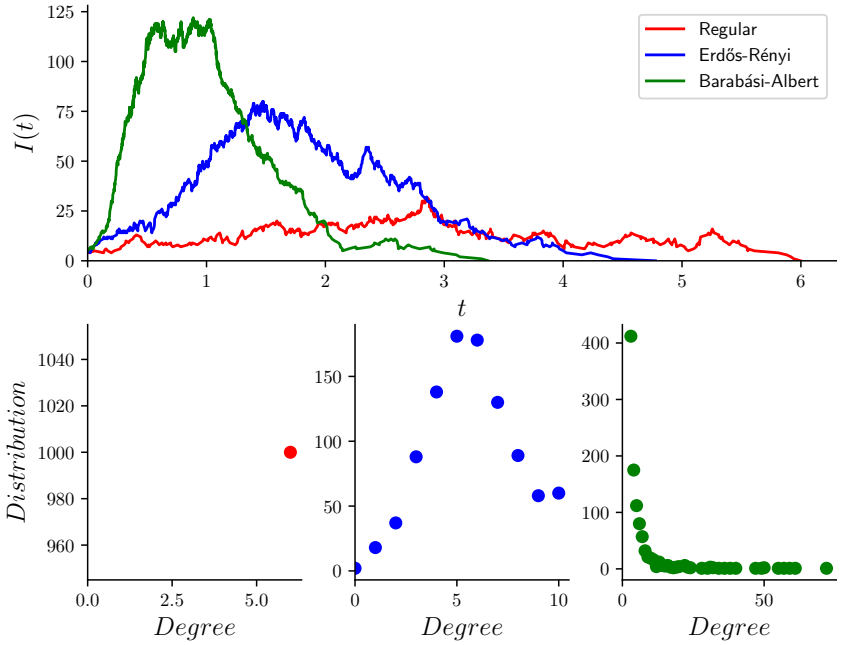

Figure 1.9: (top) Typical paths $(t, I(t))$ are shown for epidemics on various random networks of size $N = 1000$, whose degree distributions (bottom) have the same average. It is clear that epidemics on networks with heterogeneous degree distribution tend to spread faster and infect more nodes than in networks with homogeneous degree distribution.

of the degree of nodes connected by a link nodes [Badham and Stocker, 2010], and it is a quantity that measures the tendency of a node to be neighbour to other nodes with similar degree. Epidemics on assortative networks tend to grow faster than epidemics on disassortative networks - networks where nodes of different degrees are more likely to be connected. Going a step further, one might look at higher-order joint distributions [Orsini et al., 2015]. Another important quantity is the presence of communities within the network, that is, subsets of nodes that share more nodes amongst each other than with the remaining nodes, see figure 1.10. Networks with well-defined communities may exhibit multiple peaks during an epidemic, as epidemics reach different communities.

When we analyze a network, we think of it as a well-defined entity. However, most real-world networks are not fully known or specified, so there must be a degree of randomness implicit in their description. Then, it is useful to think of networks as random structures. With this approach, a single network is a random sample from a more general network class, which is a probability distribution over all the possible networks in this class. This implies that we can gain insights from a particular

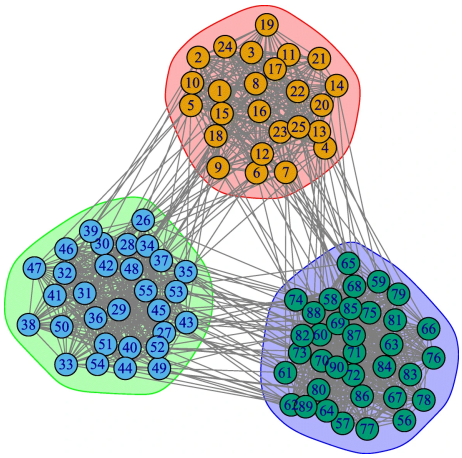

Figure 1.10: A network with three colour-coded distinct communities (the three coloured areas), taken from [Lee and Wilkinson, 2019]. Nodes have many stubs connected to other nodes in the same community, and only a few stubs connecting them to other communities. A most favoured generative algorithm to generate random networks with community structure is the so-called Stochastic Block Model [Holland et al., 1983].

network by studying the whole network class and its properties. A few well-known network models are used extensively in this thesis, and are described in this section. In terms of numerically generating such networks, Python's *networkx* [Hagberg et al., 2008] was the library of choice through my work. The following three network models are presented in order of increasing heterogeneity in degree distribution.

Regular Networks

Regular random networks are networks in which each node has exactly the same number of edges, and are connected randomly. The number of edges per node is called order; for instance, a regular network of order 3 means a network in which every node has exactly three neighbours. A network of N nodes with degree k has a total of $\frac{Nk}{2}$ edges, so two necessary conditions for a regular network to be realisable is that Nk is even and $N \geq k$. These conditions are also sufficient for regular networks to exist. Of particular interest is the regular network of order $k = N - 1$, in which every node is connected to every other node in the network. This is also known as the *complete* network or *fully-connected* network. Regular networks are well studied because of their simple degree distribution, which makes them the ideal benchmark to derive analytical results. Historically, regular networks are known for

their importance in statistical Physics [Dorogovtsev et al., 2002].

Erdős-Rényi networks

Erdős-Rényi networks [Bollobás, 2001] are perhaps the most known class of random networks, and many researchers refer to this class with the term random network. In an Erdős-Rényi network, the degree distribution is binomial $\mathcal{B}(N-1, p)$: each unique pair of nodes is connected with probability $0 < p \leq 1$. It is interesting to notice that, when $p = 1$, Erdős-Rényi networks become fully connected, and that in some sense, for high values of p, Erdős-Rényi networks tend to become similar to Regular networks. Erdős-Rényi networks have been extensively studied in Epidemiology, as they offer good degree heterogeneity and depend on only one parameter; they have therefore become a useful benchmark in many studies [Pastor-Satorras et al., 2015].

Barabási-Albert networks

Although the Erdős-Rényi model produces heterogeneous networks, the binomial degree distribution does not often reproduce what is observed in real-life networks. Further, the degree distribution is based on a predefined set of nodes, meaning that links are added to a pre-existing set of nodes. In many real-world networks that is hardly the case; rather, networks are ever-evolving structures that are, in many cases, driven by two concepts: growth and preferential attachment. The former term means that new nodes are added as time goes by, while the latter means that new nodes connect more readily to nodes that have more links. These two phenomena are at the basis of the so-called scale-free networks, which are very common contact structures in social networks, and explain intuitively why real-world networks have fat-tailed degree distributions. The Barabási-Albert network model proposes an algorithm to generate networks with this property [Barabási and Albert, 1999]. One starts with $m + 1 \leq N$ nodes that are fully connected to each other. Then, a new node is added, with m open stubs to be connected to m different nodes already in the network with probability proportional to their degree k_i, so the probability that a node i gets a new connection is

$$p_i = \frac{k_i}{\sum_i k_i}.$$

New nodes are added with this rule until the desired number of nodes N is reached. At the end of the algorithm, a few nodes will have a huge number of links (such nodes are called hubs), whereas the vast majority will have $\sim m$ links. The degree distribution of a Barabási-Albert random network can be calculated in the limit $N \to \infty$ and is a power-law $P(k) \sim k^{-3}$. Simple modifications of this rule can shift the power-law exponent of the degree distribution.

Configuration models

The most flexible model for a network, in terms of the degree distribution, is the configuration model [Barabási, 2013]. For this model, one needs to define a degree sequence, and assign to each node of the network a degree k_i - that is, a number of open node stubs to be connected at random. In general, this allows for self-loops or multi-edges, which are to be removed after the generation of the network.

Limitations of using these models

Although these models can reproduce, in terms of the degree distribution, a wide variety of real-world networks, they have some limitations that need to be clearly spelled out.

First, as mentioned at the beginning of these sections, these networks are static both in the number of nodes and in the links, meaning two things (a) there is no mechanism to remove or add nodes after the network is generated and (b) two nodes that are neighbours of each-other, remain neighbours forever. In realistic contact structures, new individuals join communities, and in general people tend to form and break contacts with others as time goes by. Further, some contacts are more at risk than others (for example, a student who lives with their family has more frequent contacts with their relatives than with their friends). This can be at least partially solved by assigning a weight distribution to the links on the network, and including the weights into the epidemic models. This means modifying the elements of the adjacency matrix $\mathcal{G}_{ij} = (\omega_{ij})$, $i, j \in [1, N]$, with $0 < \omega_{ij} < \infty$. Links that are more important for the dynamics considered will have a higher weight.

In terms of clustering, all the models above tend to have a small global clustering coefficient. This might be problematic as many real world-networks instead show

clusters of densely connected connected nodes. In other words, the triangle structure -that is, three nodes connected to each other- is under-represented in these network classes. This structure is so common in real contact networks because of the well-known paradigm: "My friends tend to be friends of each other".

Another potential issue of these network models is the lack of mesoscopic structures, such as communities of densely connected nodes. Real world structures, such as social networks, instead show a strong presence of groups of individuals (such as families, neighbourhoods, friendships). This has important implications in terms of the evolution of an epidemic, which in this case might present many different peaks, associated with outbreaks in different communities [Stegehuis et al., 2016].

Although all of these aspects have potentially important consequences on the evolution of epidemics on networks, these networks still have their place in the field; first, because they offer a good range of benchmarks for many modellers; second, because the degree distribution is by far the most important quantity to be studied when it comes to epidemic processes; lastly, because, although they do not take into consideration many real-world characteristics, these models are a major improvement over mass-actions models, while still allowing us to derive analytical results which show how the mean degree, variance of the degree distribution and clustering enters expressions such as $\mathcal{R}_0$ and final epidemic size.

1.2.2 Epidemics on Networks

Stochastic models

Many epidemic models start from a stochastic formulation rather than a deterministic one. This is because compartmental models fail to account for statistical fluctuations that are observed in many real-world epidemics. When a virus has infected very few people, for instance, there is still a non-negligible probability that the outbreak does not evolve into a sustained epidemic, due to the intrinsic stochastic nature of disease transmission. Ordinary differential equations (ODE) models instead predict that, as soon as one infected individual is introduced in a fully susceptible population, and if $\mathcal{R}_0 > 1$, epidemics are bound to happen. In reality, a typical epidemic undergoes two distinct phases: the first one, at the beginning, is dominated by stochastic fluctuations, in which infection events and recoveries hap-

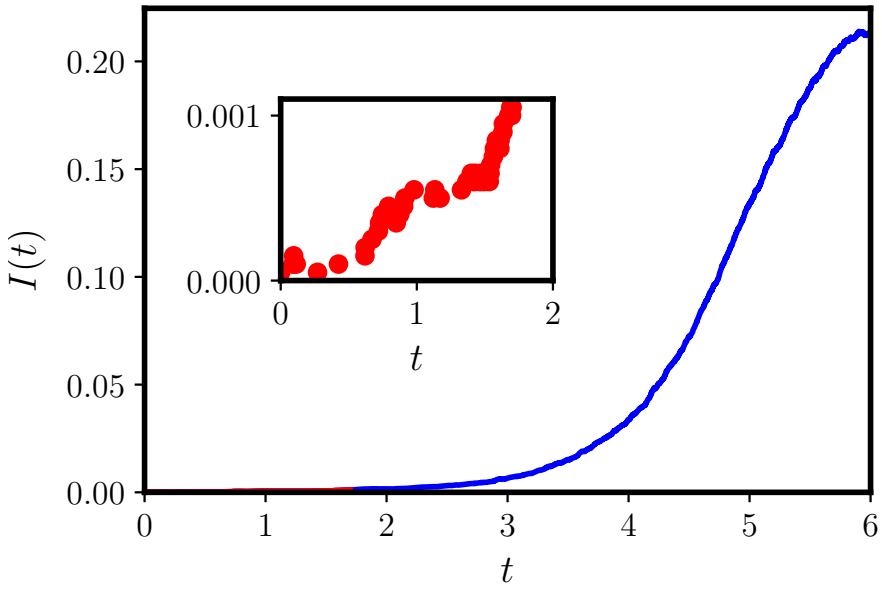

Figure 1.11: A typical realisation of an epidemic on a Erdős-Rényi network, with a focus on the first few events (in the inset), where the stochastic fluctuations dominate. Parameters are $N = 1000$, average degree $\langle k \rangle = 7$, $\tau = 0.5$, $\gamma = 1.2$.

pen with the same frequency (this is particularly true for epidemic with $\mathcal{R}_0 \sim 1$). This phase cannot be captured by ODE models. If the stochastic fluctuations bring the system to a point where the rate of infections events outgrows that of recoveries, the system enters in a new phase, in which the epidemic sustains itself. ODE models are able to describe this phase. An example of the different phases of the epidemic can be found in figure 1.11. In the following, we give examples for SIR models, but the framework is rather general.

The assumptions behind stochastic epidemic models are:

- the contact structure of the population is a network of size N, described by the adjacency matrix $\mathcal{G}$;
- each infected individual transmits the disease to their susceptible neighbours at rate $\tau = \frac{\beta}{N}$;
- each infected individuals recovers at constant rate γ independently of other individuals.
- the process is fully Markovian and it is completely determined by the initial condition.

Master equations

The state space in this description consists of all the possible arrangements of m states over N nodes, that is m^N possible states. An exact description requires an equation for each state. For a SIR model, this is 3^N equations needed to describe how the epidemic evolves at a node level. The resulting Kolmogorov (or master) equation would be of the form

$$\frac{d}{dt}\mathbb{P}\left(X(t)=\alpha\right)=\sum_{\beta}\mathcal{G}_{\alpha\beta}\mathbb{P}(X(t)=\beta)-\sum_{\beta}\mathcal{G}_{\beta\alpha}\mathbb{P}(X(t)=\alpha), \tag{1.4}$$

where $X(t)$ is the state of the system at time t, which can take values $\alpha \in [S_0, \ldots, S_{3^m}]$, and $\mathcal{G}_{\alpha\beta}$ is a $3^N \times 3^N$ constant matrix of (non-negative) rates that determine the possible transitions among the states of the system. The first term refers to all the possible ways to transition to a state α from other states, while the second term refers to possible ways to exit from state α and transition to state β. This equation can be solved for very modest network sizes, as shown in [Kiss et al., 2017]. However, realistic systems have a number of nodes that easily goes above the thousands, or even the tens of millions; it would simply require too much computational capacity to get numerical solutions. Therefore, instead of focusing on the detailed description of the system, it is more convenient to study aggregate statistics. In particular, the focus is usually on the expected number of infected nodes in the system. In more formal terms, we consider a partition of the state space in $N(N+1)/2$ disjoint sets $\mathcal{C}_{ij}$, with $0 \leq i, j \leq N$, that include all the states in which i nodes are susceptible and j infected. Ideally, one would like to give equations for the variables defined by this new state space, but unfortunately, although the underlying system is fully Markovian, this property is lost when considering this reduced state space. There is a notable exception, that is fully connected networks. These networks are symmetric under any relabeling of the nodes, that is the permutation group $\mathbb{S}_N$: all states with i susceptible nodes and j infected can be lumped together. This allows to define a partition on the set of possible states, say C_{ij}, that groups all the states with the same number of susceptible and infected nodes. For instance, $C_{N,0} = \{(S, S \ldots, S)\}$ includes only the state in which all the nodes are susceptible; $C_{N-1,1} = \{(I, S \ldots, S), \ldots, (S, \ldots I, \ldots S) \ldots, (S, \ldots, S, I)\}$, instead includes all the states where all the nodes but one are susceptible. We can now define a new set of

variables

$$p_{ij}(t) = \sum_{k \in C_{ij}} X_k(t),$$

where k is an index that moves across the elements of C_{ij}, and $X_k(t)$ is the probability of the k-th element. This quantity represents the probability of having i infected and j susceptible nodes at time t. When the network is fully connected, all the states that share the same number of infected and susceptible nodes also have the same rate of transition from or into other states. The reason for this is essentially that the fully connected network is symmetric under any relabeling of the nodes - that is, the permutation group, so all the states with a specified number of infected and susceptible are identical from the point of view of the dynamics. Then, it is easy to derive an exact Master equation for this particular case

$$p_{i,j}(t) = \tau(i-1)(j+1)p_{i-1,j+1}(t) + \gamma(j+1)p_{i,j+1}(t) - (\tau j + \gamma i)\, p_{i,j}(t), \qquad (1.5)$$

where τ is the per-contact rate of transmission of the disease and γ the rate of recovery. Without going into the details of the derivation (that can be found in [Kiss et al., 2017]), the meaning is clear. To get into a state with i susceptible and j infected nodes, one starts either from $i-1$ susceptible and $j+1$ infected, and an infection happens, or from i susceptible and $j+1$ infected and a recovery happens. Once in this state, the probability of exiting comes either from infection events or from recoveries. Because of the symmetry, it does not matter which node is infected or susceptible, because the rate at which events happen is determined by the number of $S-I$ links, i.e. it depends only on i and j in this case. Unfortunately, we cannot derive an explicit master equation for a general network, the reason being that, in general, the way infected and susceptible nodes are positioned on the network determines the number of $S-I$ links, and therefore the rates at which transitions happen. A reduced master equation for the general network then is completely out of reach, as its rates would be themselves random variables, that depend on the topology and on the whole history of the disease.

Not everything is lost though, because if one looks at the expected quantities $[A] = \sum_{i=1}^{N} \mathbb{P}(x_i(t) = A)$, where $\mathbb{P}(x_i(t) = A) = \sum_{j=1}^{3^N} \mathbb{P}(S_j(t))\mathcal{I}_{(S_j)_i = A}$, and $\mathcal{I}_x$ is

the indicator function, then the following system of equations holds:

$$\begin{aligned}
[\dot{S}](t) &= -\tau[SI](t), \\
[\dot{I}](t) &= \tau[SI](t) - \gamma[I](t), \\
[\dot{R}](t) &= \gamma[I](t),
\end{aligned} \tag{1.6}$$

where $[SI](t)$ is the expected number of $S-I$ links in the system at time t, that is $[SI] = \sum_{i=1}^{N}\sum_{j=1}^{N} g_{ij}\mathbb{P}(x_i = S, x_j = I)$, where (g_{ij}) is the adjacency matrix of the network, see [Kiss et al., 2017] for a rigorous derivation. Although the derivation of this system is not immediate, the meaning is quite important: epidemics are driven by the term $[SI]$, which describes how many contacts infected nodes have with susceptible nodes. The problem is that we do not have an expression or an equation for the term $[SI](t)$. To recap, one starts from an exact system of equations (1.4), that involves solving m^N differential equations. To reduce the dimensionality of the problem, the focus is shifted on aggregate statistics - i.e. the number of infected and susceptible at time t, without information on whom is in either state. For the fully connected network, this reduces to an explicit system of $(N+1)$ equations. Unfortunately, for the general network, we cannot derive an exact system of equations, unless we focus on expected values, as in equation (1.6). The problem now is that, although elegant, to solve it one needs an equation for the value of $[SI](t)$ as well. There are two options: (1) to approximate $[SI]$ as a function of $[S]$ and $[I]$ only, a method known as *closing* the system or (2) to consider a differential equation for $[SI]$.

Mean-Field models

The problem of approximating the expected value of the $S-I$ links can be approached from many directions (the book by [Kiss et al., 2017] summarises the state of the art in this area). Here we briefly explore the method of closures. The goal is to write an approximate expression for $[SI]$. If the average degree of the network is n, the *homogeneous mean-field model* prescribes that $[SI] \sim \frac{n}{N-1}[S][I]$. The intuition is the following: as a first approximation, we consider that each susceptible nodes has exactly n neighbours, each one infected with probability $\frac{[I]}{N-1}$, therefore

the closed system becomes:

$$
\begin{aligned}
[\dot{S}](t) &= -\frac{\tau n}{N-1}[S](t)[I](t), \\
[\dot{I}](t) &= \frac{\tau n}{N-1}[S](t)[I](t) - \gamma[I](t), \\
[\dot{R}](t) &= \gamma[I](t).
\end{aligned}
$$

Note how the dynamics at this approximation level is driven by the term τn, that is the product of a term that accounts for the biological characteristics of the pathogen and a term that accounts for the average number of contacts a typical node has. This means that, in principle, it is not possible to distinguish between a very transmissible pathogen spreading on a very sparse network and a weaker pathogen on a denser network, at least in the first order. This approximation ignores both the degree heterogeneity and any higher order network property. Even on homogeneous networks, the mean-field model tends to over-estimate $[SI](t)$. This is because of the epidemic dynamics itself: to become infected, a node needs to have at least an infected neighbour, meaning that the neighbourhood of an infected node likely contains other infected nodes, while this approximation implies that infected nodes are uniformly distributed on the network. In terms of $\mathcal{R}_0$, linear stability analysis around the disease-free state $[N, 0, 0]$ shows that epidemics are possible only when $\frac{\tau n N}{N-1} > \gamma \rightarrow \mathcal{R}_0 \simeq \frac{\tau n}{\gamma}$. With respect to the value $\mathcal{R}_0$ obtained from eq (1.1), this expressions explicitly shows how the density of contacts impacts the reproduction number: denser networks lead to higher values of $\mathcal{R}_0$.

To improve on this approximation, we can shift the closure one term further, that is, deriving equations for the terms $[SI](t)$, $[SS](t)$. Note that the count is directional, so in principle $[IS]$ and $[SI]$ are two distinct quantities, although their value is the same. The rigorous derivation can be found in [Kiss et al., 2017]. The exact equations that $[SI]$, $[SS]$ satisfy are

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

To explain how these terms arise, we can focus on the term $[SI](t)$; the depletion of $S-I$ links depends is driven by three distinct mechanisms: 1) recovery of the infected node ($-\gamma[SI]$) 2) infection of the susceptible node by the other node ($-\tau[SI]$) 3)

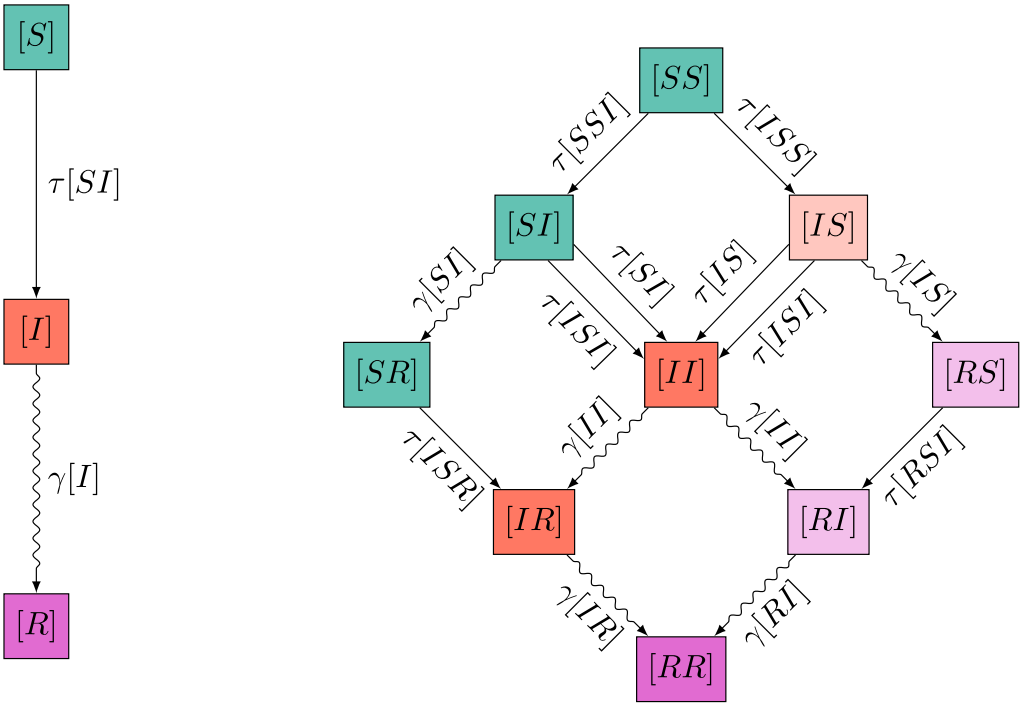

Figure 1.12: Flow diagrams showing the flux between compartments of singles (left) and compartments of pairs (right) in the SIR pairwise model (taken from [Kiss et al., 2017]). On the right, straight arrows denote infections caused by a node within the pair or from outside the pair (in this case, the rate depends on a triple). Curved arrows indicate recoveries. The colours refer to the first node status of the pair.

infection of the susceptible node by one of its other infected neighbours ($-\tau[ISI]$). Finally, the rate at which $[SI](t)$ pairs are created is given by infection of either of the nodes of a $[SS]$ pair by one of their infected neighbours, that is $[SSI]$. Note that the term $[SSI]$ is counted directionally, meaning that triples of the form $[ISS]$ are not part of the term $[SSI]$.

The equation for the pairs involve the triples, then the idea is to make a closure at that level, as depicted in figure 1.12: start from the total number of edges starting from susceptible nodes, that is $n[S]$. The total number of $S-I$ links is $[SI]$, hence $\frac{[SI]}{n[S]}$ is the proportion of edges from susceptible to infected. Similarly, for $S-S$ links, we get $\frac{[SS]}{n[S]}$. For the $[SSI]$ triples, iterate this to consider three nodes, i.e. start from a susceptible node, consider two of its neighbours. The probability that one is susceptible and the other is infected is, in the first order, $[SS][SI]/n^2[S]^2$. Then, considering that a susceptible node has $n(n-1)$ different couples of neighbours, we get that as a first approximation that

$$[SSI] \simeq \frac{n-1}{n}\frac{[SS][SI]}{[S]}.$$

Similarly for the $[ISI]$ triples

$$[ISI] \simeq \frac{n-1}{n}\frac{[SI]^2}{[S]}.$$

Putting all the equations together with these closures, we get

$$\begin{aligned}
[\dot{S}] &= -\tau[SI], \\
[\dot{I}] &= \tau[SI] - \gamma[I], \\
[\dot{R}] &= \gamma[I]. \\
[\dot{SS}](t) &= -2\tau[SSI] = \frac{n-1}{n}\frac{[SS][SI]}{[S]}, && (1.7)\\
[\dot{SI}](t) &= -\gamma[SI] + \tau\left([SSI] - [ISI] - [SI]\right) && (1.8)\\
&= -\gamma[SI] + \tau\left(\frac{n-1}{n}\frac{[SS][SI]}{[S]} - \frac{n-1}{n}\frac{[SI]^2}{[S]} - [SI]\right), && (1.9)
\end{aligned}$$

where the explicit dependence on time has been omitted for readability. In this case, the value of $\mathcal{R}_0$ can be found with linear-stability analysis around the disease-free state, leading to

$$\mathcal{R}_0 = \tau\frac{(n-2)}{\gamma}$$

This expression is different from the previous one, and in general has a lower value. This means that the "$I - I''$ correlations that mean-field models failed to account, have an impact in this more refined approximation.

When the network is not homogeneous, we need to break the approximation down by degree. Instead of considering $[S]$, $[I]$, $[S - I]$ and so on, we need to consider $[S_k], [I_k]$, that is, the expected number of susceptible and infected nodes with degree k. Similarly, for the couples $[S - I]$, we need to consider $[S_k - I_j]$, that is, the couples in which the first node is susceptible and has degree k and the second node is infected and has degree j. The equations obtained in this way lead to the so-called heterogeneous mean-field models [Kiss et al., 2017]. Clustering and assortativity can be considered in this approximations as well, leading to clustered pairwise models and heterogeneous pairwise models, respectively. Without showing explicitly the equations for these cases, which are derived in Chapter 5, the ideas can be briefly summarised in this way. In a clustered network, it is important to distinguish between closed triples and open triples. This means that the closures for terms such as $[SSI]$ need to be modified to account for the fact that the third node might be connected to the first one. To do so, one starts from an $S - I$ link and consider the $(n - 1)$ other neighbours of the susceptible node. The average number of those neighbours that are connected to the infected node, if C is the clustering coefficient as defined in 1.3, is $(1-C)(n-1)$, while the average number of neighbours not connected to it is $C(n-1)$. For the former, the closure is identical to the original one, that is $[SSI]_{open} = (1 - C)\frac{n-1}{n}\frac{[SS][SI]}{[S]}$. For the neighbours that are connected instead, we know that the correlation between $[S]$ and $[I]$ nodes plays an important role, so we scale the previous quantity with it $[SSI]_{closed} = \frac{n-1}{n}\frac{[SS][SI]}{[S]}C\frac{N[SI]}{n[S][I]}$; the intuition is that knowing that the three nodes form a closed triple changes the probability of the states of the neighbours. Then, the closure for $[SSI]$ becomes:

$$[SSI] = \frac{n-1}{n}\frac{[SS][SI]}{[S]}((1 - C) + C\frac{N[SI]}{n[S][I]}),$$

and similarly for other closures. Concerning assortativity instead, one can implement it directly via initial conditions such as $[A_k B_j][0]$. This can be done choosing the heterogeneous pairwise model.

To summarise the state of the art on this topic, It should be clear that one major problem of epidemics in general, and on networks in particular, is the forbiddingly

high dimensionality of the system. Tackling this is perhaps one of the main goals of Mathematical Epidemiology. Mean-field approximations, pairwise models, edge-based compartmental models [Miller et al., 2012], percolation approaches [Miller, 2016] are just a few approximation methods to handle the dimensionality of the problem. In this sense, the first part of my thesis (Chapters 2,3) tackles the problem of dimensionality reduction from a novel perspective. The approach starts from the master equation (1.4). For a general network, we know that the rates are random variables that cannot be described exactly without knowing the exact state of the system. A possible solution to this problem is to define a new master equation, in which rates of infections are chosen so that the resulting $p_k(t)$ approximate that of the true process. This would provide a master equation that is analogous to that of Birth-and-Death processes **??**. The benefit of this approach with respect to the method of closures is that, while the latter results in a deterministic trajectory of the epidemic, the master equation gives the probability distribution of observing k infected at time t and it is thus able to account for the stochastic fluctuations typical of the true process. The second part of my thesis (Chapters 4,5,6) instead focuses on the application the well-known theoretical advancements in this field described in this section for epidemic control. Linking mean-field models to epidemic control is a step-up in complexity that is much required in pandemic response, that often relies either on simpler models such as the SIR or on theoretically intractable agent based models.

A note on Gillespie simulations on a network

The assumptions described at the beginning of this section allow immediately to formulate epidemics on networks in terms of algorithms. The reason for introducing such algorithms now is that all the analysis of mean-field models can be validated only by means of simulations on a computer, where the assumptions underlying the formulation of epidemic on network are explicitly guaranteed. This means that simulations *are* realisations of the true process, and all the theory and approximations regarding the true process can be tested and numerically validated on simulations, although their scope is to formalise a set of assumptions about a real-world system. The most known algorithm to simulate the true process is called

Gillespie [Gillespie, 1977] (it is not the only one; for instance, the Sellke construction works even with non-Markov epidemics [Sellke, 1983]). The algorithm follows directly from the following assumptions: at any given point, we have many concurrent processes, recoveries and infections. As soon as an event happens at time t, thanks to the Markovian property of the process, the next event depends only on the state of the system at time t. For each infected node one extracts a single sample from a $\exp(\gamma)$ distribution, which determines the time of the node's recovery. Then, for each connection between an infected and a susceptible node, a sample from an $\exp(\tau)$ distribution is drawn, which determines the timing of when the susceptible node turns infected. Then, the event with the earliest sampled time is resolved. This is interesting, because it is well known that if $T_1, T_2, \ldots, T_k$ are k random variables exponentially distributed with parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ respectively, then $T = \min(T_1, T_2, \ldots, T_k) \sim \exp(\sum_{i=1}^{k} \lambda_i)$. This offers a compact way to define a pseudocode: compute all the rates - that is, count how many infected nodes and how many $S - I$ links there are on the network, extract the time to next event distributed as $\exp(\tau[S-I] + \gamma I)$, where $[S-I]$ is the number of $S-I$ links, and finally select the event at random but proportionally to its rate (the higher the rate, the larger the probability of getting selected). Then repeat this until there are no more events possible. Pseudocode 1 shows a simple version of this algorithm. Note that this is only a version of the Gillespie algorithm that solves all the events at a node-level, but does not inform on who-infects-who, which may be important for the modeller. Other (more efficient) versions of the Gillespie algorithm, such as the event-driven approach, can be found in the Appendix of [Kiss et al., 2017].

Algorithm 1: Gillespie simulations

```
Result: Times, Events
A ← Adjacency matrix;
X ← list of node states;
X ← X_0 initial condition;
R ← vector of length X full of zeros;
while no more infected do
    for node in X do
        if X infected then
            R[node] = γ;
        end
        if X is susceptible then
            for neighbour in A.neighbours(node) do
                R[neighbour] += τ;
            end
        end
        else
            R[X] =0
        end
    end
    T = random_exp(sum(R)) time of next event;
    E = random_sample(R) extract node that undergoes event with
     probability proportional to rate;
    X ← update;
    Times.append(T);
    Events.append(E);
end
return Times,Events;
```

Chapter 2

Network Inference from Population-Level Observation of Epidemics

2.1 Introduction

Networks are an important tool for modelling systems with many interacting parts such as epidemics spreading within a population or neuronal activity in the brain. Indeed, the intricate interplay of many individual well-defined units can be captured by the links of a network, and this can be done with an unprecedented level of detail [Newman, 2003b; Keeling and Eames, 2005; Danon et al., 2010; Porter and Gleeson, 2016; Kiss et al., 2017]. For instance, directed, weighted or temporal links can all be considered within this modelling paradigm. A main feature of network epidemic models is that the topology of the contact structure is treated separately from the characteristics of the pathogen (such as infectivity and typical recovery time), in contrast to mass action models such as Kermack–McKendrick [Kermack and McKendrick, 1927]. The transmission dynamics of epidemic spreading on networks can be modelled as a continuous time Markov Chain process on a network [Pastor-Satorras et al., 2015]. Unfortunately, the literature only show few exact results, and these are mainly related to specific cases or small networks. Therefore, approximations are often introduced to simplify the exact model and derive quantitative results. Most notably, well-known and widely used theoretical approaches include mean-field

and higher order approximations [Gomez Rodriguez et al., 2010; Kiss et al., 2017], edge-based dynamics [Miller et al., 2012; Wang et al., 2018], percolation [Moore and Newman, 2000; Miller, 2009; Gleeson, 2009] and generating functions [Newman, 2003b; Pastor-Satorras et al., 2015].

These approaches have led to the realisation that the structure of the network has a profound impact on how diseases invade, spread and how to best control them. This impact is particularly well understood for degree heterogeneity and assortativity/disassortativity, and to a lesser extent, for clustering, the propensity of nodes that share a common neighbour to be connected [Pastor-Satorras et al., 2015; Kiss et al., 2017].

However, depending on the field of application, the precision to which the underlying network is known can vary greatly, from absolute (when full description is available) to absent (when a description is entirely lacking). For example, whereas some technological networks can be mapped out to a great degree of detail, social networks can be challenging to query [Brugere et al., 2018]. This has resulted in a significant amount of research aimed to develop methods for link prediction (for a survey, see [Brugere et al., 2018]). Instead of assuming the availability of explicit information about nodes and edges, these methods rely on 'observables' from dynamical processes taking place on the network, under the assumption that these provide latent information about the missing underlying network structure. In the framework of epidemics on networks this suggests that it is possible to get insights about the structure of the network by observing quantities of interest at node and perhaps population level. Indeed, the inverse problem of inferring networks from epidemic data has been the subject of great scrutiny.

In particular, in the context of statistical inference, this task has been approached by either formulating it as a likelihood optimisation problem [Gomez Rodriguez et al., 2010; Netrapalli and Sanghavi, 2012; Myers and Leskovec, 2010; Du et al., 2012; Gomez Rodriguez et al., 2014] or using Bayesian inference [O'Neill and Roberts, 1999; Britton and O'Neill, 2002; Groendyke et al., 2011; Stack et al., 2013; Dutta et al., 2018]. Compared to maximum likelihood optimisation methods (e.g. independent cascade model [Gomez Rodriguez et al., 2010]), the Bayesian inference is usually based on a smaller number of observations of the epidemic [Britton and

O'Neill, 2002; Groendyke et al., 2011; Dutta et al., 2018]. However, both network inference approaches (explicit link inference and inferring parameters of a known network model) lead to good estimates for the network and parameters of the epidemic dynamics. Moreover, there is an interesting tradeoff between them. The former is able to identify the adjacency matrix, but requires the observation of a large number of cascades, whereas the latter can only infer some structural parameters (such as the probability of a link between two nodes), but relies on fewer observations. Recently, it has been conjectured [Prasse and Van Mieghem, 2018] that an exact (link-by-link) reconstruction of networks might not be feasible due to requiring a subexponentially increasing number of observations and an exponentially increasing computation time with respect to the number of nodes in the network.

A common feature of the above mentioned work is their reliance on the availability of detailed data at node level, such as the complete temporal knowledge of all cascade trees in [Gomez Rodriguez et al., 2010] or the observation of all the removal/infection times in the Bayesian framework of [Britton and O'Neill, 2002]. However, in most real-world scenarios, such detailed information is unlikely to be available. A more reasonable expectation is to be provided with population-level observations, that is, the number of infected nodes in the whole network at various times. Our aim in this paper is to establish the feasibility of inferring the class of the underlying network from population-level observations. Whilst a very recent paper [Ma et al., 2019a] provides an algorithm to infer properties of a given network-type from prevalence data, we are not aware (for a survey, see [Danon et al., 2010]) of any research that specifically addresses the problem of network class inference based purely on population-level observations. We do so within the framework of continuous-time SIS epidemics on networks when only population-level data from a single realisation of the epidemic are available.

We treat this problem as an inverse problem and adopt a Bayesian approach which involves the following steps:

(a) propose a parametric forward model that reproduces network/population-level dynamics and reflects network structure;

(b) build a prior distribution for these model parameters on a network class basis;

(c) use the posterior measure to identify the most likely network class.

A complete description of the SIS dynamics on a network with N nodes requires to solve 2^N equations, one per possible state. The distribution of population-level statistics in time can be described via the count of the number of infected nodes in this dynamics; however, this process scales exponentially with the size of the network. Here, we take a different route and choose to use a surrogate model to represent the evolution of the count of infected nodes in the population. A reasonable candidate for this is a Birth-and-Death process (BD), see [Nagy et al., 2014], characterised by only $N+1$ equations and $2(N+1)$ free parameters, the rates of infection and recovery, that need to be tuned to best represent the exact model. Whilst the rates of recovery are network independent and known exactly, the rates of infection in the surrogate model are more challenging to define.

In this work, the rates of infection in the surrogate model are provided by a simple parametric model, together with an estimation procedure based on extensive and detailed simulations of epidemics on three classes of well-known random networks: Regular, Erdős-Rényi and Barabási-Albert. This procedure leads to distinct rate models for the three classes of networks. These observations are encapsulated in a prior distribution for the rates of the BD process.

Finally, when one observes a single epidemic through population-level data, our prior and forward model can be used within a Bayesian model selection framework to identify the most likely underlying network class. It is worth noting that this framework is versatile enough to be used in conjunction with any set of population-level epidemic data, as it will still output the most likely network class, that is, the closest class (in terms of heterogeneity of the degree distribution) to that of the true underlying network.

The paper is structured as follows. In Section 2.2 we describe the BD surrogate forward model together with a three-parameter model for its rates of infection. Section 2.3 includes all aspects of the Bayesian approach we used, from building priors to model selection and model validation/stress testing. We conclude with a discussion and further research directions in Section 5.5.

2.2 The forward model

A population of N individuals is considered with the contact structure between individuals described by an undirected network with adjacency matrix $G = (g_{ij})_{i,j=1,2,\ldots,N}$ where $g_{ij} = 1$ if nodes i and j are connected and zero otherwise. Self-loops are excluded, so $g_{ii} = 0$ and $g_{ij} = g_{ji}$ for all $i, j = 1, 2, \ldots N$. The standard SIS epidemic dynamics on a network [Kiss et al., 2017] is considered, which is driven by two type of events: (a) infection and (b) recovery from infection. Infection can spread from an infected and infectious node (I) to any of its susceptible neighbours (S) and this is modelled as a Poisson point process with per-link infection rate τ. Infectious nodes recover at constant rate γ, independently of their neighbours and become susceptible again. Initialization is made by randomly choosing I_0 nodes to be infected at the initial time, the others being susceptible. The resulting model is a continuous-time Markov Chain, and to fully specify its state we need an equation for each arrangement of length N with entries being either S or I, resulting in a state space of 2^N elements. While this is easy to formalise and write down theoretically, the numerical integration of the system becomes intractable even for modest values of N [Simon et al., 2011; Danon et al., 2010; Simon and Kiss, 2013; Kiss et al., 2017]. This motivates us to use a surrogate model, offering sufficient flexibility to approximate the time evolution of the number of infectious nodes in the network.

2.2.1 Birth-and-death approximation of SIS epidemics

We use a BD process, a continuous-time Markov chain with state space $\{0, \ldots, N\}$ and transitions of unit size, as the surrogate model. The up-jumps or infections are described by rates a_k, that is, the rates of infection in the presence of k infected nodes and encode the network structure. The down-jumps or recoveries are described by rates $c_k = \gamma k$. To understand why, we first observe that recoveries are independent events (since once a node is infected, its status no longer depends on other nodes). An infected node recovers after a time drawn from an exponential distribution with rate γ. If k nodes are infected, the first recovery is going to happen according to the minimum of all recovery times, which is again exponentially distributed with rate γk. Hence, the transition probabilities of the surrogate process are given by

the following forward Kolmogorov (or Master) equation:

$$\forall k \in \{0, \ldots, N\},\ \dot{p}_{k_0,k}(t) = a_{k-1}p_{k_0,k-1}(t) - (a_k + c_k)\, p_{k_0,k}(t) + c_{k+1}p_{k_0,k+1}(t), \quad (2.1)$$

together with $a_{-1} = c_{N+1} = 0$ and an initial condition $k_0 \in \{0, \ldots, N\}$, with $p_{k_0,k}(t)$ the probability of being in the state with k infected at time t, given initial k_0 infected. The solutions of Eq. (2.1) and the rates of infection will be denoted by $p^{\alpha}_{k_0,k}$ and a^{α}_{k}, respectively, when the dependence on additional parameters α needs to be enforced.

The quality of the surrogate model, i.e., how well it approximates the exact model, depends strongly on the choice of infection rates a_k. The way a_k depends on k is determined by the underlying network structure. An analytic formula for a_k is only available for the fully connected network, namely: $a_k = \tau k(N - k)$, that is the number of S-I links (i.e., links connecting susceptible and infected nodes) in the network multiplied by the per-contact rate of infection τ.

In fact, in a stochastic simulation of the epidemic on a network, the rate of going from k to $k+1$ infected nodes is exactly $\tau \times$ #S-I links. Hence, during a simulation it makes sense to keep track of the number of infected nodes, the number of S-I links and the time spent in each respective state. Further important observations can be made. The number of S-I links is a random variable and given a fixed number of infected nodes, say k, the number of S-I links can take different values. This is simply due to the stochasticity in how the infected nodes are laid out in the network. Thus a plausible choice for the rate a_k may be simply the average of the number of S-I links when there are exactly k infected nodes. However, some states are longer lived than others and this needs to be accounted for. Combining all the above, an empirical average rate of infection emerges, that is

$$\hat{a}_k = \tau \frac{\sum_i i t_{i,k}}{\sum_i t_{i,k}},\ 1 \leq k \leq N, \quad (2.2)$$

where $t_{i,k}$ is the lifetime of a state with k infected nodes and i S-I links. We will use the notation $\hat{a}_k^{\theta,\tau,\gamma}$ to indicate the resulting estimate given the network class $\theta \in \Theta$:

$$\Theta := \{\text{Reg, E-R, B-A}\}.$$

where we use Regular (Reg), Erdős-Rényi (E-R) and Barábasi-Albert (B-A) network classes. There are a number of reasons for this choice. First, these three classes are perhaps the most popular random network models, so they provide a good

benchmark to test our model. Second, they can produce rich topologies in terms of degree heterogeneity, and therefore allow us to test the flexibility of our framework. Finally, the absence of higher-order structures (such as communities, or clustering) enables us to simplify the problem of fitting the (k, a_k) curves and thus to focus more specifically on the problem of network inference.

Hence, we can calibrate the infection rates a_k through a statistical analysis based on stochastic simulations of the SIS epidemics on networks. Namely, for a network class (with given average degree) and given disease parameters (τ, γ), we run 50 outbreaks on 50 different realisations of the network. We keep track of the states that the process visits along with the number of infected nodes, number of S-I links and lifetime of the states. This data feeds into Eq. (2.2) and leads to the value of $\hat{a}_k$ for all $0 \leq k \leq N$. To cover the entire range, $0 \leq k \leq N$, half of the outbreaks are started from $k_0 = 5$ infected nodes, chosen uniformly at random, and the other from $k_0 = N$ infected nodes. The former allows us to explore the curve up to the steady state, while the latter, although an artificial scenario, allows us to explore the curve from the steady state to N. Typical $(k, \hat{a}_k^{\theta,\tau,\gamma})$ curves are shown in Fig. 2.1. In what follows we assume that these rates are 'optimal' and that they lead to a surrogate model that agrees well with the exact one. This choice is motivated by the heuristics presented above which is further validated through extensive numerical simulations for three network classes and a large set of disease parameter values (see Section 2.3).

2.2.2 Three-parameter model of infection rates

Consistent with results in [Nagy et al., 2014], we notice that, although estimated $\hat{a}_k$ curves are distinct for different network classes, they all share some common features: specifically, they all satisfy $\hat{a}_0 = \hat{a}_N = 0$ and exhibit a single maximum in $[0, N]$. Perhaps the most important features that change between the three distinct network classes are the flatness and skewness of the $\hat{a}_k$ curves (see Fig. 2.1). It is clear that high heterogeneity in the degree distribution (i.e. Reg $\rightarrow$ E-R $\rightarrow$ B-A, displaying respectively no $\rightarrow$ medium $\rightarrow$ high heterogenity) increases the left skew.

The intuitive reason for these differences in the $(k, \hat{a}_k)$ curves is that epidemics on such different networks spread with distinct enough characteristics. In scale-free

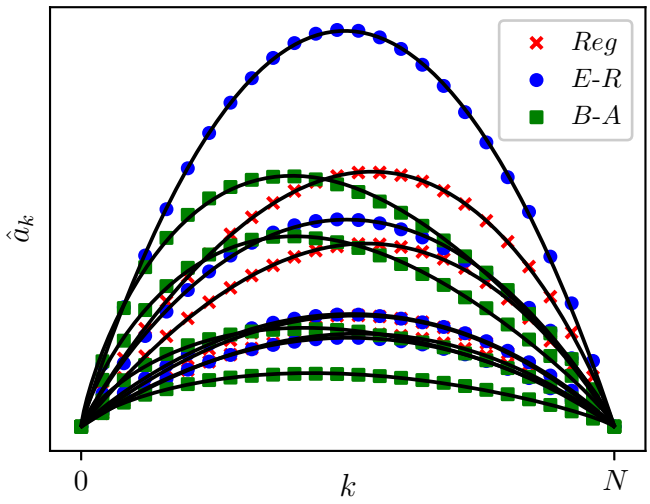

Figure 2.1: $\hat{a}_k$ curves (markers) along with the best fits from the C, a, p model (plain lines) on 12 different combinations of network classes and epidemic parameters. Parameters of the simulations considered are, from top to bottom: Reg (crosses), $(\langle k \rangle, \tau, \gamma) = \{(12, 1.43, 5.69)(6, 9.46, 4.23)(8, 4.47, 8.62)(13, 1.56, 9.18)\}$; E-R (circles), $(\langle k \rangle, \tau, \gamma) = \{(7, 5.96, 8, 07), (13, 5.8, 9.06), (6, 3.08, 7.61), (16, 0.99, 8.5)\}$; and B-A (squares), $(\langle k \rangle, \tau, \gamma) = \{(6, 3.09, 7.61)(8, 5.99, 7.01)(12, 0.79, 8.96)(16, 2.18, 5.81)\}$.

networks for example, the most exposed nodes are the hubs, so they get infected early on. This skews the $(k, \hat{a}_k)$ curve to the left, because once infected these hubs generate a disproportionately large number of S-I links. On the contrary, when all nodes have similar degrees, the $(k, \hat{a}_k)$ curves are more symmetric. Concerning E-R and Reg networks, the most important difference is that the former allows for some degree heterogeneity, whereas the latter does not. Degree heterogeneity plays an important role when it comes to disease transmission so it is no surprise that epidemics on E-R networks can affect a higher proportion of nodes in the initial stage of an outbreak when compared to epidemics on Reg networks [Kiss et al., 2017].

This suggests that $\hat{a}_k$ curves could be parametrised with a low dimensional model. The departure from the fundamental assumption of homogeneous random mixing in epidemiological and ecological models has led to a myriad of models where bi-linear transmission terms proportional to $\sim I \times S$ or $\sim I \times (N - I)$ have been replaced by non-linear infection terms such as $I^p S^q$ [Liu et al., 1986; Hethcote and van den

Driessche, 1991; Roy and Pascual, 2006]. In particular it is noted that, in the context of classical compartmental and mean-field models, such terms can be inferred from the number of S-I links taken from simulation and that they can lead to more exotic model behaviours. In the same spirit, we put forward the following model for the rates:

$$\forall k \in \{0, \ldots, N\},\ a_k := a_k^{(C,a,p)} = Ck^p \left(N-k\right)^p \left(a\left(k - \frac{N}{2}\right) + N\right), \tag{2.3}$$

where the three parameters C, a and p offer flexibility to adapt to various networks and epidemics of different severity. This choice is motivated by the heuristic thinking of how the epidemic unfolds on the network. The parameter $C > 0$ gives a general scaling, dealing with different infection intensities, $a \in [-2, 2]$ helps to shift the peak from the centre (e.g. $a < 0$ shifts the peak to the left), and $p > 0$ allows for different flatnesses (smaller p values leading to flatter curves). Note that, when $a = 0$, the model results in a particular case of the previously mentioned non-linear models. Immediately, one can note that this model fulfils a number of desirable properties: (a) it is low dimensional/parsimonious, (b) the model satisfies $a_0 = a_N = 0$ by construction, (c) it includes the complete network when $a_k = \tau k(N-k)$ and finally, (d) it has a single maximum within $[0, N]$.

The C, a, p values are obtained using a non-linear least-square fit (using a particle swarm algorithm [Kennedy and Eberhart, 1995]):

$$e(C, a, p; \mathcal{S}) = \sum_{k,\ \sum_i t_{i,k} > 0} \left(a_k^{(C,a,p)} - \hat{a}_k\right)^2. \tag{2.4}$$

Fig. 2.1 showcases the flexibility of the model in fitting $\hat{a}_k$ curves coming from different network classes and confirms our observations about the rates being more left-skewed with increasing heterogeneity in node degree.

In the same figure, curves based on the (C, a, p) model are compared to the $(k, \hat{a}_k)$ curves. Systematic numerical investigations (not all plots shown) demonstrate that the proposed parsimonious three-parameter model fits the $(k, \hat{a}_k)$ curves well for all considered network classes, particularly Reg and E-R. For B-A networks small discrepancies between the $(k, \hat{a}_k)$ curves and the (C, a, p) model are possible.

2.2.3 Dataset

Proving that the behaviour of the exact system of 2^N equation is well approximated by our proposed system of $(N+1)$ ordinary differential equations (2.1) is still an open question. Therefore, the validations that we provide in this paper are entirely based on extensive numerical simulations. Here, we discuss briefly the synthetic dataset $\mathcal{S}$ underpinning those numerical validations. For each network class, we varied the average degree ($5 \leq \langle k \rangle < 20$). This covers a large number of scenarios and the networks remain relatively sparse. Regarding the epidemic parameters, we varied the infection and recovery rates ($(\tau, \gamma) \in (0, 10] \times (0, 10]$). Values for the rates were chosen via Latin hypercube sampling [McKay et al., 1979]. By doing so, we could observe many unique scenarios, providing a solid base upon which to test our methods.

However, there may be situations where the epidemic does not spread. Indeed, the behaviour of an epidemic is determined by the characteristics of both the network class and epidemic dynamics. The former includes quantities such as the average degree and higher-order moments, the latter includes per-link infection and recovery rates. All of this is captured by the reproduction number [Kiss et al., 2017], R_0, which is the number of secondary infections caused by a typical infectious individual introduced into a fully susceptible population:

$$R_0 = \frac{\tau}{\gamma + \tau} \frac{\langle\, k^2 - k \,\rangle}{\langle\, k \,\rangle}. \tag{2.5}$$

If $R_0 \leq 1$ the infection will die out. However, if $R_0 > 1$, then an outbreak is expected. Since R_0 depends directly on the sampled network and disease parameters, we accepted only situations where $1 < R_0 \leq 10$. This led to 360 valid choices (network class, $\langle k \rangle, \tau, \gamma$), with 120 per network class. For all the 360 scenarios, data from the simulations were used to determine network class- and disease-parameter specific infection rates $\hat{a}_k^{\theta,\tau,\gamma}$ and the corresponding (C, a, p) models.

2.2.4 Numerical validation of the forward model

To validate our claim that the BD process is a good approximation of the true epidemic behaviour, we numerically integrated the master equation (2.1) with rates $a_k = \hat{a}_k$ and $c_k = \gamma k$, where $\hat{a}_k$ are the estimated rates via Eq. (2.2), for all 360

scenarios in $\mathcal{S}$. The master equation was also numerically integrated with rates given by the (C, a, p) model. The expected number of infected nodes from the numerical solution of both master equations was then compared to the average number of infected nodes based on simulations. Four representative examples of epidemics for each network class are shown in Fig. 2.2. For the vast majority of the tested cases (not all shown), the agreement between simulation and the (C, a, p) model is good to excellent. It is worth noting that, in the case of B-A networks there are a few parameter combinations where the agreement between the master equation with the rates given by the (C, a, p) model and simulation results is poorer, see Fig. 2.2(c). This is despite the seemingly small discrepancy between $(k, \hat{a}_k)$ curve and the corresponding (C, a, p) model (not shown). However, the master equation with the $\hat{a}_k$-rates still leads to good agreement with simulations as shown in Fig. 2.2(c) (markers versus continuous line). Even so, it is reassuring to see that even when the agreement between the master equation with the (C, a, p) model breaks down, the agreement with the $\hat{a}_k$ holds. In [Nagy et al., 2014], a similar surrogate model was used and the authors obtained good agreement between the BD model and simulations for an even wider range of network classes. This gives us confidence that the surrogate model is a viable model.

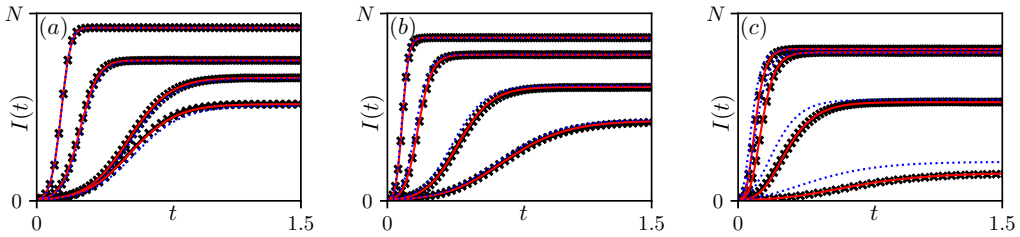

Figure 2.2: Average number of infected nodes from simulations (markers) and the numerical solutions of system (2.1) with rates a_k given either by the raw data $\hat{a}_k$ (continuous curve) or by the (C, a, p) model (dotted curves), with initial condition $k_0 = 5$. Three network classes are reported, each with $N = 1000$ nodes, from left to right, ordered by increasing heterogeneity, from Reg (a) and E-R (b) to B-A (c) networks. Networks and epidemic parameters are the same as in Fig. 2.1.

2.3 Bayesian inference of network class from single epidemics

In the framework presented so far, we proposed a surrogate model which approximates the evolution of the total number of infected nodes in a SIS epidemic on a network. The rates of infection in this forward model (i.e. $\hat{a}_k$) are parametrised by a three-parameter model (C, a, p) as detailed in equation (2.3). Early investigation shows that the $(k, \hat{a}_k)$ curves (thus the associated C, a, p triple) are distinct across the three different network classes that we considered. Hence, one may expect that discrete observations taken from a single epidemic spreading on a unknown network carry sufficient information to identify its most likely class.

We consider the following task: to infer the most likely network class, given one set of discrete-time observations of one realisation of the true process on a network. To do so, we adopt the (C, a, p) model to approximate the true process at a population level. Then, the likelihood of the observed data can be written in terms of the solution operator of a Birth-and-Death model of rates (C, a, p) and γ. Since the task is to infer the most likely network class rather than the actual (C, a, p) triplet, we set up a Bayesian framework where priors over different network classes are built by fitting the $(k, \hat{a}_k)$ rates of the surrogate model on many realisations of the true process over different network classes and epidemic parameters. In figure 2.3 a simple scheme to describe the inference model is provided.

To be more precise, we consider a population-level dataset $y = (k_1, \ldots, k_n)$ where $k_j \in \{0, \ldots, N\}$ for any $j = 1, \ldots, n$ is the number of infected nodes in the network at time $t_j \in [0, T]$, and we define the vector $s = (t_1, \ldots, t_n)$. Our objective is to predict the class $\theta \in \Theta$ of the underlying network from y. Figure 2.4 illustrates 10 distinct data sets for each of the three network classes. These data are obtained directly from Gillespie simulations [Gillespie, 1976, 1977] of the SIS epidemic on the respective networks. Observations are taken at regular times from the start of the epidemic to the point where the quasi steady-state is approached.

For each value of θ (that is a network class), we build a distribution $\pi_{0,\theta}$ over the parameters C, a, p based on offline simulations of SIS epidemics for a range of networks in each given class θ (see Section 2.3.1). By looking at the outcomes

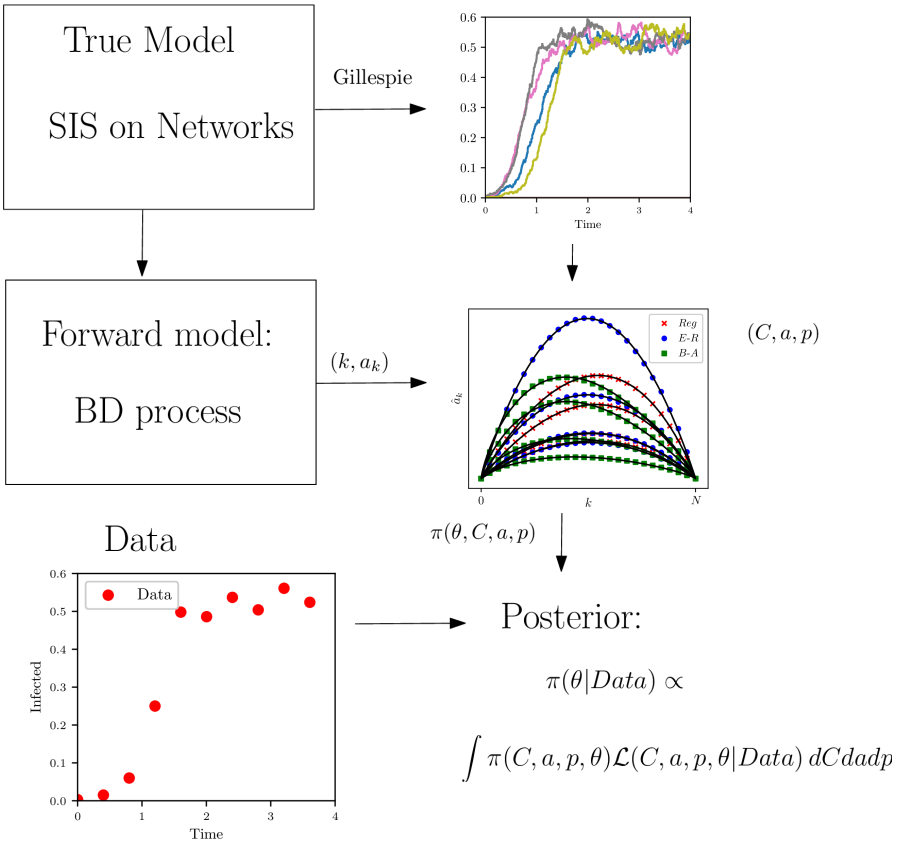

Figure 2.3: Scheme of the inference framework. The groud truth is a SIS process on a randomly generated network, simulated through the Gillespie algorithm [Gillespie, 1977]. At each event, we record the number of $S-I$ links and the number of infected nodes $I(t)$. We repeat the simulation many times, possibly starting from $I(t=0) = N$, to characterise how the $S-I$ links are distributed for each I on a given network class. From this, by using the forward model and the (C, a, p) function, we can build priors over different network classes θ. When a new dataset comes in, in the form of discrete population-level observations of an outbreak, we can infer the probable network class thorugh the likelihood (2.6).

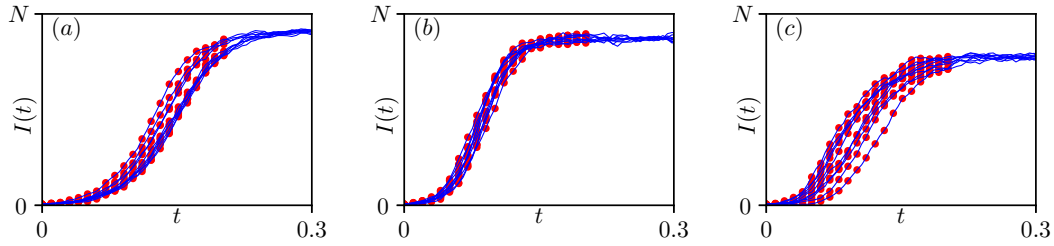

Figure 2.4: 3 examples on different network classes of 10 average epidemic paths, taken from dataset $\mathcal{S}$. Continuous curves represent the evolution of infectious counts and dots the observations y. Network and epidemic parameters for each panel are, from left to right, Regular (a) with $\langle k \rangle = 17$, $\tau = 2.62$, $\gamma = 4.03$, Erdős-Rényi (b) with $\langle k \rangle = 13$, $\tau = 5.80$, $\gamma = 9.06$, and Barabási-Albert (c) with $\langle k \rangle = 6$, $\tau = 8.16$, $\gamma = 8.23$. In each realisation, $k_0 = 5$ randomly selected nodes are infected at the beginning of the epidemic.

of our simulations, we observe that, for our chosen set of candidate classes Θ, the distributions $\pi_{0,\theta}(C, a, p)$, $\theta \in \Theta$, cluster in distinct regions of the (C, a, p) parameter space. This is necessary for the inference to work, and it contributes to the validation of our model of choice for the rates $\hat{a}_k$. Assuming a non-informative uniform prior for θ, we derive a prior distribution $\pi_0(C, a, p, \theta)$ in the form of a mixture:

$$\pi_0(C, a, p, \theta) = \frac{1}{3}\pi_{0,\theta}(C, a, p).$$

Our objective is the prediction of the underlying network θ given the data (y, s), which will be done using the posterior distribution $\pi(\theta|y, s)$ obtained by Bayes' rule:

$$\begin{aligned} \pi(\theta|y,s) &= \int \pi(C,a,p,\theta|y,s)dCdadp \\ &\propto \int \mathcal{L}^{C,a,p}(y,s)\pi_0(C,a,p,\theta)dCdadp, \\ &\propto \int \mathcal{L}^{C,a,p}(y,s)\pi_{0,\theta}(C,a,p)dCdadp, \end{aligned} \tag{2.6}$$

where, given C, a, p, the likelihood $\mathcal{L}^{C,a,p}$ can be expressed in terms of the solution operator of the forward model discussed above (see Section 2.2). This Bayesian classification methodology is also known as model selection, where the model is a particular class of networks. Once we have computed the posterior distribution $\pi(\theta|y, s)$ (see Section 2.3.2), we simply pick the most likely underlying network class (Maximum a posteriori estimator for θ given the data (y, s)).

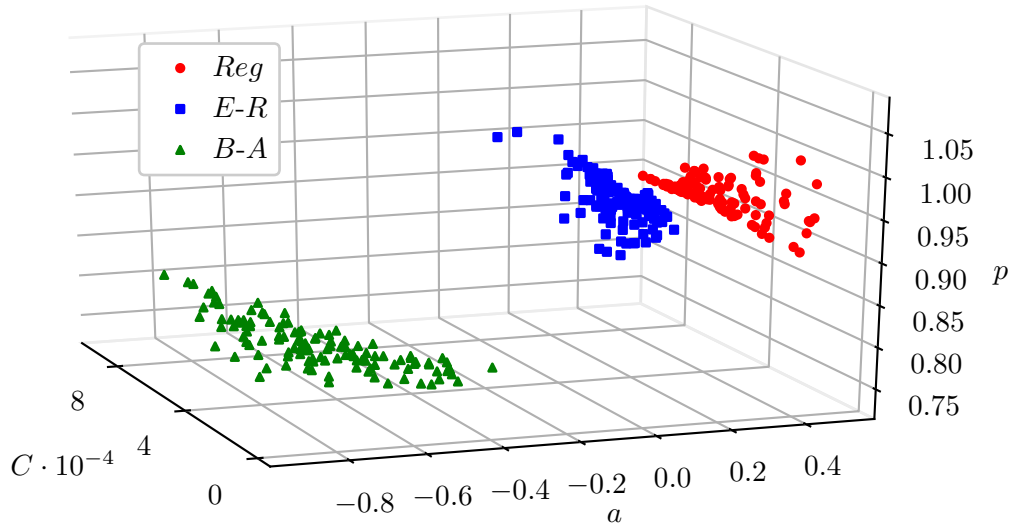

Figure 2.5: Estimated C, a, p values from the dataset $\mathcal{S}$ (360 points in total, each coming from a unique combination of (network class, $\langle k \rangle$, τ, γ)). From left to right, we observe three distinct regions corresponding to Barabási-Albert (triangles),Erdős-Rényi (squares) and Regular networks (circles) networks.

2.3.1 Prior distributions for each network class

In this work, we consider prior distributions for each network class as a different density $\pi_{0,\theta}$ over the C, a, p space. To do this, we use the very same dataset that was used for numerical validation (see Section 2.2). Given the C, a, p values of each network class (see Fig. 2.5), we choose 100 triples to estimate a probability distribution and leave 20 for testing. The (C, a, p) values associated with the training scenarios are used to infer three Gaussian kernel density estimators [Pedregosa et al., 2011] to be used as prior distributions. The bandwidth of these estimators is set by 10-fold cross-validation.

2.3.2 Numerical method for posterior marginals computations

Finally, to predict the underlying network class given a dataset (y, s), we need to compute the three marginals given in equation (2.6) (one per network class) and this is done by Monte-Carlo estimation. As already mentioned, the likelihood $\mathcal{L}^{C,a,p}(y, s)$ can be obtained using the forward operator associated with equation (2.1). Indeed,

given a (C, a, p)-triple, the likelihood of a dataset (y, s) is given by:

$$\mathcal{L}^{C,a,p}(y,s) = \prod_{i=1}^{n-1} p^{C,a,p}_{k_i,k_{i+1}}(t_{i+1} - t_i),$$

using the fact that the BD process is time homogeneous. We choose to compute each term $p^{C,a,p}_{k_i,k_{i+1}}(t_{i+1} - t_i)$ where $1 \leq i \leq n-1$ using the algorithm introduced in [Crawford et al., 2014], allowing BD transition probabilities to be computed individually. This represents a significant reduction in computational time, when compared to matrix exponential since we are working with a network of size $N = 1000$ nodes.

Once we have an efficient numerical method to compute the likelihood, we use the corrected Arithmetic Mean estimator, recently introduced in [Pajor, 2017] for the Monte-Carlo estimation of all marginals. Let A be a given subset of the (C, a, p) space, then it follows that:

$$\int \mathcal{L}^{C,a,p}(y,s)\pi_{0,\theta}(C,a,p)dCdadp = \frac{\pi_{\theta,0}(A)}{\pi_{\theta|(y,s)}(A)} \int_A \mathcal{L}^{C,a,p}(y,s)\pi_{A,\theta}(C,a,p)dCdadp, \tag{2.7}$$

where $\pi_{A,\theta}$ is the prior density of network class θ, conditional on $\theta \in A$. Each marginal is then estimated using the following procedure:

1. Find
$$(C^*, a^*, p^*) = \arg\max_{C,a,p} \left(\log \mathcal{L}^{C,a,p}(y,s) + \log \pi_{\theta,0}(C,a,p)\right).$$
This is done via a combination of global/local optimisation routines.

2. Sample the distribution $\pi_\theta(C, a, p|y, s)$ using a Random-Walk Metropolis-Hastings algorithm starting from (C^*, a^*, p^*) and denote the samples by $(C_i, a_i, p_i)_{1\leq i\leq K}$ with $K = 500$.

3. Let H be the Fisher information evaluated at (C^*, a^*, p^*) and let $d(C, a, p)$ be defined as
$$d(C,a,p) := \langle (C_i,a_i,p_i) - (C^*,a^*,p^*), H\left[(C_i,a_i,p_i) - (C^*,a^*,p^*)\right]\rangle.$$
We then take $A := \{(C, a, p)|d(C, a, p) \leq r\}$ where $r = \max_{1\leq i\leq K} d(C_i, a_i, p_i)$. This choice was already suggested in [Pajor, 2017]. In particular, it leads to $\pi_{\theta|(y,s)}(A) \approx 1$, which simplifies the right-hand-side of equation (2.7).

4. Use a Gaussian distribution $\mathcal{N}((C^*, a^*, p^*), H^{-1})$ to estimate both $\pi_{\theta,0}(A)$ and the integral term on the right-hand-side of equation (2.7) by importance sampling.

Our complete Python implementation of this routine is available online [1].

2.3.3 Network class inference

In this section, we provide numerical results assessing the overall quality and applicability of our approach. We start by inferring networks from a testing dataset, where all data are simulated from either Regular, E-R or B-A networks, see Section 2.3.3. We then consider networks outside of our framework, namely synthetic networks with negative binomial degree distributions (Section 2.3.3) and real-world networks (Section 2.3.3). In all cases, we provide posterior probabilities for each network class across independent repetitions of the datasets to quantify uncertainty.

Inference based on the testing set

During the construction of the prior, we deliberately set aside 60 estimated (C, a, p) values to build a test set (20 per network class taken at random), meaning that they were not used in the calibration of the prior. In this section, we use this set to check if we can infer the known underlying network class.

The inference was performed as follows. For each of the (C, a, p) parameters in the testing set, we used the known underlying network and disease parameters (network class, $\langle k \rangle$, τ, γ) to simulate a dataset (y, s) with Gillespie's algorithm. We only generated a single network from the appropriate class and simulated a single epidemic. However, we generated 10 independent datasets, as shown in Fig. 2.4, and ran our inference model on each of them separately. The second step was to compute the 3 posterior probabilities corresponding to the different network classes, as detailed earlier. We thus obtained 3 posterior probabilities for each of the 60 elements in our test set and predicted the most likely underlying network class. To assess the uncertainty due to data sampling, we considered the results across all the independent datasets.

[1] https://github.com/BayIAnet/NetworkInferenceFromPopulationLevelData

The quality of the inference is shown by the confusion matrix (Table 2.1), which provides the averaged posterior probabilities along with their standard deviation for each of the possible outcomes. The level of accuracy achieved in our tests is remarkable, with a score as high as 95% for Barabási-Albert, and a minimum of 79% for Erdős-Rényi. This also shows that there can be a moderate confusion between the Regular and Erdős-Rényi network classes, as their characteristics are quite similar w.r.t. (C, a, p) values (see Fig. 2.5) whereas Barabási-Albert is rarely miss-classified. Further, the standard deviations show that these scores are stable across different data realisations, suggesting that our approach is consistent.

True/Predicted	Regular	Erdős-Rényi	Barabási-Albert
Regular	85.5% (7.9%)	14.5% (7.9%)	0.0% (0.0%)
Erdős-Rényi	21.5% (10.7%)	78.5% (10.7%)	0.0% (0.0%)
Barabási-Albert	0.0% (0.0%)	5.0% (5.0%)	95.0% (5.0%)

Table 2.1: Averaged confusion matrix based on the test dataset (standard deviation is brackets).

To get a more precise description of the classification results, we computed the average posterior probability for each of the 60 test elements, see Fig. 2.6. This revealed that the average posterior probability varies within each of the network class, probably due to differences in network or disease parameters. In some sense, this shows that for some network and disease parameters, the similarity between Regular and Erdős-Rényi is significant. For example, when the epidemic spreads fast and infects many nodes early on, the structure of the network is less important as the infection will be transmitted on. This means that the average degree is more important than the degree distribution. Nevertheless, our inference methodology returns a good classification in most cases. In fact, these tests show that our approach can successfully recover the network class from as little as 21 observations of a single epidemic.

Finally, we detail specificity and sensitivity for the 10 repetitions of the classification, offering per network class and global statistics in Fig. 2.7. We note that each marker has 10 occurrences but in some cases these are superimposed. Here again, one can see the stability and high efficiency of our approach for Barabási-Albert,

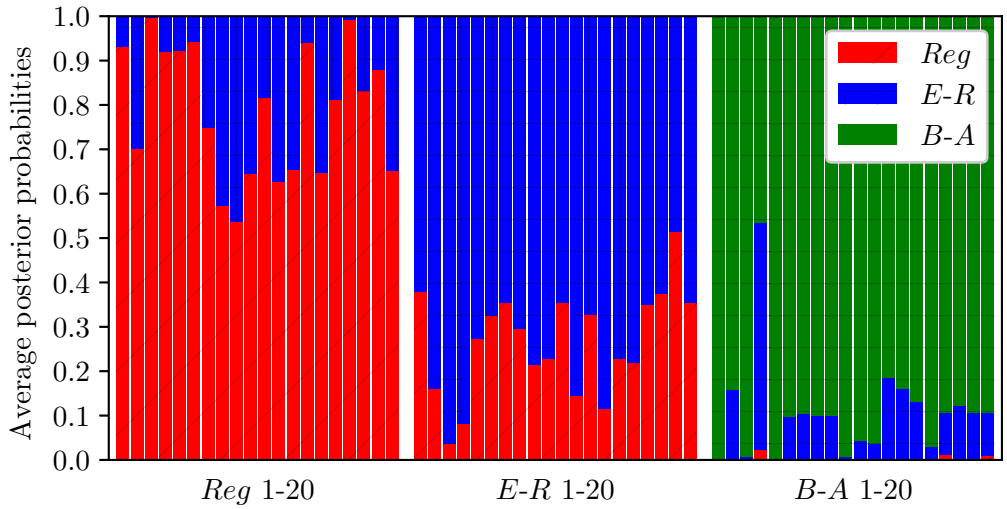

Figure 2.6: Average posterior probability over the 60 tests (20 per network class and 10 realisations).

with more confusion for Erdős-Rényi.

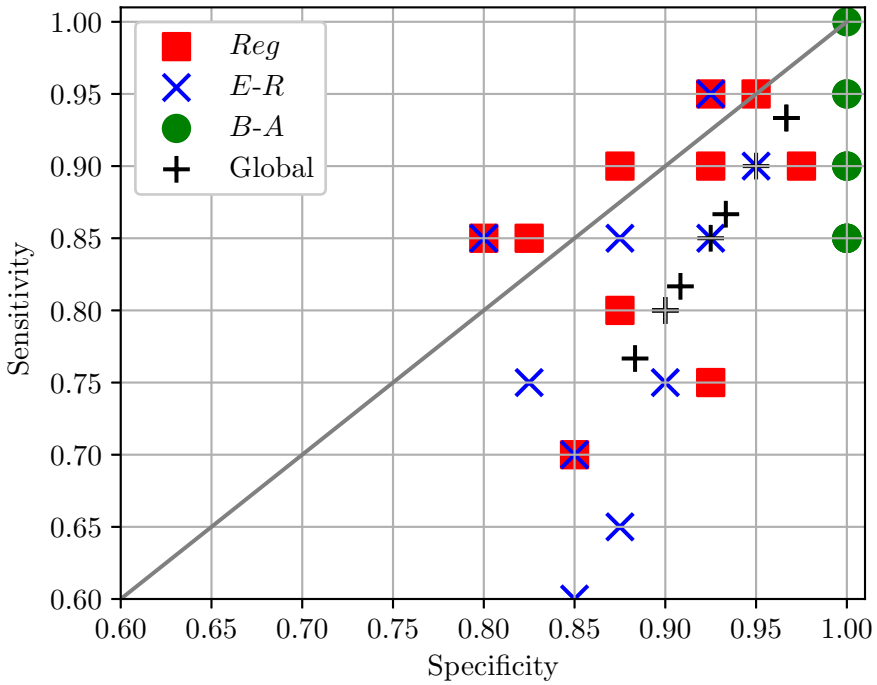

Figure 2.7: Specificity and sensitivity of the 10 independent classification at global and per-network levels.

Inference of synthetic networks

We have shown that our methodology performs well when applied to the data generated on the networks that it was trained on. In this Section, we consider alternative network types for two reasons: (a) to stress-test the classification by using networks whose degree distributions do not come from the models used to build priors, and (b) to study the extent to which it can distinguish between different levels of heterogeneity in degree distribution.

To do this, we generated three synthetic networks using the configuration model [Newman et al., 2001] and a negative binomial degree distribution with parameters (p, n):

$$\forall k \in \{0, n\},\ P(k) = \binom{k+n-1}{k} p^k (1-p)^{n-k}, \tag{2.8}$$

where p is the probability of success and n the number of failures. This choice is motivated by both the simplicity and the flexibility of this distribution. The average degree in all three networks was identical (i.e. fixed at $\langle k \rangle = 6$) but with different levels of heterogeneity depending on the variance, see Fig. 2.8(a). To avoid the possibility of having disconnected components, the degree distributions were shifted so that the minimum was greater or equal to 3. Here, the degree distributions were chosen to exhibit different levels of heterogeneity, from low to a level comparable to those achieved in B-A networks. We then ran 10 independent epidemics with parameters $\gamma = 1$ and $\tau = 0.5$, starting from 5 infected nodes. As in Section 2.3.3, the inference was based on a dataset with 21 equally-spaced observations of the number of infected nodes. The results are shown in Fig. 2.9, and confirm that our inference scheme is able to distinguish between networks with high/low levels of degree heterogeneity. In particular, by looking at Fig. 2.8(a) it is reasonable to expect that the first and third networks are going to be classified as E-R and B-A networks, respectively. Indeed, Fig. 2.9 shows that the first network in Fig. 2.8(a) is identified as E-R 80% of the time, whereas the third network in Fig. 2.8(a) is correctly classified as B-A for every single epidemic realisation.

When the degree distribution of the test network is such that its variance falls between typical variances observed in E-R and B-A networks (see the second network in Fig. 2.8-a) our results are more sensitive to the individual realisation of the epidemic. However, even in this case, the network is identified to the closest type in

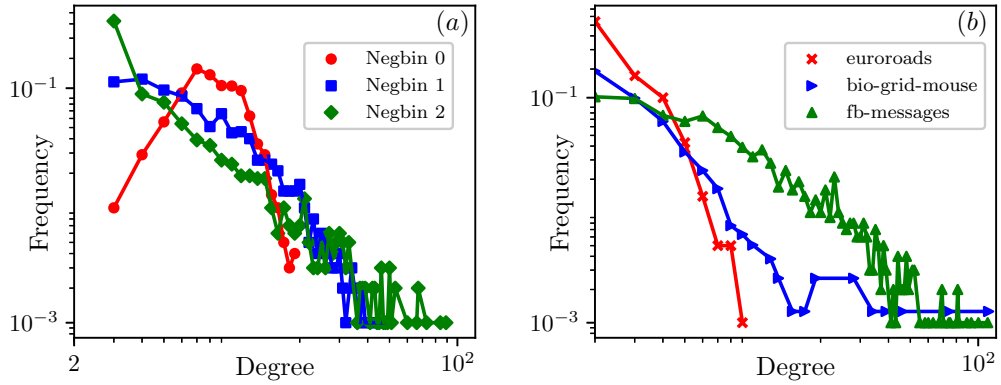

Figure 2.8: Degree distributions, ordered by variance, of three single negative binomials (a) following equation (2.8) and of the three real networks (b) used for the stress test. For (a), the average degree is $\langle k \rangle = 6$ for all networks. From low to high variance we have $\sigma = 8$ (Negbin 0), $\sigma = 40$ (Negbin 1), $\sigma = 120$ (Negbin 2). The values of (p, n) are $(0.23, 20), (0.85, 1), (0.95, 0.3)$, respectively. For (b) the basic metrics of these networks are $\{\langle k \rangle, \sigma^2, \text{Assortativity}, \text{Clustering}\} = \{2.53, 5.24, 0.102, 0.02\}, \{2.77, 40, -0.21, 0.04\}, \{12.30, 268.90, -0.08, 0.09\}$, respectively.

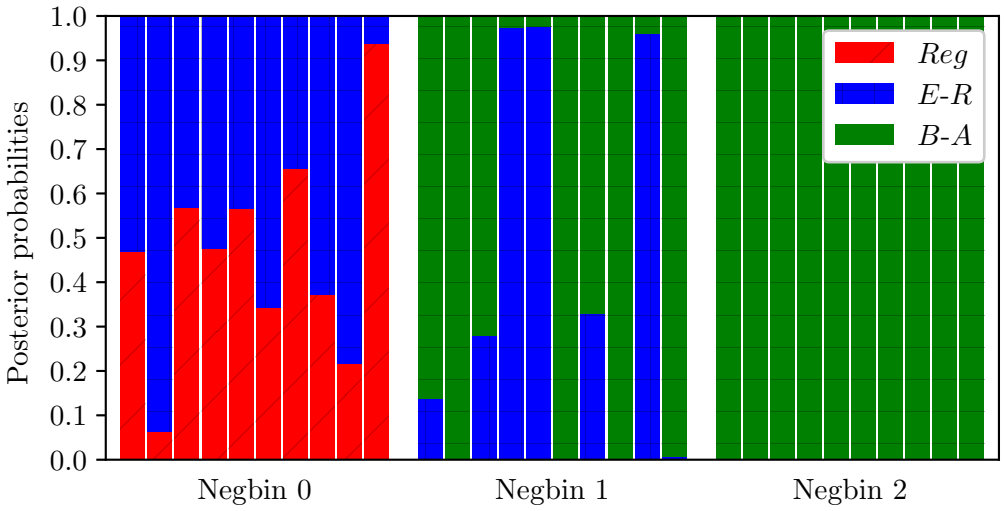

Figure 2.9: Posterior probabilities for the 10 repetitions on each synthetic network.

terms of degree distribution. Moreover, heuristically at least, the B-A network seems to be favoured, which seems reasonable upon inspecting the degree distribution of the test network.

Inference of real-world networks

Finally, the last test we conducted was based on real world networks, which can exhibit higher-order structure beyond degree heterogeneity. We chose three real networks: the first is labelled euroroads and is part of the KONECT collection [Kunegis, 2017], the second and third, bio-grid-mouse and fb-messages, are part of the network data repository Networkrepository [Ryan and Nesreen, 2015]. The euroroads is an infrastructure network, bio-grid-mouse is a protein-protein network whilst fb-messages is based on the interactions of an online community of students at University of California. In Fig. 2.8(b) the degree distributions of these networks are shown. To keep the number of nodes equal to $N = 1000$, we only considered the largest connected component, and then, where necessary, removed peripheral, low-degree nodes such that the resulting network was still connected.

In line with Section 2.3.3, we fixed $\gamma = 1$, and ran 10 distinct epidemics on each network in order to generate data for the inference. Values for the infection parameter were $\tau = \{1.5, 2.5, 0.4\}$ for euroroads, bio-grid-mouse and fb-messages, respectively. The posterior probabilities obtained from our approach are reported in Fig. 2.10 and are in line with our expectations based on the inspection of the respective degree distributions: the infrastructure network is very homogeneous, whilst the other two are scale-free, and hence correctly classified as B-A.

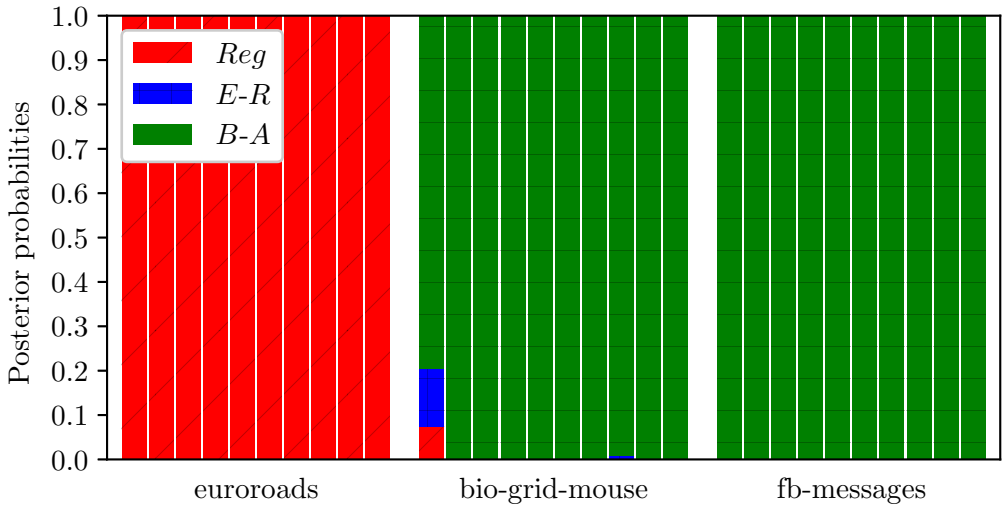

Figure 2.10: Posterior probabilities for the 10 repetitions on each real-world network.

2.4 Discussion

In this paper, we proposed a new inference scheme that uses population-level incidence data at discrete regular times to infer the most likely network class over which the epidemic has initially spread. This is a challenging task because the exact epidemic model on a given network is forbiddingly high-dimensional meaning that even a numerical solution is out of reach. The key to carry out the inference is the approximation of the exact epidemic model by a BD process, whose rates not only encode the structure of the networks but also allow us to distinguish between the different network classes through a parsimonious three-parameter model. Whilst we have successfully numerically validated this surrogate model over a number of network classes and different values of disease parameters, with further evidence in [Nagy et al., 2014], a mathematical characterisation of the relation between the exact and this surrogate model remains an open problem.

Our analysis has focused on three classes of random networks: Regular, Erdős-Rényi and Barabási-Albert. This choice is motivated by the fact that such classes are well-known, simple to describe (depending only on one parameter) while differing in terms of degree heterogeneities. For each network class, the rates of infection in the corresponding BD approximation was obtained by using the time-weighted mean of S-I link counts. Despite these rates being network class-dependent, they all share some common features. This in turn allowed us to propose a parsimonious three-parameter model (C, a, p) that works across all network classes and, at the same time, can capture the differences in the rates of the approximating BD process. In addition to being robust to different values of τ, γ and average degree, these parameters exhibit a clear distinction between the three different network classes when plotted in the 3-dimensional (C, a, p) space. This knowledge is then encoded into prior distributions, constructed using kernel density estimators over the (C, a, p) space. Our Bayesian model selection procedure then consists in the numerical estimation of the relative marginal probabilities. Our results show that the inference scheme has good specificity and sensitivity, despite the simplicity of the model. These encouraging results lead to a number of questions and remarks. First of all, our choice of classes of random networks means that the main feature of the networks is their degree heterogeneity. We have yet to consider more complex

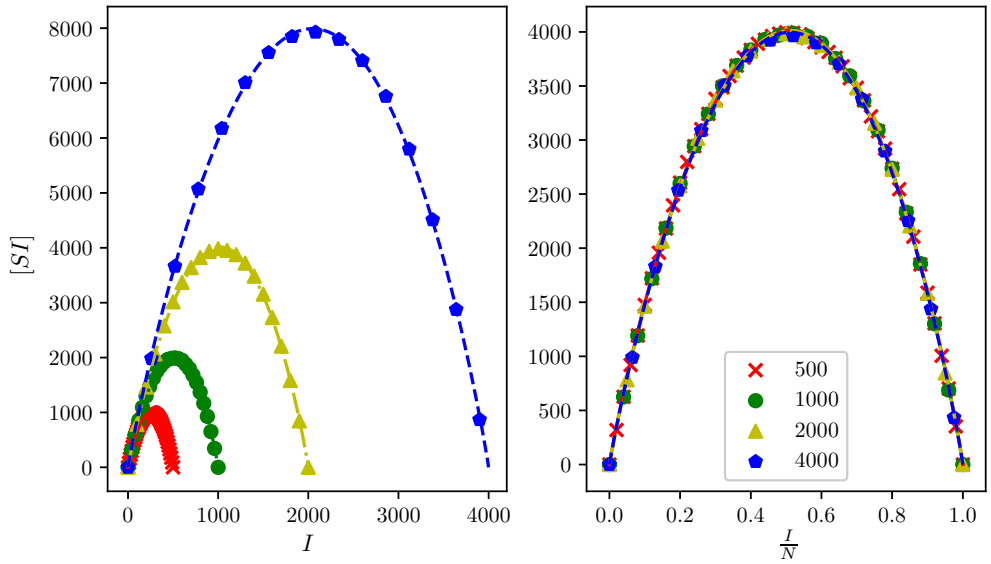

Figure 2.11: (Left) $(k, \hat{a}_k)$ curves based on Erdős-Rényi networks with $\langle k \rangle = 5$, $\tau = 1.793$, $\gamma = 3.785$ and $N = 500, 1000, 2000, 4000$. (Right) Scaled $(k, \hat{a}_k)$ curves relative to the $N = 2000$ case. Scaled version are obtained by plotting $(k/N, \frac{2000}{N}\hat{a}_k)$.

networks, such as those exhibiting clustering or community structure. This would certainly lead to $(k, \hat{a}_k)$ curves of different shapes, potentially having other features such as multiple peaks for networks with multiple communities, and thus requiring either a more sophisticated or non-parametric model. Nevertheless, considering epidemics in terms of an approximate BD process appears to be a powerful approach if a tractable likelihood is desired. Moreover, once the most likely network class has been identified, one could continue and estimate τ, γ and the average degree.

We have used a fixed number of nodes ($N = 1000$) in all our numerical experiments. We do not expect major changes when the number of nodes is different. Preliminary numerical tests, see Fig. 2.11, suggest that there is a good degree of universality such that the $(k, \hat{a}_k)$ curves only differ by a scaling factor when the number of nodes changes, all other parameters being fixed. In this respect, our methodology could easily be adapted by directly considering the scaled epidemic (on $[0, 1]$) and repeating our tests for different values of N. Fortunately, our numerical method [Crawford and Suchard, 2012] scales well with N, since the transition probabilities in the likelihood are computed individually (with deeper continued fractions). The question of the limiting behaviour in the limit of large N can also be further

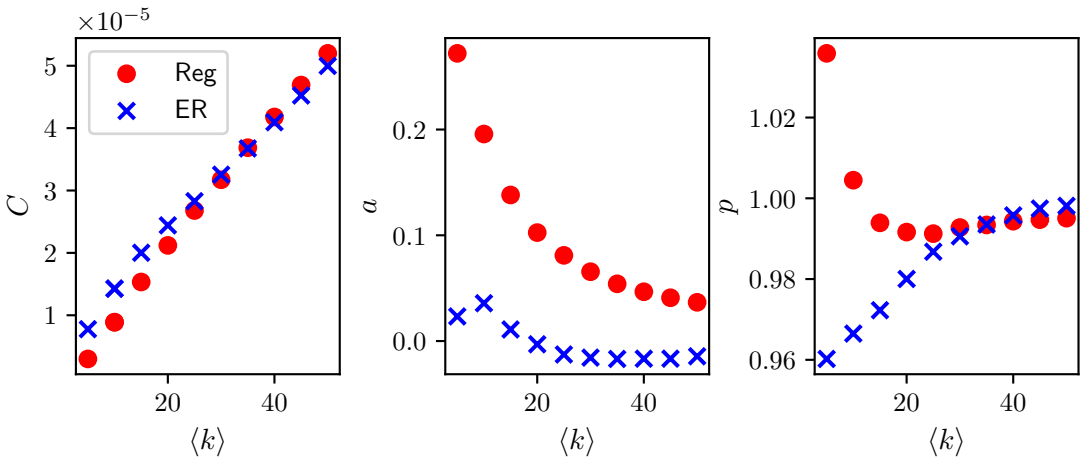

Figure 2.12: Behaviour of C, a, p when $\langle k \rangle$ increases for regular networks (circles) and Erdős-Rényi (crosses), all other parameters being equal. The maximum value of $\langle k \rangle$ explored is $\langle k \rangle = 50$. $\tau = 1, \gamma = 5$. The fully-connected limit, reached when $\langle k \rangle = N - 1$, is $C = \frac{1}{N}$, $a = 0$, $p = 1$; however, even at $\langle k \rangle = 50$, we can see how un-identifiability emerges between regular and Erdős-Rényi networks.

investigated.

An interesting open question is that of the extent to which different network families are mapped onto distinct regions in the C, a, p space if networks are weighted, i.e., if the adjacency matrix has entries of magnitude $0 \leq g_{ij} \leq 1$. While a comprehensive answer to this query would require extensive simulations beyond the scope of this paper, there are a couple of points worth making. First, we already see that Regular and Erdős-Rényi classes are only really distinguishable when networks are sparse. If we keep τ, γ fixed and increase the average degree $\langle k \rangle$, we observe that both tend to the fully-connected network limit, where $C = \frac{\tau}{N}$, $a = 0$, $p = 1$. This is because the fully-connected network can be seen both as a regular network with degree $\langle k \rangle = N - 1$ and as an Erdős-Rényi with $p = 1$, see figure 2.12. This means that there is some degree of un-identifiability when network classes generate networks that are topologically almost identical to one another. Further, any unweighted network can be seen as a weighted fully-connected network, with weights either 0 or 1. For instance, an Erdős-Rényi network is a weighted complete network such that the element g_{ij} has weight 0 with probability p, 1 otherwise. With this consideration in mind, the question can be rephrased as: is it possible to use this framework to infer the weight distribution over a fully-connected network? Our con-

jecture is that the answer is yes. Provided that the weight distributions are distinct enough and the k, a_k curves can be captured by a model such as the C, a, p, we do expect to find similar scenarios to those shown by figure 2.5.

So far, we have used discrete data taken on a regular time grid covering the epidemic from its early stage (a few infectious nodes) up to its steady state. Increasing the frequency of data or restricting data to the very beginning of the epidemic are of significant practical interest. In the former case, one expects the discrete likelihood to converge to the simpler continuous one, enabling faster and easier analysis. In the latter case, it would lead to a model that does not require describing the whole epidemic as we currently do. Focusing on the initial stages of the epidemic, the most critical period in many cases, and upon solving a potential un-identifiability problem, such an approach could have an important real-world impact, making it possible to predict and control more accurately yet-to-be epidemics.

Finally, the proposed inference scheme could be improved by using more sophisticated models for the infection rates and by learning a larger number of different network classes, leading to a wide portfolio of data which can then be used for estimation. Of course, there is a trade-off in terms of what we can infer about networks using population-level discrete data. We cannot infer individual links for example but this is to be expected since the data we use for inference is not at the link- or node-level. Nevertheless, we believe that our approach could have practical implications, as the inference scheme is based on the kind of data that is most likely to be available in real-world scenarios (e.g. the number of infected people every day or week). Where such data is available but little is know about the contact pattern, our inference scheme may be able to provide some high-level information about the properties of the network which in turn could be exploited in the planning or implementation of control, in particular during the early stages of an epidemic.

2.5 Acknowledgments

All authors acknowledge support from the Leverhulme Trust for the Research Project Grant RPG-2017-370.

Chapter 3

PDE limits of stochastic SIS epidemics on networks

3.1 Introduction

An epidemic is a complex phenomenon that arises from a pathogen spreading over the contact structure of a population. Similar spreading phenomena occur in various disciplines, from biology and social sciences to engineering. Unsurprisingly, much modelling effort has been put into studying spreading processes on networks, as they offer a natural framework to mimic real-life contact patterns [Brauer and Castillo-Chavez, 2012] and the important heterogeneities within these. The use of networks is extremely intuitive with each individual encoded as a node, and all its contacts (to other individuals) as links. Unfortunately, the resulting exact probabilistic model does not scale well with the size of the network, N. Even when relatively simple models, such as susceptible-infected (SI) or susceptible-infected-susceptible (SIS), are considered, the exact model has 2^N equations, which quickly becomes intractable.

To address the difficulty posed by high dimensionality, mathematical descriptions often focus on population-level statistics (e.g. expected number of infected people at any given time). This has led to a number of so-called mean-field models [Kiss et al., 2017; Porter and Gleeson, 2016; Pastor-Satorras et al., 2015], offering a good first approximation of the evolution of some population-level or averaged quantities. These include pairwise models based on moment closure techniques [Keeling, 1999; Kiss et al., 2017], effective-degree [Lindquist et al., 2011], edge-based-

compartmental [Miller et al., 2012] models and even PDE models [Silk et al., 2014]. All such mean-field models share a number of caveats [Roberts et al., 2015]. For example, (i) in general the agreement between these and the exact stochastic model breaks down close to the epidemic threshold, (ii) there are very few cases where it is possible to prove mathematically that the mean-field model is the limit of the exact stochastic process (this has only been done for SIR epidemics and configuration networks [Decreusefond et al., 2012; Janson et al., 2014]) and (iii) they give no estimate of the variability observed in the exact process. It is also well-known that such mean-field models only work for a limited class of networks; epidemics on clustered networks are not well-understood, except for idealised clustered networks, i.e. networks with non-overlapping triangles or other clustering-inducing subgraphs.

There are ongoing efforts to try to understand and answer rigorous mathematical questions when it comes to analyse or approximate dynamical processes on networks, see [Rempala et al., 2019] for a recent summary. Progress in this area is usually achieved by bringing in and combining results and techniques from different areas of mathematics. One particularly promising prospect for SIS epidemics on networks is to consider them as Birth-and-Death (BD) processes. In a recent paper [Di Lauro et al., 2020a], we conjectured and confirmed numerically that SIS epidemics are well captured by BD processes, whose rates encode characteristics of both the network structure and the epidemic dynamics. This was tested on Regular, Erdős-Rényi and Barabási-Albert networks. This choice was motivated by the intuition that epidemic spread is driven by the 'birth' of new infected nodes. However, this occurs at a rate which is proportional to the number of S-I (active) links, and these are readily observable during explicit stochastic simulations of the epidemic on networks.

In this paper we build on the above observation and take the next natural step, that is, to consider the large N limit of the BD process, i.e., the one-dimensional PDE (Fokker-Planck equation). We extend the repertoire of network models and consider Regular, Erdős-Rényi and Barabási-Albert networks and 2D lattices with periodic boundary conditions and show that the resulting rates in the BD process are density-dependent such that the limit is well defined in the sense of [Kampen, 2007]. We compute the rates numerically and also provide a parametric form for them (with the exception of the lattice). We show that the resulting PDE agrees

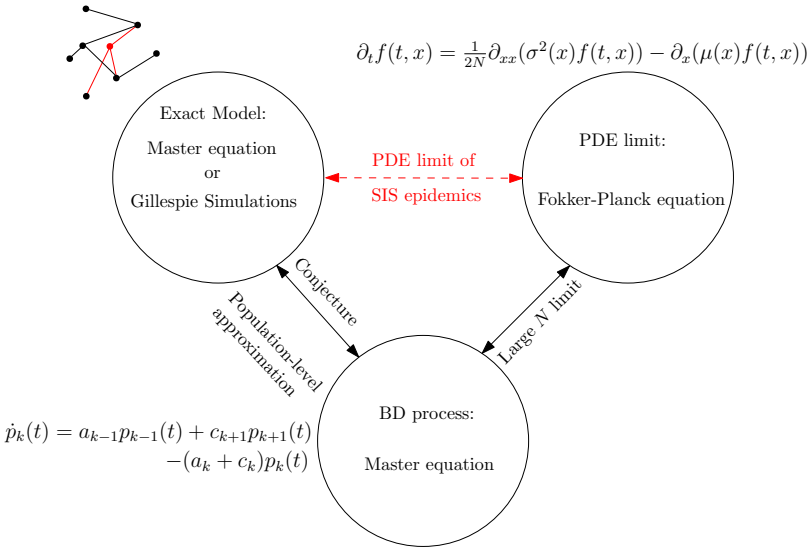

Figure 3.1: Schematic illustration of various approximations of the exact stochastic SIS epidemics on networks. The PDE limit comes as a result, and further confirms the validity, of the Birth-Death approximation conjectured in [Di Lauro et al., 2020a].

well with the output of explicit simulations of stochastic epidemics on networks. The existence of the PDE limit has multiple advantages. First, it reduces further the dimensionality of the system. Second, it gives us the opportunity to compute an epidemic threshold even in an implicit form. Finally, it provides the means to get a handle on the variability of the stochastic process with the solution of the PDE providing a likelihood that can be computed cheaply and efficiently for inference purposes. Finally, the good agreement between the PDE and the exact process provides further evidence that the BD model may indeed serve as a valid approximation of the exact process (the relation between the exact, BD and PDE-limit model is illustrated in figure 3.1) and that a formal proof of this observation may be possible.

This paper is structured as follows. In Section 3.2 we briefly outline the Birth-and-Death approximation of SIS epidemics as in [Di Lauro et al., 2020a]. In Section 3.3 we numerically test and prove that the conditions for the existence of the PDE limit, as $N \to \infty$, are met for different network topologies and epidemic parameters. We then show that the solutions of the partial differential equations agree well with the empirical distributions based on simulations of the true process. In Section 3.4 we draw some conclusions and outline further research directions.

3.2 Methods

3.2.1 Birth-and-Death Approximation of SIS Epidemics

We briefly describe the model proposed in [Di Lauro et al., 2020a] which conjectured that exact stochastic SIS epidemics on networks can be approximated by BD processes. A standard SIS model on an undirected unweighted network $\mathcal{G}$ with N nodes is considered, where each node is either susceptible (S) or infected (I). Infected nodes spread infection to their neighbours with constant per-link rate τ and recover with rate γ (independently of the network). This stochastic process results in a continuous time Markov chain on a state space of cardinality 2^N, which forbids analysis even for relatively small values of N. Instead, we consider the population-level count of infected nodes, defined as $k(t) = \sum_{i=1}^{N} \mathcal{I}_i(t)$, where $\mathcal{I}_i$ is an indicator function equal to 1 if node i is infected at time t and 0 otherwise. $k(t) \in [0, N]$, where $k(t) = 0$ indicates the state where no infection is present in the network. Given the stochasticity of the process, $k(t)$ is itself a random variable taking values on state space of cardinality $(N + 1)$. This reduction in dimensionality makes computations much more tractable. We note that each time an infection/recovery occurs, the value of $k(t)$ changes by discrete jumps of size ± 1, respectively. This has led to the conjecture [Di Lauro et al., 2020a] that the population-level count process can be approximated by a Birth-and-Death process, governed by the following master equation:

$$\dot{p}_k(t) = a_{k-1}p_{k-1}(t) + c_{k+1}p_{k+1}(t) - (a_k + c_k)p_k(t), \tag{3.1}$$

where $p_k(t)$ is the probability of having k infected nodes at time t, $c_k = \gamma k$ is the global recovery rate when k nodes are infected, and a_k is the rate at which the population goes from k to $k + 1$ infected individuals.

The approximation is exact in the case of complete or fully connected networks, where the a_k rates are given by the expression $a_k = \tau k(N - k)$. In the general case, the a_k's are random variables themselves, since the rates at which infections happen are the product of τ times the total number of $S - I$ links in the network, a random variable itself that reflects the topology of the network and the way in which the epidemic positions the k infected nodes on the network. This means that the epidemic at population-level is not Markovian, making an exact treatment much

more difficult and still out of reach.

However, by using the master equation, we can recast this process as a Markovian one using a suitable approximation of each rate a_k. A natural proposal is a quantity that captures the average rate of infection, weighted by the time spent in the observed states, that is:

$$a_k = \frac{\tau \sum_{\xi_k} \xi_k t_{\xi_k}}{\sum_{\xi_k} t_{\xi_k}}, \tag{3.2}$$

where ξ_k are the observed counts of the number of S-I links on the network when k infected nodes are present and t_{ξ_k} is the lifetime of this particular state. This quantity is responsible for driving the epidemic: The higher the number of S-I links, the larger the rate of generating more infected nodes. The a_k's can be obtained by averaging over many realisations of the epidemic on different realisations of networks from the same family. This can also be interpreted as averaging out stochasticity at link-level and transferring it to population-level. Hence, the variability in epidemic paths will be due to the stochasticity of the master equation itself, guaranteeing the Markov property of the Birth-Death process. The solution of equation (3.1) with these proposed rates has been shown to be in excellent agreement with the average from many simulations for various network models and epidemic parameters [Di Lauro et al., 2020a].

3.2.2 Fokker Planck equation as a limit of the Birth-Death process

Master equation (3.1) can be used as a starting point to build its continuous (in space) limit, i.e., the Fokker-Planck equation [Gardiner, 2004; Kiss et al., 2017]. The idea is to approximate the solution $p_k(t)$ by considering it as a discretisation of a continuous function $f(t, x)$ in the interval $[0, 1]$, defined as

$$f\left(t, x = \frac{k}{N}\right) = p_k(t).$$

For the large N limit to exist, it is known [Kiss et al., 2017; Kurtz, 1970; Ethier and Kurtz, 2009; Nagy et al., 2014; Batkai et al., 2013] that the rates of the master equation need to satisfy the following density-dependence condition (with a slight

abuse of notation):

$$\frac{a(k)}{N} = A\left(\frac{k}{N}\right), \qquad \frac{c(k)}{N} = C\left(\frac{k}{N}\right), \tag{3.3}$$

where A and C are not necessarily the same functions as a and c. It is worth noting that condition (3.3) is not guaranteed to hold for every network model, and must therefore be validated on networks of interest.

In the density-dependent case, it can be shown [Kiss et al., 2017; Nagy et al., 2014; Batkai et al., 2013] that $f(t,x)$ satisfies the following forward Fokker-Planck equation:

$$\partial_t f(t,x) = \frac{1}{2N}\partial_{xx}\big(\sigma^2(x)f(t,x)\big) - \partial_x\big(\mu(x)f(t,x)\big), \tag{3.4}$$

with initial condition $f(0,x) = \delta(x - x_0)$, where the diffusion coefficient $\sigma^2(x)$ and the drift $\mu(x)$ are related to the a_k and c_k rates via:

$$\begin{aligned} \sigma^2(x) &= \frac{1}{N}(A(x) + C(x)), \\ \mu(x) &= A(x) - C(x). \end{aligned} \tag{3.5}$$

Boundary conditions are naturally emerging from two considerations: (1) if the process hits $k = 0$ at some time (disease-free state) it will stay there forever, and (2) the number of infected nodes cannot be greater than N at any given time. In this framework, such physical constraints translate naturally into Dirichlet and Robin boundary conditions:

$$\begin{cases} f(t, x = 0) = 0, & \text{absorption in } x = 0, \\ \frac{1}{2}\partial_x(\sigma^2(x)f(t,x))|_{x=1} - (\mu(1)f(t,1) = 0, & \text{reflection in } x = 1. \end{cases}$$

Fokker-Planck equations of this kind have been extensively studied numerically, especially in the biological context of population random genetic drifts [Duan et al., 2018; Chen et al., 2014; Applegate, 2013; Cacio et al., 2012], as well as analytically [Feller, 1954; Trabelsi and Naouara, 2017; Kovacevic, 2018]. In particular, in [Kovacevic, 2018], this equation is studied in the limit of large t to characterise the so-called quasi-steady state [Méléard and Villemonais, 2012; Collet et al., 2013] (where the only steady state possible is absorption at 0), whereas in [Cacio et al., 2012; Chen et al., 2014] various numerical schemes to solve such equations are employed and compared in terms of numerical instabilities and performance. In Appendix 3.6 we describe our numerical scheme of choice, which is an adaptation of a finite volume method (FVM) scheme already described in [Chen et al., 2014].

Networks	$\langle k \rangle$	τ	γ	R_0
Regular	9	1	6	1.28
	7	2.5	8	1.66
	8	3.5	7	2.65
Erdős-Rényi	8	1	5	1.50
	10	1	4.5	2.44
	7	4	7	2.90
Barabási-Albert	10	0.9	3.5	2.12
	4	2	5	3.72
	18	0.55	6.2	5.38
2D lattice	4	1.2	2	1.12
	4	2	2	1.5
	4	8	2	2.4

Table 3.1: Values of network and epidemic parameters for the benchmark scenarios chosen to test the PDE limit of large networks. R_0 has been computed on networks of size $N = 10000$ via the formula $R_0 = \frac{\langle k^2 - k \rangle}{\langle k \rangle} \frac{\tau}{\tau+\gamma}$, as described in [Kiss et al., 2017].

3.3 Results

3.3.1 Validation of the density dependence condition

In order to use eq. (3.4) we need to verify that the rates of the BD process satisfy condition (3.3). Recoveries are independent of the network, therefore, the condition is automatically satisfied as the expression for these rates is $c_k = \gamma k$. Infection rates, instead, need to be inspected more closely, as their values are dependent on the topology of the network. As an example, even fully-connected networks with $a_k = \tau k(N - k)$ violate condition (3.3). This can be corrected by requiring that τ scales as $\tau/N = ct$ in the limit of large N. This case is well-known in the literature [Gray et al., 2011; Allen, 2017], albeit in a SDE formulation, so we limit our treatment of it to reporting the exact Fokker-Planck equation for the fully connected network, i.e.

$$\partial_t f(t,x) = \frac{1}{2N} \partial_{xx} \left[(\beta x(1-x) + \gamma x) f(t,x) \right] - \partial_x \Big[(\beta x(1-x) - \gamma x) f(t,x) \Big],$$

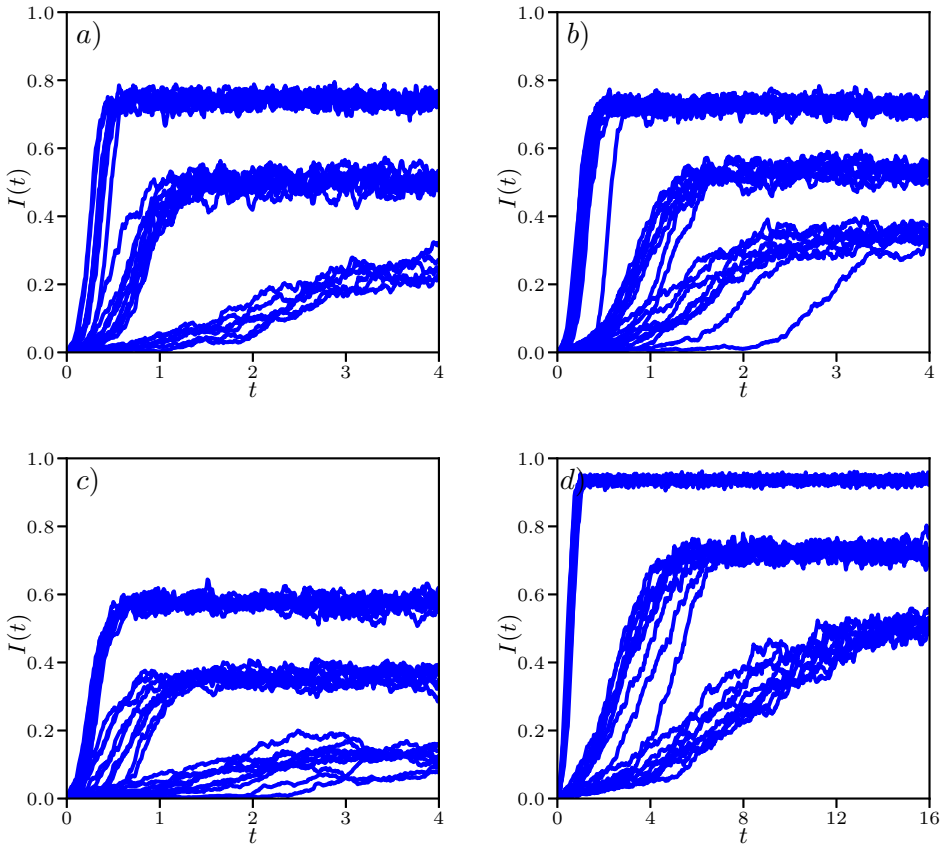

Figure 3.2: Typical realisations of SIS epidemics on (a) Regular, (b) Erdős-Rényi, (c) Barabási-Albert and (d) 2D lattice networks, for the parameter values shown in Table 3.1 and with $N = 1000$ (for the lattice, this number is 1024). In each panel 10 realisations of the epidemics are plotted,. The parameters used to generate such networks are also reported in Table [3.1], higher prevalence corresponds to higher values of R_0.

where $\beta = \frac{\tau}{N^2}$.

Since degree heterogeneity and higher-order stucture in networks have a marked effect on epidemics we explore Regular, Erdős-Rényi and Barabási-Albert networks and lattices. To test the scaling hypothesis, the infection rates, based on eq. (3.2), are computed on different networks and for different values of N (typically from $N = 10^2$ to $N = 10^5$ with slight variations for lattices). The resulting (k, a_k) curves are plotted in figures 3.3 (Regular) and 3.4 (Erdős-Rényi) (see also figures 3.11 and 3.12 corresponding to Barabási-Albert networks and lattices, respectively, in Appendix 3.7). Using eq. (3.3), these rates are rescaled and plotted again in the same figures confirming that they define a universal rate.

In figure 3.5, the universal curves (based on the highest N explored) are compared

for the four different network types in order to highlight how the topology of the network impacts the shape of the (k, a_k) curves. As expected on a lattice, the a_k's grow linearly with k. On all other networks, the curves are parabola-like but higher degree heterogeneity leads to a more pronounced left skew in the location of the maximum point of the rate curve. This is because nodes with many links are likely to be infected early on in the epidemic, meaning that even for low k values a_k can be high (if a hub is infected then this leads to many new S-I links) compared to networks with milder degree heterogeneity. Once most of the highly connected nodes are infected (typically only a small proportion of N), the epidemic will unfold on the less well connected nodes, therefore fewer links, and thus the parabola decreases early on as shown by the left skew. We note that some variability between the scaled (k, a_k) curves emerges and this is likely due to finite size effects where stochastic variability is accentuated. However, the difference is so small that the Fokker-Planck equation and its solution appear insensitive to the exact choice of the universal rates.

3.3.2 Comparing PDE and simulations

Since the limit of large N is of interest, it is beneficial to have a continuous function that fits the discrete a_k rates (3.2). In [Di Lauro et al., 2020a], the following three-parameter model was proposed:

$$a_k^{(C,\alpha,p)} = Ck^p(N-k)^p\left(\alpha\left(k-\frac{N}{2}\right)+N\right). \tag{3.6}$$

This model can be fitted to the a_k curves via a least-square approach, by minimizing the following cost:

$$e = \sum_{k,\sum t_{\xi_k}>0} \left(a_k^{(C,\alpha,p)} - a_k\right)^2. \tag{3.7}$$

In [Di Lauro et al., 2020a], we showed that this approach leads to good agreement with simulations from different network classes, in particular, Regular and Erdős-Rényi networks. The fit for Barabási-Albert networks is acceptable and the fit breaks down for lattices.

In the following, we make use of this function to model the infection rates of master equation (3.1). However, using a simple function to model the complexity of

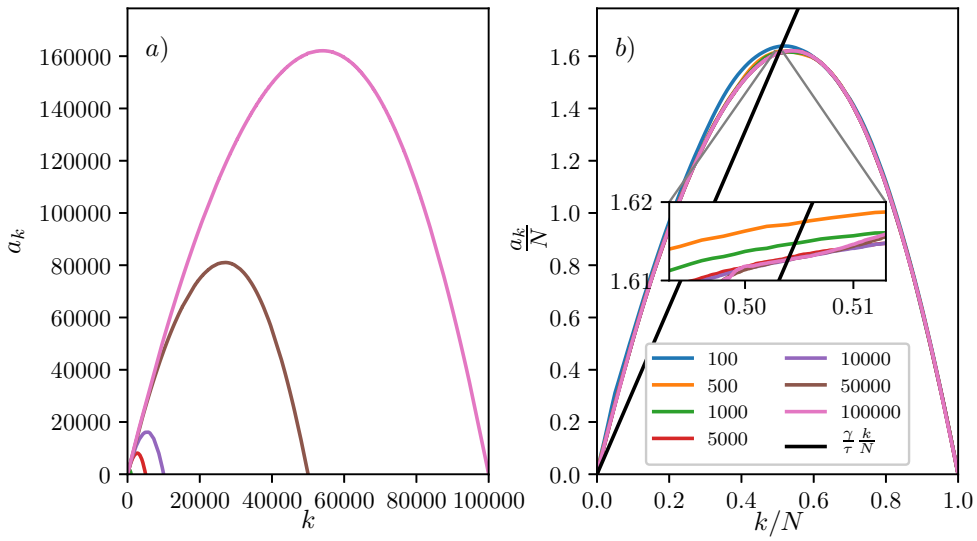

Figure 3.3: Scaling for regular networks using parameters given in the first row of Table (3.1). (a) Unscaled (k, a_k) curves for values of N ranging from $N = 100$ to $N = 100000$. Each curve is obtained by simulating 10000 realisations of the epidemic across 50 realisations of the network, half of the epidemics starting from $k_0 = 1$, the other half from $k_0 = N$. (b) Corresponding scaled rate $(k, \frac{a_k}{N})$ curves. The scaling hypothesis can be checked by noticing that the higher the values of N, the closer the scaled curves get to the limiting universal curve. As N increases, the differences between scaled rates decrease. In the inset, the small mismatch between curves with $N \geq 1000$ are highlighted using a 30x zoom. For completeness, the $(k/N, \gamma k/N)$ curve is provided (in black); it intercepts the scaled curves around the steady state.

the a_k rates adds an additional layer of approximation to our approach. Therefore, in addition to eq. (3.6) we also consider a cubic spline of the a_k rates, as it provides an even better fit to the rates based on eq. (3.2) and therefore yields better results. This is particularly apparent for lattices, where the (C, α, p) model fails for obvious reasons, and to some extent for Barabási-Albert networks. To summarise, the rates of infection are first found based on simulations via eq. (3.2). As this approach produces a discrete function that cannot be used as is in the Fokker-Planck equation, we propose two alternatives: (a) the (C, α, p) model, eq. (3.6), and (b) a spline. The PDE is considered with both rates (except for lattices), and the numerical solution of the PDE is computed via a Finite Volume Method (several other numerical schemes [Mohammadi and Borzì, 2015; Cacio et al., 2012; Chen et al., 2014] were

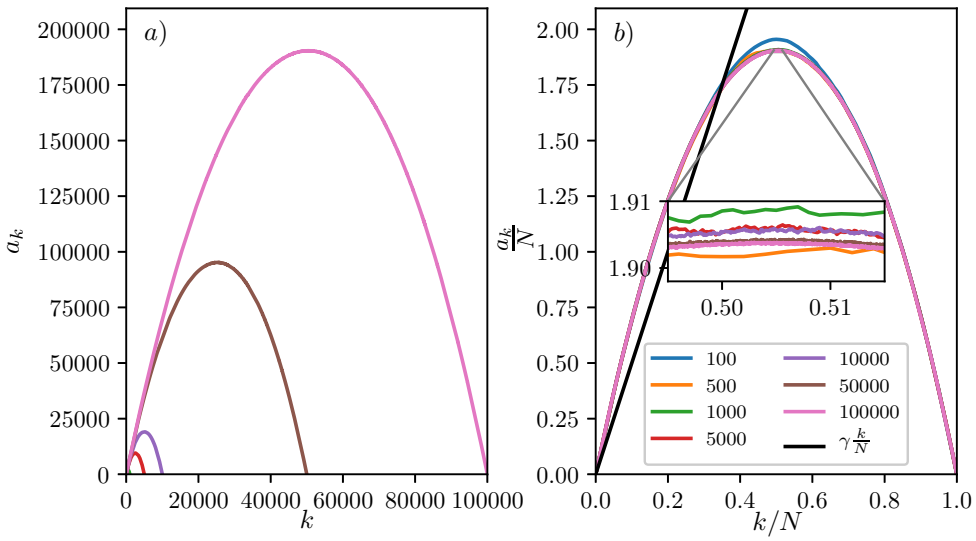

Figure 3.4: Same scenario as in figure 3.3 but for Erdős-Rényi networks using parameter values from the sixth row of Table (3.1). (a) Unscaled (k, a_k) curves for values of N ranging from $N = 100$ to $N = 100000$. Each curve is obtained by simulating 10000 realisations of the epidemic across 50 realisations of the network; half of the epidemics starting from $k_0 = 1$, the other half from $k_0 = N$. (b) Corresponding scaled $(k, \frac{a_k}{N})$ curves.

tested) as it guarantees that the solution remains non-negative and preserves mass, see Appendix.

To show the agreement between the Fokker-Planck equation (3.4) and results from simulations on networks, we selected twelve (three for each network model) combinations of network and epidemic parameters, as described in Table 3.1. We tuned the parameters such that for each family we could get epidemics with different characteristics, i.e., different transient and quasi-steady state. To show this, in figure (3.2) we illustrate a few realisations of epidemics on networks of size $N = 1000$ for each scenario (for the 2D-lattice network, the size is $N = 1024$). We also report the computation of R_0 as described in [Kiss et al., 2017], see Table 3.1. Note that the initial condition for the PDE is always taken as in the simulations, so $k_0 = 1$. In the simulations, at every run of the epidemic we select a node at random to be the initially infected one. This, however, does not prevent setting initial conditions with a higher number of infected nodes. The initial condition in such cases should be based on measurements taken from the simulations.

Parameters were chosen so that, for each family, the three quasi-steady states

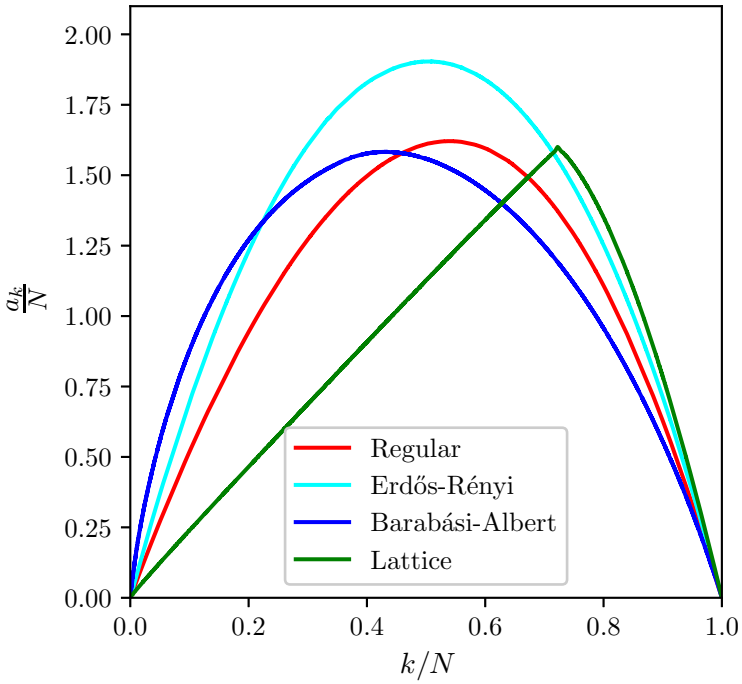

Figure 3.5: Comparison of different scaled (k, a_k) curves produced by different network models for large N. The (k, a_k) curves are scaled by a factor N. The parameters to generate data for each curve are provided in the second, the fourth, the eighth and the last rows of table 3.1, respectively. The lattice network is of size 316×316, and the other networks have $N = 10^5$.

showed different levels of prevalence. To find the (k, a_k) curves via minimisation of (3.7), we generated data as follows: for each scenario, we created 50 realisations of the network, and on each we ran 200 realisations of the epidemic, half of which started from $k_0 = 1$, the other half from $k_0 = N$. This was done in order to obtain observations over the whole range of possible values of the infected nodes. Indeed, when epidemics start from low k_0 values, they only very rarely reach a prevalence much higher than the quasi-steady-state.

The numerical solutions of equation (3.4) are compared with results based on Gillespie simulations [Gillespie, 1976, 1977], see figures 3.6, 3.7, 3.8 and 3.9. Excellent agreement holds for all scenarios we tested, as long as the size of the network is ≥ 1000. For small networks, there is a finite-size effect that does not allow for as good a fit. Interestingly enough, although there are small differences between different a_k curves, as long as $N \geq 1000$, the exact choice of N has little

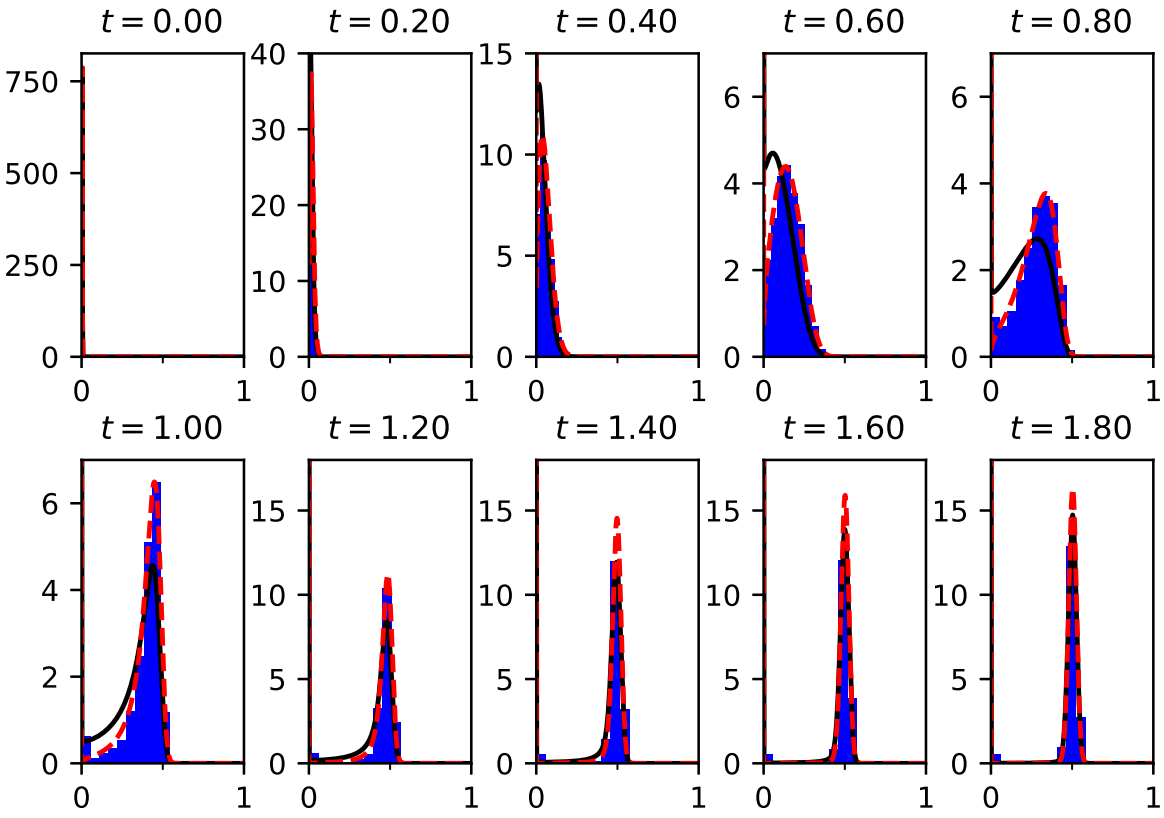

Figure 3.6: Temporal evolution of the probability distribution $p_{x=\frac{k}{N}}$ (blue histogram) sampled from 25000 realisations of epidemics across 100 realisations of regular networks (2nd row of Table 3.1), with $N = 1000$. Lines are the numerical solutions to the Fokker-Planck equation (3.4) computed from two different a_k rates: best (C, α, p) fit (continuous curve) and cubic spline of the raw a_k computed as in eq. (3.2) (dashed line). The first panel shows the initial condition ($t = 0$), which for all simulations is $k_0 = 1$, while the last panel shows the quasi-steady state distribution.

impact on the numerical solutions of the PDE. This supports our conjecture that there is indeed a large N limit and, therefore, a universal scaled a_k curve which is approached as N increases. As can be seen, the spline consistently leads to a better approximation. This is simply due to a tighter fit to the discrete data compared to the fit based on the (C, α, p) model. We note, however, that the (C, α, p) model captures the trend of the epidemic and the quasi-steady state is fitted well. Of course, in the case of the lattice we only use the spline as the (C, α, p) model cannot capture the linear rise.

To realise the comparisons provided in figures 3.6, 3.7, 3.8 and 3.9 we proceeded

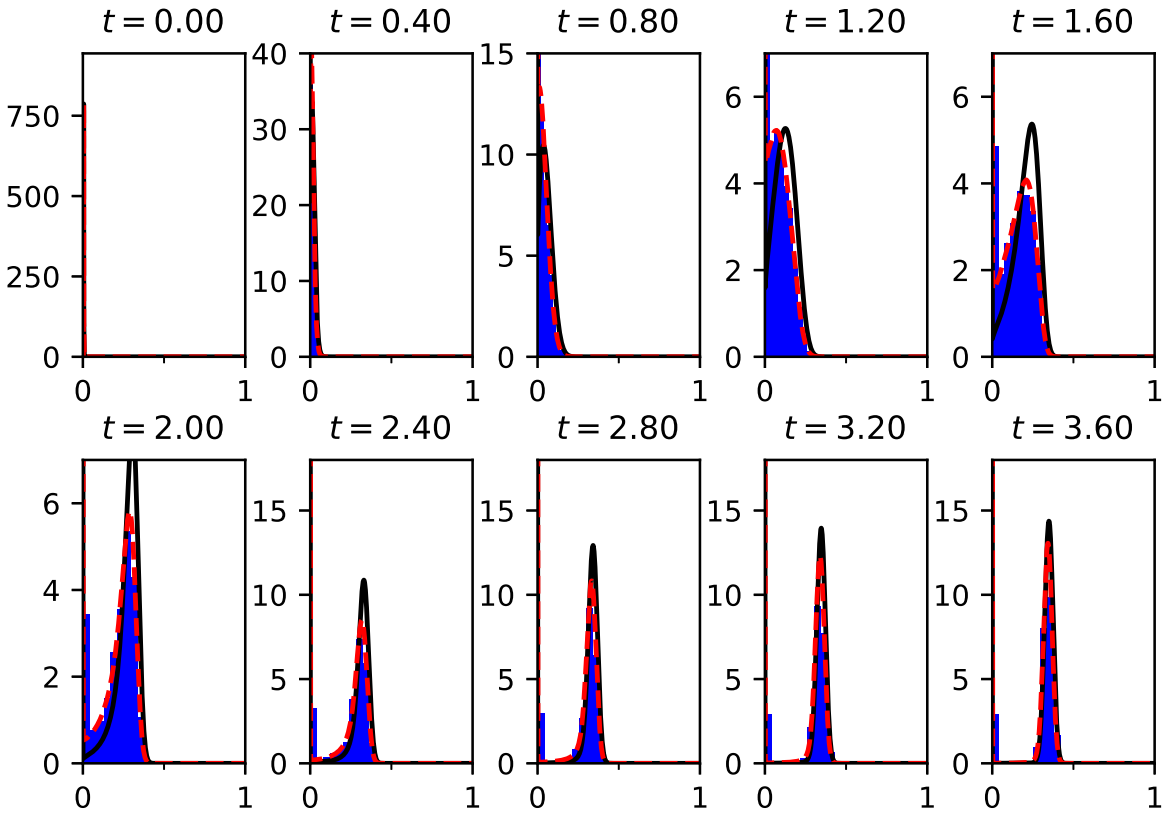

Figure 3.7: Same scenario as in figure 3.6, but using the parameters given in the sixth row of Table 3.1, i.e., the first parameter configuration for Erős-Rényi networks.

as follows. We considered the same network realisations and epidemic parameters used to find the (k, a_k) rates. We fixed the initial condition to be $k_0 = 1$ and ran 200 simulations on each realisation. Each individual path was then sampled at regular times in order to build the empirical distribution $p(x, t)$. Note that all simulations were kept, even those that died out early. This is because the numerical scheme preserves the total probability and can account for these early extinctions.

The PDE with the (C, α, p) model is

$$\partial_t f(t,x) = \frac{1}{2N}\partial_{xx}\left[\left(CN^{2p}\left(x^p(1-x)^p\right)\left(\alpha\left(x-\frac{1}{2}\right)+1\right)+\gamma x\right)f(t,x)\right] + \\ -\partial_x\left[\left(CN^{2p}\left(x^p(1-x)^p\right)\left(\alpha\left(x-\frac{1}{2}\right)+1\right)-\gamma x\right)f(t,x)\right]. \quad (3.8)$$

Our three-parameter model, (C, α, p), can be used to derive the epidemic threshold. In terms of the PDE, see equation (3.8), and as figures 3.6, 3.7, 3.8 and 3.9 show, an epidemic is supercritical when the drift term is positive. This implies that the

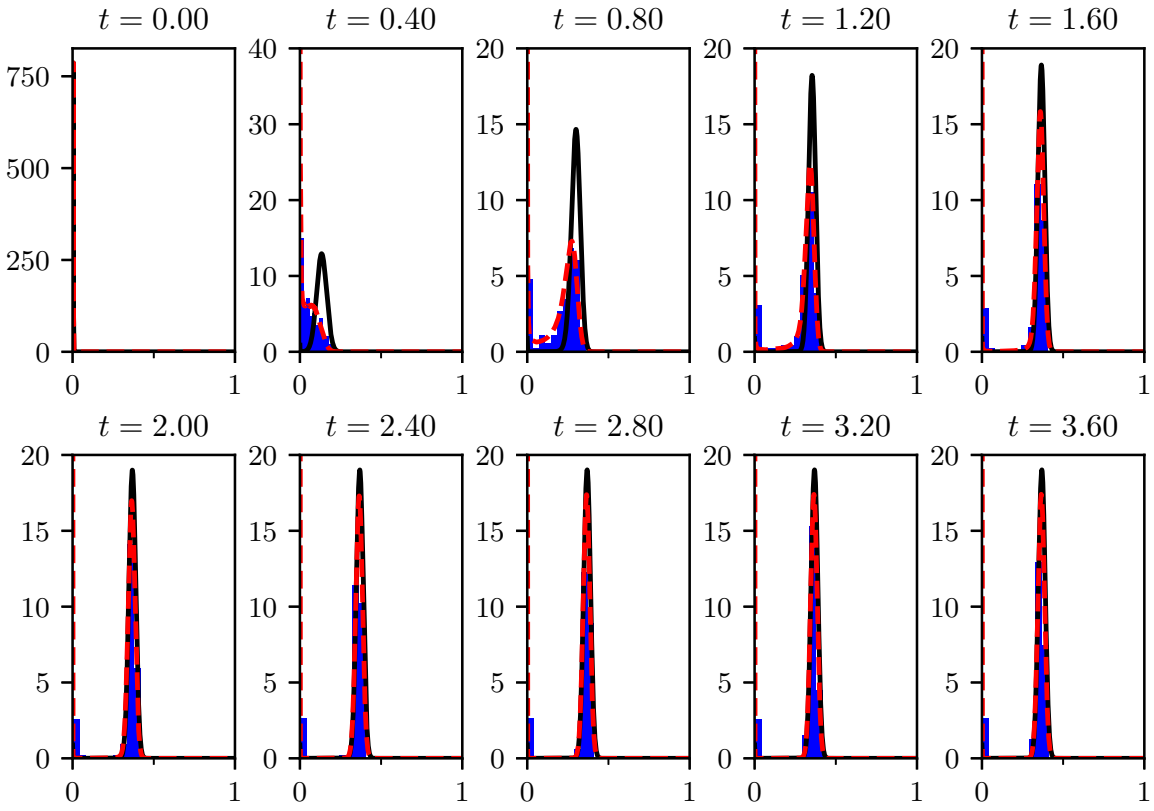

Figure 3.8: Same scenario as in figure 3.6, but using the parameters given in the eighth row of Table 3.1, i.e., the second parameter configuration for Barabási-Albert networks.

epidemic threshold is equivalent to

$$CN^{2p}\left(x^p(1-x)^p\right)\left(\alpha\left(x-\frac{1}{2}\right)+1\right)-\gamma x \geq 0,$$

at the start of the epidemic, that is $x \simeq 0$. As shown in [Di Lauro et al., 2020a], for Regular and Erdős-Rényi networks, $p \simeq 1$ and this leads to

$$x\left[CN^2(1-x)\left(\alpha\left(x-\frac{1}{2}\right)+1\right)-\gamma\right] \geq 0,$$

taking the limit $x \to 0$ in the expression within the brackets above leads to

$$CN^2\left(1-\frac{\alpha}{2}\right) > \gamma.$$

This expression reduces to the well-known condition $R_0 = \frac{\tau}{\gamma} \geq 1$ for fully connected networks. Indeed, scaled rates for such networks can be computed exactly to be $a(x) = \frac{\tau}{N^2}x(1-x)$, meaning that $C = \frac{\tau}{N^2}$, $a = 0$ and $p = 1$ in this case.

This equation is implicit, as, of course, both C and α depend on the network and epidemic parameters in a non-trivial way. Therefore it cannot be used as it is, but it

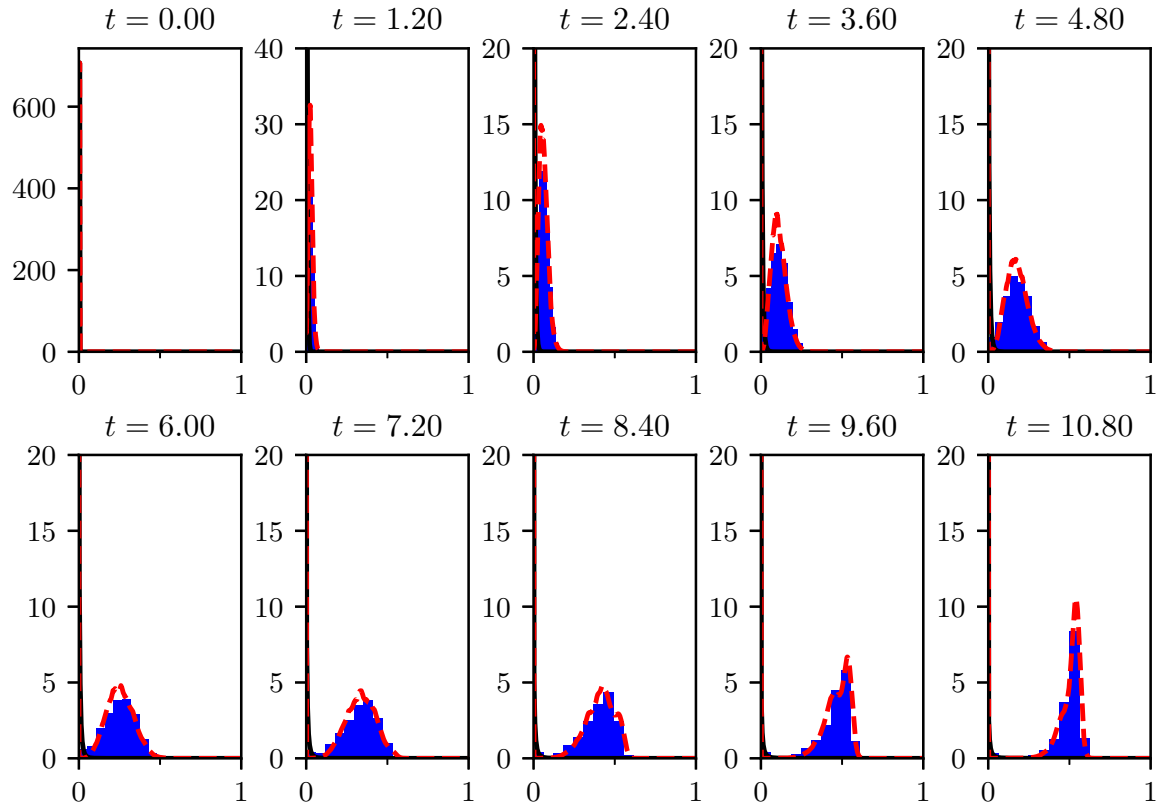

Figure 3.9: Same scenario as in figure 3.6, but on a 2D lattice with periodic boundary conditions, using the parameters given in the 10th row of Table 3.1, i.e., the first parameter configuration for 2D lattices networks.

offers an interesting interpretation since α determines whether the (k, a_k) curves are left- or right-skewed, see figures 3.5 and [Di Lauro et al., 2020a]. Furthermore, the topology of the underlying network plays an important role in determining the shape of this curve; for example, Barabási-Albert networks lead to (k, a_k) curves with a left skew [Di Lauro et al., 2020a]. Thus, all else being constant, networks with high degree heterogeneity are more likely to see the threshold go past the critical value.

3.3.3 Inference of infection rates using the Fokker-Planck approximation

In this last section, we provide a simple example of the usefulness of the Fokker-Planck approximation: inferring epidemic and network parameters given data. Specifically, we consider the case in which a single trajectory of BD (or Gillespie simu-

lation of the epidemic on an explicit network) process is observed at discrete time-steps, i.e.:

$$y = \{(t_1, k_1), \dots, (t_n, k_n)\},$$

where $(t_1, \dots, t_n) \in [0, T]^n$ $(0 \leq t_1 < \cdots < t_n \leq T)$ and $(k_1, \dots, k_n) \in \{0, \dots, N\}^n$ are the sets of times and states, respectively. To set up the inference, we express the likelihood using the transition probability function of a BD process as follows (using the independence of increments and time homogeneity):

$$\mathcal{L}_{BD}(y; a, c) = \prod_{i=1}^{n-1} \mathbb{P}\left(X(t_{i+1} - t_i) = k_{i+1} | X(0) = k_i; a_k, c_k\right).$$

Unfortunately, for a large state space, these transition probabilities are numerically expensive to compute. Additionally, inferring the full set of rates a_k's and c_k's may not be efficient. Instead, we recast this problem as that of inferring the C, α, p parameters of the Fokker-Planck approximation, as in eq. (3.8). This is a much more tractable numerical problem, that can still provide useful information about the underlying network and epidemic, as showed in [Di Lauro et al., 2020a]. In terms of the computational complexity of solving the PDE versus solving the ODE system, we argue as follows: the system of ODEs requires to solve exactly $(N + 1)$ equations for each time-step (in a naive explicit Euler scheme), meaning that the complexity scales as $O(N)$. The Finite Volume Method we are using requires us to invert a matrix at each time-step. The size of the matrix is $V \times V$, where V is the size of the volume mesh. The computational complexity of this operation is $O(V^\alpha)$, $\alpha \geq 1$. The advantage of PDE and FVM is that we can choose $V = O(N^\beta)$ such that $\beta\alpha \leq 1$. Notice that a viable choice of β is $\beta = 0$, in which case the grid size is constant with respect to N. This results in a trade-off between space resolution and speed of our solver, while the ODE does not offer this degree of flexibility. This leads us to replace the previous likelihood function with the following:

$$\mathcal{L}_{FP}(y; C, \alpha, p) = \prod_{i=1}^{n-1} f(t_{i+1} - t_i, x_{i+1}; x_i, C, \alpha, p), \tag{3.9}$$

where $f(t, x; x', C, \alpha, p)$ is the transition probability density obtained from equation (3.4), the coefficients are given by the C, α, p model, $x_i = \frac{k_i}{N}$ for all $i \in [1, n]$ and the initial data is a Dirac delta at location $x' \in (0, 1)$. In this example, f is computed numerically using the finite-volume numerical scheme described in Appendix 3.6. To

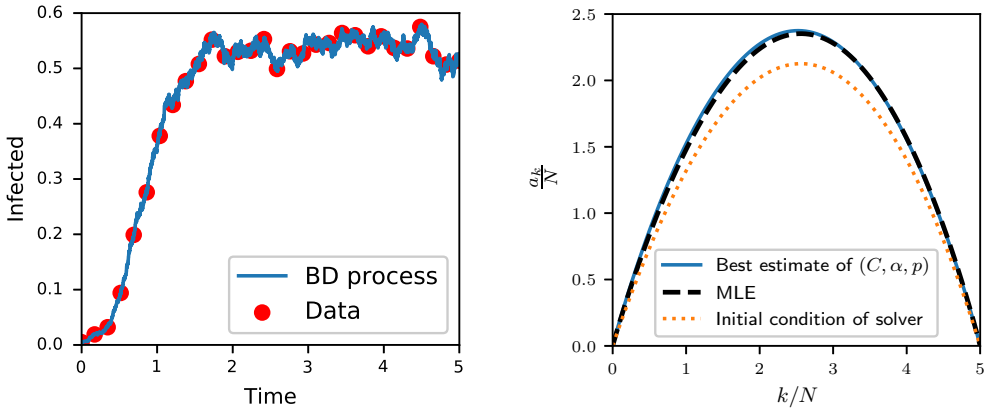

Figure 3.10: (left) Data generated from a single realisation of an SIS process on an Erdős-Rényi network with $N = 1000$, $\langle k \rangle = 10$, $\tau = 1$, $\gamma = 4.5$ via the Gillespie algorithm. the curve was sampled regularly to get 30 data-points over 5 units of time. (right) (C, α, p) function obtained by maximising the logarithm of the likelihood 3.9 (black dashed line) compared to the $(\hat{C}, \hat{\alpha}, \hat{p})$ function obtained by fitting the (k, a_k) curve obtained for $N = 1000$ by exploring the full curve with continuous observations of 10000 epidemics across 50 network realisations, as in figure 3.4 (blue continuous line). The initial condition inputed to the locally bounded gradient-descent solver is shown by the orange dotted line.

illustrate the accuracy of this approach, we consider a set of parameters from the choices of Table 3.1 (figure 3.10 shows the behaviour of the system when parameters are those on the 5th row of Table 3.1, i.e. $C = 1.36e-05, \alpha = 3.44e-2, p = 9.7e-1$) and generate a trajectory from a single realisation of the SIS epidemic on a Erdős-Rényi network of size 1000, via Gillespie algorithm. This dataset is shown in figure 3.10 and consists of $n = 30$ distinct and equally spaced data points taken from the epidemic curve. These are then scaled to $[0, 1]$ (taking $x_i = \frac{k_i}{N}$ for all $i \in [1, n]$). The dataset is then used to find a Maximum Likelihood Estimator (MLE) by simply maximizing the likelihood function from equation (3.9) with respect to C, α, p, that is finding

$$(\hat{C}, \hat{\alpha}, \hat{p}) = \arg\max \mathcal{L}_{FP}\left(y; C, \alpha, p\right).$$

To show that this method provides a good estimate, we simply plot the MLE rates, $(\hat{C}, \hat{\alpha}, \hat{p})$, against the rates obtained by fitting the (k, a_k) curve directly from data from continuous observations of multiple realisations of epidemics on networks of size $N = 1000$, we call this the best estimate, as in figure 3.4. The best estimate and the rates based on $(\hat{C}, \hat{\alpha}, \hat{p})$ are indeed in good agreement. We repeated the

inference scheme for all benchmark cases in Table 3.1 (not shown), with the exception of lattices. The agreement was similar to that shown in figure 3.10 (right panel).

It is worth noting that the goodness of inference depends on how many points the dataset contains and also how much of the transient and the quasi-steady state is captured. In the transient, the drift dominates and the process is more stochastic. On the other hand, in the quasi-steady state, the drift coefficient tends to zero and fluctuations around it are mainly due to diffusion. Hence, both regimes are needed if drift and diffusion are to be inferred correctly. Data in the transient or in the quasi-steady state alone can lead to sub-optimal inference as different parameter combinations that provide good fit can be found. The example reported in figure 3.10 is an illustration of how useful the PDE limit of epidemics on finite networks can be in a network and epidemic inference setting. Further, the approximation that we provide can also be used in a Bayesian approach, by first setting a prior over the parameters C, α, p for instance.

3.4 Conclusions

In this paper we conjectured and showed numerical evidence for the existence of PDE limits for exact SIS epidemics on Regular, Erdős-Rényi and Barabási-Albert networks and 2D lattices with periodic boundary conditions. The key to our approach is to use a BD approximation which then has a PDE limit provided that the coefficients of the BD process are density-dependent. Hence, one of the main challenges was to verify, at least numerically, that this was the case. What is common between all the networks that we considered is that simply increasing the number of nodes in the network will not change what a node experiences locally, e.g. the number or distribution of neighbours. In fact for Erdős-Rényi networks we made sure this is the case by choosing the probability of connection p such that $p = \langle k \rangle/(N-1)$. The same argument seems to hold for scale free networks where the average degree stays constant and nodes at any scale in networks of any size experience the same type of neighbourhood. Of course, this is not the case for fully connected networks since the number of neighbours of a node increases with network size. Furthermore, we note that a_k's are random variables with some distribution around a well-defined

mean. This spread/variance in a_k's is in some sense due to higher-order moments in the network. The variance of these distributions is larger at the beginning of the epidemic and it decreases with time or as k increases - meaning that the system tends to reach an equilibrium where the higher order moments in the network are not significant enough to produce a real effect. Based on these arguments, we expect our method to extend readily to configuration networks [Molloy and Reed, 1998] whose degree distribution does not depend on network size.

Of further interest will be to test and, if it works, extend our approach to clustered and/or networks with community structure. This is a difficult task as clustering and community structure can be introduced in different ways; for example clustered networks can be generated by using Big-V rewiring or by using a family of clustering inducing subgraphs (e.g. triangles, four fully connected nodes, and other motifs) [Ritchie et al., 2017]. For networks with community structure, a good choice could be the stochastic block model [Holland et al., 1983]. However, as our analysis shows the (C, α, p) model struggles to capture the infection rate curves for all four network models that we considered. This suggests that a more flexible model is needed, possibly a non-parametric one, expecially when networks with more complex topologies are considered.

To solve the PDE numerically, we employed a second order in time finite volume method whose stability was proven in [Chen et al., 2014]. We compared such numerical solutions to probability distribution sampled from the Gillespie simulation. The agreement is in general good and, as expected, it becomes excellent as N increases. The existence of the PDE limit is not surprising, given that the coefficients of the BD process are density-dependent. However, it is important to note that the agreement between the solutions of the PDEs and empirical distributions based on simulations provides strong support for the validity of the BD process, strengthening the evidence provided in [Di Lauro et al., 2020a], and thus closing the loop illustrated in figure 3.1.

A PDE perspective on epidemics provides several efficiency gains. The first is to do with computational efficiency and the possibility to quantify variability. More importantly perhaps, the solution of the PDE serves as a likelihood which can be very efficiently computed/evaluated and can form the basis of many networks and

epidemic inference models, see Section 3.3.3. This is in contrast with approaches where the networks are explicitly modelled [Ma et al., 2019b] and computational complexity can make inference out of reach.

At least three separate avenues of future research emerge. First, and perhaps most importantly, a theoretical justification for the Birth-and-Death approximation is still needed, if indeed that is possible. Second, there is a need to investigate the extent to which this method can be extended to other network families and epidemic dynamics. Thirdly, there is scope to consider how the approximate master equation could be used to look into the impact of the network on quantities such as time to extinction. Nevertheless, given that handling exact epidemic models on networks is still challenging even for networks of modest size, we believe that proposing new ways to approximate epidemics is worthwhile and may contribute new modelling and analysis perspectives.

3.5 Acknowledgments

All authors acknowledge support from the Leverhulme Trust for the Research Project Grant RPG-2017-370. The authors acknowledge useful discussions with Dr M. Dashti during the development of this research.

3.6 Appendix: numerical method for solving the PDE

In this section we detail the numerical method and algorithms used to solve eq. (3.4). The algorithm employed is an adaptation and modification of the finite volume method (FVM) named FVM3 in [Chen et al., 2014]. First, we write the Fokker-Planck equation in the form

$$\frac{\partial f(x,t)}{\partial t} + \frac{\partial j(x,t)}{\partial x} = 0,$$

where in our case the current term is $j(x,t) = -\frac{1}{2N}\frac{\partial \sigma^2(x) f(x,t)}{\partial x} + \mu(x) f(x,t)$, while initial and boundary conditions are:

$$\begin{cases} f(x,0) = \delta(x - x_0), & \text{initial condition,} \\ f(0,t) = 0, & \text{absorption in } x = 0, \\ \frac{\partial f(x,t)}{\partial x}|_{x=1} = 0, & \text{reflection in } x = 1. \end{cases}$$

In our case, both $\mu(x)$ and $\sigma(x)$ vanish at 0, indicating that the only possible steady state is absorption [Kovacevic, 2018]. Therefore, the solution to this equation is such that $\lim_{t\to\infty} f(x,t) = \delta(x)$. Further, since the solution should provide a probability density function, we require that $f(x,t) \geq 0$ everywhere (positivity) and that $\int_0^1 f(x,t)\,dx = 1$ for any $t > 0$ (conservation of mass). Finite Volume Methods are a class of numerical methods to solve PDEs [Eymard et al., 2000] in which the constraints described above are explicitly satisfied, therefore FVM is the natural candidate for this type of problems. Following notation of [Chen et al., 2014] we consider a uniform grid, with spacing $h = \frac{1}{M}$ and grid points $x_i = ih$, $0 \leq i \leq M$. Similarly for time, we consider a uniform grid with spacing τ and grid points $t_i = n\tau$, $0 \leq n \leq n_{max}$. We define j_i^n and f_i^n to be the numerical approximations of $j(x_i, t_n)$ and $f(x_a, t_n)$, respectively. The control volume $\mathcal{D}_i = \{x \; s.t. \; x_{i-\frac{1}{2}} \leq x \leq x_{i+\frac{1}{2}}\}$ is associated to each inner point x_i, whereas two control domains $\mathcal{D}_0 = \{x \; s.t. \; 0 \leq x \leq x_{\frac{1}{2}}\}$ and $\mathcal{D}_M = \{x \; s.t. \; x_{M-\frac{1}{2}} \leq x \leq 1\}$ are reserved for boundary points, and $x_{i+\frac{1}{2}}$ is defined as $(i + \frac{1}{2})h$.

The discretisation of the time derivative can be done with a first-order scheme (as in [Chen et al., 2014]) or a higher-order scheme. We opted for a second-order

scheme for time, for which, in general

$$\frac{\partial f(x_i, t_n)}{\partial t} \approx \frac{3f_i^n - 4f_i^{n-1} + f_i^{n-2}}{2\tau},$$

and the first iteration is done with the first order time scheme $\frac{\partial f(x_i,t_{n+1})}{\partial t} \approx \frac{f_i^{n+1}-f_i^n}{\tau}$. The reason for this choice is that in our case the current term contains both first and second order space derivatives, so to balance out the required space precision, we matched it with a second-order discrete time derivative. This improved the stability of the solution.

The discretised Fokker-Planck equation then becomes:

$$\frac{3f_i^n - 4f_i^{n-1} + f_i^{n-2}}{2\tau} + \frac{j_{i+\frac{1}{2}}^n - j_{i-\frac{1}{2}}^n}{h} = 0 \tag{3.10}$$

We now define the numerical equations imposed by the boundary conditions. The stability of the numerical scheme (in particular, conservation of mass) is influenced by the boundary condition at $x = 0$. Naturally, this condition would be $f(0, t) = 0$ (absorption), as we already discussed. However, changing it to be a zero-current condition (i.e. $j(x, t) = 0$) results in a numerical solution that is more stable. This change of condition does not influence the solution, as the discretised process is never evaluated at $x = 0$. Therefore, we use the following boundary conditions:

$$j(0, t) = j(1, t) = 0, \tag{3.11}$$

which translates to:

$$\frac{3f_i^n - 4f_i^{n-1} + f_i^{n-2}}{2\tau} + \frac{j_{\frac{1}{2}}}{h/2} = 0, \quad \frac{3f_i^n - 4f_i^{n-1} + f_i^{n-2}}{2\tau} - \frac{j_{M-\frac{1}{2}}^n}{h/2} = 0. \tag{3.12}$$

The difference between instances of FVMs is how the current term is discretised. In [Chen et al., 2014], several different schemes are explored. In particular, the FVM that performed better was the so-called central scheme FVM, in which rather than discretise each component of the current term, we discretise the current as a whole, as follows:

$$j_{i+\frac{1}{2}}^n = -\frac{1}{2N}\frac{\sigma^2(x_{i+1})f_{i+1}^n - \sigma^2(x_i)f_i^n}{h} + \frac{\mu(x_{i+1})f_{i+1}^n - \mu(x_i)f_i^n}{2}, \tag{3.13}$$

where $\sigma^2(x_i) = a(x_i) + \gamma x_i$ and $\mu(x_i) = a(x_i) - \gamma x_i$.

The discretised PDE then becomes:

$$\frac{3f_i^n - 4f_i^{n-1} + f_i^{n-2}}{2\tau} - \frac{1}{2N}\frac{\sigma^2(x_{i+1})f_{i+1}^n - \sigma^2(x_i)f_i^n}{h^2} + \frac{\mu(x_{i+1})f_{i+1}^n - \mu(x_i)f_i^n}{2h} + \\ + \frac{1}{2N}\frac{\sigma^2(x_i)f_i^n + \sigma^2(x_{i-1})f_{i-1}^n}{h^2} + \frac{\mu(x_i)f_i^n - \mu(x_{i-1})f_{i-1}^n}{2h} = 0, \tag{3.14}$$

with the boundary conditions (3.12). Note that this is an implicit scheme, that requires at each time-step to invert a matrix of size $M \times M$.

The initial condition $f(x,0) = \delta(x - x_0)$ is approximated by a Normal distribution $f(x,0) \approx \mathcal{N}(x_0, \tilde{\sigma})$ with $\tilde{\sigma} \ll 1$. The stability of the solution with respect to the variance $\tilde{\sigma}$ is discussed in [Chen et al., 2014], and we have chosen $\tilde{\sigma} = h^2$. The mismatch that can be seen in figures 3.6 and 3.7 at 0 is due to the fact that the algorithm cannot reproduce a δ in 0, and should not be considered a problem, as the mass outside of 0 is correctly computed by the numerical solver. To test whether absorption at 0 could have been a problem for the solver, we repeated the calculation allowing for a small re-infection rate at 0 $\epsilon > 0$, without noticing differences in the results.

The numerical advantage of solving a PDE over a system of N ODEs can be discussed in the following terms. To solve explicitly a system of N ODEs one needs to compute a matrix M at each time-step, that scales proportional to N. It is worth noting that it is not possible to change N, as it is given by the system size. To solve the PDE instead, one needs to define a volume mesh V, whose size scales as $V \sim N^\beta$, for some $0 < \beta < 1$. Inverting the matrix V has a cost of V^α, for some $\alpha > 0$ which depends on the specifics of the numerical scheme. Therefore, as long as one chooses $\beta \leq 1/\alpha$, solving the PDE instead of the master equation is guaranteed to be computationally more efficient.

Our implementation is available online at https://github.com/Fdl1989/PDElimitofepidemics.

3.7 Appendix: universal infection rate curves for Barábasi-Albert networks and lattices

Below we show evidence that the (k, a_k) parameters scale favourably with network size and thus satisfy the density-dependent condition.

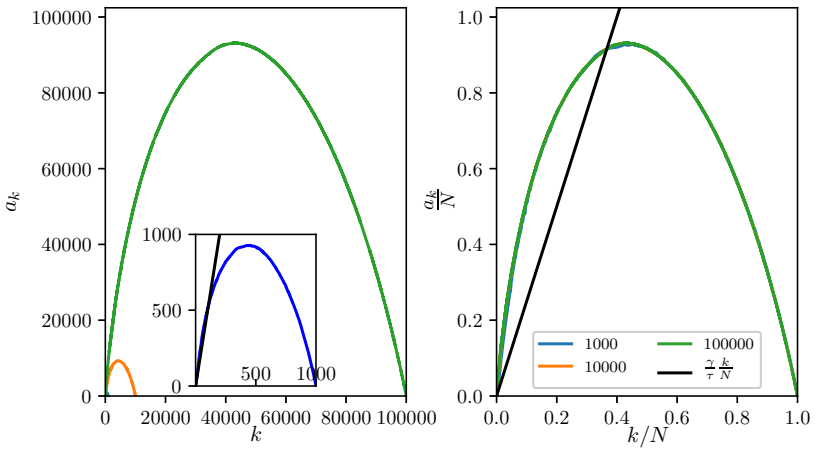

Figure 3.11: Same scenario as in figure 3.3 but for Barabási-Albert networks using parameter values from the eighth row of Table (3.1). (a) Unscaled (k, a_k) curves for values of $N = 10^3, 10^4$ and 10^5. Each curve is obtained by simulating 1000 realisations of the epidemic across 50 realisations of the network; half of the epidemics starting from $k_0 = 1$, the other half from $k_0 = N$. The inset shows a $30x$ zoom of the curve produced for $N = 1000$. (b) Corresponding scaled $(k, \frac{a_k}{N})$ curves.

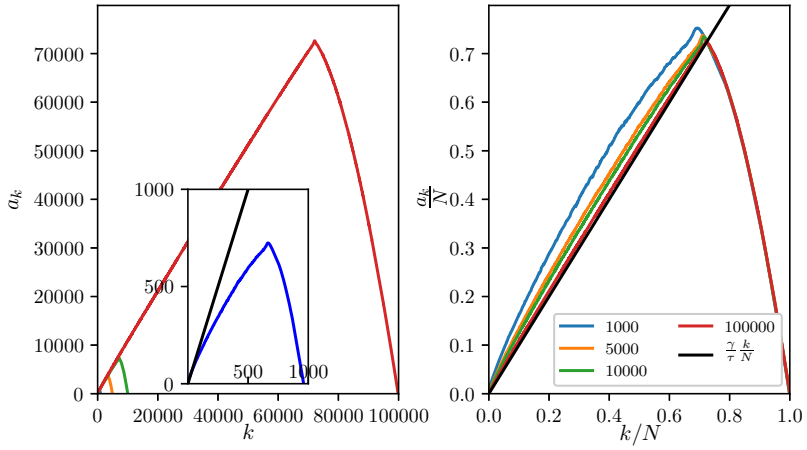

Figure 3.12: Same scenario as in figure 3.3 but for 2D lattice with periodic boundary conditions, using parameter values from the tenth row of Table (3.1). (a) Unscaled (k, a_k) curves for values of $N = 1024, 5041, 10000, 100489$. Each curve is obtained by simulating 1000 realisations of the epidemic across 50 realisations of the network; half of the epidemics starting from $k_0 = 1$, the other half from $k_0 = N$. The inset shows a $30x$ zoom of curve produced for $N = 1000$. (b) Corresponding scaled $(k, \frac{a_k}{N})$ curves.

Chapter 4

Optimal timing of one-shot interventions for epidemic control

Author Summary

Some interventions which help control a spreading epidemic have significant adverse effects on the population, and cannot be maintained long-term. The optimal timing of such an intervention will depend on the ultimate goal.

- Interventions to delay the epidemic while new treatments or interventions are developed are best implemented as soon as possible.

- Interventions to minimize the peak prevalence are best implemented partway through the growth phase allowing immunity to build up so that the eventual rebound is not larger than the initial peak.

- Interventions to minimize the total number of infections are best implemented late in the growth phase to minimize the amount of rebound.

For a population with subcommunities which would have asynchronous outbreaks, similar results hold. Additionally, we find that it is best to target the intervention asynchronously to each subcommunity rather than synchronously across the population.

4.1 Introduction

The Influenza pandemic of 1918 was one of the deadliest epidemics of infectious disease the world has ever seen. In response, many cities introduced widespread interventions intended to reduce the spread. There is evidence [Bootsma and Ferguson, 2007] that some cities which implemented these interventions later had fewer deaths. This seemingly counter-intuitive observation suggests that they were more successful by being slow to respond.

When the 2009 influenza pandemic first arrived outside of Mexico, many schools shut after the first observed infection. Once these schools reopened, and received a new introduction, the remaining susceptible population was almost as large as at the outset, so the resulting epidemic was likely to be nearly as large as the original epidemics would have been. The closure provided increased time to prepare a response and learn more about the disease, but the overall epidemic was very similar to what would have happened without the closure. In contrast, evidence suggests that summer holidays altered the final outcome of the influenza pandemic (at least in the UK), significantly reducing the total number of infections by splitting the epidemic into two smaller peaks [Eames, 2014].

This phenomenon can be explained by noting that epidemics rely on two things to spread: infected individuals and a supply of susceptible individuals. If the intervention is too early, the number infected may fall, but there will be enough susceptibles available that it can re-establish and grow again. When it returns to the original size, the remaining susceptible population will be effectively the same size as it was the first time. Thus, the intervention primarily delays the spread; the resulting epidemic is comparable to what would have been seen before. However, if the intervention occurs once the susceptible population has been noticeably depleted, then the number of infections falls and when the intervention is relaxed, the depleted susceptible population makes the rebound smaller or even nonexistent.

To make this explanation more robust, we note that is well-known that after an unmitigated epidemic, the total number of infections exceeds the number of infections required to achieve the "herd immunity threshold" (the level of immunity required to reduce the effective reproduction number below one) [Cobey, 2020]. We refer to this extra level of infection as the "overshoot". It is a consequence of the

fact that when the effective reproduction number (in absence of intervention) finally falls to 1, the population reaches the "herd immunity threshold" and incidence no longer increases. However, because the epidemic is at its peak, this is the time at which those who have escaped infection so far face the highest force of infection. As the number infected falls, significant transmission still happens and the epidemic overshoots the herd immunity threshold. In the absence of further intervention, the size of the overshoot is determined by the number infected when the herd immunity threshold is reached.

When we think about this in terms of a temporary intervention, the option to minimize the total number of infections becomes clearer. A short intervention that ends with the effective reproduction above one would see a rebound and would see a larger overshoot than a slightly later intervention that ends with the effective reproduction number equal to 1.

This underlies the explanation of [Bootsma and Ferguson, 2007] for why temporary interventions are generally more effective if introduced later in the epidemic (but not too late). Similar, more detailed theoretical results have been found by [Hollingsworth et al., 2011; Anderson et al., 2020]. Most studies of these effects are focused on a single population, and they do not carefully consider the tradeoffs between competing goals of delaying infections, reducing the peak prevalence, or reducing the total size.

In the ongoing COVID-19 pandemic, China introduced drastic control measures very early. These significantly reduced transmission, apparently reducing the effective reproduction number (the number of new infections per infected individual) below one [Kucharski et al., 2020; You et al., 2020], although it took a very long time for new cases to stop. Despite quite significant interventions in Italy in force for a long period of time, the rate of new cases was slow to fall [De Flora and La Maestra, 2020].

Many other places have turned to aggressive control of infection in an attempt to keep transmission suppressed [Iwasaki and Grubaugh, 2020; La et al., 2020; Hoang et al., 2020; Wang et al., 2020; Wu et al., 2020; Baker et al., 2020; Kim et al., 2020].

In places which have have nearly eliminated the disease, the threat of re-emergence requires constant vigilance. In places which have failed to contain transmission, the

pervasive interventions that would be required to get transmission low would impose significant costs through the entire population, and such extensive interventions are unlikely to be maintained long term. Thus policy-makers face challenges about whether or when to implement such restrictive interventions.

Motivated by ongoing decisions facing policy makers for the COVID-19 pandemic, we develop mathematical models which allow us to explore how to time short-term interventions in response to an emerging epidemic. We will refer to these temporary interventions as "one-shot" interventions, meaning that the intervention cannot be maintained indefinitely or repeated. We are particularly interested in how the timing might affect the total fraction infected and the peak prevalence, but we are also interested in the resulting delay of infections.

We must exercise care in determining that a given intervention cannot be sustained. In the initial phase of the COVID-19 pandemic there was significant uncertainty in the fatality rate. With this in mind [Xu, 2020], the tolerance of the population for drastic interventions could be significant. What might appear to be an unsustainable intervention given one set of assumptions about severity may in fact be sustainable under another set of assumptions. We assume perfect information and focus on choosing the time at which a given strategy will be more effective. A separate, but related line of research focuses on whether (and how long) we should hold an intervention in reserve while we learn more about the disease: sometimes the greatest expected benefit comes from learning more before choosing the intervention [Baxter and Possingham, 2011; Ludkovski and Niemi, 2010; Thompson et al., 2018]. For an epidemic that grows quickly (like the early stage of the COVID-19 pandemic), there is effectively no time to learn about the disease before a decision is needed, and so these strategies would not be relevant until strong enough interventions are in place to suppress transmission.

We model an infection spreading in an initially fully susceptible population. We will model the spread within a single well-mixed population and a population made up of several weakly-coupled subcommunities (a *metapopulation*). We will investigate the impact of intervention on the attack rate (the final fraction infected), the peak prevalence, and the timing of infections, and in the metapopulation model we will additionally consider whether it is better to have a synchronized intervention

or to have the intervention timed separately for each subcommunity. The important question of whether disease can be eliminated locally is beyond our scope.

Our goal is not to provide predictions for a specific population, but rather to demonstrate the generic impact of delaying a one-shot intervention, to show its robustness, and to provide intuition and some guiding principles which will apply to more complex scenarios.

Our results have important implications for the ongoing COVID-19 pandemic. If an intervention cannot be sustained for an extended period of time but new interventions or treatments are being developed, it is likely to be best to perform the intervention sooner to delay potential infections until other methods are available to treat or further delay infection (e.g., masks distributed, contact tracing implemented, healthcare capacity increased, therapeutic treatments identified, or even vaccine produced). However, if no other intervention or treatment improvement is likely to emerge, then it is best if the intervention is "held in reserve" until depletion of susceptibles has reduced the effective reproductive number enough that the intervention will have maximal impact on the total number of infections (by preventing the overshoot) or the peak prevalence.

We completed and released this research early in the initial stages of the COVID-19 pandemic (at the end of February 2020) [Di Lauro et al., 2021], but did not immediately pursue it further due to other pressing questions. In the interim, a number of other papers have emerged studying related questions, including [Haushofer and Metcalf, 2020; Gevertz et al., 2020; Della Rossa et al., 2020; Giordano et al., 2020; Morris et al., 2021]. It is clear from much of this work that nonpharmaceutical interventions are an important part of epidemic control. In particular, the timing of an intervention, be it in a single population or over different communities, has a major impact on its effectiveness and overall outcome.

Our results provide insights into ongoing discussions of "circuit-breaker" interventions: in particular, such an intervention is particularly valuable because it can delay infections while other interventions are brought into place, and it can keep the infection count low enough that interventions that cannot scale well can remain effective. However, if there is no significant effort to increase other interventions, then a repeated sequence of such "circuit-breakers" may be needed or the circuit-breakers

should be delayed.

In this paper, we first introduce the mathematical models we use to explore the impact of a one-shot intervention against an infectious disease in a single well-mixed population and in a metapopulation made up of several distinct subcommunities. Then we discuss results from those mathematical models. Finally we discuss the implications of these results. In the Appendix we develop some mathematical theory explaining the mechanism underlying the effect in more detail.

4.2 Methods

In this section we introduce mathematical models for an "SIR" (Susceptible–Infected–Recovered) epidemic in a single well-mixed population and in a metapopulation made up of several subcommunities. We assume that the intervention is initiated at a specific time t^* (typically once the cumulative number of infections $I + R$ reaches some threshold), and that the intervention lasts for a fixed duration D. It reduces the transmission rate by a "strength factor" c. We explore the impact of the threshold, duration and strength of intervention. In the metapopulation model, we compare outcomes when the intervention is implemented in all populations at the same time or in each individual population separately. In both models we measure time in multiples of the typical infection duration.

We will measure the impact of interventions on three quantities of interest:

- the *attack rate* or *final size*: the total fraction infected $R(\infty)$,
- the *peak prevalence* or maximum value of $I(t)$, and
- the *average time of infection*, $\bar{t}$, the average time (or date) at which individuals become infected. It is given by $\int_0^\infty -t\dot{S}\,dt / \int_0^\infty \dot{S}\,dt = \int_0^\infty t\beta IS\,dt/R(\infty)$.

In general the (often conflicting) goals of our intervention are to reduce $R(\infty)$, reduce $I_{\max}$, and increase $\bar{t}$.

The value of minimizing the attack rate is clear as it minimizes the number of infections. The value of minimizing the peak prevalence is highlighted by the struggles that many health systems have faced during the early phase of the COVID-19 pandemic. The importance of increasing the average date at which infections

occur is somewhat less clear. However, early in an epidemic, medical knowledge about the disease, health care capacity, and testing/contact tracing capacities are likely to be limited. Important knowledge about the transmission mechanisms may be missing. In this early stage where knowledge is increasing, any intervention that delays the bulk of infections until later is likely to help increase both the quality of medical care provided and the effectiveness of interventions that may prevent those infections altogether in the future. This may be particularly important for interventions such as contact tracing which take time to put in place and lose effectiveness when there are many infections.

4.2.1 Well-mixed population

To study an intervention in a well-mixed population, we use the standard SIR model [Anderson and May, 1991].

$$\begin{aligned} \dot{S} &= -\beta IS, \\ \dot{I} &= \beta IS - \gamma I, \\ \dot{R} &= \gamma I, \end{aligned} \tag{4.1}$$

where S, I, and R denote the susceptible, infected and recovered fractions of the population with $S + I + R = 1$, and the dot denotes differentiation with respect to time. There are a few important quantities to consider.

- The *basic reproduction number* $\mathcal{R}_0$: The average number of infections an infected individual causes early in the epidemic in the absence of intervention and the absence of any depletion of susceptibles. This is $\mathcal{R}_0 = \beta/\gamma$.
- The *effective reproduction number* $\mathcal{R}_e$: As depletion of susceptibles occurs or interventions are put into place, the number of infections an infected individual causes is reduced. When $\mathcal{R}_e < 1$, the number of infections declines.

By measuring time in multiples of the typical infection duration, we impose that $\gamma = 1$, and so $\beta = \mathcal{R}_0$.

If $\mathcal{R}_0 > 1$ the typical behavior of an epidemic without an intervention is that at $t = 0$ we have $S \approx 1$, I is very small and $R = 0$. As time increases, I and R grow and S decreases. The reduction in S reduces the effective reproduction

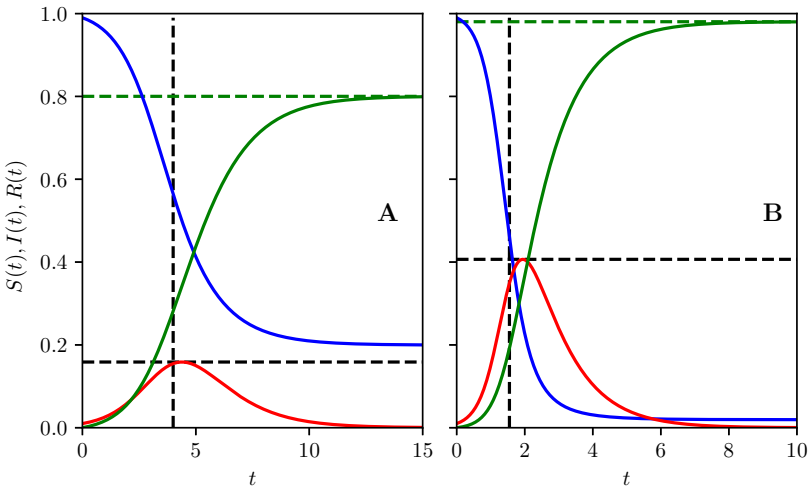

Figure 4.1: The time-evolution of S, I and R for epidemics with no control. $\mathcal{R}_0 = \beta = 2$ (left) and B $\mathcal{R}_0 = \beta = 4$ (right) with $\gamma = 1$ in both. Horizontal and vertical dashed black lines indicate the peak prevalence $I_{\max}$ and average time of infection $\bar{t}$ respectively, while green dashed horizontal lines show the attack rate $R(\infty)$ found by numerically solving $R(\infty) = 1 - S(0)e^{-\mathcal{R}_0 R(\infty)}$.

number: $\mathcal{R}_e = \mathcal{R}_0 S$. Once $S < 1/\mathcal{R}_0$, I begins to fall because recoveries outweigh new infections: $I \to 0$. Some fraction remains uninfected: $S(\infty) > 0$ and $R(\infty) = 1 - S(\infty)$ [Anderson and May, 1991; Ma and Earn, 2006; Miller, 2012]. See Fig 4.1 for typical profiles of S, I, and R in time.

We assume that at some time $t = t^*$, an intervention that reduces the transmission rate is introduced for a duration D. The intervention reduces β by some factor c. So from time $t = t^*$ to time $t = t^* + D$ the transmission rate $\beta = \mathcal{R}_0$ is replaced by $\beta = (1 - c)\mathcal{R}_0$. During the intervention, the effective reproduction number is $\mathcal{R}_e = S(1 - c)\mathcal{R}_0$. After time $t = t^* + D$ the transmission rate returns to $\beta = \mathcal{R}_0$, and $\mathcal{R}_e = S\mathcal{R}_0$.

We will typically assume that t^* is chosen based on the cumulative number of infections $I(t)+R(t)$ crossing some threshold. We choose a monotonically increasing measure $I + R$ because this lets us choose any t^*, which would not be possible if we focused on prevalence (I) or instantaneous rate of infection ($-\dot{S}$).

4.2.2 Weakly-coupled Metapopulation model

We will also investigate the effectiveness of interventions in a metapopulation made up of distinct subcommunities that do not have synchronized epidemics. The most obvious reason for this setup would be geographically separated populations. However there could be stratification by age, religion, ethnicity or socio-economic status. We are particularly interested in whether it is better to time interventions to the dynamics within each subcommunity separately or for the intervention to be synchronized even through the respective epidemics are not.

It is well-known that if the subcommunities have strong enough coupling, the epidemics in all subcommunities are effectively synchronised [Ball et al., 2015; Dickison et al., 2012]. In this case there is little distinction between asynchronous interventions for each subcommunity or interventions synchronized across all subcommunities. Thus to compare the results from synchronized interventions with asynchronous interventions targeted to each subcommunity, we need to explore a population with weak coupling. We use a standard meta-population model [Anderson and May, 1991], allowing most transmission to be within a subcommunity and some cross interactions between the subcommunities.

$$\begin{aligned}
\dot{S}_j &= -\sum_i \beta_{ij} I_i S_j, \\
\dot{I}_j &= \left(\sum_i \beta_{ij} I_i S_j\right) - \gamma I_j, \\
\dot{R}_j &= \gamma I_j,
\end{aligned}$$

where $0 \leq S_i \leq 1$, $0 \leq I_i \leq 1$ and $0 \leq R_i \leq 1$, with $S_i + I_i + R_i = 1$ for all t, represent the fraction of susceptible, infected and infectious and recovered individuals in subcommunity i, where $i = 1, 2, \ldots, N$.

To simplify the presentation, all subcommunities are of equal size. The recovery rate γ is identical for all subcommunities. As before we measure time in multiples of the typical infectious period, so we set γ to 1. The cross-infection between subcommunities is modelled by $B = (\beta_{ij})_{i,j=1,2,\ldots N}$, where β_{ij} represents the rate at which infectious contacts are made from subcommunity i towards susceptible individuals in subcommunity j.

We implement a weak coupling by joining the population in a linear fashion:

population i is only connected to population $(i-1)$ and $(i+1)$. The first and the last populations only connect to the second and the pen-ultimate population, respectively. The entries for the coupling/mixing matrix are generated as follows. On the main diagonal, the β_{ii} values are set to $2+(Unif(0,1)-0.5)$ where $Unif(0,1)$ produces a random number chosen uniformly between 0 and 1. Off-diagonal entries are set to $Unif(0,1)(\beta^*/10)$ ($\beta^* = \max_{i=1,2,\ldots,N} \beta_{ii}$) and represent a scaled and randomised version of the largest entry on the main diagonal. This yields an $\mathcal{R}_0$ above 2, comparable to current estimates for COVID-19 [Kucharski et al., 2020; Liu et al., 2020].

We will use this model to explore whether it is better to implement an intervention in a synchronized fashion across all subcommunities or to implement it in each subcommunity. In particular, we will consider the following scenarios:

- track $I_i + R_i$ in each subcommunity and as soon as $I_i + R_i > T$ for some threshold T, a one-shot control is deployed in the corresponding subcommunity,

- track $I + R = \frac{1}{N}\sum_{i=1}^{N} I_i + R_i$ globally and as soon as $I + R > T$, a one-shot control across all subcommunities, and

- track each subcommunity and deploy the one-shot control is deployed across *all* subcommunities as soon as $I_i + R_i > T$ for the first subcommunity.

One-shot control in a subcommunity is understood to mean reduction in the internal, incoming, and outgoing rates of infection with a factor of $(1-c)$, where $0 \leq c \leq 1$ denotes the intervention strength (we assume that the strength is the same in each subcommunity, and if two communities are both acting, then the movement between them is scaled by $(1-c)^2$). This reduction lasts for a duration D and, as soon as the control is over, the transmission rates for that subcommunity are restored to the starting levels.

In our results, we will present the average outcome of simulations across 100 distinct populations whose mixing matrices are chosen stochastically based on the rules described above.

4.3 Results

We use our mathematical models to explore how the timing of a one-shot intervention can impact

- total attack rate,
- peak prevalence, and
- average time of infection.

These are expected to be good proxies of the total impact on the population or the burden on the health services.

We find that one-shot interventions that begin at the first sign of transmission have the most impact on delaying the epidemic, but they have little impact on the attack rate or the peak prevalence. This is because only a few individuals are infected when the intervention is implemented so not many transmissions are blocked. When the restrictions are lifted, almost as many transmissions end up happening: the disease spreads in an almost fully susceptible population, and its trajectory is very nearly the same, just delayed. In contrast if the intervention is delayed until a non-negligible fraction of the population has been infected it will have more impact on the epidemic's shape.

For the weakly-coupled metapopulation model, the subgroups are likely to have somewhat asynchronous epidemics. In this case it is better to implement the one-shot interventions based on a local threshold rather than a global threshold. If the coupling is stronger, the epidemics are closely synchronized and there is little difference between the strategies.

4.3.1 Well-mixed population

We can think of a strong, but temporary, intervention as dividing the overall epidemic into two phases. We allow an epidemic to spread until the intervention is started. The intervention resets I to a small value (that is, the intervention shifts the epidemic to a new trajectory with a similar S, but a smaller I). Depending on how long the epidemic was allowed to spread prior to the intervention, we have some new value of $S(t^* + D)$. Then a new epidemic happens starting from the new initial

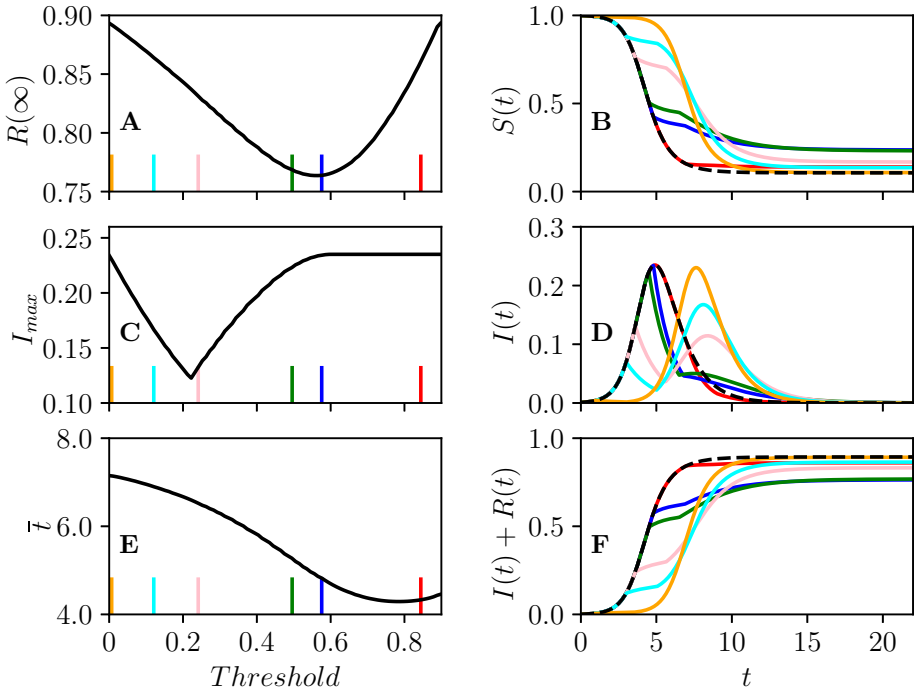

Figure 4.2: Illustration of the impact of one-shot intervention in a population with $\mathcal{R}_0 = 2.5$. The intervention has $c = 0.8$ for a duration of $D = 2$ time units. This intervention is introduced at different times as determined by a range of *Threshold* values. The impact of the threshold ($I+R > Tr$) for implementing the intervention is shown for A the attack rate $R(\infty)$; B $S(t)$; C peak prevalence $I_{\max}$; D $I(t)$; E average time of infection $\bar{t}$; and F plots of $I(t) + R(t)$. In (B,D,F), the no-control case is plotted as a dashed line. The vertical lines in (A,C,E) correspond to the threshold for cumulative infections $I + R$ which yields the intervention leading to the corresponding color in (A,C,E).

state, spreading as if a fraction $1 - S(t^* + D)$ were vaccinated. The longer we allow the first phase epidemic to spread, the smaller the value of $S(t^* + D)$, and so the smaller the second phase will be. An early intervention truncates the first phase, but a later intervention reduces the second phase.

For a well-mixed population we find that the timing of a one-shot intervention has an important impact on the total epidemic. If the intervention is put in place very early, then the impact is to simply delay the epidemic. Because $S(t^*+D) \approx S(0)$ the second phase is effectively like the first phase without an intervention, but delayed. The delay is somewhat larger than D because it takes some time for I to grow back to $I(t^*)$.

Fig 4.2 shows the impact of an intervention in a population with $\mathcal{R}_0 = 2.5$ and an intervention of strength $c = 4/5$ (it prevents 4 of every 5 transmissions), and duration 2 (time units measured in multiples of the typical infection duration). The figure focuses on the impact of varying the threshold value of $I + R$ at which the intervention is introduced.

Fig 4.3 shows how the optimal threshold changes as the parameters of the disease or intervention change.

Impact on attack rate

The impact on the attack rate (the total number infected) can be understood by a mental model of the intervention as a way to shift from the current epidemic trajectory to a new epidemic trajectory with a similar number of susceptibles, but fewer infected (this is made more rigorous in Appendix 4.5.1).

If the intervention is introduced early on, it will have an immediate impact. However, when the intervention is lifted, the epidemic rebounds until the number of infections is the same as the original value. The number susceptible is relatively unchanged, and so a similar epidemic happens with an almost identical epidemic curve once it rebounds, except with a shift to later time. So an early intervention has little impact on the attack rate $R(\infty)$.

In Fig 4.2A we see that if the intervention is introduced later, there is clear improvement in $R(\infty)$, up to a threshold of $I + R$ of 0.6, which is close to where the peak prevalence occurs in the epidemic without intervention. This is because when the epidemic peaks, $\mathcal{R}_e = 1$, and so if we immediately and dramatically reduce the number infected at this point the epidemic quickly dies out.

As new infections do happen during the intervention, this mental model is only an approximation. It can be made more precise by recognizing that to reduce the attack rate $R(\infty)$, the intervention is most effective if it is timed to directly block as many transmissions as possible. So we want to time the intervention to maximize the number of infected individuals present while it is in place (mathematically we want to maximize $\beta \int_{t^*}^{t^*+D} I(\tau) d\tau$ given D, c, and $\mathcal{R}_0$).

Thus the ideal timing to reduce the total number of infections is not at the first hint of transmission (when there are not enough infected individuals to cause many

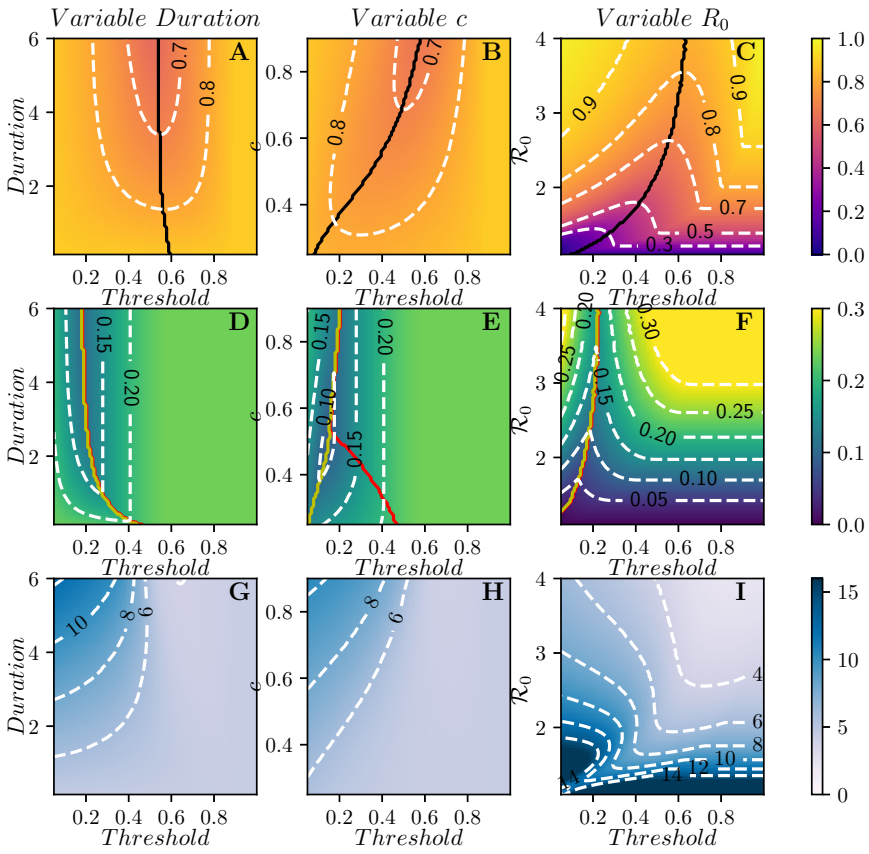

Figure 4.3: Contour plots for $R(\infty)$ (top), $I_{\max}$ (middle) and the mean time of infection $\bar{t}$ (bottom) as a function of parameters for the well-mixed population. We explore different threshold values of $I + R$ for the intervention to start, from a minimum of 0.05 to a max of 0.9. In the first column duration varies from $D = 0.1$ to $D = 6$, holding $\beta = 2.5$ and $c = 0.8$. in the second column, intervention duration is $D = 4$ and c ranges from 0.2 to 0.9. Finally, in the third column, $c = 0.8$ and $D = 4$, and the values of $\beta = \mathcal{R}_0$ vary from 1 to 4. In all cases $\gamma = 1$. In the first row, the black curve denotes the threshold for which $\mathcal{R}_e = 1$ when the intervention completes. In the three regions defined by the two lines in the panels of the second row, the peak prevalence is observed after the intervention has ended (from left to yellow curve), during intervention (area between the curves), or before intervention (from red curve to the end of the figure). Where the two curves align, the prevalence decays as soon as the intervention is implemented and then recovers to the pre-intervention peak.

transmissions), but rather, a little before the peak, and if the intervention is perfect ($c = 1$), then at the peak. This suggests that the more effective an intervention is, the closer we should be to the peak before implementing it. It also suggests that for an intervention of a longer duration, we can implement it somewhat sooner, but not significantly sooner. For a more infectious disease, the need to begin the intervention near the peak implies that the threshold value of $I + R$ will need to be larger (though the time t^* at which it is implemented is smaller).

These predictions are borne out by observations of the first column of Fig 4.3 which shows how the optimal threshold value of $I + R$ for implementing the intervention changes as the strength c, the duration D, or the reproductive number $\mathcal{R}_0$ change. The earliest interventions have the most impact on the average time of infection, while somewhat delayed interventions affect the peak prevalence the most, and later interventions (near the epidemic peak) affect the final attack rate.

Impact on peak prevalence

As in the attack rate case, an early intervention primarily delays the epidemic curve. It does not significantly alter the shape. Thus the peak prevalence remains effectively the same unless the intervention is delayed until S is noticeably depleted.

If the susceptible population has been sufficiently depleted prior to the elimination of the intervention, then once the intervention is stopped, the epidemic rebound will be muted. Moving the intervention later makes the rebound smaller still. However, it means that the number of infections prior to the intervention is larger. There comes a threshold at which the phase before and the phase after the intervention have the same maximum. This is the time that minimizes the peak prevalence [Morris et al., 2021]. Delaying the intervention past this value results in a larger pre-intervention peak, while doing it sooner results in a larger post-intervention peak.

Fig 4.2C shows the optimal threshold to reduce the peak prevalence occurs sooner than to reduce the attack rate. We can understand this intuitively, because for optimizing peak prevalence a moderate rebound is less of a concern than for optimizing the attack rate. For the purpose of reducing peak prevalence, Fig 4.2D shows that the optimal time to introduce the intervention is when the current prevalence

matches the peak prevalence that would occur once the disease rebounds.

We can crudely estimate the threshold necessary for minimizing the peak prevalence. If we know a population's reproductive number $\mathcal{R}_0$ and its initially immune fraction R^* and susceptible fraction $S^* = 1 - R^*$, we can determine the peak prevalence (There is an analytic formula for peak prevalence $1 - \frac{1}{\mathcal{R}_0} - R^* - \frac{\ln(S^*\mathcal{R}_0)}{\mathcal{R}_0}$ but for our purposes we just need to recognize that $\mathcal{R}_0$, R^* and S^* are sufficient to determine it). In the limit of a very long ($D \to \infty$) and strong intervention ($c \to 1$), at the end of the intervention $S(t^* + D) \approx S(t^*)$ and $R(t^* + D) \approx I(t^*) + R(t^*)$. If D is long, but c is small enough that we cannot ignore transmissions occurring during the intervention, then we need to correct for the fact that $S(t^* + D)$ may be somewhat smaller than $S(t^*)$. Accounting for these transmissions further reduces the size of the second peak.

We can use this to estimate when $I(t^*)$ will approximate the rebound. So long as duration is not too long, and c is not too much smaller than 1, this is not strongly dependent on duration or c. This explains why the optimal threshold for peak prevalence does not vary much in Fig 4.3B and 4.3E.

It is worth highlighting that in related recent work [Morris et al., 2021] showed that the penalty for making a small error in the *timing* of the intervention is larger if it is too late compared to too early. As we see in our figure, the error as a function of the threshold $I+R$ appears roughly symmetric, but because the optimal intervention time often occurs while the epidemic growth is increasing, this means that being a little too late means a larger error in $I + R$ than being a little too early.

Impact on timing of infections

The impact on the timing of an emerging epidemic is an additional factor that plays an important role. If we anticipate rapid development of new treatments or interventions, then this may be more important than reducing the anticipated peak prevalence or total fraction infected.

As we noted for the attack rate and the peak prevalence discussions, for very early interventions (for which $I + R$ is very small), the entire epidemic curve shifts in time. Of course if the disease is eliminated locally which is more likely with a small threshold, then the next peak depends on frequency of reintroduction which

we do not consider. However, as the threshold increases and we start to see an impact on the final attack rate $R(\infty)$ and peak prevalence $I_{\max}$, we also see an additional impact on the average time of infection. Unlike the other targets, a later intervention tends to have an decreased impact because more of the infections have occurred earlier.

In a real-world context, we anticipate that the model may overstate the delay from a very early intervention if there is significant transmission outside the population of interest. In a setting where the disease is spreading outside the population, the reduction of infections within the population during the intervention may be immediately negated by new transmissions from outside, which are likely to be increasing. So the effect to delay the epidemic is largest if accompanied by reduction in transmissions from outside. However, in a setting where the disease is not well-established outside the population (as occurred in China early in the COVID-19 pandemic), or travel from outside can be restricted, a major effort at early time may significantly delay the eventual epidemic.

4.3.2 Weakly-coupled metapopulation model

We now consider a more realistic population which consists of coupled subcommunities, effectively a metapopulation model. We again consider one-shot interventions that either target the entire population at once (synchronous interventions) or that target individual subcommunities at different times (asynchronous interventions). If they were strongly coupled, the epidemics would be synchronous [Ball et al., 2015; Dickison et al., 2012]. So the single well-mixed population results would carry over. Our focus is on weakly-coupled subcommunities.

A typical plot of the prevalence level in each subcommunity is shown in Fig 4.4 in the absence of intervention. The epidemic starts in subcommunity two but it then spreads to the others. The entries of the cross-infection/mixing matrix are generated following the description in Section 4.2.2, and the specific mixing parameters are given in the Appendix.

As before we consider the impact of intervention on attack rate, peak prevalence, and peak timing. The overall effect of interventions is qualitatively similar to that of the single-population model. However, we find that asynchronous interventions

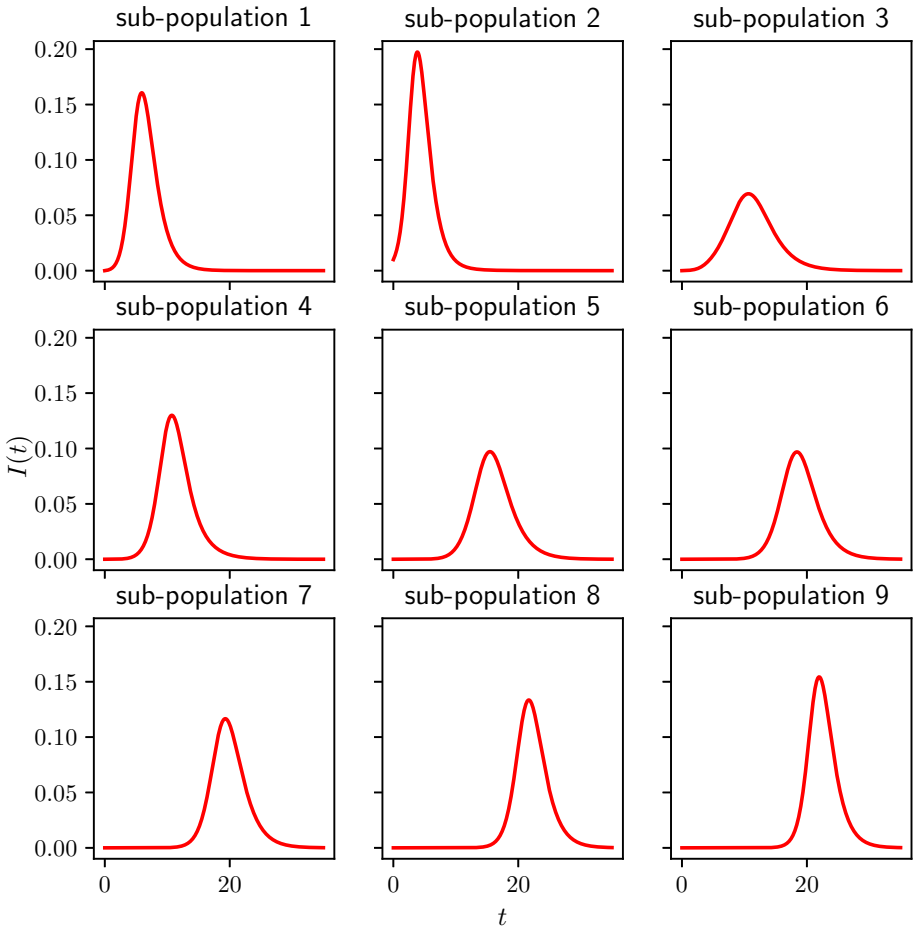

Figure 4.4: Example of an epidemic spreading across 9 subcommunities with different contact rates (see the Appendix 4.5.2 for the precise mixing matrix B). The epidemic starts from subcommunity 2 and it is run for $T = 35$ units of time. $\gamma = 1$ for all subcommunities. With no control the attack rate or final epidemic size is 0.744.

that separately target each subcommunity significantly outperform synchronized interventions that begin when either the first subcommunity reaches a threshold or the global infection crosses a threshold.

For synchronized interventions, the overall impact is smaller, and the best outcomes are not driven by the actual threshold value. Rather they result from the intervention being timed to have significant impact on multiple communities, or optimally delaying the spread between communities. Consequently, the ideal times for this will depend on the parameters for between-community transmission, and are likely to be population-dependent.

Because the epidemics may not be synchronized across subcommunities, when a synchronous intervention is applied some may have already completed their epi-

demic, while others have not yet begun. Interventions that are based on the first population to reach a threshold may not be valuable if the particular intervention is most effective if it disrupts patterns that do not appear until the first subcommunity has effectively completed its epidemic.

In our results, we consider 100 simulated populations consisting of 9 subcommunities, whose contact structure is generated from the random process described in Section 4.2.2. For most results we present only the average behavior. We note that this aggregation may hide important behavior from individual simulations [Juul et al., 2020]. However, except where noted, our averaged results are qualitatively similar to what is happening in the subcommunities. Where we look at an individual simulation, we use the specific population of Fig 4.4.

Impact on attack rate

Our primary observation about the attack rate is that interventions acting at different times for each subcommunity are substantially more effective than synchronized interventions.

The smallest values of the attack rates are achieved when control acts independently in each subcommunity meaning that as soon as $I_i + R_i$ crosses a threshold, the one-shot control is switched on in subcommunity i. This is done independently of whether the efficacy or duration of control is kept fixed, while the other is varied, see Fig 4.5A and 4.5D. Typically, as in the case of a single population, there seems to be a clear optimal threshold value which leads to the smallest attack rate. Applying the control too early or too late leads to higher attack rates. Fixing the threshold value and increasing the duration of control, see Fig 4.5A, or the strength of control, Fig 4.5D, leads to smaller attack rates. Both strength and duration of control have no significant impact on the attack rate if the intervention is too early or too late.

Fig 4.6 shows how the best one-shot control works when the optimal threshold for fixed control efficacy and duration is implemented. As expected, this plot confirms that intervention happens close to the peak of the epidemic in each subcommunity so secondary waves of infection are heavily suppressed.

The impact of the synchronized intervention based on the global level of $I+R$, see Fig 4.5C and 4.5F, or on the first subcommunity to reach a threshold, see Fig 4.5B

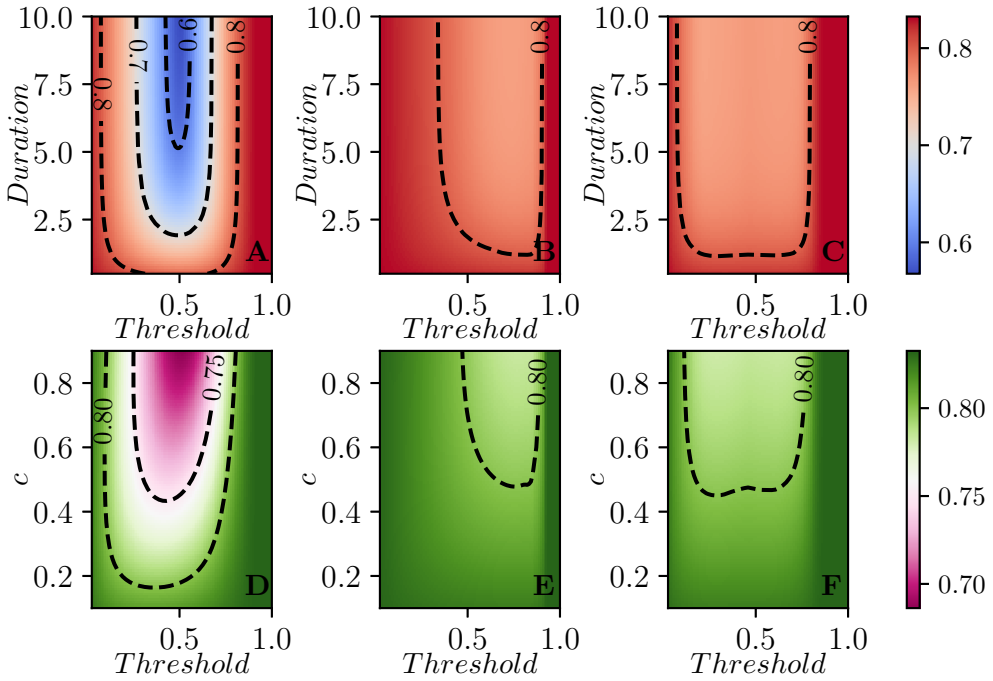

Figure 4.5: Contour plots showing the average attack rate (final epidemic size) over 100 simulated populations for each set of parameter values. In the first row c is fixed and the duration of control varies on the vertical axis, while in the second row duration is fixed and c varies. Each column corresponds to one of the three strategies: A,D intervention in each subcommunity based on that subcommunity reaching a threshold, B,E global intervention when the first subcommunity breaches the threshold, and C,F global intervention at global threshold for a population consisting of 9 subcommunities. In each plot, the $x-axis$ shows the values that the threshold for intervention can take (from a minimum of 0.05 to a maximum of 0.8). In the first row $c = 0.8$ is constant, while the duration of control varies from a minimum of $T = 1$ to a maximum of $T = 10$. On the second row instead, the duration of control is kept fixed at $T = 2$, and the values of c varies from $c = 0.1$ to $c = 0.9$. The recovery rate is $\gamma = 1$ for all subcommunities. In all cases, if the threshold is set too large the intervention is never implemented. The two synchronized interventions can be approximately mapped to one another by noting the largest $I_i + R_i$ at the time the global $I+R$ reaches a given threshold. The subcommunity threshold gives more resolution at small values while the global threshold gives more resolution at large values.

and 4.5E, are much smaller than asynchronous interventions. This is because when it is implemented in the synchronous case, some communities have already completed their epidemic while others have not yet begun. So there is less overall impact (see the asynchrony in Fig 4.4).

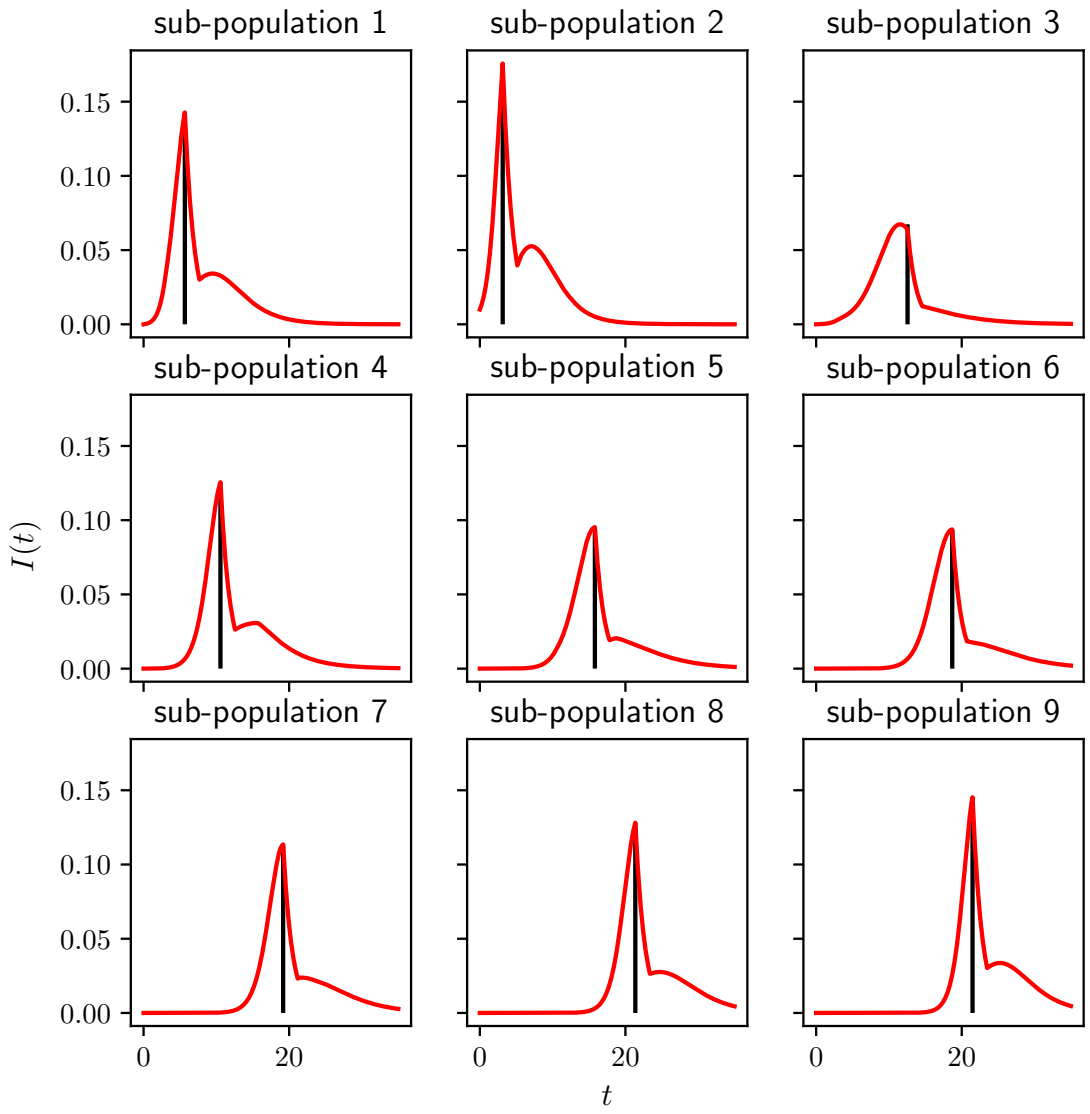

Figure 4.6: Illustration of best control strategy (i.e. smallest attack rate) (controlling subcommunities individually but using the same threshold for each) when efficacy and duration of control are fixed at $c = 0.8$ and $D = 2$, respectively. It turns out that the optimal threshold is close to (0.4). This combination represents the point $(0.4, 2)$ in Fig 4.5A, or equivalently the point $(0.4, 0.8)$ in D. With this strategy, we find that $R(\infty)$ goes from $R(\infty) = 0.75$ to $R(\infty) = 0.63$. If we increase control duration from 2 to 10 we would achieve a further reduction to $R(\infty) = 0.44$. The vertical black lines show the onset of control.

When the intervention is based on the first time a subcommunity crosses a threshold, we find that the optimal thresholds are at relatively large values. This suggests that the value of the synchronized intervention comes from disrupting transmission when the disease is spreading in multiple subcommunities.

Under the synchronous intervention scenario, we also see some surprising behavior where there are multiple local maxima for the specific metapopulation used in Fig 4.4 (not shown in Fig 4.5). This effect is because the timing aligns with different outbreaks. If we intervene at one time, we may have a big impact on one subcommunity, and if we miss that window, it is best to wait until another subcommunity begins to have an outbreak. This effect disappears in the aggregated data of Fig 4.5 because the specific ideal timing is a consequence of the randomly chosen parameters of each population.

Impact on peak prevalence

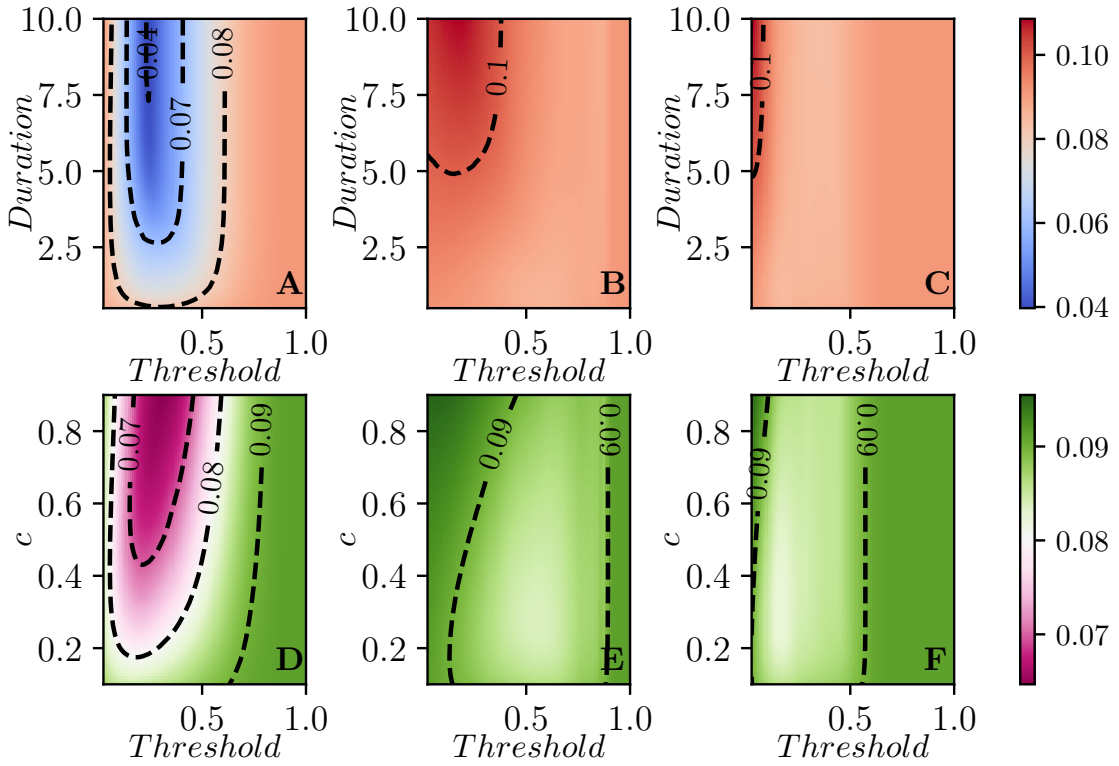

Figure 4.7: Contour plots of the peak prevalence $I_{peak} = \max_t \left(\frac{1}{9}\sum_i I_i(t)\right)$, averaged across 100 simulated populations each with 9 subcommunities. Control strategies and setup the same as in Fig 4.5.

Here we look at the effect of the intervention on the peak prevalence, that is the maximum value of $I(t) = \frac{1}{N}\sum_i I_i(t)$ during the time course of the epidemics. As with the attack rate, our primary observation for the peak prevalence is that it is significantly reduced by targeting based on the individual subcommunity.

Fig 4.7 shows the average of the peak prevalence across the same 100 populations

as Fig 4.5. Perhaps not surprisingly, Fig 4.7 is qualitatively similar to Fig 4.5. The most impact is through having interventions occurring when the individual populations reach a threshold. The optimal choices for intervention come earlier in the epidemic. We still observe that if the intervention is too soon or too late then there is no significant reduction in peak prevalence.

In Fig 4.7A,D the optimal threshold for intervention is relatively early, this is in line with the trend observed in Fig 4.2 for the single population case. We should wait until some immunity builds up before intervening, so that the rebound in each population is muted.

For our two synchronized strategies, the effectiveness is much less, because the overall peak prevalence is related to how the individual subcommunities' peaks align, and different details of intervention timing, combined with the random parameters of the simulation, can make the individual peaks align or not.

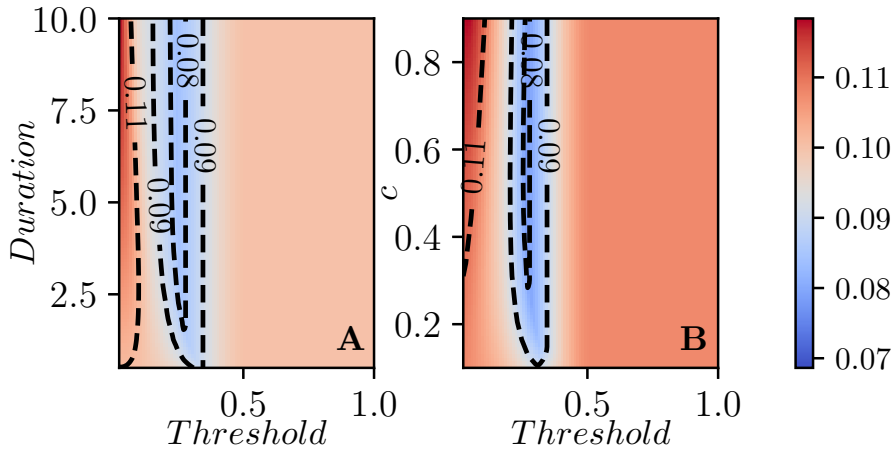

Figure 4.8: Contour plots of the peak prevalence, I_{peak}, that is the maximum value achieved by $I(t) = \frac{1}{N}\sum_i I_i(t)$ during the time-course of the epidemic for the population used in Fig 4.4, with the intervention occuring when the global infection count reaches a threshold (as in C and F in earlier figures).

Interestingly, if we look at an individual simulations, there are thresholds which yield significantly larger improvements in the peak prevalence than we see in the aggregated data. This is because in each simulated population, the relative timing of the epidemics subpopulations depends on the system parameters. In a weakly-coupled metapopulation model with relatively few subcommunities, the global peak

prevalence is likely to occur when multiple subcommunities happen to be aligned. We observe this in Fig 4.8 which shows the equivalent of Fig 4.7C and 4.7F for the same population as in Fig 4.4 (described in the appendix). If we do the same calculation for other populations chosen from the same distribution, we find that the optimal time can shift dramatically. This is because in a weakly-coupled metapopulation model with relatively few subcommunities, the global peak prevalence is likely to occur when multiple subcommunities happen to be aligned. This is highly sensitive to parameters, and so the optimal intervention time will vary between different realizations of the population.

Impact on epidemic timing

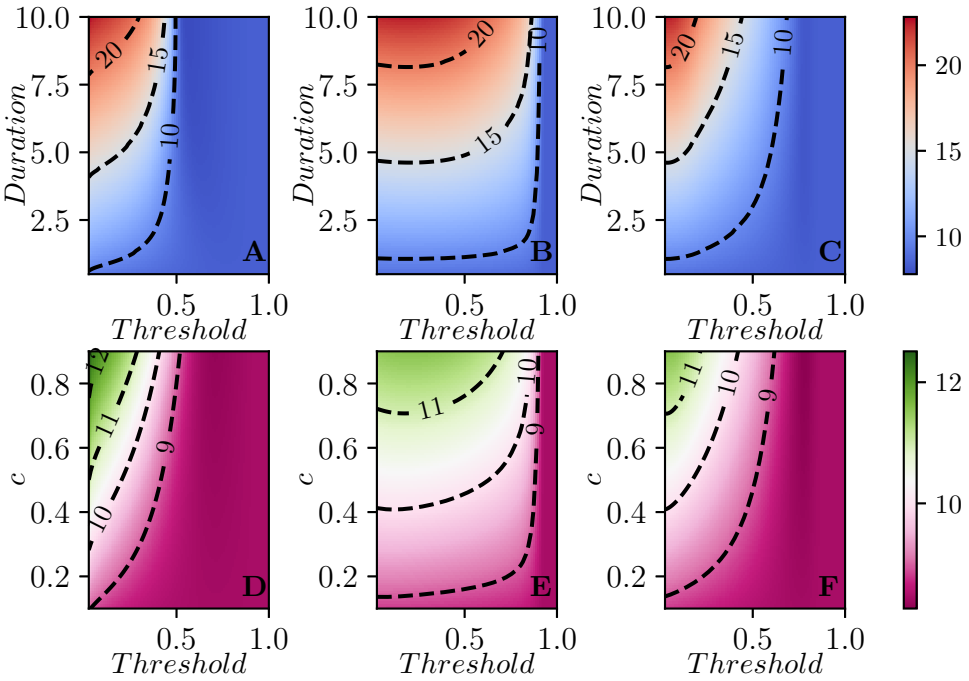

Figure 4.9: Contour plots of the global mean infection time, defined as $T = -\frac{1}{R(\infty)} \int_0^\infty t \sum_i \frac{dS_i}{dt} dt$, averaged over 100 simulations. In terms of control strategies and parameter values the same setup as in Figs 4.7 and 4.5 are used.

When we investigate the average time of infection, we see in Fig 4.9 that targeting the intervention at each subcommunity is again the most effective. In general the interventions need to be implemented very early in the epidemic.

Most of the impact comes from slowing the epidemic in the initial subcommunity. A delay in the initial place of introduction results in a delay in all subcommunities.

Once the intervention is no longer in place in the initial subcommunity it begins to grow and spill over into other communities. If other subcommunities wait until then to begin their response, they gain some benefit. However, once they stop, they face rapid reseeding from the initial subcommunity. So the main benefit comes from the initial subcommunity's actions. When we use a synchronized intervention, the effect is somewhat smaller, but it is not significantly smaller.

In fact, most of the benefit comes from the initial subcommunity engaging in preventative measures. There is relatively little impact on the average time of infection to be gained from the other subcommunities acting early, unless they can maintain a very small spillover rate for a long period through extensive travel restrictions or similar interventions. This suggests that significant benefit may come from a hybrid strategy which focuses on delaying infections out of the initial subcommunity while other locations focus their interventions on optimizing peak prevalence or attack rate.

4.4 Discussion

We have considered the impact of a single one-shot limited duration intervention on the spread of an infectious SIR disease, in both a single well-mixed community and in a weakly-coupled metapopulation model.

We have found that in a single well-mixed population, an intervention at the first hint of infection is best for delaying infections. An intervention that waits until the epidemic is well-established but still well short of the peak is best to reduce the peak prevalence of the epidemic. An intervention (whose duration would be D) that starts a little less than D units of time before the peak would otherwise be reached is best to reduce the total number of infections.

In a weakly-coupled metapopulation model, we find qualitatively similar results. If the goal is to reduce the total number of infections or to reduce the peak prevalence, the best strategy times the interventions asynchronously. The intervention is applied when the subcommunity reaches a threshold rather than being synchronized to when the average reaches a threshold or the first subcommunity reaches a threshold. This is because once sufficient infection has entered a community, the

dynamics are driven by internal transmissions rather than external introductions.

For delaying the average time of infection, the most important detail is that the subcommunity with the first introduction responds as quickly as possible. Whether the other subcommunities respond immediately or delay their response until more infections are present within the subcommunities would have a smaller effect.

4.4.1 Limitations

Our results are somewhat limited by the assumptions we have made to produce a tractable problem.

We assume that behavior responds immediately to changes in interventions. In reality, behavior may change prior to an intervention being implemented. Additionally adherence may drop as the intervention continues, and some adherence to the intervention may remain even once the intervention is removed.

The assumption that the individuals are largely homogeneous may lead to pessimistic predictions of when the herd immunity threshold is reached. In the presence of significant population heterogeneity, there is evidence suggesting that the herd immunity threshold would be reached earlier, and the epidemic could proceed significantly faster [Gomes et al., 2020; Britton et al., 2020]. Our qualitative predictions remain robust, but the timings would need to move sooner.

We must think critically about what constitutes a one-shot intervention. Whether an intervention can be maintained may depend on context. Early estimates of case fatality rate (not to be confused with infection fatality rate) of COVID-19 ranged from 0.7% in China outside of Hubei province to around 2% in much of the world, to around 5.8% in Wuhan [World Health Organisation, 2020]. These estimates were affected by the proportion of cases identified (leading to uncertainty in the denominator), and whether the health system was over capacity (which would increase the death rate leading to uncertainty in the numerator). True infection fatality rates appear to lie between 0.5% and 1%, with many estimates closer to 1% than 0.5% [Meyerowitz-Katz and Merone, 2020b; Ward et al., 2020; Pastor-Barriuso et al., 2020]. With such high fatality rates, our tolerance for drastic interventions should increase. Thus an intervention that would be considered one-shot for the 2009 H1N1 pandemic which had a significantly lower fatality rate [Wong et al., 2013] might be

considered sustainable for the COVID-19 pandemic.

In deciding whether an intervention is sustainable, policy makers could formulate an answer to this question: "Assume you impose the intervention now, and infection rates remain the same or higher in the future, and would increase if they were dropped, would you be willing to maintain the intervention in place?" If so, then the intervention is sustainable. If the answer is "no", and the intervention will be abandoned at some future time regardless of the new infection profile, then this is a one-shot intervention, and it should be held in reserve until it will have maximal impact.

We have ignored logistical challenges that might be associated with implementing the intervention separately for each subcommunity. On a large scale (e.g., states within a country or cities within a state) we anticipate that this is logistically feasible, while on a small scale (e.g., suburbs in a city) it is more likely that the epidemics will be synchronized and this benefit is small compared to the logistical challenges.

4.4.2 Policy Implications

Our observations have a number of important policy implications for an epidemic which is sufficiently established that elimination is not a goal. Primarily:

- To reduce the attack rate of an epidemic, a one-shot intervention should be introduced shortly before the epidemic peak.
- To reduce the peak size of an epidemic, the intervention should be late enough to allow significant immunity to develop, but early enough to allow a substantial rebound after the intervention.
- To delay infections as much as possible, the intervention should be implemented early on.
- In a population made up of many weakly-coupled subcommunities, interventions should be asynchronous.

Because they require different timings, the three goals we have considered are in conflict. The benefits of reducing the total number of infections are clear, and if health care capacity is threatened, the benefits of minimizing the peak prevalence

are also clear. It is less obvious that delaying infections may help. However when there is an expectation that improved treatments and improved inteventions may be developed, a delay is likely to be the best available option.

An additional benefit of delaying the epidemic is observed when we have an intervention, such as testing & tracing, which does not scale well. As the number of infections grows beyond the testing capacity, the effectiveness per infection goes down. At low infection levels these interventions may be enough to suppress transmission. However, if levels get too high, then a quick intervention that causes infections to drop may be more effective than an intervention that waits until the optimal time to minimize the size.

Although we have focused on three distinct goals which lead to different optimal strategies, the ideal goal is likely to be a combination of these effects. So the timing will need to respond to different pressures. If, for example, the goal is to keep the peak prevalence below a certain value while minimizing the maximum prevalence, then the ideal strategy would let as much immunity develop in the population as possible before the prevalence limit is reached and then intervene (this is of course only feasible if there is a time that can keep prevalence below that level). Delaying the intervention until this point would mean that the second peak would be lower. If the goal is instead to keep the peak prevalence below some level while maximizing the delay of infections, then the intervention would be sooner and timed so that the second peak would reach the target prevalence.

We finally note that in a population made up of weakly-coupled subcommunities whose epidemics will not be synchronized, the ideal intervention might be to react strongly and immediately in the first subcommunity where the infection begins to spread. This can provide protection to the other communities and significantly delay the spread. Once it spreads beyond that initial subcommunity then the focus may turn to minimizing the peak prevalence or the attack rate.

4.5 Mathematical Analysis

In this section we provide mathematical analyses of the single population model to support our results for reducing attack rate and peak prevalence.

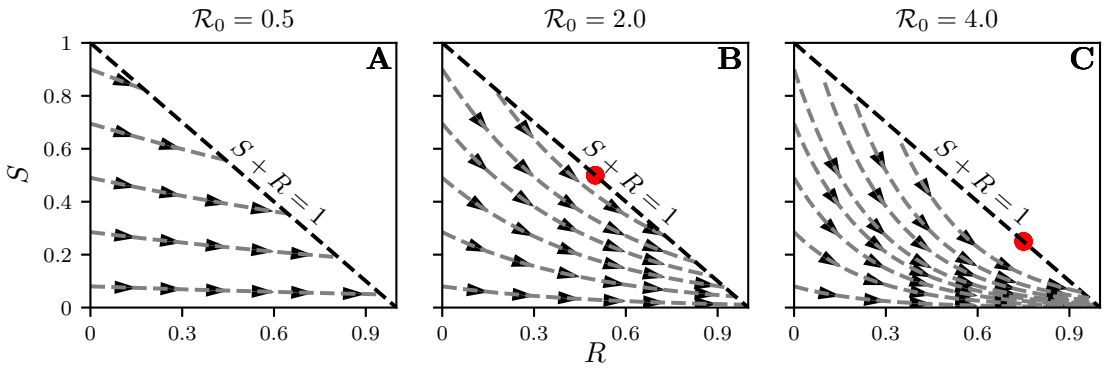

Figure 4.10: We plot $S(t)$ versus $R(t)$ for $\mathcal{R}_0 = 0.5$ A, 2 B, and 4 C. For given $S(t)$ and $R(t)$, the proportion infected is $I(t) = 1 - S(t) - R(t)$, which equals the vertical or horizontal distance from the point $(R(t), S(t))$ to the line $S + R = 1$. The curves and arrows show how a solution to System (4.1) evolves in time. At points $S > 1/\mathcal{R}_0$ (which occurs only for $\mathcal{R}_0 > 1$) curves move farther from the diagonal, representing an increase in I. Note that the velocity a curve is traversed varies depending on location, and goes to zero close to $S + R = 1$. Red dots in B–C indicate the point $(S = \frac{1}{R_0},\ I = 0,\ R = 1 - S)$.

4.5.1 A phase-plane based analysis

Because $S+I+R = 1$, we can fully specify the current state and the future dynamics by knowing S and R, in which case $I = 1 - S - R$. It will be useful to use this to explore the dynamics of an epidemic and the impact of an intervention.

In Fig 4.10 we show how $S(t)$ and $R(t)$ evolve together in time for three values of $\mathcal{R}_0$ (0.5, 2, and 4) and for several different initial conditions. For a given point $(S(t_0), R(t_0))$, the value of $I(t_0)$ is given by the horizontal (or vertical) distance to the diagonal line $S + R = 1$.

In the figure, we can see that if $S > 1/\mathcal{R}_0$ (which is only possible if $\mathcal{R}_0 > 1$), then the horizontal distance from the curve to the $S+R = 1$ line is increasing as the curve moves forward. In other words, I is increasing. Once $S < 1/\mathcal{R}_0$, the distance decreases and eventually goes to 0.

Using these curves, we can investigate the impact of an intervention, as shown in Fig 4.11. We follow S and R along a curve. When we turn on the one-shot intervention at time t^*, it no longer follows the original curve. Instead the curve temporarily follows the paths we would find for $(1 - c)\mathcal{R}_0$, starting from the point $(R(t^*), S(t^*))$. It follows this curve until reaching $(R(t^* + D), S(t^* + D))$ when the

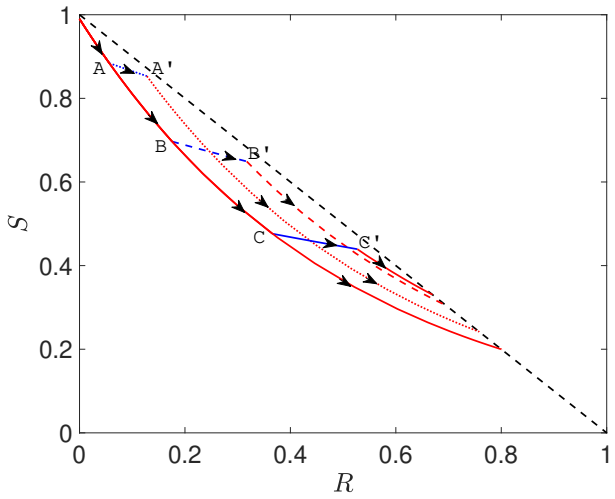

Figure 4.11: (S,R) phase portrait (arrows indicate growing time) based on an SIR model in a single population with $\beta = 2$, $\gamma = 1$ (giving $\mathcal{R}_0 = 2$) and initial condition $I(0) = 0.01$. The plot shows a trajectory with no control (continuous red line) as well as three other trajectories where $\beta = 0.5$ for a time period of length $D = 2$ but with the intervention setting in only once $I + R$ goes past 0.1 (partially dotted line), 0.3 (partially dashed line) and 0.5 (continuous broken line), respectively. Control for the three different scenarios sets in at the points denoted by A, B and C and control ends at A', B' and C', respectively.

intervention is halted. It then follows the curves for the original $\mathcal{R}_0$, but starting from this new point. Note that there is a point $(R, S) = (1 - 1/\mathcal{R}_0, 1/\mathcal{R}_0)$ at which separates the points on the line $R + S = 1$ from which an epidemic could start from the points at which epidemics finish. The closer a curve is to this point, the smaller the attack rate.

So for $\mathcal{R}_0 = 2$, a temporary intervention gives us a way to move from one curve in the $\mathcal{R}_0 = 2$ plot to another. We see this in Fig 4.11. The timing of the intervention determines which of the curves the system lands on.

In this context the goal of reducing the attack rate is equivalent to ensuring that the intervention shifts the curve to a curve as close as possible to $(R, S) = (1 - 1/\mathcal{R}_0, S = 1/\mathcal{R}_0)$. Reducing the peak prevalence is equivalent to ensuring that the curve remains as close as possible to the line $S + R = 1$.

Attack rate

If our goal is to minimize the number of infections, we accomplish this by having the curve $(R(t), S(t))$ land on a curve that is as close to $(R, S) = (1 - 1/\mathcal{R}_0, 1/\mathcal{R}_0)$ as possible given the constraints on the intervention.

Typically we have to wait until the curve has moved closer to the desirable curves before implementing the intervention. Implementing the intervention early, see the dotted line in Fig 4.11 means that at the end of the intervention there is still a large pool of susceptibles which are at risk of becoming infected. Crossing from A to A' simply puts the epidemic on a slightly different trajectory but the attack rate is very close to the case with no control. An intervention at a later stage, see dashed line, improves the final outcome resulting in an attack rate that is smaller when compared to the case of no control. Finally, the continuous broken line shows an almost optimal intervention with a further small reduction in the attack rate.

In general, the intervention that will get us closest to the optimal value occurs when the original curve is close to, but has not yet reached, the largest value of I, which occurs when $S = 1/\mathcal{R}_0$. As the effectiveness of the intervention increases, the curves it follows during the intervention become more horizontal. For very effective interventions, this suggests we should wait until very close to the epidemic peak, while for less effective interventions (which will slope downwards more), we will want to implement them somewhat sooner.

Peak prevalence

For peak prevalence, the goal is to keep the curve as close as possible to $S + R = 1$. The longer we wait to implement the intervention, the closer the final curve is to $S + R = 1$, but the farther the original curve moves from the line. With this in mind it becomes clear that the optimal t^* to reduce peak prevalence is smaller than the optimal value to reduce attack rate.

4.5.2 The mixing matrix

The cross-infection between subcommunities is modelled by $B = (\beta_{ij})_{i,j=1,2,\ldots N}$, where β_{ij} represents the rate at which infectious contacts are made from subcommunity i towards susceptible individuals in subcommunity j.

We implement a weak coupling by joining the population in a linear fashion: population i is only connected to population $(i-1)$ and $(i+1)$. The first and the last populations only connect to the second and the pen-ultimate population, respectively. The entries for the coupling/mixing matrix are generated as follows. On the main diagonal, the β_{ii} values are set to $2 + (Unif(0,1) - 0.5)$. Off-diagonal entries are set to $Unif(0,1)(\beta^*_{ii}/10)$ ($\beta^* = \max_{i=1,2,\ldots,N} \beta_{ii}$) and represent a scaled and randomised version of the largest entry on the main diagonal. This yields an $\mathcal{R}_0$ above 2, comparable to current estimates for COVID-19.

We choose random values because the coupling parameters determine the timing of epidemics in the subcommunities. The optimal timing of interventions for a given realization may depend on how well-synchronized the epidemics are.

To avoid having our results heavily influenced by the particulars of a single realization of the population, we use randomly assigned mixing parameters in multiple simulations. For most of our results we aggregate over 100 distinct simulated populations. However, in Figs 4.4, 4.6, and 4.8 we use a single realization of the metapopulation model. For this we take the mixing matrix

$$B = \begin{pmatrix} 1.917 & 0.059817 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.062024 & 2.2203 & 0.03117 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0094413 & 1.5001 & 0.0043357 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.070055 & 1.8023 & 0.076213 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.01146 & 1.6468 & 0.049723 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.02948 & 1.5923 & 0.054573 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.07709 & 1.6863 & 0.045981 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.015262 & 1.8456 & 0.015462 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.098061 & 1.8968 \end{pmatrix}. \tag{4.2}$$

4.5.3 The impact of aggregation

In our figures showing the impact of interventions in the weakly-coupled metapopulation model, we showed the average across many realizations. However, the synchronization of epidemics in the individual subpopulations would vary depending on the mixing parameters. Most notably, in the global interventions, the overall effectiveness will depend on the relative timing of the epidemics in the subpopulations. This is seen in Fig 4.12. As a result of this, we must have caution in interpreting the optimal thresholds for global interventions from the averaged figures. See also [Juul et al., 2020].

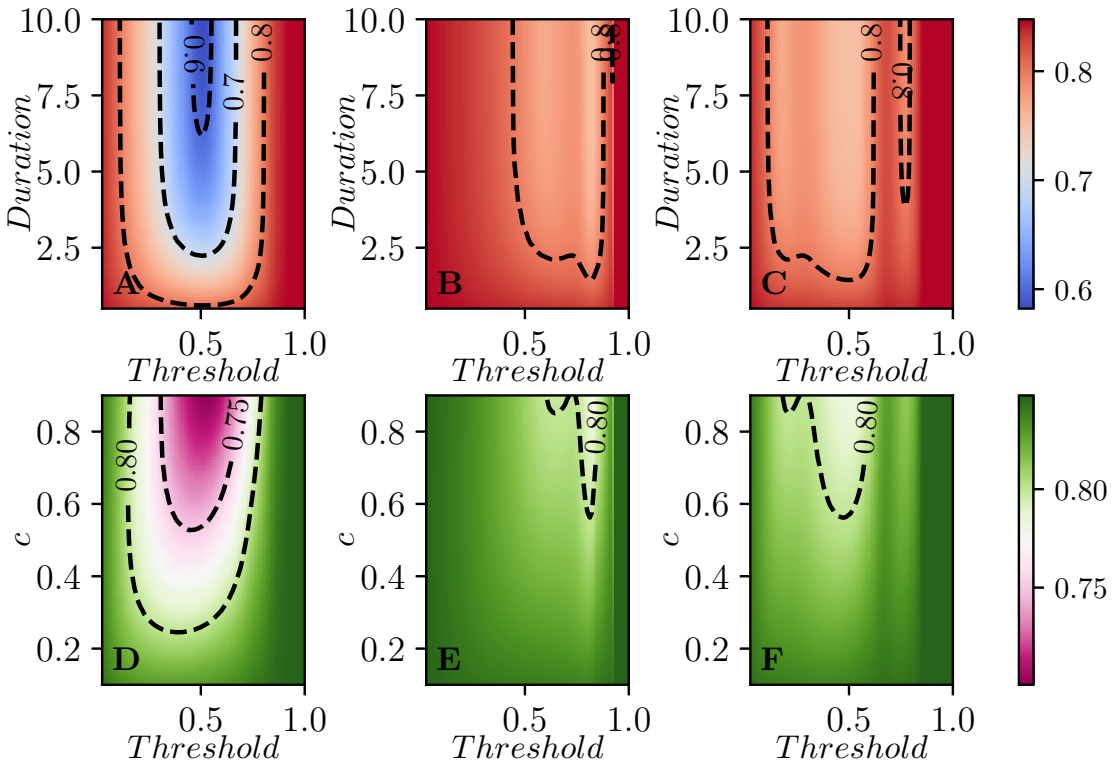

Figure 4.12: Contour plots of R_∞ for a particular realization of the mixing matrix. In C we see that there can be multiple peaks in the optimal time [this is also present in B but it is too small to see]. This is because the effectiveness of the interventions depend on the timing of epidemics in the different subpopulations and these are asynchronous.

Chapter 5

The impact of network properties and mixing on control measures and disease-induced herd immunity in epidemic models: a mean-field model perspective

5.1 Introduction

The recent emergence of SARS-CoV-2 and the associated disease COVID-19 has had worldwide impact. Many cities have had large outbreaks and brought them under control through major interventions. Once those interventions are lifted, in absence of effective vaccination, a homogeneous model of infection spread would predict that as long as less than $1-1/\mathcal{R}_0$ of the population was infected, there is always a threat of resurgence.

Despite large epidemics, cities such as New York remain well below the threshold expected to be be required to achieve herd immunity [Stadlbauer et al., 2020].

To avoid the significant economic and health costs associated with continued interventions, it is natural to consider the so-called "herd immunity" strategy. This strategy allows infection to spread with restrictions in place so that the outbreak finishes and interventions are lifted when the herd immunity threshold is reached.

Typically in an uncontrolled epidemic the herd immunity threshold is reached at the peak of the epidemic, and many additional infections occur as the outbreak slowly wanes. The additional infections are sometimes referred to as the *overshoot* [Handel et al., 2007]. By calibrating the intervention so that there are no (or almost no) infections when the herd immunity threshold is reached, interventions can be removed with minimal overshoot [Di Lauro et al., 2021; Morris et al., 2021].

The severity of the epidemic in many places whose seroprevalence is still very low has led many to suggest that the herd immunity strategy is not tenable. However, recent papers [Britton et al., 2020; Gomes et al., 2020] suggest that immunity acquired through infection may be distributed through the population in such a way as to achieve herd immunity at a lower fraction affected than a homogeneous model would predict. This is because the initial wave of infections preferentially affects those most at risk. Thus it acts like a targeted vaccination, removing the people who are most likely to transmit infection from the susceptible pool. Generally, the time course of a real epidemic involves the following stages: (i) short period of unconstrained transmission, (ii) significant control or lockdown, and (iii) relaxation of control measures. Typically, during lockdown some spread persists and one pertinent question is whether relaxing the lockdown will lead to a second wave. In [Britton et al., 2020], this question was explored by looking at prevalence at the end of lockdown in scenarios where lifting the lockdown did not lead to resurgence. For the purposes of this paper, we refer to this fraction as the disease-induced herd immunity (DIHI). This means that if DIHI is higher in one scenario than in another, then in the former a higher prevalence is required than in the latter to achieve herd immunity through the disease.

These papers make some simplifying assumptions about population structure that may not hold. In particular they do not consider the fact that existing interventions tend to affect some contacts more than others. For example, the transmission rates of household contacts are not significantly reduced (and in fact may increase) during interventions focused on reducing movement. Moreover, many of the highest risk positions in the disease network are in fact roles (health care workers, delivery drivers, teachers, etc) that would need to be maintained in most forms of a lockdown, meaning that some community links may be increased even as others are limited,

and the changes are distributed heterogeneously through the population.

In this paper we will use several network-based models to improve our understanding of how population structure may affect the DIHI threshold. Modelling epidemics on networks and the analysis of the resulting systems, be they stochastic or deterministic, is a well studied research area [Kiss et al., 2017; Pastor-Satorras et al., 2015; Porter and Gleeson, 2016]. The main reason for using networks is twofold. First, networks are intuitive and easy to understand and use, and second, they allow us to represent contact structure within a population to a high degree of accuracy. One of the most frequently studied questions in this area is how to understand and quantify the impact of various network properties on the invasion, spread and control of infectious diseases. Epidemics are usually modelled as stochastic process unfolding on the network. For example, the susceptible-infected-recovered/removed (SIR) epidemic on a network is modelled by the two separate processes: infection and recovery. In the simplest case both are Poisson processes such that an infectious node u connected to a susceptible node v transmits at per-contact rate τ, causing v to convert to infected. Nodes recover at rate γ to an immune state independently of the network.

On a network with N nodes, the SIR model leads to a continuous-time Markov Chain over 3^N states. Even for small values of N such a system becomes tedious to handle (numerically or otherwise). This has led to a myriad of approximations in the form of mean-field models which focus on some average quantities to reduce the dimensionality of the resulting system [Kiss et al., 2017]. Other approaches include percolation [Miller, 2016; Moore and Newman, 2000] and message-passing models [Bianconi et al., 2020].

The mathematical analysis of epidemic models on networks (from regular and Erdős-Rényi to scale-free networks such as the Barabási-Albert preferential attachment model) has led to many elegant analytical / explicit results regarding the impact of network properties (e.g., degree heterogeneity, assortativity and clustering) on the epidemic threshold, final epidemic size or endemic equilibrium, optimal vaccination strategies [Porter and Gleeson, 2016; Pastor-Satorras et al., 2015; Chen and Sun, 2014; Holme and Litvak, 2017], as well as the fact that certain mean-field SIR models are exact in the limit of large configuration model networks [Miller and

Volz, 2013; Miller and Kiss, 2014]. Many such models are mathematically tractable and help us gain important intuition about how an epidemic spreads and which properties of the contact network affect this. While they may not directly inform policy making, they remain useful tools to develop intuition, highlight important population features and test various scenarios quickly and effectively.

In this paper we use four different mean-field models [Kiss et al., 2017] to approximate exact epidemics on networks: (a) degree-based heterogeneous mean-field, (b) clustered and unclustered pairwise, (c) a new edge-based-compartmental model that allows us to distinguish between household/local contacts and community/global contacts, and (d) an age-structured compartmental model parameterised with realistic age structure and contact matrices. The primary aim is to investigate how heterogeneity in model structure impacts DIHI. More importantly perhaps, we challenge the way lockdown has been implemented in many models, namely, by a simple reduction in $\mathcal{R}_0$ or the transmission rate while keeping the contact network or mixing matrix the same. We build a new edge-based-compartmental model able to distinguish between household and community transmissions and use this to implement lockdown by either intervening on both types of connections or only on the community-based ones. In the same spirit, we use an age-structured compartmental model in which we implement lockdown either as a simple scaling of the entire mixing matrix or, more realistically, a set of age-specific structural changes.

The paper is structured as follows. In section 5.2 we describe all models including the underlying network types and relevant model and epidemic parameters. Section 5.3 contains the results for the network models, whilst section 5.4 provides the results for the age-structured model. Finally, in section 5.5 we discuss the implications of our findings. Additional technical details are given in the Appendix.

5.2 Methods

We consider a set of mean-field models that capture different levels of detail of the network structure. This helps reveal the underlying mechanisms where qualitatively different outcomes occur.

The first two are approximations of the exact stochastic SIR model on networks

with heterogeneous degree distribution, without and with clustering. This is followed by a new edge-based-compartmental model with household structure, i.e., with the ability to distinguish between household and community contacts. Finally, we consider an age-structured SEIRD model based on realistic age-structure and mixing matrices. For the network-based mean-field models we choose a flexible degree distribution with good control of the mean and variance, i.e., the negative binomial. We loosely order the models by their relative complexity, corresponding to gradually incorporating more features of the underlying network or population contact structure. While the last model is not explicitly network-based, it uses realistic mixing matrices over an age-structured model based on UK data. One can consider this an extrapolation from (or an approximation of) an explicit network model where individual-level interactions are averaged out over groups of interest from an epidemiological view point.

5.2.1 Contact structure and epidemic model

We consider a SIR/SEIRD epidemic spreading in a closed population of size $N = 6.65{\cdot}10^6$ (loosely the population size of the UK) with a well-defined contact structure. For illustrative purposes, we assume that the probability of an individual having k contacts follows a negative binomial distribution

$$P_{n,p}(k) = \binom{k+n-1}{n-1} p^n (1-p)^k. \tag{5.1}$$

The reason for this choice is that we want to highlight how heterogeneities in the contact structure play a central role in determining the DIHI. To illustrate this point, we consider three different scenarios for the degree distribution of the population. In all cases, we fix $\langle k \rangle = n(1-p)/p$ and we use the remaining free parameter to tune the variance. To avoid individuals with degree 0, the degree distribution is shifted by a constant m, thus making the effective average degree $\langle k \rangle = m + n(1-p)/p$. For normal-like and scale-free-like distributions we take $m = 1$, and for the delta-like distribution we take $m = 9$ (see Table 5.1).

The parameters chosen are reported in Table 5.1, and represent degree distributions of increasing variance, see figure 5.1, moving from almost no variance (delta-like degree distribution) to a degree distribution with a longer tail, akin to a scale-free

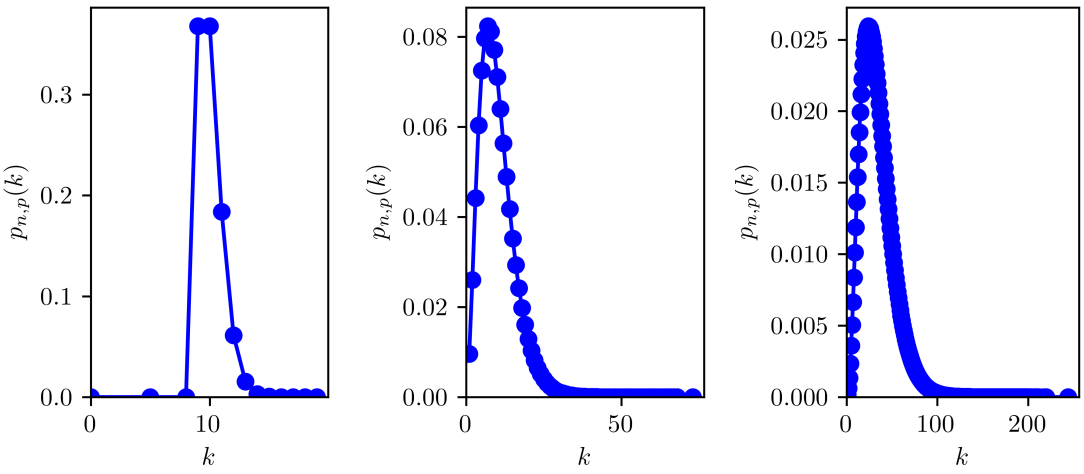

Figure 5.1: The three degree distributions described in Table 5.1.

network.

For all network-based models, we consider the simple susceptible-infected-recovered model (SIR) on a network. The infection is driven by a per-link transmission rate τ and a recovery rate γ. Mostly, we focus on the corresponding mean-field models with increasing level of accuracy in terms of what network features are incorporated.

Name	n	p	$\langle k \rangle$	σ^2	τ
delta-like	1	0.99	10	1	0.016
normal-like	3.86	0.3	10	30	0.016
scale-free-like	1.07	0.107	10	300	0.016

Table 5.1: The three degree distributions considered. The delta-like distribution is shifted by 9, as its mean would be 1 otherwise. The reason for this choice is the fact that in negative binomials the variance cannot be lower than the mean. Normal-like and scale-free like distributions instead are shifted by one, so that the minimum degree is 1. The resulting degree distributions are shown in figure 5.1.

5.2.2 Degree-based mean field model

In the degree-based mean field model (also called Heterogeneous Mean Field [Pastor-Satorras et al., 2015]) we denote by $[S]_k(t)$ the expected number of susceptibles with degree k at time t, similarly for $[I]_k$ and $[R]_k$. We define $[S] = \sum_{k=1}^{\infty}[S]_k$, similarly $[I]$ and $[R]$. The closure is made at the level of individuals, meaning that the infection

pressure across a link is simply averaged across the entire spectrum of infected nodes. The resulting ODEs are

$$\begin{aligned} [\dot{S}_k] &= -\tau k[S_k]\pi_I, \\ [\dot{I}_k] &= \tau k[S_k]\pi_I - \gamma[I_k], \\ [\dot{R}_k] &= \gamma[I_k], \\ \pi_I &= \frac{\sum_{\ell=1}^{M} \ell[I_\ell]}{\sum_{\ell=1}^{M} \ell N_\ell}, \end{aligned} \tag{5.2}$$

where $N_\ell = P_{n,p}(\ell)N$ is the number of nodes with degree ℓ. This system effectively keeps track of degree and heterogeneity in it, but mixing between nodes of different degrees happens at random but proportionally to degree [Pastor-Satorras et al., 2015; Kiss et al., 2017], with clustering (the tendency of nodes to form connected triples) playing no role.

The degree-based mean field model can be derived exactly, under the assumption that individuals with degree k reselect their partners very rapidly, so at every moment, the status of a node is independent of the status of its current partners. In reality, many edges are long-lasting, and so correlations build up: a randomly selected infected node is more likely to connect to another infected node than a randomly selected susceptible node. More complex models are needed to correct this.

5.2.3 Heterogeneous pairwise without and with clustering

In the heterogeneous pairwise model, we also count pairs: for example $[A_k B_\ell]$ is the expected number of links connecting a node of degree k in state A to a node of degree ℓ in state B [House and Keeling, 2011; Kiss et al., 2017]; likewise for triples of the form $[A_k B_\ell C_m]$. The closure is done at the level of pairs (i.e. triples are approximated by singles and pairs), and hence an approximation for the triples are needed. These are given by

$$[A_k B_\ell C_m] = \frac{\ell - 1}{\ell}\left((1-\varphi)\frac{[A_k B_\ell][B_\ell C_m]}{[B_j]} + \varphi\frac{[A_k B_\ell][B_\ell C_m][C_m A_k]}{[A_k][B_\ell][C_m]}\right), \tag{5.3}$$

where φ is the global clustering coefficient in the network. For the un-clustered case we simply set $\varphi = 0$. The resulting ODEs are,

$$\begin{aligned}
[\dot{S_k}] &= -\tau \sum_\ell [S_k I_\ell], \\
[\dot{I_k}] &= \tau \sum_\ell [S_k I_\ell] - \gamma [I_k], \\
[\dot{R_k}] &= \gamma [I_k], \\
[\dot{S_k I_\ell}] &= -\gamma [S_k I_\ell] + \tau \left(\sum_\alpha [S_k S_\ell I_\alpha] - \sum_\alpha [I_\alpha S_k I_\ell] - [S_k I_\ell] \right), \\
[\dot{S_k S_\ell}] &= -\tau \left([S_k S_\ell I] + [I S_k S_\ell] \right), \\
[\dot{I_k I_\ell}] &= -2\gamma [I_k I_\ell] + \tau \left(\sum_\alpha [I_\alpha S_k I_\ell] + \sum_\alpha [I_k S_\ell I_\alpha] + [S_k I_\ell] + [I_k S_\ell] \right), \quad (5.4)
\end{aligned}$$

where triples are closed using equation (5.3). The system can be significantly simplified for the $\varphi = 0$ case. When $\varphi > 0$, the closures become more complicated and present further challenges when implemented numerically (see notes in Appendix 5.6.1).

The number of equations in the heterogeneous pairwise model grows very large if the network has degrees of many different types (e.g., because there is an equation for $[\dot{S_k I_\ell}]$ for every k, ℓ pair). Generalizing this to more complex structures can become unwieldy. Edge-based compartmental models provide an alternative and are discussed next.

5.2.4 Edge-based compartmental model with household structure and community transmission

It is interesting to consider models which explicitly distinguish between links that happen within the households and those that happen elsewhere, as lockdowns act mostly on inter-household contacts. To consider household structures, we take advantage of the edge-based compartmental modelling (EBCM) framework [Miller, 2011; Miller et al., 2012], adapting the model in [Volz et al., 2011] to build a model that (i) has a more realistic contact structure with households, and (ii) can distinguish between within household and community transmission, see figure 5.2. This model keeps the number of equations tractable.

We assume that individuals are divided into households of size 4. Within house-

holds, there is complete mixing. In addition, each individual has a number of contacts outside the household, which allow for community transmission. The equations for this model are given in Appendix 5.6.2.

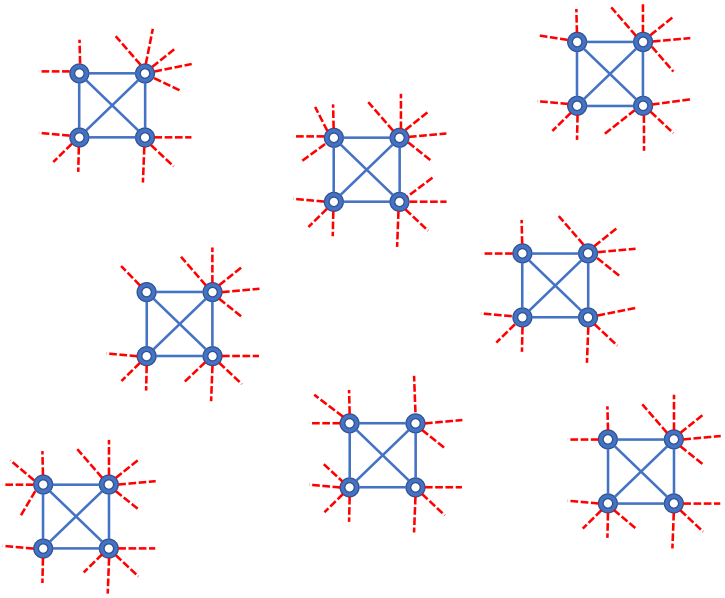

Figure 5.2: Caricature of the network model with households of size four and community stubs due to be connected up following the configuration model.

5.2.5 Age-structured compartmental model

We use a version of the SEIRD compartmental model by Alvarez et al. [Álvarez et al., 2015] adapted to remove any built-in control measures (originally modelled as a Hill repression function modulating the number of daily contacts in response to control measures) and to include age-structured interactions in the population. The model is as follows:

$$\begin{aligned}
\dot{S}_i &= -\beta S_i \sum_{j=1}^{n} \mathbf{C}_{ij} I_j / N_j, \\
\dot{E}_i &= +\beta S_i \sum_{j=1}^{n} \mathbf{C}_{ij} I_j / N_j - \gamma_E E_i, \\
\dot{I}_i &= \gamma_E E_i - \gamma_I I_i, \\
\dot{R}_i &= +\left(1 - m_i\right) \gamma_I I_i, \\
\dot{D}_i &= +m_i \gamma_I I_i,
\end{aligned} \tag{5.5}$$

where β, γ_E and γ_I are age-independent parameters denoting infectivity, rate at which exposed individuals become infected (inverse of incubation period) and rate at which infected individuals recover or die (inverse of disease duration) respectively (note that γ_I in this model corresponds to γ in the above models). The m_i are age-dependent mortality probabilities and control the fraction of those infected individuals who die. Susceptible individuals become exposed proportionally to a force of infection defined as the product of the contagion matrix with the prevalence by age. The contagion matrix is simply the product of the intrinsic infectivity of the epidemic and the daily contacts of individuals in age group i with individuals from age group j, $\mathbf{C}_{ij}$. Finally, n is the number of age groups considered and N_i is the count of individuals in age group i.

5.2.6 Epidemic parameters

Ling et al. [Ling et al., 2020] reported that the median time from symptoms onset to first negative RT-PCR in oropharyngeal swabs of convalescent patients was around 10 days, and further evidence [Wei et al., 2020] suggested that pre-symptomatic infection could happen $1-3$ days before the first symptoms appear. Accordingly, we set $\gamma = 1/14$ (i.e. an average of two weeks before recovery). Before setting τ, we summarise the expression for R_0 for the various models that we consider. For the heterogeneous degree-based mean-field model [Kiss et al., 2017] we have

$$R_0 = \frac{\tau\langle k^2\rangle}{\gamma\langle k\rangle}. \tag{5.6}$$

For the heterogeneous pairwise model, we use

$$R_0 = \frac{\tau}{\tau+\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}. \tag{5.7}$$

For the edge-based model, we set the in-household infection parameter β_h to be $3-5$ times bigger than the inter-households infection parameter β_c.

The basic reproduction number of the edge-based and the age structured models are given by the leading eigenvalues of the following two next-generation matrices [Diekmann et al., 2010]. First, for the edge-based compartmental model, based on [Pellis

et al., 2012], we have

$$A = \begin{bmatrix} \tilde{\mu}_c\mu_0 & 1 & 0 & 0 \\ \mu_c\mu_1 & 0 & 1 & 0 \\ \mu_c\mu_2 & 0 & 0 & 1 \\ \mu_c\mu_3 & 0 & 0 & 0 \end{bmatrix}, \tag{5.8}$$

where

$$\tilde{\mu}_c = \frac{\beta_c}{(\beta_c+\gamma)}E[\tilde{D}], \quad \mu_c = \frac{\beta_c}{(\beta_c+\gamma)}E[D], \tag{5.9}$$

with $\tilde{D}$ and D the distribution of the excess degree (i.e. the distribution of the left-over edges attached to a node reached by following one random edge, see [Newman, 2018]) and degree distribution of the network, respectively. μ_0, μ_1, μ_2 and μ_3 are the expected number of infected in generation 0, 1, 2 and 3 in a household of size 4. These values can be found in [Britton et al., 2019] on page 222/223, with $\varphi_I(\beta_h) = \gamma/(\beta_h+\gamma)$.

Finally, the R_0 for the age-structured compartmental model is given by the largest eigenvalue of:

$$\frac{\beta}{\gamma^I}\begin{bmatrix} \frac{N_1}{N_1}\mathbf{C}_{11} & \frac{N_1}{N_2}\mathbf{C}_{12} & \cdots & \frac{N_1}{N_n}\mathbf{C}_{1n} \\ \frac{N_2}{N_1}\mathbf{C}_{21} & \frac{N_2}{N_2}\mathbf{C}_{22} & \cdots & \frac{N_2}{N_n}\mathbf{C}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{N_n}{N_1}\mathbf{C}_{n1} & \frac{N_n}{N_2}\mathbf{C}_{n2} & \cdots & \frac{N_n}{N_n}\mathbf{C}_{nn} \end{bmatrix}.$$

where β is the intrinsic infectivity, γ^I is the rate at which infected individuals either recover or die, N_i is the size of the population in age group i, and $n = 18$ is the number of age groups in the model. $\mathbf{C}$ is the age-mixing matrix and the normalisation factor $\frac{N_i}{N_j}$ comes from the fact that at t_0, there are only susceptible individuals in the population of each age group and therefore $S_i = N_i$ in the partial derivative with respect to I of the r.h.s of the second equation in system (5.5).

5.3 Results

5.3.1 Validation of mean-field models

Before proceeding with the study of how contact heterogeneity and the choice of model impacts DIHI levels, we first test the validity of our proposed mean-field

models on synthetic networks. This is a well-studied and challenging problem, especially for clustered networks where results are limited to networks in which clustering is introduced in a very specific way (e.g., no overlapping motifs/subgraphs for example [Newman, 2003a; Kiss and Green, 2008; Newman, 2009]). Another way of introducing clustering in a network is to use so-called BigV rewiring [House and Keeling, 2010], which does not enforce the non-overlap between motifs/subgraphs but distributes these at random. This is the approach that we used to generate clustered networks for validation purposes. In fact, exact results only exist for SIR epidemics on configuration networks (with no clustering) with certain conditions on the moments of the degree distribution. In [Decreusefond et al., 2012; Janson et al., 2014], it is proven that the edge-based compartmental model is exact in the limit of such large graphs. Therefore, here, our aim is not to give a full justification of mean-field models but rather to show that the proposed mean-field models perform well compared to the simulated exact epidemic model.

Concretely, for each distribution of Table 5.1, we generated a single instance of a network with $N = 10000$ nodes by drawing the degree of each node from that distribution. The network was then built using the configuration model. The number of infected nodes at time $t = 0$ was set to 10 nodes (these were identical in each run to ensure that the initial conditions for simulations and mean-field models were correctly matched). We also tallied the number of all possible pairs of type A_iB_j, where $A, B \in \{S, I, R\}$ and i, j took values in the set of all possible degrees, in order to inform the initial conditions of the heterogeneous pairwise models. Then, we ran 100 realisations of the SIR epidemic on the generated networks using the Gillespie algorithm, implemented in the EoN software package [Miller and Ting, 2019]. We then compared the stochastic trajectories with the numerical solution of the homogeneous mean-field, heterogeneous mean-field and heterogeneous pairwise unclustered/clustered (see figure 5.3) models. When comparing the outcome of many realisations of the epidemic to results from a mean-field model most often averages are taken. This works well especially when the initial stochasticity is small, as it is the case for R_0 well above the epidemic threshold. A time-shifted average is also possible, or using confidence intervals based on many simulations [Juul et al., 2020; Lofgren, 2012]. For our purposes we chose the average, to have a single curve to

compare against deterministic models. It is well known, [Juul et al., 2020; Lofgren, 2012] and Appendix A.2 of [Kiss et al., 2017], that individual realisations may differ from the average. However, after the initial stochastic phase, all epidemics grow almost deterministically and in line with the proposed mean-field models.

Not surprisingly given that edge-based models have already been shown to be equivalent to heterogeneous pairwise models [Kiss et al., 2017], figure 5.3 shows good agreement. Figure 5.4 shows the good agreement for clustered pairwise models on networks with low degree heterogeneity and clustering at 0.1 and 0.2. In other settings (medium and high degree heterogeneity), as we will see in section 3.3, the impact of clustering is dominated by that of degree heterogeneity.

We found that in the medium- and high-heterogeneity cases, homogeneous mean-field models with the parameters chosen predicted no epidemics at all. For this reason, they were no longer used in the following section. Unsurprisingly, the model that performed best, was the heterogeneous pairwise, consistent with the fact that this model captures higher-order structure. Finally, it should be noted that in the low-heterogeneity case, the homogeneous mean-field model performed better than the heterogeneous mean-field. The reason for this is that, by construction, the heterogeneous mean-field model does not account for the fact that an edge can only transmit once, and therefore, overestimates the growth of the epidemic. On the other hand, the homogeneous mean-field model ignores degree heterogeneity, leading to an under-estimation. However, figures 5.3 and 5.4 show that our chosen models perform well.

5.3.2 The impact on DIHI

Most of our scenarios are concerned with determining the impact of model and demographic heterogeneities on the DIHI levels. In well-mixed homogeneous populations, each individual contributes equally to spreading, and therefore DIHI is a well-defined quantity, independent of whom has been infected during the first wave. In models with degree heterogeneities, nodes with higher degree contribute to the spreading of the disease much more than nodes with fewer links. This means that depending on whom has been infected, we can observe different levels of DIHI. This effect, however, does not show in our results since the DIHI is based on the infections

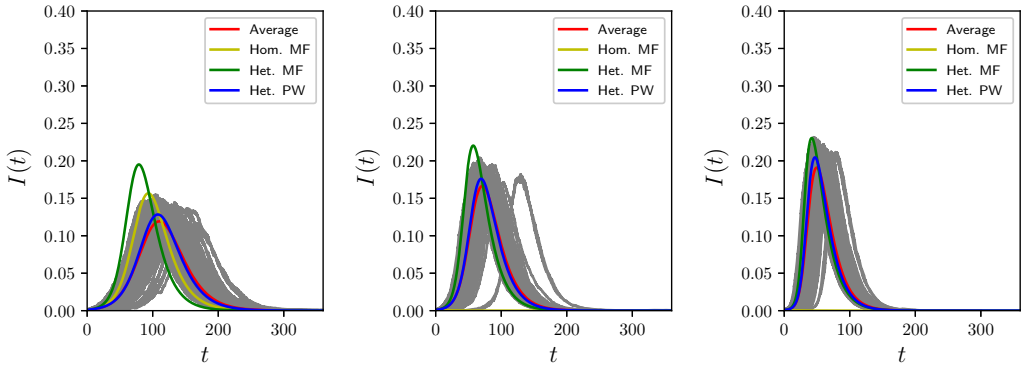

Figure 5.3: Comparison of 100 simulations of epidemics on a configuration model network of size $N = 10000$ (gray lines) and various mean-field models (Hom. stands for homogeneous, MF for mean-field, PW for pairwise). Each panel shows results for a different degree distribution of Table 5.1. Epidemic parameters are $\tau = 0.016$ and $\gamma = 0.07$ in all figures, and the initial number of infected nodes is 10. Note that in the second and third panels, the homogeneous mean-field approximation predicts no epidemic at all.

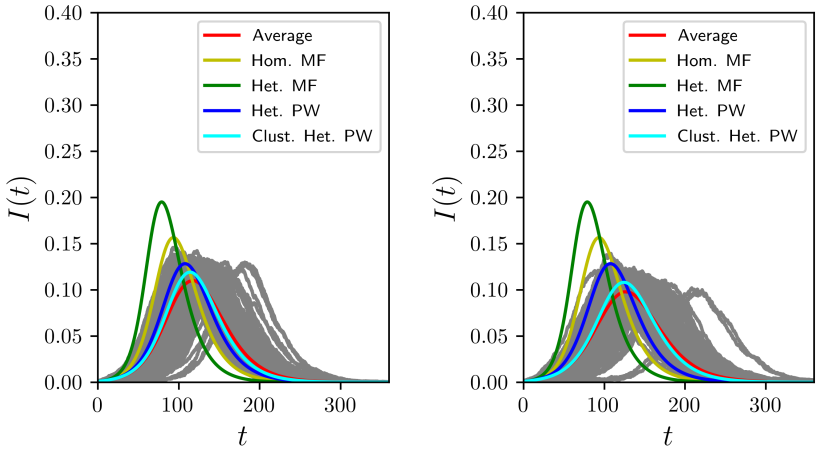

Figure 5.4: Comparison of 100 simulations of epidemics on a configuration model network of size $N = 10000$ (gray lines) and various mean field models, for two levels of clustering (left) 10% (right) 20%. Clustering is achieved by means of the BigV algorithm, as outlined in [House and Keeling, 2010]. Each panel shows results for the first degree distribution of Table 5.1 - low heterogeneity. Epidemic parameters are $\tau = 0.016$ and $\gamma = 0.07$ in all figures, and the initial number of infected nodes is 10.

accumulated during the first wave as determined by our equations.

It is well known that in the simple compartmental model herd immunity at time $t = 0$ is achieved as long as at least $1-1/R_0$ of the susceptible individuals are removed or immunised. In line with [Britton et al., 2020], our general setup is that an initial epidemic spreading freely for a short time is intervened upon by implementing a lockdown period of fixed duration. Afterwards, all parameters immediately return to their pre-lockdown values. In the most basic case this is done by keeping the network or the mixing matrix the same and multiplying the transmission rate by a constant $0 < \alpha < 1$. Crucially, however, we also explore the implications of how the lockdown is modelled, that is, we investigate the difference between reducing the transmission rate whilst keeping the network the same and changing the contact network, the latter being more in tune with what happens in reality. In the more realistic age-structured model, we compare a reduction of all entries in the mixing matrix with a number of scenarios involving school closure and work distancing.

For the edge-based compartmental model we focus on final epidemic size but still under the assumption of a lockdown period. This is because we want to compare how the two different strategies affect the eventual outcome of the epidemic, rather than how the optimal α varies between the two strategies, a comparison that would be difficult to interpret.

There is an extremely strong relation between the speed or rate of spread of the uncontrolled epidemic and the timing and length of the lockdown. In fact this can be visualised in terms of the 'flattening of the curve' argument. A reduction of the transmission rate during lockdown leads to a flattening of the epidemic curve with a reduced peak and an extended duration which ideally should fit within the control period. This means that if the epidemic grows quickly and the lockdown period is short two outcomes are possible. First, a fast growing epidemic with a short lockdown period needs to be met with a significant reduction in the transmission rate, i.e., small values of α. This will lead to a reduced epidemic which does not have enough time to unfold and the lifting of lockdown is followed by a full blown epidemic. Second, if the reduction is not strong enough (i.e., larger values of α), then a significant epidemic will occur during the lockdown itself with no further peak after lifting control [Di Lauro et al., 2021], see also figure 5.5.

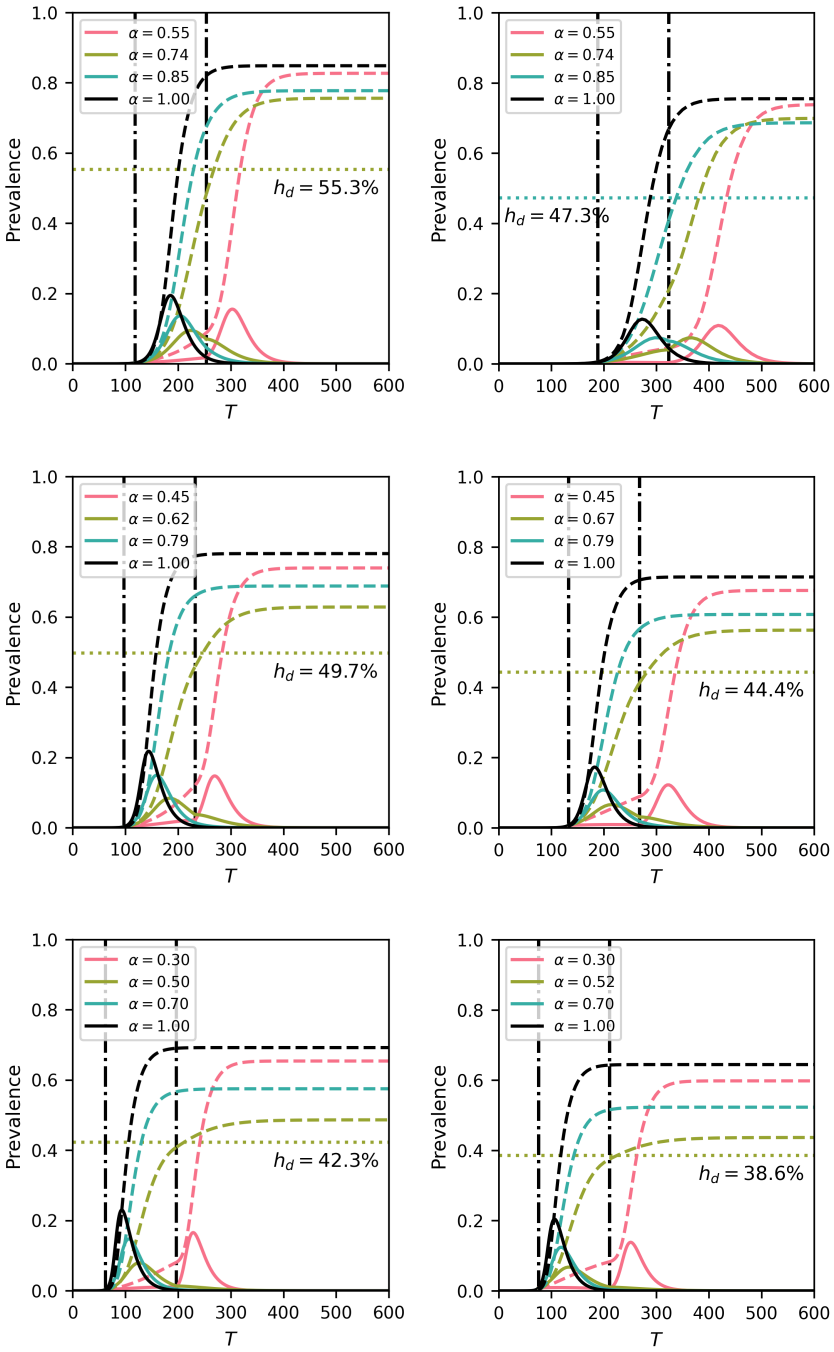

Figure 5.5: Optimal α (see legend) and DIHI (denoted as h_d in figure legends and axis labels) in delta-like (first row), normal-like (second row) and scale-free-like (third row) networks using the heterogeneous mean-field (first column) and heterogeneous pairwise model with $\varphi = 0$ (second column). Continuous curves indicate $[I](t)$, while dashed curves indicate $[R](t)$. The two vertical curves represent the beginning and the end of the lockdown. Duration of lockdown is 130 days. Finally, the horizontal line and the corresponding percentage reported are the cumulative prevalence at the end of lockdown for the best strategy that does not allow for a second wave.

5.3.3 Contact heterogeneity and clustering

The main focus here is to investigate the impact of degree-heterogeneity and clustering on herd immunity induced by a first wave of the epidemic. In networks with heterogeneous degrees, one can preferentially target high-risk individuals and this will reduce the number needed to achieve DIHI. Such heterogeneity can be exploited in many ways: for example, targeted immunisations [Albert et al., 2000] and acquaintance immunisation [Cohen et al., 2003]. Equally, the epidemic itself typically finds the high-risk groups first and thus 'removes' important individuals or risk groups. In line with [Britton et al., 2020; Gomes et al., 2020], we exploit this fact and consider different levels of degree-heterogeneity and different mean-field models to explore what happens in the wake of the lockdown period when some level of spreading is possible.

In each scenario considered, we seeded nodes with degree $k = 10$ with $[I_k](0) = 5$ infected individuals, the rest of the population being fully susceptible. We let the epidemic run until the cumulative number of infected people reached 0.5% of the population. Then, a one-shot intervention lasting exactly $T = 130$ days kicked in. During lockdown, control measures made $\tau \to \tilde{\tau}_0 = \alpha\tau$ (by acting on τ). Afterwards, lockdown was lifted and τ immediately returned to its pre-lockdown value. For the edge-based-compartmental and age-structured models, lockdown also involved preferential interventions on community or household links and modulation of the mixing matrix, respectively.

Figure 5.5 shows results from the degree-based mean field model (left column) and heterogeneous pairwise model without clustering (right column) for networks with increasing levels of degree heterogeneity (from top to bottom). In each case, we find the optimal α (a simple down scaling of the transmission rate without change to the network) and report the number of infections required to achieve DIHI (i.e., total of infected and recovered nodes at the end of lockdown such that the the epidemic after lockdown is subcritical). Several observations can be made. First, a higher value of DIHI also means a higher value of the cumulative incidence, but the two quantities are different. Secondly, for both models, aggressive control (low value of α) leads to a second wave. Equally, if the control is too weak (high value of α) the epidemic will still run its course during the first wave with some reduction in the

final size. Hence, there is an optimal value of α for which the final epidemic size is smallest and the epidemic post-lockdown is subcritical.

Both models clearly show that the value of DIHI decreases with variance of the degree distribution. Despite displaying the same overall trends, the two models are quantitatively different. In a like-for-like comparison, the degree-based heterogeneous mean-field model leads to larger overall epidemics. This is to be expected since this model does not keep track of the links explicitly and thus over estimates what would happen in a true simulation on an explicit network [Kiss et al., 2017]. The pairwise model, however, accounts for links and correlations and leads to epidemics that are typically less potent. Furthermore, the heterogeneous pairwise model leads to smaller values of DIHI showing that the accuracy with which the network structure is accounted for matters. This demonstrates that model choice is important as the precise levels of DIHI matter in a real-world scenario.

The effect of clustering is illustrated by figure 5.6. Typically, clustering lengthens the duration of the epidemic and lowers the peak when compared to the unclustered case (see also [Volz et al., 2011]). The final epidemic size is also smaller. This means that there is a small amount of leeway for implementing control and that the control effort can be smaller compared to the unclustered case. It is worth noting that the final epidemic size is also smallest at the optimal α value (see also [Britton et al., 2020]).

Finally, opting for the more accurate heterogeneous pairwise model, the level of DIHI is plotted for increasing values of variance and for different clustering levels, see figure 5.6. It is clear that higher variance can drive DIHI levels to as low as 30%. Clustering in the heterogeneous pairwise model leads to even smaller values of the DIHI, although increasing variance in the degree negates the effect of clustering and the DIHI levels are very similar to those observed with $\varphi = 0.25$ and $\varphi = 0.5$. This highlights the non-trivial interactions between network properties where clustering has biggest impact in sparse networks and where high levels of degree heterogeneity can negate the effect of clustering.

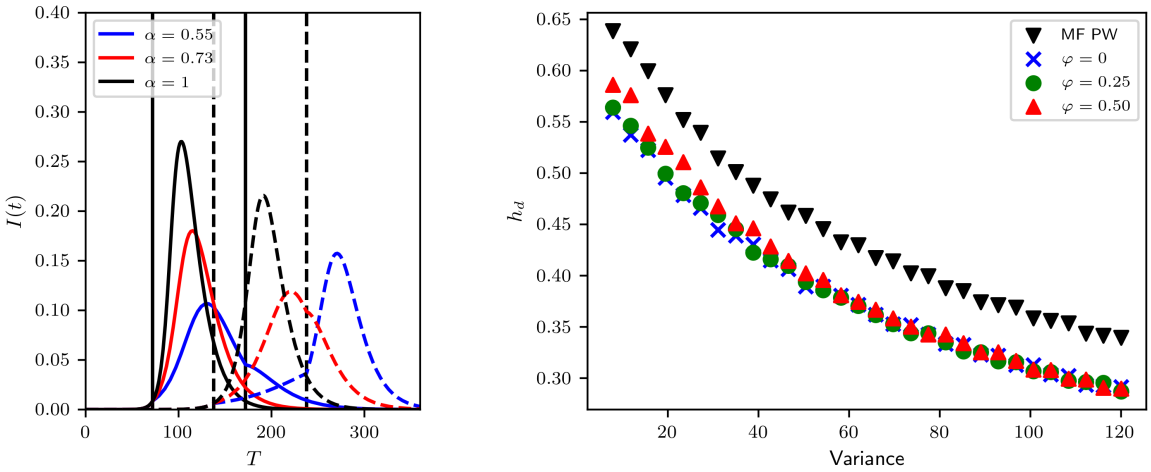

Figure 5.6: (left) Difference between control acting on un-clustered networks (continuous lines) and clustered networks (dashed lines), with clustering coefficient $\varphi = 0.5$, corresponding to the second point on the x-axis of the right panel. Vertical lines are at the beginning (continuous) and end (dashed) of control. Blue curve is optimal control for $\varphi = 0$, red for $\varphi = 0.5$. (right) Impact of variance in degree distribution on DIHI h_e, for different pairwise models with different values of φ. Average degree is $\langle k \rangle = 6$, $\tau = 0.04$ and $\gamma = 1/14$. Control duration is 100 days from the moment $I(t) + R(t) \geq 0.025$.

5.3.4 All versus community control only

Although the previous models do account for some important contact network features, they are not ideal to capture structure such as households. Being able to capture households explicitly and having the flexibility to differentiate between household and community transmission is important because many of the interventions available to us (e.g., closing schools and workplaces) affect community transmission differently from household transmission. Thus, a distinctive feature of most lockdown measures is a change in network structure, rather than a global reduction in transmission rate.

Although our model is not an exact reflection of true population structure, it allows us to investigate whether an intervention that disproportionately affects between-household transmission can be appropriately captured by a model that treats the intervention as reducing all transmission rates.

For a given start time and intervention strength, if the duration of the intervention is long enough, the number of infections becomes very small and the even-

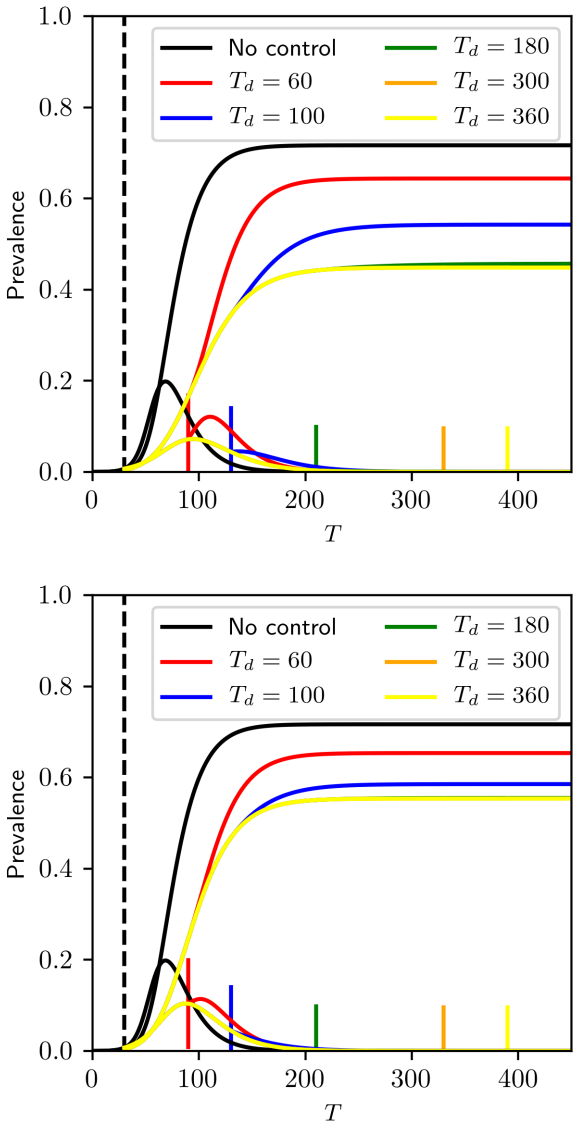

Figure 5.7: Control scenarios based on the EBCM model with intervention scaling factor of $\alpha = 0.6$ starting at $T = 60$ (dashed vertical line), and lasting for different durations (continuous vertical lines). (left) Intervention on the whole network, (right) intervention on the community structure. Parameters of the epidemic and community network are $\langle k \rangle = 4, \sigma^2 = 7.5, \beta_h = 0.045, \beta_c = 0.015$.

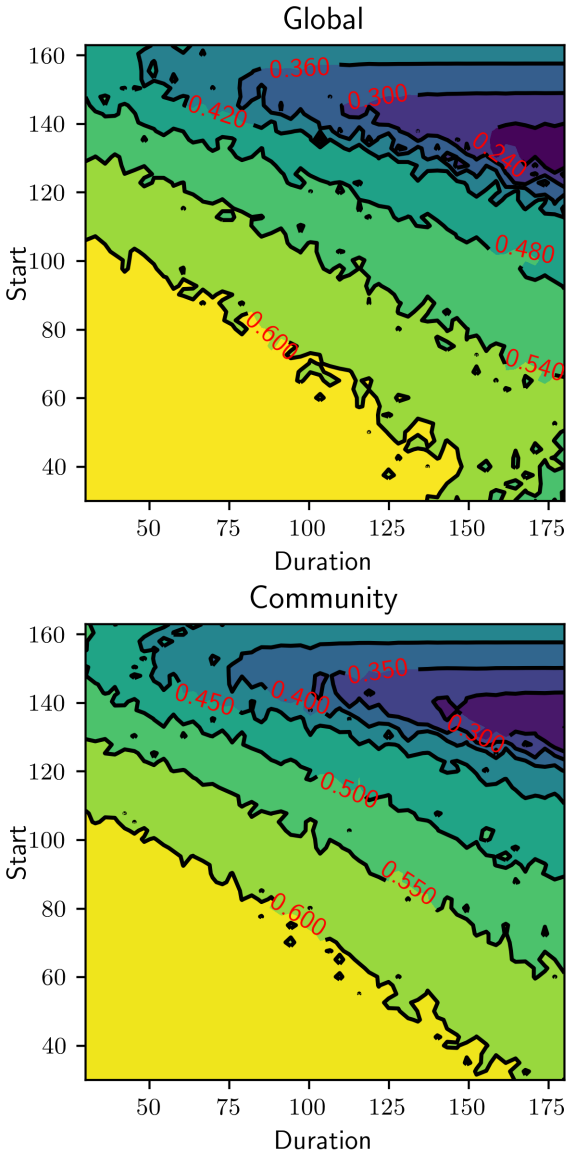

Figure 5.8: Final epidemic sizes based on the EBCM as a function of the beginning of lockdown and its duration, with two different strategies: intervention on the whole network (left) or intervention on the community links only (right). Each value is the minimum final epidemic size that can be obtained for varying α. Parameters of the epidemic and community network are: $\langle k \rangle = 4$, $\sigma^2 = 7.5$, $\beta_h = 0.045, \beta_c = 0.015$

tual rebound has the same shape, regardless of duration. The optimal intervention strength leaves the population sitting at the DIHI threshold when the epidemic dies out. When the intervention is lifted, no rebound occurs. Thus the outcomes of the optimal intervention are the same if the duration is long enough. In figure 5.7 and at the top right of figure 5.8 we notice that if the intervention is long enough, the optimal final size becomes independent of duration. To see this for earlier start times requires longer durations. Our argument to explain why this happens is as follows: suppose that for a given α, we find the smallest time such that DIHI is achieved. Prolonging the intervention for longer than said time will result in the number of actively infectious people reducing even further by end of lockdown but not change the downward trend of the epidemic, which will eventually die out.

Comparing the top right of both panels in Figure 5.8 show that the optimal community intervention allows more infections than the optimal global intervention.

The intuition behind this is based on the observation that epidemics typically exploit 'heterogeneities' in the population. For networks,, this means that high-degree nodes typically become infected early on in the epidemic. This is the main reason why a first epidemic wave in a network with very heterogeneous degree infects the highly connected nodes. In a scale-free like network, the number of such nodes is small (e.g., 20% of the nodes responsible for 80% of infection). If the epidemic progresses with strong interventions in place, it cannot spread far beyond these high-degree nodes. Once it dies out, the residual network is highly fragmented and made up of much lower degree nodes.

In this household model, community links drive degree heterogeneity. Household links alone lead to a regular network. Hence, when we effectively cut most community links, heterogeneity in degree is significantly reduced for the duration of control. This means that many high-degree nodes that would normally be infected during the first wave will now not get infected. The infection is unable to target the highest-degree nodes. When control is lifted, the high-degree nodes reactivate their links, allowing the epidemic to rebound.

However, when controlling both link-types equally, degree heterogeneity is preserved and the infection again preferentially targets the high-degree nodes. A weaker control targeted to all edges rather than just community edges may allow an initial

wave of similar size, but it will preferentially target the high-degree nodes. So this type of control acts as a very effective way of finding highly-connected nodes. This then means that at the end of lockdown, it is more likely that the most 'dangerous' nodes be removed, and with that, a smaller chance of a second wave.

Consider two epidemics with the same intervention start time, providing the optimal strategy for either the global or the community case. We would expect that the average community degree of those who have been infected in the global case would be higher than in the community case. Thus, on average, the individuals immunised by infection in the community case will be less important to disease transmission and thus more of them must be immunised to achieve the DIHI threshold. This is important because households may be able to sustain the epidemic for extended amounts of time and therefore change the outcome in DIHI levels.

5.4 Scaling versus modulating the mixing matrix model

In this section, we further explore the notion that modulating the effective R_0 of the epidemic through modifying the structure of the mixing matrix (in this case, the age-structure mixing matrix) can affect the system differently from achieving it through simply scaling each element of the mixing matrix. To do so, we began by considering three scenarios described by [Prem et al., 2017], namely, *school closure*, *school closure* and *social distancing*, and *work distancing*. In a baseline, no-intervention case, the matrix of daily contacts $\mathbf{C}$ was set to be the sum of 4 components: *school* contacts, *work* contacts, *home* contacts and *other* contacts. In what follows, we use the corresponding matrices of age-banded daily contacts in the UK produced by the POLYMOD study [Prem et al., 2017]. Briefly, *school closure* is realised by zeroing the school component of the mixing matrix); *school closure* and *social distancing* involved zeroing school contacts as well as reducing by half the number contacts at other locations between school-going individuals (first four age groups); *work distancing* is implemented by halving the contacts made at the workplace. For each of these interventions, we compared the behaviour of the system when re-scaling the matrix of total contacts so that its R_0 during the intervention was

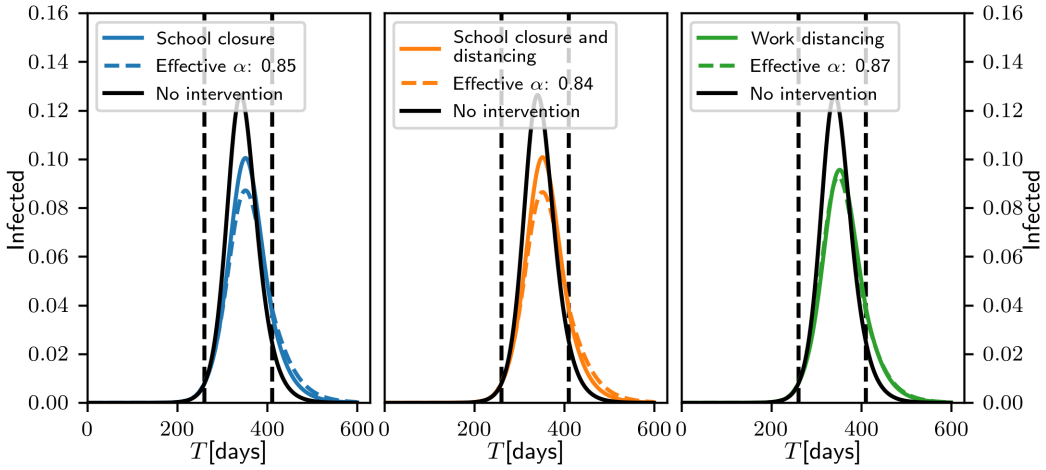

Figure 5.9: Comparison of the effect of three different control measures in the age-structured compartmental model. The three measures, *school closure*, *school closure* and *social distancing*, *work distancing* (coloured continuous lines, from left to right), act on the structure of the matrix (see text). For reference, the dashed lines result from an intervention reducing infectivity but yielding the same effective R_0 during the intervention. Epidemic values are $(\gamma_E, \gamma_I) = (1/7, 1/14)$. Vertical dashed lines indicate beginning and end of control.

the same as the R_0 of the modified mixing matrix during the intervention (namely, 2.114, 2.106 and 2.179 for scenarios 1 to 3 respectively).

Age-banded population counts (18 5-year bands) were taken from the Office for National Statistics (licensed under the Open Government Licence), pooling all age groups above 85. Mortality rates were taken from the modelling of [Verity et al., 2020], assuming that the rates in ten year age bands are the same as across two five year age bands. These rates were calculated based on cases in China in the initial outbreak and from the closed population on the Diamond Princess cruise ship. It is possible that these rates may prove to be overestimates compared to populations that do not experience overwhelming levels of hospitalisation, but they are not likely to be impacted by the interventions considered here.

To determine the evolution of the epidemic in this baseline case, we first scaled the contact matrices so that the system's R_0 was 2.5 (to maintain consistency with previous sections). In all cases, we used the same start date ($T = 260$) and duration (150 days) for the intervention. These parameters were arbitrarily chosen among the sets of possible parameters resulting in a sub-critical epidemic post-intervention.

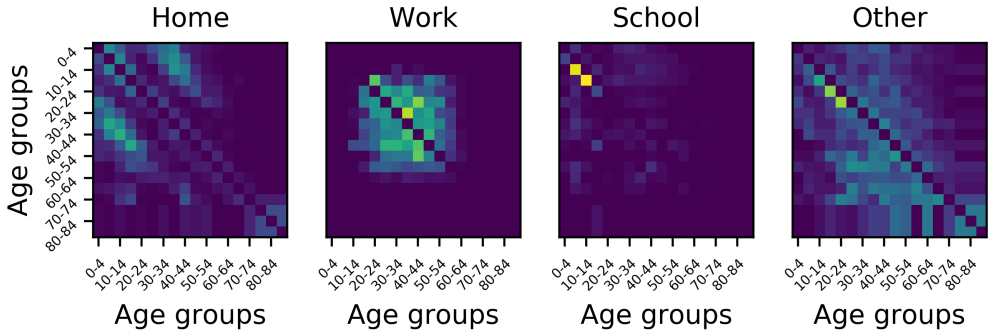

Figure 5.10: The four components of the POLYMOD age mixing matrix (after subtraction of the diagonal).

Figure 5.9 confirms that all interventions (coloured lines) result in a reduction in the number of infected individuals. However, whether this intervention is realised through modulating the contact matrix (solid lines) or through uniform scaling (dashed lines) results in substantially different outcomes ($\approx$ 16% differences for scenarios 1 and 2, $\approx$ 3% in scenario 3), even though the scaling factors used in the control interventions (dashed lines) are very similar from one scenario to the other. This is clear evidence that the structure of the contacts modulates the effects of the intervention.

Figure 5.10 provides a visual intuition as to why the *work distancing* scenario is closest to simply scaling the matrix. As pointed out by [Mossong et al., 2008], assortative mixing dominates in 3 of the components (home, school, other). Thus, *school closure* primarily affects diagonal elements of the mixing matrix (and primarily for the first 4 age groups). Intervention 2 does involve halving (some of) the contacts in the *other* component and some of those terms diffuse away from the diagonal, however, these contribute little to the overall mixing matrix. In contrast, the *work* component is the only component to feature what Mossong et al. [Mossong et al., 2008] describe as a wide contact plateau. Because this plateau accounts for more than half of the total number of contacts within the corresponding 8 age-bands, intervention 3 (social distancing) is most akin to scaling the entire matrix.

To further clarify how the structure of the contacts modulates the effects of the intervention, we carried out simulations in which two confounding factors were removed, namely heterogeneity in the number of individuals in the different age-

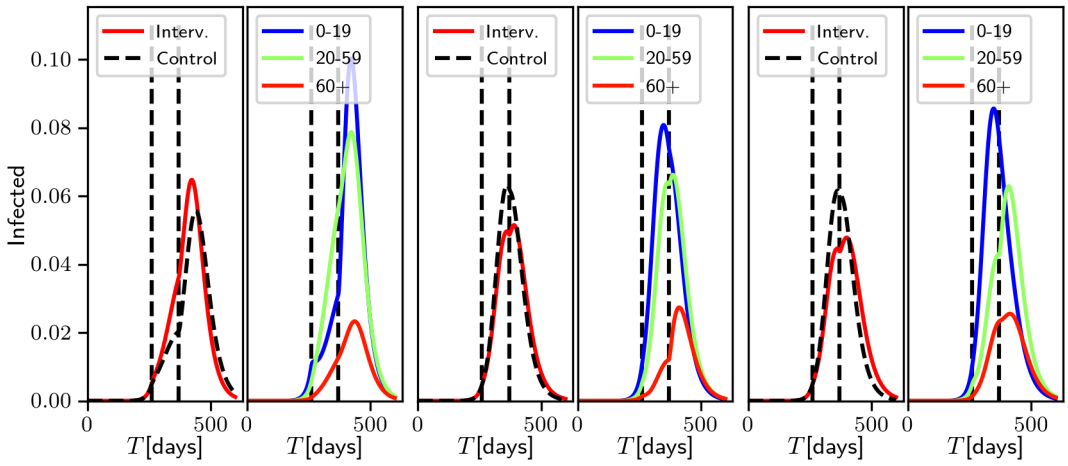

Figure 5.11: Impact of zeroing *school* (left), *work* (middle) and *other* (right) components when each age group has the same number of individuals and each component contributes the same number of contacts. For each, left panel shows the total prevalence of infected individuals in the population using the intervention (red) and the control (scaling of the entire contact matrix to achieve the same effective R_0, in black). The right panel shows the prevalence of infected in three pooled age groups chosen to reflect the target of each intervention.

bands and in the frequency of contacts by age. Whereas the former plays a key role in the calculation of the effective R_0 (see Section 5.2.6), the latter weights the impact of the intervention. For example, zeroing *school* contacts which only account for 12% of the total number of contacts will be negligible compared to zeroing *other* contacts that account for 40%. Therefore, in what follows, all age groups were set to have the same number of individuals (1/18-th of the total population) and all contact components were scaled to have the same sum of elements (arbitrarily, the sum of the original number of *other* contacts). We then systematically analysed the effect of three different interventions in which one component (*school*, *work*, or *other* was systematically scaled down by a factor taking values between 1.0 (no intervention) and 0.0 in steps of 0.1.

Figure 5.11 shows the effect of the most severe form of intervention (the zeroing of the relevant component) and clearly demonstrates that, once confounding factors are removed, changes in the mixing matrix lead to different outcomes in terms of whether such an intervention is more or less effective than simply scaling the matrix to achieve the same effective R_0. Here, zeroing the *school* component is less effective

than scaling the overall matrix (despite an overall scaling factor of 0.61). Instead, zeroing either the *work* or *other* component is more effective than scaling the overall matrix (despite larger scaling factors, 0.87 and 0.88, respectively). It should be noted that whilst zeroing the *work* and *other* components leads to similar results in terms of total prevalence, there are differences in prevalence by age which could have significant implications when age-structured mortality rates are considered.

5.5 Discussion

In this paper we explored a range of mean-field models previously used to approximate exact epidemics on networks and providing some analytical traction regarding how network properties impact epidemic invasion, final size and the efficacy of control measures. In order of increasing complexity, these models are: the degree-based heterogeneous mean-field, heterogeneous pairwise (without and with clustering), and an edge-based compartmental model which explicitly includes household structure and can distinguish between household and community transmission. While these cannot be used as they are to inform policy they still provide important insights into model selection and key features that need to be captured, or can be exploited, to identify the best possible control measures. In addition, we also tested our findings against a more realistic age-structured model with real mixing matrices.

We have shown that increased degree heterogeneity (i.e., higher variance in the degree distribution) leads to DIHI levels that are much smaller than the basic compartmental model, $1 - 1/\mathcal{R}_0$. This is in line with the findings of [Britton et al., 2020; Gomes et al., 2020]. Moreover, we quantified the extent to which the DIHI induced by the first wave depends on the variance in the degree distribution. We have shown that herd immunity in clustered networks is even lower because epidemics on clustered networks last longer and have lower peaks, allowing more flexibility regarding the start and intensity of control.

Perhaps, the most important question that we addressed regards how lockdown/-control is implemented in different models. Many models assume that during lockdown the contact network or mixing matrix is not changing but rather the transmission rate is scaled [Britton et al., 2020; Di Lauro et al., 2021; Gomes et al., 2020;

Morris et al., 2021]. We do not believe that this is appropriate because during lockdown the underlying contact structure changes. Our results with the edge-based compartmental and age-structured models have shown that these two approaches differ in outcome. Perhaps, the assumption of a non-changing contact structure during lockdown is more likely to be made in mean-field models. In models at higher than mean-field resolution (e.g. agent-based) it is much easier to explicitly modify the contact network.

We have therefore revealed a possible risk with using a model that ignores household structure to infer the level of infection required to reach the DIHI threshold. The favourable change in the DIHI threshold compared to what we predict from a homogeneous model is a consequence of heterogeneities. Where the intervention makes the population more homogeneous, the disease will no longer act like an effective intervention. In an extreme case, it will require more infections to achieve herd immunity than a random vaccination would need [Ferrari et al., 2006]. As the infection spreads along edges, we would see that at the end of the first wave, susceptible people are disproportionately in contact with other susceptible people and recovered people are disproportionately in contact with recovered people. So the residual network of susceptible nodes has more contacts than would occur if the same fraction were vaccinated randomly.

Another important observation resulting from our work is that it is extremely difficult to make general statements by extrapolating from findings based on simple models. Most models in fact ignore meso-scale structures (e.g., degree heterogeneity does well for local or micro structure whilst mixing matrices do well for macro-scale mixing) and their absence may exacerbate the impact of an intervention (either positively or negatively) leading to erroneous conclusions. In the present paper we saw that when intervention could not act on the global network of contacts, DIHI levels varied substantially, although heterogeneities still played a major role in reducing them.

Finally, it is worth noting that in all models we considered, when lockdown ended (and if DIHI was not reached), epidemics went on to grow exponentially. However, in many real-life scenarios, a prolonged phase of sub-critical spreading

(i.e. slow decay) has been observed before the exponential growing phase returned. There is a number of reasons why this behaviour is not observed in our models: (i) Deterministic models fail to capture fluctuations that dominate when the number of infected people is small, which might have a major impact on resurgence; (ii) during lockdown, the contact structure changes drastically and abruptly, but when lockdown is lifted, there is a delay/inertia in going back to pre-lockdown status, such lag having important implications on resurgence; (iii) after lockdown, some social distancing measures remain along with social awareness reducing exposure to the disease. Understanding and modelling these effects is an interesting challenge to address in future work.

In complementing our study of network-based mean-field models, we used a model with no explicit contact structure, where instead contact structure was alluded to via age-related mixing patterns. In the present situation with Covid-19, such models are appealing because they explain some of the structure of the population and provide a rationale for establishing a lockdown. Moreover, with higher mortality rates among older people, age-structured models are of interest in their own right. There is some tension between which model is most apt for describing population experience of the infection, the control measures effected and the outcomes for the population. The ideal solution may lie between some of the options presented here.

However, age-structured models cannot say much that is explicit about the structure of contacts within the population. While we have discussed ways in which lockdowns can be implemented in such models, the formulation is arguably less intuitive than in the network case. How to bridge the gap is an interesting question. An age-related network might be one labelled with age classes, with analysis focused on understanding the positions in the network occupied by individuals of a given age. This could be coupled to household models of network formation, one with variable sizes and smaller households more likely to feature older individuals. Once the general structure of such networks is known, adapting the models here would be simple enough. These considerations are beyond the presentation here but would be a fruitful avenue for future discussions, particular if additional lockdowns are required.

5.6 Appendix

5.6.1 Pairwise equations

We can account for clustering in the pairwise model, by introducing a clustering factor φ. The equations are similar to (5.4), but when closures are implemented we have to consider both open and closed triples.

In the case of open triples, the equations for the closures are the same as in (5.4), scaled by $(1-\varphi)$ to account for clustering, i.e.

$$\begin{aligned}[A_\ell S_k I]^o &= (1-\varphi)\tfrac{k-1}{k}\tfrac{[A_\ell S_k][S_k I]}{[S_k]},\\ [I S_k A_\ell]^o &= (1-\varphi)\tfrac{k-1}{k}\tfrac{[I S_k][S_k A_\ell]}{[S_k]},\end{aligned} \tag{5.10}$$

where the superscript ^{o} indicates open triples. For closed triples, we first study $[A_\ell S_k I_m]^c$. This quantity represents triples in which the first node is in state A, the middle node is in state S with degree k, and the third node is in state I with degree m. The triple is closed, therefore the third node is connected to the first one. This introduces correlations when we write the triple in terms of pairs for the closure, which we indicate with $C_{S_\ell I_M}$.

$$[A_k S_\ell I_m] = \varphi[A_k S_\ell](\ell-1)\frac{[S_\ell I_m]}{\ell[S_\ell]} C_{A_k I_m}. \tag{5.11}$$

The correlation can be written in terms of the ratio between realized pairs $[A_k I_m]$ and possible pairs $[A_k I_m]$ in a well-mixed population:

$$C_{A_k I_m} = \frac{[A_k I_m]}{k[A_k]\frac{m[I_m]}{\langle k\rangle N}}. \tag{5.12}$$

Hence, the closure is

$$[A_k S_\ell I_m] = \varphi\frac{(\ell-1)}{\ell}\frac{\langle k\rangle N}{km}\frac{[A_k S_\ell][S_\ell I_m][A_k I_m]}{[A_k][S_\ell][I_m]}. \tag{5.13}$$

In a similar manner, we can write the expression for $[I_m S_k A_\ell]$. The resulting system

is therefore

$$\begin{aligned}
[\dot{S_k}] &= -\tau[S_k I] \\
[\dot{I_k}] &= \tau[S_k I] - \gamma[I_k] \\
[\dot{R_k}] &= \gamma[I_k] \\
[\dot{S_k I_\ell}] &= -\gamma[S_k I_\ell] + \tau\left[\sum_m ([S_k S_\ell I_m] - [I_m S_k I_\ell]) - [S_k I_\ell]\right] \\
[\dot{S_k S_\ell}] &= -\tau\left[\sum_m \Big([S_k S_\ell I_m] - [I_m S_k I_\ell]\Big)\right],
\end{aligned} \tag{5.14}$$

where

$$[A_k S_\ell I_m] = \varphi[A_k S_\ell]\frac{\ell-1}{\ell}\frac{[S_\ell I_m]^2}{\langle k\rangle [S_\ell]^2 \frac{[I_m]}{N}} + (1-\varphi)\frac{\ell-1}{\ell}\frac{[A_k S_\ell][S_\ell I_m]}{[S_\ell]}.$$

5.6.2 Edge-based compartmental model

We consider a network with $4N$ nodes partitioned into households of size 4. Apart from the within household community, each node has a number of links to nodes outside the household, according to the degree distribution $P_{n,p}(k)$. The within household per-contact-transmission is denoted by β_h while the community transmission by β_c. The resulting system is given below,

$$\dot{\varphi}_{SSS}(t) = -3A(t)\varphi_{SSS}(t), \tag{5.15}$$

$$\dot{\varphi}_{SSI}(t) = 3A(t)\varphi_{SSS}(t) - \gamma\varphi_{SSI}(t) - 2(A(t)+\beta_h)\varphi_{SSI}(t) - \beta_h\varphi_{SSI}(t), \tag{5.16}$$

$$\dot{\varphi}_{SSR}(t) = \gamma\varphi_{SSI}(t) - 2A(t)\varphi_{SSR}(t), \tag{5.17}$$

$$\dot{\varphi}_{SIR}(t) = 2A(t)\varphi_{SSR}(t) + 2\gamma\varphi_{SII}(t) - \gamma\varphi_{SIR}(t) - \beta_h\varphi_{SIR}(t), \tag{5.18}$$

$$\dot{\varphi}_{SRR}(t) = \gamma\varphi_{SIR}(t) - A(t)\varphi_{SRR}(t), \tag{5.19}$$

$$\dot{\varphi}_{IRR}(t) = A(t)\varphi_{SRR}(t) + 2\gamma\varphi_{IIR}(t) - \gamma\varphi_{IRR}(t) - \beta_h\varphi_{IRR}(t), \tag{5.20}$$

$$\dot{\varphi}_{RRR}(t) = \gamma\varphi_{IRR}(t), \tag{5.21}$$

$$\dot{\varphi}_{IIR}(t) = -2\gamma\varphi_{IIR}(t) - 2\beta_h\varphi_{IIR}(t) + 3\gamma\varphi_{III}(t), \tag{5.22}$$

$$\dot{\varphi}_{III}(t) = (A(t)+2\beta_h)\varphi_{SII}(t) - 3\gamma\varphi_{III}(t) - 3\beta_h\varphi_{III}(t), \tag{5.23}$$

$$\dot{\varphi}_{SII}(t) = -(A(t)+2\beta_h)\varphi_{SII}(t) - 2\beta_h\varphi_{SII}(t) + 2(A(t)+\beta_h)\varphi_{SSI}(t) - 2\gamma\varphi_{SII}(t), \tag{5.24}$$

$$\dot{\Theta}(t) = -(\beta_h\varphi_{SSI}(t) + 2\beta_h\varphi_{SII}(t) + 3\beta_h\varphi_{III}(t) + \beta_h\varphi_{SIR}(t) + 2\beta_h\varphi_{IIR}(t) + \\ +\beta_h\varphi_{IRR}(t)), \tag{5.25}$$

$$\dot{\theta}(t) = -\beta_c\varphi_I(t), \tag{5.26}$$

$$\varphi_I(t) = \theta(t) - \gamma(1-\theta(t))/\beta_c - (1-\epsilon)\frac{\varphi'(\theta(t))}{\langle k\rangle}\Theta(t), \tag{5.27}$$

$$A(t) = \frac{\psi'(\theta(t))}{\psi(\theta(t))}\beta_c\varphi_I(t), \tag{5.28}$$

$$S(t) = (1-\epsilon)\Theta(t)\psi(\theta(t)), \tag{5.29}$$

$$\dot{R}(t) = \gamma I(t), \tag{5.30}$$

$$I(t) = 1 - S - R, \tag{5.31}$$

with the following initial conditions:

$$\varphi_{SSS}(0) = (1-\epsilon)^3, \tag{5.32}$$

$$\varphi_{SSI}(0) = 3\epsilon(1-\epsilon)^2, \tag{5.33}$$

$$\varphi_{SII}(0) = 3\epsilon^2(1-\epsilon), \tag{5.34}$$

$$\varphi_{III}(0) = \epsilon^3, \tag{5.35}$$

$$\theta(0) = 1, \tag{5.36}$$

$$\Theta(0) = 1, \tag{5.37}$$

$$S(0) = 1-\epsilon, \tag{5.38}$$

$$I(0) = \epsilon, \tag{5.39}$$

$$\tag{5.40}$$

with all other variables set to zero at time $t = 0$.

5.7 Acknowledgments

F. Di Lauro, L. Berthouze and I.Z. Kiss acknowledge support from the Leverhulme Trust for the Research Project Grant RPG-2017-370. J.C Miller acknowledges start-tup funding from La Trobe University.

Chapter 6

Covid-19 and Flattening the Curve: a Feedback Control Perspective

6.1 Introduction

Defining and implementing social distancing protocols (SD) is a significant challenge with economical, political, and scientific considerations. The definition of a clear or optimal goal remains unclear. As an example, consider the direct reduction of deaths by Covid-19. Imposing this goal requires the harshest measures possible, for an indefinite period of time. According to the available models [Kiss et al., 2017] a monotonic relationship exists between this cost function and the SD level. Yet, this strategy has many potential drawbacks. First, extreme levels of lockdown are unsustainable in the long run, due to the vast range of pernicious secondary effects (e.g. poverty [Goolsbee and Syverson, 2020], mental illnesses [Bhuiyan et al., 2020]) which in turn are themselves associated with a rise in mortality. Additionally, relaxing or lifting control after a harsh lockdown may lead to a second wave, possibly more critical than the first one [Xu and Li, 2020]. Another strategy would be to let the epidemic spread freely (red curve in Fig. 6.1) to get herd immunity as fast as possible. This is also hardly acceptable, as it would lead to higher mortality [Armstrong et al., 2020], and to a prolonged stress of the health care system. The "*flattening the curve*" strategy provides a third option, which promises to combine

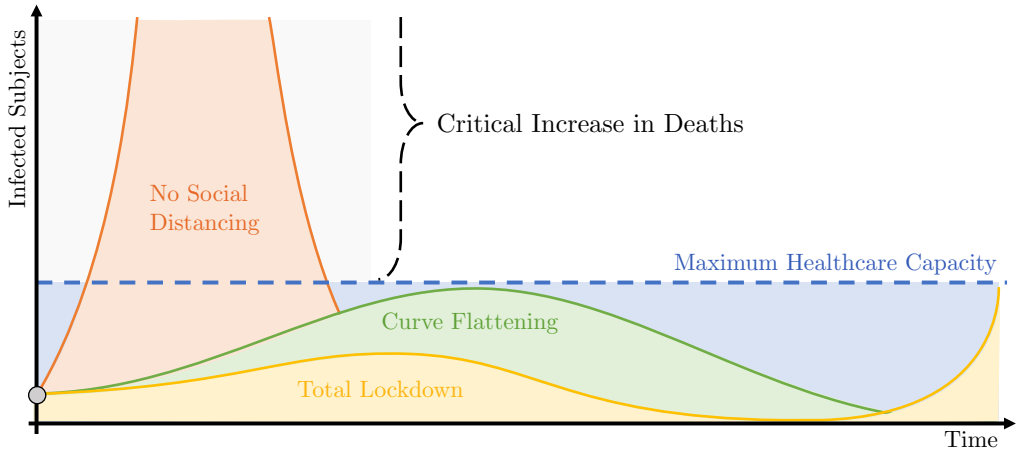

Figure 6.1: The aim of this work is to devise a control strategy that achieves the curve flattening goal, which should result in a curve similar to the green one. The two alternative extreme cases are shown as comparison: the result of no SD is shown in red, and of full lockdown in yellow.

the benefits of the two extremes [Thunstrom et al., 2020]. The key idea (of which Fig. 6.1 provides a visual representation) is to allow some level of disease spreading, while ensuring that people seeking medical assistance can access the health care system.

A vast pre-Covid-19 pandemic literature [Nowzari et al., 2016] on designing controllers for dealing with epidemics already exists. However, none of these works tackled the curve flattening goal, since no pandemics before threatened to overburden the healthcare system on such a large scale. In the context of Covid-19, open loop optimal control is proposed in [Di Lauro et al., 2021] for selecting the optimal timing of a time-limited lockdown, and in [Djidjou-Demasse et al., 2020] the authors find a trade-off between number of deaths and damage to the economy. Yet, feed-forward strategies are quite prone to uncertainties naturally associated with epidemics [Di Lauro et al., 2020b]. More robust strategies have been proposed, relying on feedback control. A linear controller is proposed in [Giordano et al., 2020]. A fast switching strategy with duty cycle selected through a slow feedback is discussed in [Bin et al., 2020]. In [Köhler et al., 2020], the loop is closed by periodically re-planning the optimal action, in a model-predictive-control fashion. An explicit formulation of curve flattening is instead provided in [Morris et al., 2021], where an open loop strategy is devised so to optimally reduce the infectious peak. An interesting alternative is discussed in [Charpentier et al., 2020], where a trade-off

between the health care and the socio-economic cost of the pandemic is proposed, and the limited capacity level of intensive care units is imposed as a constraint. Both these solutions are open loop.

This letter investigates the use of feedback control theory as a tool for engineering an effective curve flattening strategy. We wish to design a simple rule that can be implemented on a local level, without the need of accessing specialized facilities to run complex optimization routines. We perform extensive simulations of epidemics on networks [Pastor-Satorras et al., 2015; Kiss et al., 2017], with conditions inspired by real Covid-19 scenarios. This is as far as we know the first time that such analysis is carried out for Covid-19 control related research. We remark that the acceptable level of "curve flattening" is to be decided by policy makers, based upon cost-benefit analysis. However, once an optimal curve has been identified, this letter offers a novel, theoretically-backed strategy that guarantees that the goal of controlling the epidemic curve is achieved.

6.2 Background: Model of the Epidemics with Dynamic Interventions

Consider a fixed population of N individuals, and a disease spreading among them, through direct contacts. Each individual can be in either of three states: (i) susceptible, meaning that they can be infected by the pathogen; (ii) infected, meaning that they contracted the pathogen and they can now infect other susceptible people; (iii) recovered -and therefore immune, or removed. We denote with $S(t), I(t), R(t)$ the number of people at time t who are susceptible, infected or recovered, respectively. We have that $S(t) + I(t) + R(t) = N$. We can therefore neglect the study of R, as its value can always be recovered from S, I and N. If the population is well mixed, the evolution of the disease can be described by the SIR model

$$\dot{s}(t) = -\beta(t)\imath(t)s(t), \quad \dot{\imath}(t) = +\beta(t)\imath(t)s(t) - \gamma\imath(t), \tag{6.1}$$

where $s(t)$ and $\imath(t)$ are the system state, indicating respectively the number of susceptible $S(t)$ and infectious $I(t)$, divided by the total population N. Note that, despite its simplicity, the SIR model has proven able to match real data when

applied to Covid-19 [Morris et al., 2021; Thunstrom et al., 2020], and it is therefore widely used in the literature. Without loss of generality, we consider that, at $t = 0$, $s + \imath = 1$. The constant $\gamma \geq 0$ defines the transition rate from the pool of infected, to the compartment of recovered/removed. β is the rate at which an infected individual makes disease-transmitting contacts with other people. When SD policies are imposed, the value of β varies, $0 < \beta_{\mathrm{m}} \leq \beta \leq \beta_{\mathrm{M}}$, with β_{m} corresponding to total lockdown. Therefore β can vary in time, and it is the control input of (6.1).

6.3 Control Strategy

We propose here a control strategy acting on system (6.1). As shown by Fig. 6.2, this architecture is made of two components: (i) an optimal open loop action, and (ii) a feedback controller implementing trajectory tracking.

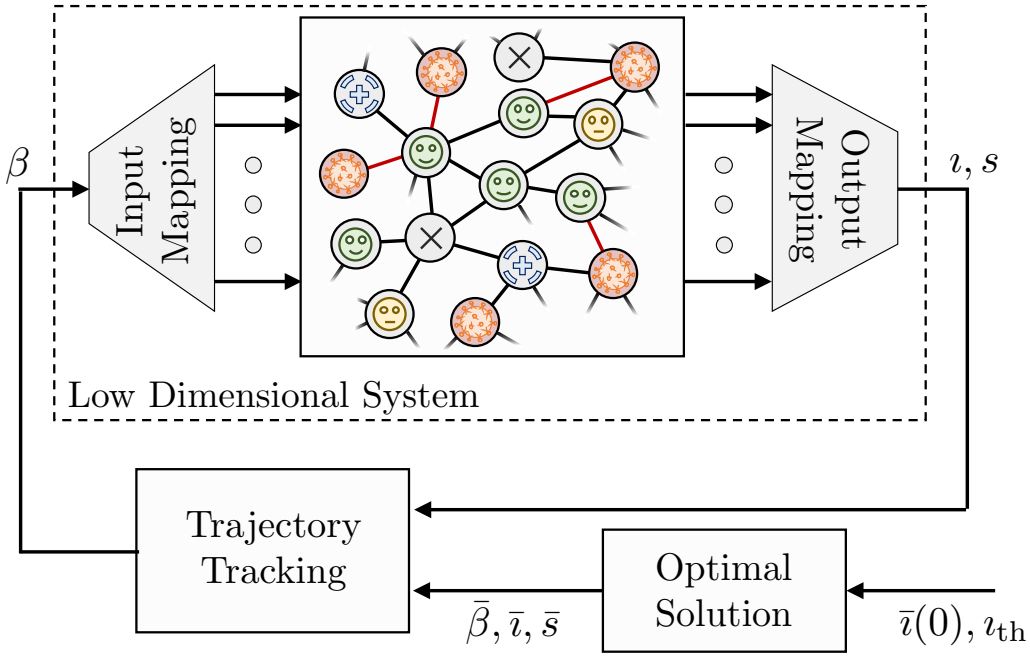

Figure 6.2: Block diagram of the strategy proposed in this paper. The input and output maps reduce the high-dimensional dynamics of the outbreak to a simpler evolution of few salient characteristics, namely the prevalence of infected and susceptible $\imath, s$, which are sensible to changes in the level of SD, modelled here as different values of the transmission rate of infection β. A nonlinear feedback controller acts within this representation implementing trajectory tracking of an optimal control policy.

6.3.1 Optimal curve flattening under nominal conditions

Our aim here is to introduce a nominal strategy ("Optimal Solution" in Fig. 6.2) for optimally flattening the epidemic curve $\imath(t)$, so to keep the number of infected people $\imath$ within the maximum capacity of the health care system, $\imath_{th} > 0$. This can, for instance, be evaluated by considering the percentage of people that will need Intensive Care Units (ICUs), which are probably the most critically limited resources. As discussed in the introduction, enforcing this constraint is of paramount importance, since exceeding it may provoke a critical failure of the healthcare system, leading to a substantial increase in the number of deaths not only from the disease, but also from uncorrelated health issues. On the other hand, we want to keep the level of restriction on the population as low as possible, to minimise secondary negative effects. Note that the curve flatting goal is the result of a careful balance between competing interests, and as such we decide to explicitly impose it as a goal. We consider the case of a constant β. This simplification is instrumental in making the optimal control problem more manageable.

We summarize the above considerations through the optimization problem

$$\max_{\beta \in \mathbb{R}} \beta, \quad \text{s.t. } 0 < \imath(t) \leq \imath_{\text{th}} \ \forall t \text{ and } (6.1). \tag{6.2}$$

We now propose a Lemma introducing a general solution to this optimal control problem.

Lemma 1 *The following is the closed form solution of* (6.2)

$$\beta = -\frac{\gamma}{1-\imath_{\text{th}}} W_{-1}\left(-\frac{1}{e}\frac{1-\imath_{\text{th}}}{1-\imath(0)}\right), \tag{6.3}$$

where W_{-1} is the Lambert W function [Corless et al., 1996], branch -1.

Proof 1 *Since the cost function is linear in the optimization parameter, the optimal value is to be found on the boundary of the feasible set, i.e. β has to be such that $\max_t \imath(t) = \imath_{th}$.*

The maximum value of $\imath$ is given by the non-trivial solution of $i(t) = 0$. Combining this condition with the second equation in (6.1) *yields $s^+ = \gamma/\beta$. Further, we can combine the two equations in* (6.1) *into $\mathrm{d}\imath/\mathrm{d}s = \gamma/(\beta s) - 1$. This nonlinear ordinary differential equation can be solved together with the initial condition*

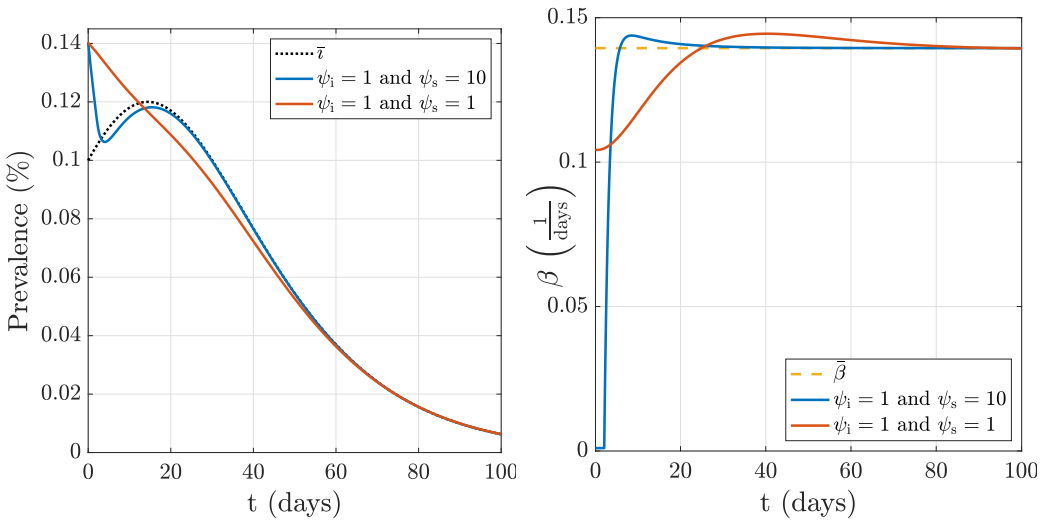

Figure 6.3: Two executions of the proposed control architecture when applied to system (6.1).(left) infected nodes, (right) control parameter β. Two different choices of control gains ψ_i and ψ_s are considered. The other parameters are $\gamma = 0.1$, $\beta_M = 0.22$, $\bar{\imath}(0) = 0.1$, $\imath_{th} = 0.12$, $\imath(0) = 0.14$. Susceptibles are not shown for the sake of space.

$s(0) = 1 - \imath(0), \imath(0)$, *to get*

$$\imath(s) = \frac{\gamma}{\beta} \ln\left(\frac{s}{1-\imath(0)}\right) - s + 1. \tag{6.4}$$

By inverting $\imath(s^+)$ *for* β, *we get the desired optimal value such that* $\max_t \imath(t) = \imath_{th}$. *The following is a solution for all integer values of* j,

$$\beta = -\frac{\gamma}{1-\imath_{\text{th}}} W_j\left(-\frac{1}{e}\frac{1-\imath_{\text{th}}}{1-\imath(0)}\right), \tag{6.5}$$

where $W_j(\bar{a})$ *is the* $j-th$ *branch of the Lambert W function [Corless et al., 1996]. Each of the branches is built as the solution of* $\bar{a} = W_j e^{W_j}$. *Among all of them, only* W_{-1}, W_0 *have domain within the real line. Moreover, it is always the case that* $W_0 > W_{-1}$, *which in turn assures that the larger value of* β *is always reached for* $j = -1$, *concluding the proof.*

It is worth noting that the argument of W_{-1} is always between $-1/e$ and 0 since $\imath(0) \leq \imath_{\text{th}}$. This is exactly the range of arguments for which the W_{-1} is well defined [Corless et al., 1996].

6.3.2 Trajectory tracking controller

The following Lemma introduces the tracking controller ("Trajectory Tracking" in Fig. 6.2) implementing the reactive change of the SD level β. Note that in principle

this controller is agnostic to the choice of the reference to be tracked, and it is introduced as such. However, in our architecture we consider the output of Lemma 1 as the reference $\bar{\beta}, \bar{\imath}, \bar{s}$.

Lemma 2 *Consider $\psi_{\mathrm{i}}, \psi_{\mathrm{s}} \in \mathbb{R}, \psi_{\mathrm{s}} > 0, \psi_{\mathrm{i}} \geq 0$. Consider also $\bar{s}, \bar{\imath}, \bar{\beta} \in \mathbb{R}$ to be a solution of (6.1) with $\bar{s}(0), \bar{\imath}(0) > 0$. Then, the feedback loop composed by the action*

$$\beta(s, \imath, t) = +\psi_{\mathrm{i}}(\bar{\imath} - \imath) - \psi_{\mathrm{s}}(\bar{s} - s) + \frac{\bar{s}\,\bar{\imath}}{s\,\imath}\,\bar{\beta}, \tag{6.6}$$

and the SIR model (6.1), is such that $(s, \imath)$ converges to $(\bar{s}, \bar{\imath})$ if $s(0), \imath(0) > 0$.

Proof 2 *Consider now the linear change of coordinates $x = -(\imath + s)/\gamma$. Adding up the two equations in (6.1), yields $\dot{\imath} + \dot{s} = -\gamma\imath$. We can therefore establish the change of coordinates*

$$\imath = \dot{x}, \quad s = -\gamma x - \dot{x}. \tag{6.7}$$

Combining the latter, with the second equations in (6.1) allows writing the following equivalent formulation of the SIR dynamics

$$\ddot{x} = -(\gamma x + \dot{x})\dot{x}\beta - \gamma\dot{x}. \tag{6.8}$$

We take the following control action

$$\beta(x, \dot{x}, t) = -\frac{\gamma\dot{\bar{x}} + \dddot{\bar{x}}}{(\gamma x + \dot{x})\dot{x}} + \alpha_{\mathrm{p}}(\bar{x} - x) + \alpha_{\mathrm{d}}(\dot{\bar{x}} - \dot{x}), \tag{6.9}$$

with $\alpha_{\mathrm{p}} > \gamma$, $\alpha_{\mathrm{d}} > 0$ being the gains of a PD-like action. Note that the denominator of the first element of the right hand side of (6.9) is equal to $s\imath$. Indeed, the hypothesis $s(0), \imath(0) > 0$ implies that $s(t), \imath(t) > 0$, $\forall t$, since (6.1) is a strictly positive dynamical system [Kiss et al., 2017]. This produces the closed loop dynamics $\ddot{e} = -(\gamma + \alpha_{\mathrm{d}}(-\gamma x - \dot{x})\dot{x})\dot{e} - \alpha_{\mathrm{p}}(-\gamma x - \dot{x})\dot{x}e$, where $e = \bar{x} - x$. As mentioned before $s, \imath > 0$, therefore $\gamma + \alpha_{\mathrm{d}}(-\gamma x - \dot{x})\dot{x} = \gamma + \alpha_{\mathrm{d}} s\imath > 0$ and $\alpha_{\mathrm{p}}(-\gamma x - \dot{x})\dot{x} = \alpha_{\mathrm{p}} s\imath > 0$, for all $t < \infty$. Thus, we propose the following Energy-like Lyapunov candidate $V(e, \dot{e}, t) = \dot{e}^2/2 + \int_0^e (\alpha_{\mathrm{p}}(-\gamma(x_{\mathrm{d}} - \epsilon) - \dot{x})\dot{x})\epsilon \mathrm{d}\epsilon$, which is positive definite. We perform time differentiation, obtaining through standard manipulations $\dot{V} = -(\gamma + \alpha_{\mathrm{d}} s\imath)\dot{e}^2 \leq 0$, since $s, \imath > 0$. This proves that $(e, \dot{e})$ are bounded. Yet, it is not enough to prove stability, since the error dynamics is time-variant. Indeed, $s, \imath$ are a combination of $x, \dot{x}, \bar{x}, \dot{\bar{x}}$, the latter two being explicit functions of time. To prove asymptotic convergence we invoke Barbalat's Lemma,

which requires to perform a further differentiation $\ddot{V} = -(\gamma + \alpha_{\mathrm{d}} s\imath)\dot{e}\ddot{e} - \alpha_{\mathrm{d}}(\dot{s}\imath + s\dot{\imath})\dot{e}^2$ *which is bounded since it is sum and products of bounded functions. Therefore,* $(e, \dot{e})$ *converges asymptotically to* $(0, 0)$*, and as a consequence* $(s, \imath)$ *converges to* $(\bar{s}, \bar{\imath})$*.*

To conclude the proof, we need to show now that (6.6) *and* (6.9) *are equivalent. First, we use* (6.8) *to obtain* $\ddot{\bar{x}} = (\gamma\bar{x} + \dot{\bar{x}})\dot{\bar{x}} - \gamma\dot{\bar{x}}$*. We then take* $\psi_{\mathrm{s}} = \alpha_{\mathrm{p}}/\gamma$ *and* $\psi_{\mathrm{i}} = \alpha_{\mathrm{d}} - \alpha_{\mathrm{p}}/\gamma$*. Finally, we combine these three equations with* (6.7) *and* (6.9)*. This leads to* (6.6)*, therefore concluding the proof.*

We want our control action to remain limited when acting on a neighborhood of $s\imath = 0$. Also, it is not meaningful to act on the system by changing β to values smaller than the one associated with total lockdown $\beta_{\mathrm{m}} > 0$, or greater than the one representing no social distancing $\beta_{\mathrm{M}} > \beta_{\mathrm{m}}$. We therefore introduce the following modification on the ideal controller

$$\beta(s, \imath, t) = \left[\psi_{\mathrm{i}}(\bar{\imath} - \imath) - \psi_{\mathrm{s}}(\bar{s} - s) + \frac{\bar{s}\bar{\imath}}{[s\imath]_{\epsilon}^{\infty}}\,\bar{\beta}\right]_{\beta_{\mathrm{m}}}^{\beta_{\mathrm{M}}}, \tag{6.10}$$

where $\epsilon > 0$ is a small constant, and $[a]_l^u$ is is equal to l or u if $a < l$ or $a > u$ respectively, and equal to a otherwise. Fig. 6.3 reports two examples of application of the algorithm to the SIR model (6.1).

6.4 Network Control

6.4.1 Network Model

We implement two important features in a refined model: (i) people interact through heterogeneous contact structures, i.e. the population is not well-mixed, and (ii) real epidemics have an intrinsic degree of stochasticity, so they cannot be exactly described by (6.1). We therefore consider stochastic epidemics on networks [Pastor-Satorras et al., 2015; Kiss et al., 2017]. A network is a pair $(V, \mathcal{E})$, where V is a set of N nodes (or vertices), and $\mathcal{E}$ is a set of edges (or links) connecting nodes, i.e. tuples $\{u, v\}$, where $u, v \in V$. A population contact structure is modelled by a network in which nodes represent individuals, and links are associated with routes of disease transmission between individuals. We consider undirected networks, such that $\{u, v\} \in \mathcal{E} \iff \{v, u\} \in \mathcal{E}$. Figs. 6.2, 6.4, 6.5, show pictorial representations of networks. Here, we focus on a particular well-known class of random networks,

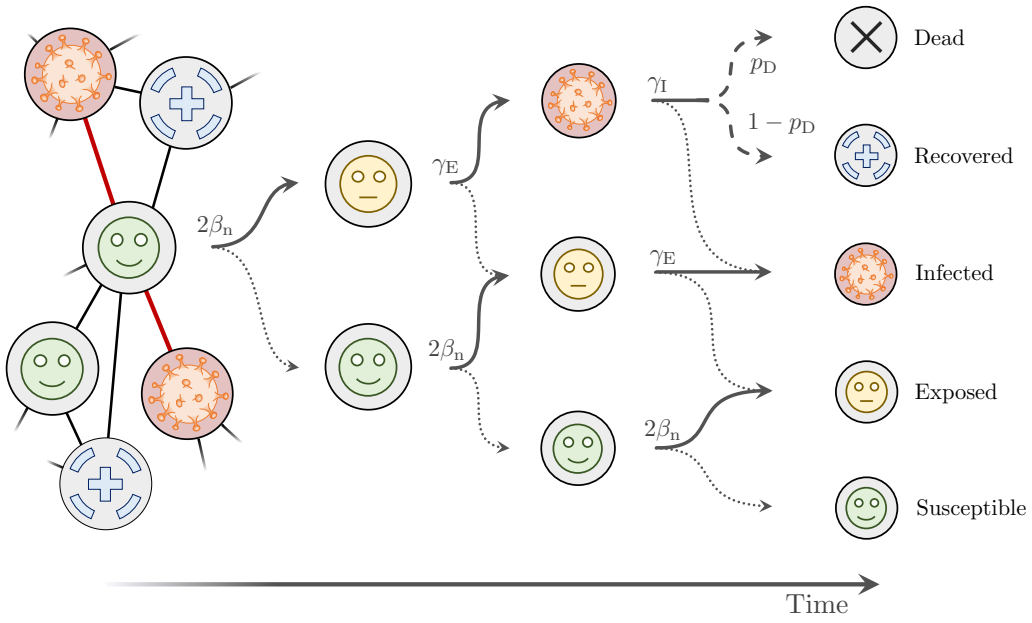

Figure 6.4: Pictorial representation of SEIRD dynamics on a network. The process is a continuous-time Markov chain. Each infected (and infectious) node spreads the disease to its susceptible neighbors at rate β_n until no longer infectious. A node that has been successfully infected, becomes first exposed, then infectious itself. Its ultimate destiny is either dying (with probability p_D), or fully recovering (with probability $1 - p_D$). The rate of each event is given on the continuous arrows.

i.e. Erdős-Rényi [Bollobás, 2001], generated as follows: start with N isolated nodes, consider each unique pair of two distinct nodes and connect them with probability $0 \leq p \leq 1$. Hence, the probability of a node having k neighbors follows a binomial distribution $\mathcal{B}(N-1, p)$, $\langle k \rangle = p(N-1)$ being the average degree. Such networks may be considered a very first order approximation of realistic contact structures, as they display sufficient heterogeneity and are easy to implement [Kiss et al., 2017].

6.4.2 Epidemic model on Network

We consider a SEIRD model for disease spreading, in which nodes are divided into comparments representing their status with respect to the disease: S (susceptible), E (exposed), I (infected/infectious), R (recovered), or D (deceased). Fig. 6.4 illustrates the possible transitions of a susceptible node that is in contact with two infectious neighbors. Compared to a SIR model (see Sec. 6.2), we add an exposed class to account for individuals who have been infected but are not yet infectious (biologically known as incubation phase). We also allow for infected individuals to either survive or die. Outbreaks are modeled as Markovian processes on the generated network, in which an infected node spreads the disease, via links, to its susceptible neighbors at a constant rate β_{n}, turning them into exposed. At a constant rate γ_E, an exposed node becomes infected, and stops infecting at a constant rate γ_I, after which two outcomes are possible: either fully recovery (transition into R), with probability $1-p_{\mathrm{D}}$, or death with probability p_{D} (transition into D). Nodes in compartments R and D play no further role in the dynamics. Further, p_{D} depends on the prevalence of the disease, to model increased mortality in case of saturation of the health care system. Control interventions in this model are implemented as changes in the value of β_{n}. At time $t = 0$, $I(0) = N\imath(0) \ll N$ randomly chosen nodes are infected. The remaining ones are initialized as susceptible. We use a Gillespie algorithm [Gillespie, 1977] adapted to networks [Kiss et al., 2017] to simulate this process. In Fig 6.5 we show a realization of an outbreak on a network of modest size, to highlight how the topology impacts the dynamics.

6.4.3 Input and Output Maps

To connect the controller to the network model, we introduce two maps, as shown

in Fig. 6.2. Such mappings are general, and they could be used in conjunction with different control techniques relying on similar input-output description of the pandemic. The output map extracts s and $\imath$ from the full state of the network by counting as s the fraction of nodes either S or E, and as $\imath$ the fraction of I. The input map provides expressions for the control input on the network level β_{n} given the output of the controller $\beta(s, \imath, t)$. With the aim of evaluating the input map, we turn to the adaptation of β to networks. From (6.1) we get

$$N\dot{\imath} = \beta N \imath s - \gamma N \imath \Rightarrow \dot{I} = \beta I \frac{S}{N} - \gamma I. \tag{6.11}$$

The term $\beta IS/N$ represents the total infectious pressure in the ODE model. This quantity drives the whole infectious process, and it is crucial that the map preserves it. On the network, the infectious pressure is given by the infectious pressure β_{n} times the number of links between infected and susceptible nodes, which is a random variable that depends on which nodes are infected/recovered and on the topology of the network. Therefore, implementing an exact mapping would require to impose a different SD level on each individual, depending on the degree of its social interactions. Although well defined in theory, this is clearly not implementable in practice. To overcome this issue, we introduce the so-called mean-field approximation [Pastor-Satorras et al., 2015; Kiss et al., 2017]. On average, an infected node is connected to $\langle k \rangle$ neighbors, of which we assume that a proportion S/N is susceptible - where $\langle \cdot \rangle$ is the expected value operator. Hence, we set the number of $S - I$ links as $\langle k \rangle IS/N$. We derive β_{n} as a simple linear function of β

$$\beta_{\mathrm{n}} I(t) \langle k \rangle \frac{S(t)}{N} \simeq \frac{\beta}{N} I(t) S(t) \Rightarrow \beta_{\mathrm{n}} \simeq \frac{\beta}{\langle k \rangle}. \tag{6.12}$$

This is a valid first-order approximation, that is known to give an upper estimate of the true $S - I$ link count (see [Kiss et al., 2017; Pastor-Satorras et al., 2015]), which in our case can only translate in a more conservative control strategy. This expression connects a SIR model (6.1) to a stochastic SIR on networks, rather than a stochastic SEIRD, as we want. Hence, we need to add an additional layer that conciliates γ_{E} and γ_{I} with γ in the SIR model. To do so, we first consider the time to full recovery (or death) of an individual who has been infected in a SEIRD model. This is a random variable exponentially distributed with rate $\gamma_{\mathrm{E}}\gamma_{\mathrm{I}}/(\gamma_{\mathrm{I}} + \gamma_{\mathrm{E}})$. We set the controller γ to this value. To find the infection rate, we use the definition of

R_0 [Kiss et al., 2017] for both models, i.e. $R_0 = \beta/\gamma$ for the SIR, and $R_0 = \tilde{\beta}/\gamma_{\mathrm{I}}$ for SEIRD (we momentarily use $\tilde{\beta}$ to distinguish it from the β in the SIR), and we impose that they are equal. This yields $\tilde{\beta} = \beta(\gamma_{\mathrm{I}} + \gamma_{\mathrm{E}})/\gamma_{\mathrm{E}}$. Finally, combining this expression with (6.12), gives β_{n} as

$$\beta_{\mathrm{n}} = \frac{\tilde{\beta}}{\langle k \rangle} = \frac{\beta}{\gamma_{\mathrm{E}}} \frac{\gamma_{\mathrm{I}} + \gamma_{\mathrm{E}}}{\langle k \rangle}. \tag{6.13}$$

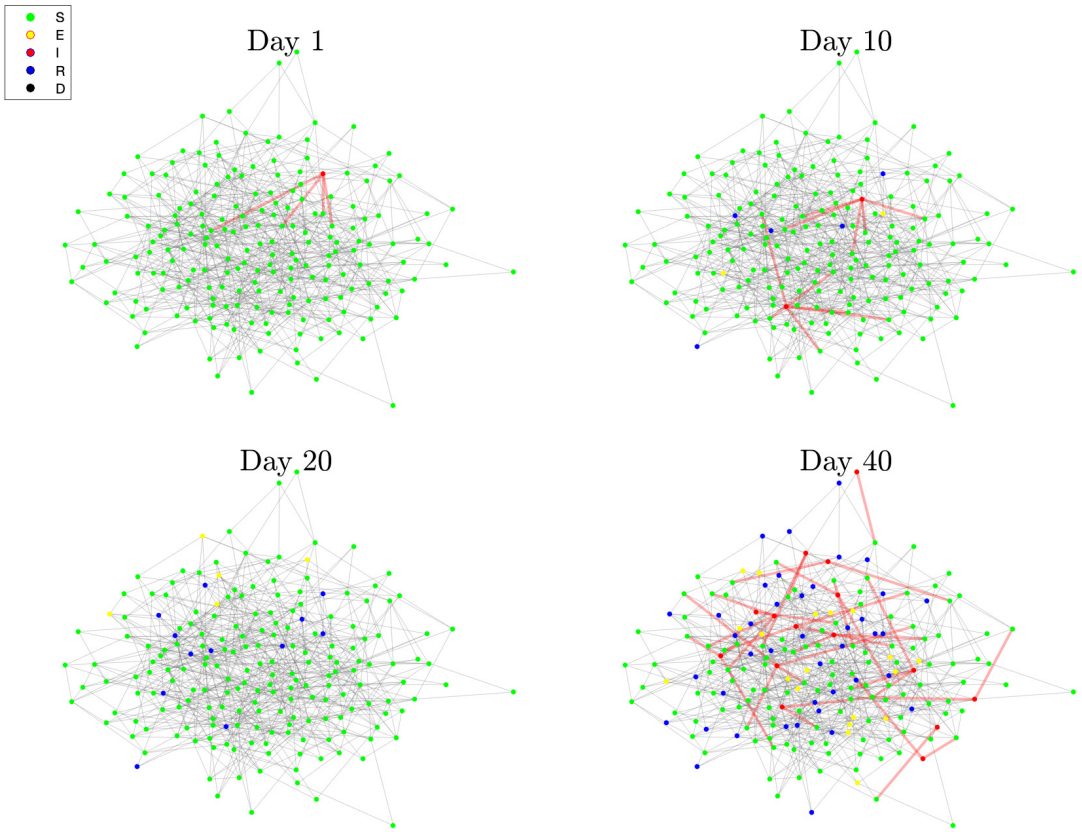

Figure 6.5: Simulation of a SEIRD outbreak on a Erdős-Rényi network of size 200, with average degree $E[k] = 7$. A single node at day 1 spreads infection to its neighbors (red edges), which in turn become first exposed, then infected, and eventually recover or die. The network is drawn in such a way that nodes with fewer links are on the periphery. The effect of the topology on the disease is particularly evident on such nodes, as only a few of them gets infected compared to central ones.

6.5 Simulations

On top of the complexity introduced by the network dynamics, we consider several non-ideal behaviors to better approximate a real-world scenario. Note that none of

Table 6.1: Parameters used for simulations in Sec. 6.5.

$\beta_{\max}$	0.0227	$\beta_{\min}$	0.0057
γ_{E}	0.25	delay (days)	$\mathcal{N}(\{3,7,20\},1)$
γ_{I}	0.1428	noise (signal)	$\mathcal{N}(0,0.1)$
N	16000	Hospitalization rate	0.02
T_f (days)	240	$\imath_{th}$ (%)	0.025
I_0	800	p_{D} if $\imath \leq \imath_{th}$	0.005
S_0	15200	p_{D} if $\imath \leq \imath_{th}$	0.02
$\langle k \rangle$	19	policy update (days)	$\{1,7,15\}$

these effects are considered in the controller design, and therefore are to be seen as uncertainties.

- Unknown random delay affects measurements, which changes every time the controller is executed. This models the difficulties in getting on-line estimates of prevalence through daily swab tests.

- Policy update is allowed at a fixed rate, to mimic real life scenarios in which policy makers are reluctant to apply different degrees of restrictions too frequently.

- Quantization of the possible levels of the network control level β_{n}, based on Italian mobility data [Pepe et al., 2020]. Policy makers can realistically implement only limited control actions. We use 5 distinct, equally spaced, levels, from $\beta_{\min} > 0$ to $\beta_{\max}$. We set $\beta_{\min} = 0.25\beta_{\max}$.

- We introduce measurement noise of the signal, proportional to its value, to model uncertainty in the estimation of the prevalence when the epidemic is out of control.

For the tuning of the model parameters we consider the case of Codogno, which has been the first city in Lombardy with a diagnosed case of Covid-19. We have used Google data for the number of people in Codogno and the hospital capacity. We considered realistic parameters for incubation period [Ling et al., 2020], infectious period [Ling et al., 2020], hospitalization rate [World Health Organisation, 2020],

infection fatality rate [Meyerowitz-Katz and Merone, 2020a], and social network connectivity [Melegaro et al., 2011]. All the parameters are reported in Table 6.1. The initial condition is set to $I_0 = 800$, to model a delayed recognition of the presence of the disease, and simulations are run for $T_f = 240$ days.

Figs. 6.6 and 6.7 show the evolution of infected $\imath$, deaths D/N, and prescribed SD $\beta(s, t, \imath)$, for the case where policy can change once a week and delay between testing and results is on average 4 days. We report the results when using the proposed feedback action $\beta(t, s, \imath)$, and, as comparison, the evolution of the uncontrolled epidemics ($\beta_{\text{n}} = \beta_{max}$) and of a one on-off intervention lasting for 60 days, during which $\beta_{\text{n}} = \beta_{\text{min}}$. Susceptible percentages s are not shown for the sake of space. We aggregate results from 100 simulations, each one run on a different network realization.

We evaluate the performances of the controller in various settings, in which we act on two main parameters, namely, the delay in knowledge of the status and the frequency at which the control policy can be changed. The former one can take values of $\{3, 7, 20\}$ days, while the latter moves between $\{1, 7, 15\}$ days. We consider all the possible combinations of these parameters. We cannot report here the complete results of our simulations, for the sake of space. We report instead some relevant performance indexes in Fig. 6.8 - namely the reduction in social distancing compared to 60 days full lockdown, and reduction in deaths with respect to not applying any strategy. We observe that the controller performs well on average even in the most extreme cases. Yet, we observe increased dispersion as we increase delays and reaction times. The use of the controller consistently induces a reduction of over 32% of deaths in the worst case, and, in the best tested case, of 63%.

6.6 Discussion

Our approach resulted in a strategy able to keep the curve below the health care capacity when uncertainty is low, with increased variability when delays and other inaccuracies in measuring become important. From this analysis, it appears clear that is crucial to have a reliable estimate of the current prevalence of the disease. This is of course the downside of closed loop strategies, i.e. that the controller

becomes less reliable as the quality of measurements deteriorates. Instead, it is worth noting that, given low delays in data, updating policies every 15 days has a limited impact on the performance of the controller. Interestingly, increasing delays (or control updates frequency), does not have a major impact on the average performance of simulations, in terms of reduction of mortality. However, this result might be misleading, because the variance between different realisations gets higher as the delay increases, meaning that the controller becomes unreliable if applied to an individual realisation. This suggests that the crucial quantity for control is on-line prevalence estimation. Therefore, this analysis confirms that, when implementing control policies based on daily testing data, policy makers should ideally have access to the exact state of the system. Clearly, this is far from being a realistic assumption. Still, our results demonstrate that periodic loop closure might be a viable solution also in a more realistic scenario.

At the same time, we observe a relevant outcome in all our simulations, namely that when control acts on an outbreak that has already reached a significant proportion of the population, the advisable strategy is to go into full lockdown until the epidemic curve is brought down to acceptable levels, and then to gradually relax and adjust control measures, according to the estimated prevalence.

6.7 Conclusions and Future Work

This preliminary work showed that a simple feedback action can improve the robustness and the effectiveness of an optimal policy for epidemic control, even in presence of quite non ideal behaviors in the system and in measuraments. The effectiveness of strategies based on control for dealing with epidemics is still an open topic, with respected academics having opposite positions [Casella, 2021; Nowzari et al., 2016]. Although our results are far from being readily applicable, they provide a new piece to this intricate puzzle. Future work will be devoted to use more reliable input maps (and possibly theoretical models for the controller), improve control design with robust and adaptive techniques, include other sources of lags and uncertainties, use more realistic network models - possibly dynamic networks, the ultimate goal being engineering a sound model that could be useful when it comes to decision making

for governments.

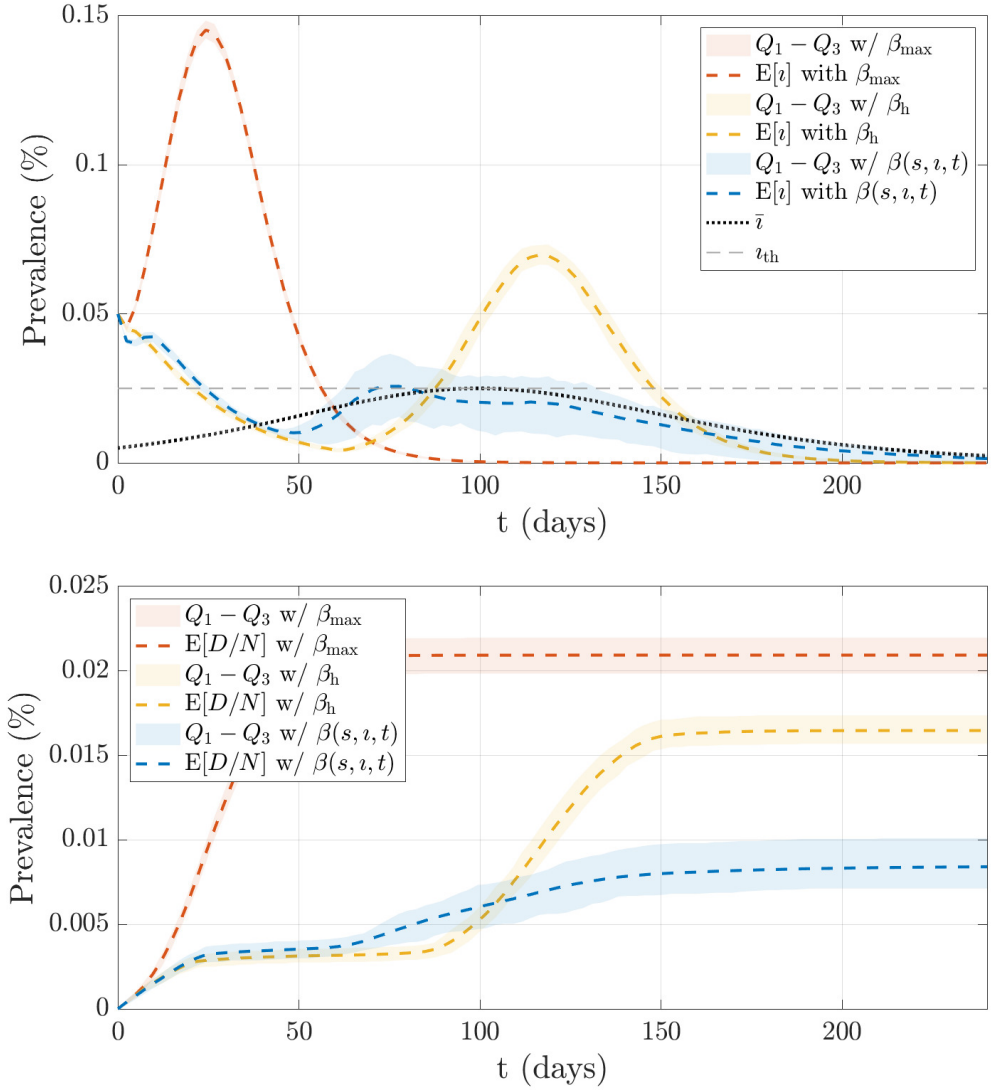

Figure 6.6: Prevalence of infected (left) and dead (right) nodes for the considered simulation scenario. It is shown here the case in which the policy changes only once every week, and the average delay in measurements is set to 3 days. All the other values are as in Tab. 6.1.

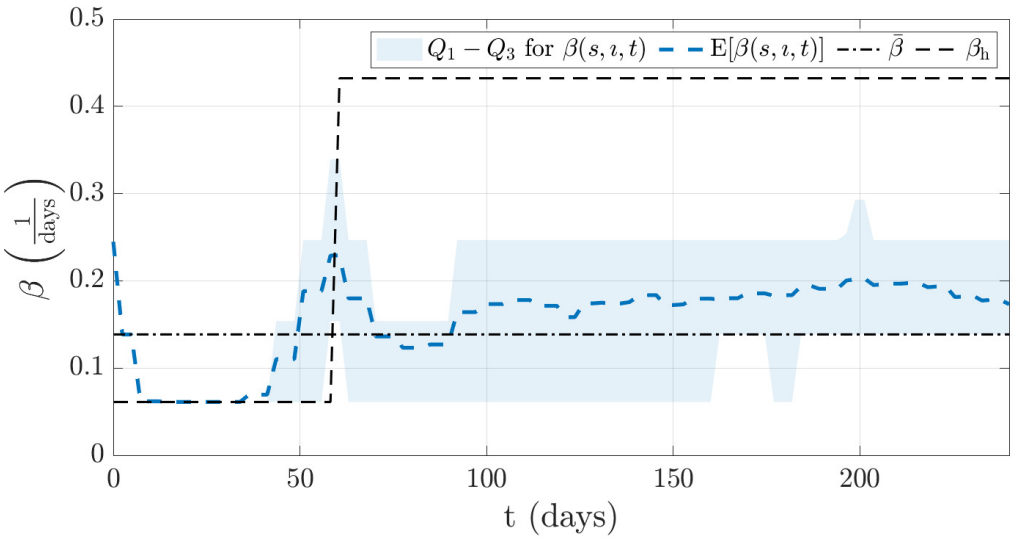

Figure 6.7: Level of SD $\beta(t, s, \imath)$ as a function of time. The average output of the controller across 100 simulations, when policy changes every 7 days and delays in data are 3 days, is shown together with its lower and upper quartiles (Q1-Q3). We also report for comparison a 60 days full-lockdown strategy, and the feedforward action $\bar{\beta}$.

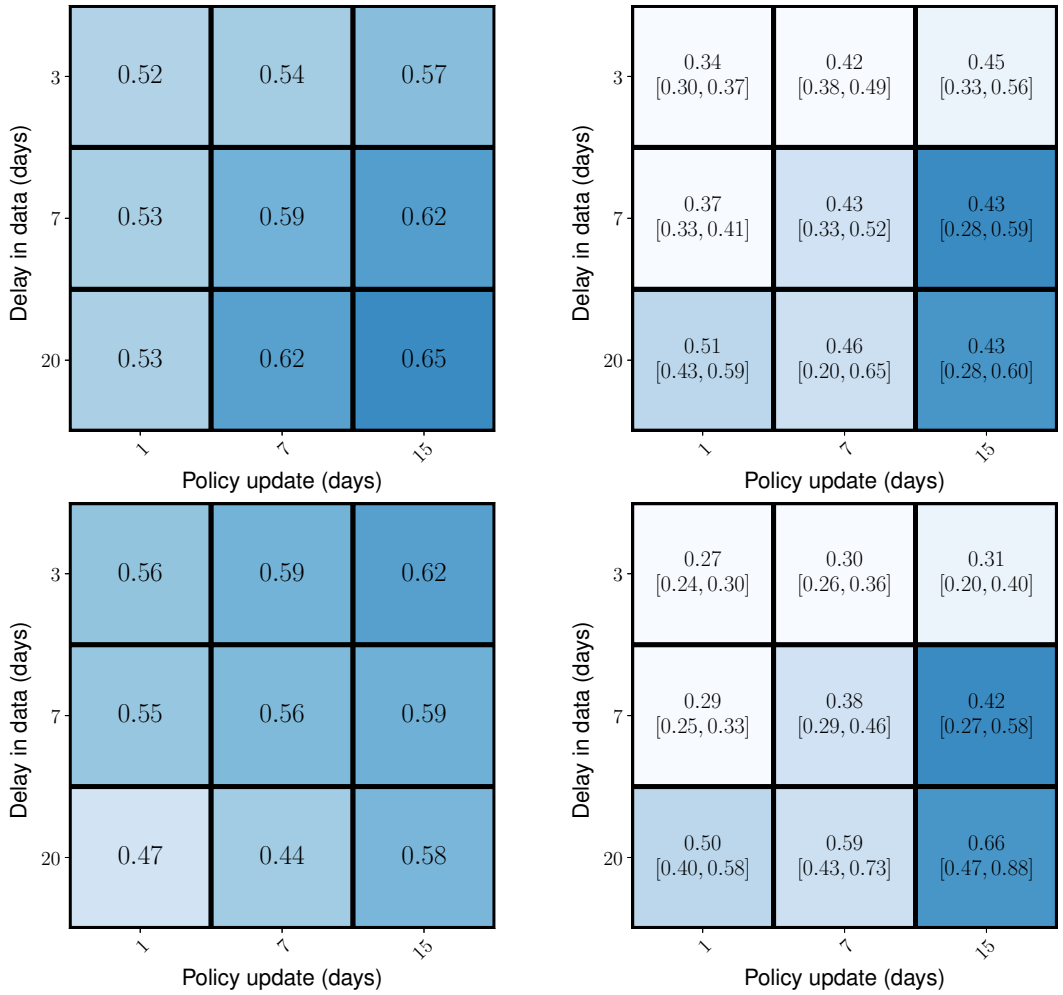

Figure 6.8: Heat maps reporting (a-c) the average reduction in β_{n}, normalised by $\int_0^{T_f} \beta_{\mathrm{h}}(t)\mathrm{d}t$ for the scenarios in which $\imath \neq \bar{\imath}(0)$ and $\imath = \bar{\imath}(0)$, respectively and (b-d) the average reduction in deaths, normalised by the average number of deaths, for the scenarios in which $\imath \neq \bar{\imath}(0)$ and $\imath = \bar{\imath}(0)$, respectively. Colors in (b) follow the width of the fist and third percentile (reported in the cells under the average). Both the indices are defined so that the smaller the better.

Chapter 7

Discussion

An epidemic spreading through a population is a complex phenomenon, whose formalisation often results in forbiddingly high dimensional models. Usually, mathematical models rely on approximations that keep the most fundamental properties of the underlying system. Although the foundation of epidemiological models were based on homogeneous mixing, i.e. they assumed that everyone is in contact with everyone, it became soon clear that knowledge of the network of contacts is fundamental to determine the evolution of epidemics; hence, networks and epidemiology are fundamentally linked. In this framework, epidemics are formalised as random processes taking places on a network, in which infection is carried through links between infected nodes and susceptible nodes. The number of equations needed to describe such systems grows exponentially with their size; hence, many models have been proposed to reduce the dimensionality of the system [Kiss et al., 2017], by focusing on aggregate statistics. Mean-field approximations are well-established methods to derive ordinary differential equations (ODEs) for the expected number of infected nodes, while explicitly including the most important properties of the underlying network. In this thesis, I have presented five pieces of research, that enter in this framework at different levels. Chapters 2,3 introduce and develop the Birth-and-Death approximation of epidemics on networks, a method to reduce the dimensionality of the problem (from 2^N equations to $N + 1$ equations) and at the same time preserve the stochasticity typical of the true process. This approximation is employed to perform network inference from population level data (Chapter 2), and it is extended to include large-networks limit, which results in a one-dimensional

Fokker-Planck equation (Chapter 3). Chapters 4,5,6 consider low-dimensional models for epidemic control, explicitly inspired by Covid-19. Different models are employed to address different key questions, and are displayed in order of complexity. Initially, we modelled one-shot interventions, such as lockdowns, as changes in the transmission rate for (metapopulation) SIR models, to address questions regarding how to optimally time intervention to control epidemics according to different goals (Chapter 4). Then, different mean-field like approximations that explicitly model the underlying network structure are considered when deriving the herd-immunity threshold for SIR diseases (Chapter 5). Finally, the flattening-the-curve strategy employed by many countries during this year of pandemic is modelled by means of feedback control theory, and tested on explicit network simulations, informed with parameters taken from the literature (Chapter 6).

Here, I will discuss the research and results presented in this thesis, as well as considering potential improvements to each piece of work and open questions.

7.1 Birth-and-Death approximation and network inference

7.1.1 Summary of contributions

Mean-field-like approximations result in systems ODEs whose solution describes the expected epidemic curves [Kiss et al., 2017]. The choice of which mean-field model to employ typically depends on which features of the underlying network are thought to have a non negligible impact on the epidemic dynamics. Reproducing the expected epidemic curves with low-dimensional models whose derivation explicitly considers the underlying topology has great value on its own, but it has a few limitations. Indeed, epidemics are intrinsically stochastic processes - especially in their initial phase; therefore it is desirable to have low-dimensional models capable of describing the variability of different realisations of the same process. Except for a few remarkable cases, such as the fully connected networks, no general theory exists to efficiently compute the evolution of an epidemic on a network.

The first piece of research, in Chapter 2, considers a novel conjecture in net-

work epidemiology. We approximate epidemics on networks by Birth-and-Death processes, characterised by a system of $(N+1)$ master equations, whose rates encode both the structure of the underlying network and the disease parameters. The solution of the master equation describes the probability distribution of the number of infected people for a SIS epidemic on a network. Extensive simulations on well-known network classes, namely Regular, Erdős-Rényi, and Barabási-Albert, showed good agreement with the master equations for the approximated model, for various choices of network and epidemic parameters. This conjecture holds for various network sizes as well, as shown in Chapter 3.

Real world epidemics spread on often only partially-known networks of contacts. This is particularly problematic because many advanced theoretical models often rely on full knowledge at least of global properties of the network. Network inference from the observation of epidemics usually is done at a local level [Britton and O'Neill, 2002; O'Neill and Roberts, 1999], meaning that it is possible only if one is able to continuously observe the status of the nodes of a network. The aim is usually to infer the whole topology, that is, the adjacency matrix of the network. In [Groendyke et al., 2011], the aim is instead to infer the generative parameter of a known network class rather than the adjacency matrix; however, the data considered rely on having precise information at a node-level. Unfortunately, it is rarely possible to observe individuals continuously in real-world scenarios, where the most common information available is often given in terms of discrete time aggregate statistics, such as the daily number of infected people. The Birth-and-Death process approximation proves particularly useful in this realistic setting. The rates of the master equations are network-class dependent, therefore techniques from inverse problems applied to Birth-and-Death process can be applied to recover information about the network class. In Chapter 2, this is done in a Bayesian framework. Priors over 3 well-known network classes (Regular, Erdős-Rényi, and Barabási-Albert) are computed numerically from simulations of many different network and epidemic parameters combinations. The likelihood can be efficiently computed from standard techniques as described in [Crawford et al., 2014], therefore the posterior for each network class can be numerically estimated, and model selection is performed with the MAP estimator. This method successfully recovers the most likely network

classes reliably, from discrete time observations of the population-level counts, and it is validated both on synthetic and real networks. In the latter case, the outcome of the inference scheme is a class among the three that we have considered, so we interpret this result as the closer network that might have generated the data.

In a follow-up research, described in Chapter 3, we considered the large N limit of epidemics on network. It is known [Janson et al., 2014] that certain mean-field approximations become exact in the large network size limit. However, such models retain the same limitations described in the previous paragraphs, and a probabilistic description of the epidemic is potentially more useful both theoretically and for applications. In this Chapter, the Birth-and-Death approximation is studied in the limit of large N, resulting in a Partial Differential Equation (PDE) limit to SIS epidemics on networks, in the form of a one-dimensional Fokker-Planck equation, whose drift and diffusion coefficients depend strongly on the underlying network structure and epidemic parameters. In order to get a PDE-limit, we verified that the rates of the Birth-and-Death process are density dependent, i.e. they can be written as a function of the prevalence $\frac{I}{N}$ only. This was numerically validated for a wide range of networks classes, including lattices and scale-free networks.

A PDE that describes the evolution of an epidemic on a large network is useful not only because it reduces greatly the numerical complexity of the system, but also because it allows to use PDE inverse problem techniques to infer information on the underlying network structure from the observation of an epidemic. Instead of the master equation, that is numerically expensive to compute for networks of large size, we showed in a fully worked out example how to efficiently compute the likelihood using the Fokker-Planck equation. The inference scheme is analogous to the one described in Chapter 2, where instead of the master equation of a birth-and-death model, we have used the Fokker-Planck equation resulting from the PDE limit.

7.1.2 Limitations and future works

This concludes the first part of my thesis, where the focus is to validate a new approximation method and apply it for network inference. There are a few limitations and potential new lines of research that are well worth of exploring. Perhaps, the most important question to answer is whether it is possible to formally prove

that Birth-and-Death processes approximate population-level epidemics. I believe that at least bounds for this approximation can be derived by studying convergence methods for random processes. In terms of numerical results, it would be important to extend and validate what has been obtained on more realistic networks - maybe using non-parametric formulations of the rates. In a non-published piece of research, we tried to extend this to stochastic block-models, and observed that rates show the presence of communities in the underlying network. This might potentially become a tool to make community inference, but unfortunately we did not have time to carry on this project. Another line of research that I hope to study in the future is to extend our results to different epidemic models, such as SIR (or even SEIR). The main problem in these cases is that, whereas on a SIS rates are one-dimensional (the population-level status is determined once the number of infected is known), in a SIR model, one needs a rate for any possible combination of S and I, that is, the number of susceptible and the number of infected. This would therefore result in a two-dimensional manifold that might prove difficult to study. In principle, the approximations remain valid, but further research needs to be carried out to ascertain this. From a technical point of view, we have used a three-parameter model that fits quite well rates on random networks such as Regular, Erdős-Rényi and Barabási-Albert. It might be desirable to use a non parametric model to fit such rates, as it would make the whole inference scheme more robust and generalisable to many other network classes. Of course, this can be done only after having answered a rather philosophical question: to carry out our research, we assumed that networks can be divided into distinct classes, so that the problem of network inference can be recasted as a problem of network class inference. Is this still true when one considers more complex network structures? For instance, does a network with, say, 4 distinct communities belong to a different class compared to a network with 5 communities? And an Erdős-Rényi network that has been modified so that its clustering coefficient is non-negligible, can be still considered belonging to that class? After all, do network classes exist in nature, or are just models to guide researchers? In this sense then, identifying a network as a member of a certain class, might simply indicate that it shares some topological properties with that class, that might or might not be enough information, depending on the problem to solve.

7.2 Mean-field models applied to epidemic control

7.2.1 Summary of contributions

The second part of my thesis (Chapters 4-5-6) considers the application of mean-field approximations for Covid-19 response. This section embraces the research produced in year 2 and year 3 of my doctorate, as well as projects resulted from collaborations outside my academic supervisor's group.

Most of the research works in this section were inspired by the urgency of matters related to Covid-19 response, so it is useful to frame my research works within the chronology of the scientific debate around Covid-19 measures. Note that all the works in this part have been produced between February, 2020 and October, 2020. The scientific consensus at the beginning of the epidemic was that epidemic control cannot rely on lockdowns or prolonged social distancing measures. Indeed Western countries initially excluded even the possibility of a lockdown, until Italy led the way [Tom Whipple, 2020]. Thus, a number of papers from early 2020 studied the timing at which time-limited lockdowns should be implemented [Morris et al., 2021; Perkins and España, 2020]. For instance, in [Morris et al., 2021], the problem of how to theoretically derive an optimal strategy for using a short lived intervention to reduce peak prevalence is answered, using a SIR model. Our approach, described in Chapter 4, enters this discussion by studying the issue from different perspectives. In particular, we consider how different goals impact optimal timings of intervention. In our research, three different strategies are considered: (i) to delay the epidemic as much as possible, (ii) to minimise the final size, (iii) to minimise the peak of the prevalence. These three questions are addressed using the SIR model, and are then extended to metapopulation SIR, to mimic an epidemic spreading on different subcommunities of a population. Additionally, we have considered how the intervention is employed, either globally or locally, and we find that it is best to target the intervention asynchronously to each subcommunity rather than synchronously across the whole population.

We explored different combinations of $\mathcal{R}_0$, control duration, and control strength, and found that results are quite general: if the goal is to delay the epidemic as much

as possible, while waiting for future pharmaceutical solutions, then the best time to introduce a lockdown is as early as possible. Instead, if the ultimate objective is to reduce the cumulative number of infected people, then the intervention needs to be employed slightly before the peak, the precise time depending on how strong the intervention can be. Finally, if the goal is to minimize the peak, then the optimal intervention is employed partway through the growth phase, to allow some immunity to build up so that the eventual rebound at the end of control is not larger than the initial peak.

More or less at the same time, it became clearer that complete eradication would be out of reach for the vast majority of countries in the World, so determining the final size of Covid-19 with precision became a hot topic in research. One key observation is that the final size depends on the value of herd-immunity threshold, which ultimately depends on the underlying model assumptions. In a simple SIR model informed with early estimates for Covid-19 ($\mathcal{R}_0 \sim 3$) the herd-immunity threshold is roughly 66% of the whole population. This is important as, if eradication is impossible, at least these many people need to be immunized before the epidemic ends - unless we consider that non-pharmaceutical interventions can be prolonged forever. Therefore, determining the exact level of herd-immunity threshold is of crucial importance. The SIR model has been deemed as too simplistic by a number of researchers [Britton et al., 2020; Gomes et al., 2020], who showed how heterogeneity in the contact structure or in the individual susceptibility of the population might lead to herd-immunity levels as low as 20%, when interventions are considered [Britton et al., 2020]. This is because the disease acts like a targeted vaccine, preferentially immunizing higher-risk individuals who play a greater role in transmission, and a limited first-wave might leave behind a residual population that cannot support a second wave. Of course, one major criticism that can be levelled at such models is that they are just slightly less simplistic than the SIR in their assumptions. Nonetheless, these results remark the importance of model choice when investigating such fundamental questions. In Chapter 5, we explore a class of epidemic models coming from network science, with the aim of showing explicitly how model selection determines strongly any result regarding herd immunity levels. We consider degree-based heterogeneous mean-field, heterogeneous pairwise,

clustered heterogeneous pairwise and edge-based compartmental models on different networks [Kiss et al., 2017]. An important finding is that the way measures are implemented in various models determines strongly the value of herd-immunity. In many works, it is assumed that during lockdown the contact network is not changing, but rather the transmission rate [Gomes et al., 2020; Morris et al., 2021; Britton et al., 2020]. In these cases, any source of heterogeneity indeed reduces the herd-immunity threshold. However, if models allow to explicitly consider the network structure to model interventions as changes in its topology, we show that the effect becomes more subtle. This is confirmed also by considering realistic scenarios taken from well-recognised institutions data and published research. In general, the more complex the model, the stronger its predictive power. The trade-off is that the information required to inform complex models might be difficult to extract from real-world data, potentially leading to more unrealistic conclusions than simplistic models informed correctly.

The paper that constitutes Chapter 6 has been conceived months later than the former two, when it became clearer that the one-time control intervention paradigm was obsolete, as many countries relied on repeated interventions to keep the prevalence below a certain threshold, as the vaccines were developed, approved, and finally distributed - a strategy known as flattening the curve. As such, the problem of controlling the epidemic through its whole course has been addressed by a number of research groups [Casella, 2021; Giordano et al., 2020; Köhler et al., 2020]. The challenge is to engineer a strategy that ensures that the healthcare system would not be overwhelmed during the whole course of the epidemic. However, as shown in Chapter 5, model misspecification can lead to wrong predictions, and therefore non-optimal control. We proposed a potential way to limit this problem by considering a feedback control approach. We formalised the problem of flattening the curve in terms of the SIR model with feedback, in which the control parameter is the rate of transmission. The ultimate goal is to reduce as much as possible the strength of control measures while ensuring that the number of infected people would, at any given time, remain under a certain pre-defined threshold. After putting forward a trajectory-tracking controller and proving that it is optimal on this simple system, we showed that it is robust by stress-testing it. In particular, a stochastic

SEIRD model on Erdős-Rényi networks informed with realistic data is considered as a ground-truth, to show that the model is robust to model misspecification. The original controller is then connected to the network model by means of two maps (for input and output, respectively), based on simple mean-field model arguments. On top of that, we considered the effect of delays and noises in the input data, as well as control acting only at a fixed rate, to mimic real life scenarios where policy makers are reluctant to apply different degrees of restrictions too frequently. The results show that using feedback control can improve dramatically the robustness and the effectiveness of an optimal policy for epidemic control, even in presence of non-ideal behaviours in the system and in presence of noisy,lagging measurements. Again, depending on how far is the model from reality, and on how precise and updated the information upon which decisions are made, the outcomes might differ a lot among different realisations. However, defining a control action that can be adapted to the actual prevalence and adjusted online proves to be robust to some degree of model misspecification and real-world delays/noises.

7.2.2 Limitations and future works

There are many open problems and future directions of research, and I believe that research aimed at better understanding Covid-19 and how to control epidemic spreading in a population will perhaps become the main topic in Epidemiology for the foreseeable future. In my work, for instance, I have not considered tracing apps, testing, quarantine, behavioural response, and finally vaccines. Each of these aspects may possibly be incorporated separately in relatively tractable epidemic models, but the complexity arising from considering all of these features and their interplay together possibly prevents any analytical result to be derived. For instance, as the awareness of the epidemic increases among the population and social distancing measures are employed by governments, the contact structure might change quite sensibly and abruptly, and it is not clear how to model this phenomenon correctly.

For this reason, many governments rely on several distinct groups, each one with its own model, to inform the epidemic response; for instance, in the UK, the SAGE group includes, among the others, independent models from Imperial College , Warwick University, and London School of Hygiene and Tropical Diseases. Such models

are more complex than the ones analysed in this work, and their aim is to consider a population as a sum of individuals that have different behaviours and respond differently to the disease - known as agent-based models (ABM) [Ferguson et al., 2020]. Although, in principle, ABM might be the most accurate models to describe the real epidemic, they require a vast amount of information to even be initialised. In the era of Big Data, this seems available and it is perhaps the most promising way forward. However, uncertainties/errors in data collecting/delays might add up and be overlooked. Predictions based upon these data might eventually not be correct. For this reason, simpler models that require less precise information such as the ones described in this thesis may work in synergy with these more complicated approaches to Epidemiology, providing bounds and benchmarks for more sophisticated models. At the same time, the mathematical theory of epidemiology can benefit from the study of epidemic models on networks, as theoretically rigorous advancements are still possible and well worth investigating.

Bibliography

R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000. 138

A. Allard, P.-A. Noël, L. J. Dubé, and B. Pourbohloul. Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics. *Physical Review E*, 79(3):036113, 2009. 11

L. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2, 03 2017. doi: 10.1016/j.idm.2017.03.001. 69

E. Álvarez, J. Donado-Campos, and F. Morilla. New coronavirus outbreak. lessons learned from the severe acute respiratory syndrome epidemic. *Epidemiology & Infection*, 143(13):2882–2893, 2015. 130

R. M. Anderson and R. M. May. *Infectious Diseases of Humans*. Oxford University Press, Oxford, 1991. 95, 96, 97

R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth. How will country-based mitigation measures influence the course of the covid-19 epidemic? *The Lancet*, 395(10228):931–934, 2020. 91

R. Applegate. An introduction to population genetics: Theory and application. *The Canadian Field-Naturalist*, 127:281, 11 2013. doi: 10.22621/cfn.v127i3.1475. 68

R. Armstrong et al. Outcomes from intensive care in patients with covid-19: a systematic review and meta-analysis of observational studies. *Anaesthesia*, 75 (10):1340–1349, 2020. 156

J. Badham and R. Stocker. The impact of network clustering and assortativity on epidemic behaviour. *Theoretical population biology*, 77(1):71–75, 2010. 9, 21

M. G. Baker, N. Wilson, and A. Anglemyer. Successful elimination of covid-19 transmission in new zealand. *New England Journal of Medicine*, 383(8):e56, 2020. 91

F. Ball, T. Britton, T. House, V. Isham, D. Mollison, L. Pellis, and G. S. Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67, 2015. 97, 105

S. Bansal, G. Chowell, L. Simonsen, A. Vespignani, and C. Viboud. Big data for infectious disease surveillance and modeling. *The Journal of Infectious Diseases*, 214(4):375–379, 11 2016. 6

A.-L. Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013. doi: 10.1098/rsta.2012.0375. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2012.0375. 6

A.-L. Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013. 24

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. 23

N. S. Barlow and S. J. Weinstein. Accurate closed-form solution of the sir epidemic model. *Physica D: Nonlinear Phenomena*, 408:132540, 2020. ISSN 0167-2789. doi: https://doi.org/10.1016/j.physd.2020.132540. URL https://www.sciencedirect.com/science/article/pii/S0167278920302694. 2

R. C. Barnard, L. Berthouze, P. L. Simon, and I. Z. Kiss. Epidemic threshold in pairwise models for clustered networks: closures and fast correlations. *Journal of mathematical biology*, 79(3):823–860, 2019. 11

A. Batkai, Havasi, R. Horvath, D. Kunszenti-Kovács, and P. Simon. Pde approximation of large systems of differential equations. *Operators and Matrices*, 9, 03 2013. doi: 10.7153/oam-09-08. 67, 68

P. W. Baxter and H. P. Possingham. Optimizing search strategies for invasive pests: learn before you leap. *Journal of applied ecology*, 48(1):86–95, 2011. 92

D. Bernoulli. Essai d'une nouvelle analyse de la mortalite causee par la petite verole. *Mem. Math. Phys. Acad. Roy. Sci.*, 1, 1766. 1

A. I. Bhuiyan et al. Covid-19-related suicides in bangladesh due to lockdown and economic factors: case study evidence from media reports. *International Journal of Mental Health and Addiction*, 2020. 156

G. Bianconi, H. Sun, G. Rapisardi, and A. Arenas. A message-passing approach to epidemic tracing and mitigation with apps, 2020. 124

M. Bin et al. On fast multi-shot epidemic interventions for post lock-down mitigation: Implications for simple covid-19 models. *arXiv preprint arXiv:2003.09930*, 2020. 157

B. Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001. doi: 10.1017/CBO9780511814068. 23, 165

M. C. Bootsma and N. M. Ferguson. The effect of public health measures on the 1918 influenza pandemic in US cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593, 2007. 90, 91

F. Brauer and C. Castillo-Chavez. *Mathematical Models in Population Biology and Epidemiology*. Springer, 2nd edition, 2012. 63

T. Britton and P. D. O'Neill. Bayesian inference for stochastic epidemics in populations with random social structure. *Scandinavian Journal of Statistics*, 29(3): 375–390, 2002. doi: 10.1111/1467-9469.00296. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9469.00296. 12, 13, 38, 39, 177

T. Britton, E. Pardoux, F. Ball, C. Laredo, D. Sirl, and V. C. Tran. *Stochastic epidemic models with inference*. Springer, 2019. 132

T. Britton, F. Ball, and P. Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science (New York, N.Y.)*, pages 846–849, jun 2020. ISSN 1095-9203. doi: 10.1126/science.abc6810. URL http://www.ncbi.nlm.nih.gov/pubmed/32576668. 2, 11, 16, 114, 123, 136, 138, 139, 148, 181, 182

I. Brugere, B. Gallagher, and T. Y. Berger-Wolf. Network structure inference, a survey: Motivations, methods, and applications. *ACM Computing Surveys*, 51 (2):24:1–24:39, Apr. 2018. ISSN 0360-0300. doi: 10.1145/3154524. URL http://doi.acm.org/10.1145/3154524. 38

E. Cacio, S. E. Cohn, and R. Spigler. Numerical treatment of degenerate diffusion equations via Feller's boundary classification, and applications. *Numerical Methods for Partial Differential Equations*, 28(3):807–833, may 2012. ISSN 0749159X. doi: 10.1002/num.20657. URL http://doi.wiley.com/10.1002/num.20657. 68, 72

L. Cai and Y. Zhu. The challenges of data quality and data quality assessment in the big data era. *Data science journal*, 14, 2015. 6

F. Casella. Can the covid-19 epidemic be controlled on the basis of daily test reports? *IEEE Control Systems Letters*, 5(3):1079–1084, 2021. doi: 10.1109/LCSYS.2020.3009912. 2, 170, 182

A. Charpentier et al. Covid-19 pandemic control: balancing detection policy and lockdown intervention under icu sustainability. *arXiv preprint arXiv:2005.06526*, 2020. 157

L. Chen and J. Sun. Optimal vaccination and treatment of an epidemic network model. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 378 (41):3028–3036, aug 2014. ISSN 03759601. doi: 10.1016/j.physleta.2014.09.002. 124

M. Chen, C. Liu, S. Xu, X. Yue, and R. Zhang. Behavior of different numerical schemes for population genetic drift problems. *arXiv preprint arXiv:1410.5527*, 2014. 68, 72, 82, 84, 85, 86

S. Cobey. Modeling infectious disease dynamics. *Science*, 368(6492):713–714, 2020. 90

R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical review letters*, 91(24):247901, 2003. 9, 138

P. Collet, S. Martínez, and J. San Martín. *Quasi-Stationary Distributions: General Results*. Springer, 2013. doi: 10.1007/978-3-642-33131-2_2. 68

R. M. Corless et al. On the lambert w function. *Advances in Computational mathematics*, 5(1):329–359, 1996. 160, 161

F. W. Crawford and M. A. Suchard. Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution. *Journal of Mathematical Biology*, 65:553–580, 2012. doi: 10.1007/s00285-011-0471-z. 60

F. W. Crawford, V. N. Minin, and M. A. Suchard. Estimation for general birth-death processes. *Journal of the American Statistical Association*, 109(506):730–747, 2014. doi: 10.1080/01621459.2013.866565.Estimation. 52, 177

L. Danon, A. Ford, T. House, M. Keeling, et al. Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives On Infectious Diseases*, 2011:28, 2010. 37, 39, 41

S. De Flora and S. La Maestra. Growth and decline of the covid-19 epidemic wave in italy from march to june 2020. *Journal of Medical Virology*, 2020. doi: 10.1002/jmv.26499. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26499. 91

L. Decreusefond, J.-S. Dhersin, P. Moyal, and V. C. Tran. Large graph limit for an SIR process in random network with heterogeneous connectivity. *The Annals of Applied Probability*, 22(2):541–575, 2012. 64, 133

F. Della Rossa, D. Salzano, A. Di Meglio, F. De Lellis, M. Coraggio, C. Calabrese, A. Guarino, R. Cardona, P. DeLellis, D. Liuzza, et al. Intermittent yet coordinated regional strategies can alleviate the covid-19 epidemic: a network model of the italian case. *arXiv preprint arXiv:2005.07594*, 2020. 93

F. Di Lauro, J.-C. Croix, M. Dashti, L. Berthouze, and I. Z. Kiss. Network inference from population-level observation of epidemics. *Sci Rep*, 10, November 2020a. doi: https://doi.org/10.1038/s41598-020-75558-9. xviii, 64, 65, 66, 67, 71, 77, 78, 79, 82

F. Di Lauro, M. Dorey, J. C. Miller, L. Berthouze, and I. Z. Kiss. The impact of network properties and mixing on control measures and disease-induced herd immunity in epidemic models: a mean-field model perspective. *arXiv preprint arXiv:2007.06975*, July 2020b. 157

F. Di Lauro, I. Z. Kiss, and J. C. Miller. Optimal timing of one-shot interventions for epidemic control. *PLoS Comput Biol.*, March 2021. doi: https://doi.org/10.1371/journal.pcbi.1008763. 93, 123, 136, 148, 157

M. Dickison, S. Havlin, and H. E. Stanley. Epidemics on interconnected networks. *Physical Review E*, 85(6):066109, 2012. 97, 105

O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of The Royal Society Interface*, 7(47):873–885, June 2010. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2009.0386. 131

Djidjou-Demasse et al. Optimal covid-19 epidemic control until vaccine deployment. *medRxiv*, 2020. 157

S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Ising model on networks with an arbitrary distribution of connections. *Physical Review E*, 66(016104), 2002. 11, 23

N. Du, L. Song, M. Yuan, and A. J. Smola. Learning Networks of Heterogeneous Influence. *Advances in Neural Information Processing Systems 25*, pages 2780–2788, 2012. ISSN 10495258. 38

C. Duan, C. Liu, C. Wang, and X. Yue. Numerical complete solution for random genetic drift by energetic variational approach. *ESAIM: Mathematical Modelling and Numerical Analysis*, 03 2018. doi: 10.1051/m2an/2018058. 68

R. Dutta, A. Mira, and J. Onnela. Bayesian inference of spreading processes on networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2215):20180129, 2018. 38, 39

K. Eames. The influence of school holiday timing on epidemic impact. *Epidemiology & Infection*, 142(9):1963–1971, 2014. 90

S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009. 67

R. Eymard, T. Gallouët, and R. Herbin. Finite Volume Methods. In J. L. Lions and P. Ciarlet, editors, *Solution of Equation in Rn (Part 3), Techniques of Scientific Computing (Part 3)*, volume 7 of *Handbook of Numerical Analysis*, pages 713–1020. Elsevier, 2000. doi: 10.1016/S1570-8659(00)07005-8. URL https://hal.archives-ouvertes.fr/hal-02100732. 84

W. Feller. Diffusion processes in one dimension. *Transactions of the American Mathematical Society*, 77(1):1–31, 1954. ISSN 00029947. URL http://www.jstor.org/stable/1990677. 68

N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. Van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. Gt Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, and A. C. Ghani. of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand. *Report 9: Imperial College Covid-19 Response Team*, 2020. doi: 10.25561/77482. URL https://doi.org/10.25561/77482. 8, 184

M. J. Ferrari, S. Bansal, L. A. Meyers, and O. N. Bjørnstad. Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society B: Biological Sciences*, 273(1602):2743–2748, 2006. 149

C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*, volume 13 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, third edition, 2004. ISBN 3-540-20882-8. 67

J. Gevertz, J. Greene, C. H. S. Tapia, and E. D. Sontag. A novel covid-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. *medRxiv*, 2020. 93

D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403 – 434, 1976. ISSN 0021-9991. doi: https://doi.org/10.1016/0021-9991(76)90041-3. URL http://www.sciencedirect.com/science/article/pii/0021999176900413. 48, 74

D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, pages 1–6, 2020. 93, 157, 182

J. P. Gleeson. Bond percolation on a class of clustered random networks. *Physical Review E*, 80(3):036107, 2009. 38

M. G. M. Gomes, R. Aguas, R. M. Corder, J. G. King, K. E. Langwig, C. Souto-Maior, J. Carneiro, M. U. Ferreira, and C. Penha-Goncalves. Individual variation in susceptibility or exposure to sars-cov-2 lowers the herd immunity threshold. *medRxiv*, 2020. 114, 123, 138, 148, 181, 182

M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 1019–1028, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0055-1. doi: 10.1145/1835804.1835933. 38, 39

M. Gomez Rodriguez, J. Leskovec, D. Balduzzi, and B. Schölkopf. Uncovering the structure and temporal dynamics of information propagation. *Network Science*, 2(1):26–65, 2014. ISSN 20501250. doi: 10.1017/nws.2014.3. 13, 38

A. Goolsbee and C. Syverson. Fear, lockdown, and diversion: comparing drivers of pandemic economic decline 2020. Technical report, National Bureau of Economic Research, 2020. 156

A. Gray, D. Greenhalgh, L. Hu, X. Mao, and J. Pan. A stochastic differential equation sis epidemic model. *SIAM Journal of Applied Mathematics*, 71(3):876–902, 2011. ISSN 0022-3999. 69

C. Groendyke, D. Welch, and D. R. Hunter. Bayesian inference for contact networks given epidemic data. *Scandinavian Journal of Statistics*, 38(3):600–616, 2011. doi: 10.1111/j.1467-9469.2010.00721.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2010.00721.x. 38, 39, 177

A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*, volume 2008, pages 11–16, 2008. 22

A. Handel, I. M. Longini, and R. Antia. What is the best control strategy for multiple infectious disease outbreaks? *Proceedings of the Royal Society B: Biological Sciences*, 274(1611):833–837, mar 2007. ISSN 14712970. doi: 10.1098/rspb.2006.0015. URL /pmc/articles/PMC2093965/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2093965/. 123

J. Haushofer and C. J. E. Metcalf. Which interventions work best in a pandemic? *Science*, 368(6495):1063–1065, 2020. 93

H. W. Hethcote and P. van den Driessche. Some epidemiological models with nonlinear incidence. *Journal of Mathematical Biology*, 29(3):271–287, Jan 1991. ISSN 1432-1416. doi: 10.1007/BF00160539. URL https://doi.org/10.1007/BF00160539. 44

V. M. Hoang, H. H. Hoang, Q. L. Khuong, N. Q. La, and T. T. H. Tran. Describing

the pattern of the covid-19 epidemic in vietnam. *Global health action*, 13(1): 1776526, 2020. 91

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983. xv, 22, 82

T. D. Hollingsworth, D. Klinkenberg, H. Heesterbeek, and R. Anderson. Mitigation strategies for pandemic influenza a: balancing conflicting policy objectives. *PLoS computational biology*, 7(2), 2011. 91

P. Holme and N. Litvak. Cost-efficient vaccination protocols for network epidemiology. *PLoS Computational Biology*, 13(9), sep 2017. ISSN 15537358. doi: 10.1371/journal.pcbi.1005696. URL http://arxiv.org/abs/1612.07425http://dx.doi.org/10.1371/journal.pcbi.1005696. 12, 124

P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, 2012. 6

T. House and M. Keeling. The impact of contact tracing in clustered populations. *PLoS Computational Biology*, 6, 2010. xxiv, 133, 135

T. House and M. Keeling. Insights from unifying modern approximations to infections on networks. *Journal of the Royal Society Interface*, 8(54):67–73, 2011. 128

A. Iwasaki and N. D. Grubaugh. Why does japan have so few cases of covid-19? *EMBO Molecular Medicine*, 12(5):e12481, 2020. 91

S. Janson, M. Luczak, and P. Windridge. Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Structures & Algorithms*, 2014. 11, 64, 133, 178

J. L. Juul, K. Græsbøll, L. E. Christiansen, and S. Lehmann. Fixed-time descriptive statistics underestimate extremes of epidemic curve ensembles. *Nature Physics*, pages 1–4, 2020. 107, 120, 133, 134

N. V. Kampen. *Stochastic processes in physics and chemistry*. North Holland, 2007. 64

M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266 (1421):859–867, 1999. 63

M. J. Keeling and K. T. D. Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005. 37

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4, Nov 1995. doi: 10.1109/ICNN.1995.488968. 45

W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Royal Society of London Proceedings Series A*, 115:700–721, Aug. 1927. xiii, 1, 3, 37

J.-H. Kim, J. A.-R. An, P.-k. Min, A. Bitton, and A. A. Gawande. How south korea responded to the covid-19 outbreak in daegu. *NEJM Catalyst Innovations in Care Delivery*, 1(4), 2020. 91

I. Z. Kiss and D. M. Green. Comment on "properties of highly clustered networks". *Physical Review E*, 78(4):048101, 2008. 133

I. Z. Kiss, J. C. Miller, and P. L. Simon. *Mathematics of Epidemics on Networks: from exact to approximate models*. Springer, 2017. ISBN 978-3-319-50806-1. xii, xvi, 2, 6, 11, 27, 28, 29, 30, 31, 33, 35, 37, 38, 41, 44, 46, 63, 67, 68, 69, 73, 124, 125, 128, 131, 134, 139, 156, 158, 162, 163, 165, 166, 167, 175, 176, 182

M. Kivelä, A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014. 6

J. Köhler et al. Robust and optimal predictive control of the covid-19 outbreak. *arXiv preprint arXiv:2005.03580*, 2020. 157, 182

R. M. Kovacevic. Stochastic contagion models without immunity: their long term behaviour and the optimal level of treatment. *Central European Journal of Operations Research*, 26(2):395–421, jun 2018. ISSN 16139178. doi: 10.1007/s10100-018-0526-y. 68, 84

A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, and the LSHTM CMMID COVID-19 working group. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*, 2020. 91, 98

J. Kunegis. Konect – the koblenz network collection. *arXiv:1402.5500*, 2017. 58

T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of applied Probability*, 7(1):49–58, 1970. 67

V.-P. La, T.-H. Pham, M.-T. Ho, M.-H. Nguyen, K.-L. P Nguyen, T.-T. Vuong, T. Tran, Q. Khuc, M.-T. Ho, Q.-H. Vuong, et al. Policy response, social media and science journalism for the sustainability of the public health system amid the covid-19 outbreak: The vietnam lessons. *Sustainability*, 12(7):2931, 2020. 91

C. Lee and D. J. Wilkinson. A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1):1–50, 2019. xv, 22

J. Lindquist, J. Ma, P. van den Driessche, and F. H. Willeboordse. Effective degree network disease models. *Journal of Mathematical Biology*, 62(2):143–164, 2011. 63

Y. Ling, S. B. Xu, Y. X. Lin, D. Tian, Z. Q. Zhu, F. H. Dai, F. Wu, Z. G. Song, W. Huang, J. Chen, B. J. Hu, S. Wang, E. Q. Mao, L. Zhu, W. H. Zhang, and H. Z. Lu. Persistence and clearance of viral RNA in 2019 novel coronavirus disease rehabilitation patients. *Chinese medical journal*, 133(9): 1039–1043, may 2020. ISSN 25425641. doi: 10.1097/CM9.0000000000000774. URL /pmc/articles/PMC7147278/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7147278/. 131, 168

W. Liu, S. A. Levin, and Y. Iwasa. Influence of nonlinear incidence rates upon the behavior of sirs epidemiological models. *Journal of Mathematical Biology*, 23(2): 187–204, Feb 1986. ISSN 1432-1416. doi: 10.1007/BF00276956. URL https://doi.org/10.1007/BF00276956. 44

Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*,

02 2020. ISSN 1195-1982. doi: 10.1093/jtm/taaa021. URL https://doi.org/10.1093/jtm/taaa021. taaa021. 98

E. T. Lofgren. Visualizing results from infection transmission models: A case against "confidence intervals". *Epidemiology (Cambridge, Mass.)*, 23(5):738, 2012. 133, 134

M. Ludkovski and J. Niemi. Optimal dynamic policies for influenza management. *Statistical Communications in Infectious Diseases*, 2(1), 2010. 92

J. J. Ma and D. J. D. Earn. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*, 68(3): 679–702, 2006. 96

L. Ma, Q. Liu, and P. Van Mieghem. Inferring network properties based on the epidemic prevalence. *Applied Network Science*, 4(1):1–13, dec 2019a. ISSN 23648228. doi: 10.1007/s41109-019-0218-0. URL https://link.springer.com/articles/10.1007/s41109-019-0218-0https://link.springer.com/article/10.1007/s41109-019-0218-0. 39

L. Ma, Q. Liu, and P. Van Mieghem. Inferring network properties based on the epidemic prevalence. *Applied Network Science*, 4(1):93, 2019b. 83

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979. 46

S. Méléard and D. Villemonais. Quasi-stationary distributions and population processes. *Probability Surveys*, 9(1):340–410, 2012. ISSN 15495787. doi: 10.1214/11-PS191. 68

A. Melegaro et al. What types of contacts are important for the spread of infections? using contact survey data to explore european mixing patterns. *Epidemics*, 3(3):143 – 151, 2011. ISSN 1755-4365. doi: https://doi.org/10.1016/j.epidem.2011.04.001. URL http://www.sciencedirect.com/science/article/pii/S1755436511000247. 169

G. Meyerowitz-Katz and L. Merone. A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates. *International Journal of Infectious Diseases*, sep 2020a. ISSN 12019712. doi: 10.1016/j.ijid.2020.09.1464. 169

G. Meyerowitz-Katz and L. Merone. A systematic review and meta-analysis of published research data on covid-19 infection-fatality rates. *medRxiv*, 2020b. 114

J. C. Miller. Percolation and epidemics in random clustered networks. *Physical Review E*, 80(2):020901(R), 2009. 38

J. C. Miller. A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of Mathematical Biology*, 62(3):349–358, 2011. ISSN 0303-6812. doi: 10.1007/s00285-010-0337-9. 129

J. C. Miller. A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology*, 74(9):2125–2141, 2012. doi: 10.1007/s11538-012-9749-6. 96

J. C. Miller. Equivalence of several generalized percolation models on networks. *Phys. Rev. E*, 94:032313, Sep 2016. doi: 10.1103/PhysRevE.94.032313. URL https://link.aps.org/doi/10.1103/PhysRevE.94.032313. 34, 124

J. C. Miller and I. Z. Kiss. Epidemi spread in networks: Existing methods and current challenges. *Mathematical modelling of natural phenomena*, 9(02):4–42, 2014. 125

J. C. Miller and T. Ting. Eon (epidemics on networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, 4(44):1731, 2019. doi: 10.21105/joss.01731. URL https://doi.org/10.21105/joss.01731. 133

J. C. Miller and E. M. Volz. Model hierarchies in edge-based compartmental modeling for infectious disease spread. *Journal of Mathematical Biology*, 67(4):869–899, 2013. doi: 10.1007/s00285-012-0572-3. 124

J. C. Miller, A. C. Slim, and E. M. Volz. Edge-based compartmental modelling for infectious disease spread. *Journal of the Royal Society Interface*, 9(70):890–906, 2012. 11, 34, 38, 64, 129

M. Mohammadi and A. Borzì. Analysis of the chang–cooper discretization scheme for a class of fokker–planck equations. *Journal of Numerical Mathematics*, 23, 09 2015. doi: 10.1515/jnma-2015-0018. 72

M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, probability and computing*, 7(03):295–305, 1998. 82

C. Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678, 2000. 38, 124

D. H. Morris, F. W. Rossine, J. B. Plotkin, , and S. A. Levin. Optimal, near-optimal, and robust epidemic control. *Commun Phys*, 4(78), 2021. 93, 103, 104, 123, 149, 157, 159, 180, 182

J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine*, 5(3):e74, 2008. doi: 10.1371/journal.pmed.0050074. 146

S. Myers and J. Leskovec. On the Convexity of Latent Social Network Inference. *Advances in Neural Information Processing Systems 23*, pages 1741–1749, 2010. ISSN 0007-4551. doi: 10.1016/j.bpc.2007.06.013. URL citeulike-article-id:8241227{%}5Cnhttp://arxiv.org/abs/1010.5504{%}5Cnhttp://papers.nips.cc/paper/4113-on-the-convexity-of-latent-social-network-inference.pdf. 38

N. Nagy, I. Z. Kiss, and P. L. Simon. Approximate Master Equations for Dynamical Processes on Graphs. *Mathematical Modelling for Natural Phenomena*, 9(2):43–57, 2014. 40, 43, 47, 59, 67, 68

I. Nåsell. *Extinction and quasi-stationarity in the stochastic logistic SIS model*. Springer, 2011. 11

P. Netrapalli and S. Sanghavi. Learning the graph of epidemic cascades. *SIGMETRICS Performance Evaluation Review*, 40(1):211–222, June 2012. ISSN

0163-5999. doi: 10.1145/2318857.2254783. URL http://doi.acm.org/10.1145/2318857.2254783. 38

M. Newman. *Networks*. Oxford university press, 2018. 132

M. E. Newman. Properties of highly clustered networks. *Physical Review E*, 68(2): 026121, 2003a. 133

M. E. J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003b. 37, 38

M. E. J. Newman. Random graphs with clustering. *Physical Review Letters*, 103 (5):058701, 2009. 133

M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64:026118, Jul 2001. 56

C. Nowzari, V. M. Preciado, and G. J. Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *Control Systems, IEEE*, 36(1):26–46, 2016. 157, 170

C. Orsini, M. Mitrovic Dankulov, P. Colomer-de Simon, A. Jamakovic, P. Mahadevan, A. Vahdat, K. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli, S. Fortunato, and D. Krioukov. Quantifying randomness in real networks. *Nature Communications*, 6, 10 2015. doi: 10.1038/ncomms9627. 21

P. D. O'Neill and G. O. Roberts. Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):121–129, 1999. doi: 10.1111/1467-985X.00125. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-985X.00125. 12, 38, 177

A. Pajor. Estimating the Marginal Likelihood Using the Arithmetic Mean Identity. *Bayesian Analysis*, 12(1):261–287, 2017. doi: 10.1214/16-BA1001. 52

R. Pastor-Barriuso, B. Perez-Gomez, M. A. Hernan, M. Perez-Olmeda, R. Yotti, J. Oteo, J. L. Sanmartin, I. Leon-Gomez, A. Fernandez-Garcia, P. Fernandez-Navarro, et al. Sars-cov-2 infection fatality risk in a nationwide seroepidemiological study. *medRxiv*, 2020. 114

R. Pastor-Satorras, A. Vespignani, et al. Epidemics and immunization in scale-free networks. *Handbook of Graphs and Networks, Wiley-VCH, Berlin*, 2003. 9

R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, aug 2015. ISSN 15390756. doi: 10.1103/RevModPhys.87.925. URL https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.87.925. 6, 11, 23, 37, 38, 63, 124, 127, 128, 158, 163, 166

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 51

L. Pellis, F. Ball, and P. Trapman. Reproduction numbers for epidemic models with households and other social structures. i. definition and calculation of r0. *Mathematical biosciences*, 235(1):85–97, 2012. 131

E. Pepe et al. Covid-19 outbreak response, a dataset to assess mobility changes in italy following national lockdown. *Scientific data*, 7(1):1–7, 2020. 168

T. A. Perkins and G. España. Optimal control of the covid-19 pandemic with non-pharmaceutical interventions. *Bulletin of mathematical biology*, 82(9):1–24, 2020. 180

M. A. Porter and J. P. Gleeson. *Dynamical Systems on Networks: A Tutorial*. Springer International Publishing, 2016. 6, 37, 63, 124

B. Prasse and P. Van Mieghem. Exact network reconstruction from complete sis nodal state infection information seems infeasible. *IEEE Transactions on Network Science and Engineering*, 2018. 39

K. Prem, A. R. Cook, and M. Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, 13 (9):e1005697, 2017. 144

T. Pueyo. Coronavirus: The hammer and the dance. https://tomaspueyo.medium.com/coronavirus-the-hammer-and-the-dance-be9337092b56, 2020. Accessed: 2021-09-04. 17

G. Rempala, H. Koeppl, R. Durrett, and H. van der Hofstad. 19w5071 - scaling limits of dynamical processes on random graphs. *https://www.birs.ca/events/2019/5-day-workshops/19w5071*, 2019. 64

M. Ritchie, L. Berthouze, T. House, and I. Z. Kiss. Higher-order structure and epidemic dynamics in clustered networks. *Journal of Theoretical Biology*, 348: 21–32, 2014. 9, 20

M. Ritchie, L. Berthouze, and I. Z. Kiss. Beyond clustering: Mean-field dynamics on networks with arbitrary subgraph composition. *Journal of Mathematical Biology*, pages 1–27, 2015. 9

M. Ritchie, L. Berthouze, and I. Z. Kiss. Generation and analysis of networks with a prescribed degree sequence and subgraph family: higher-order structure matters. *Journal of complex networks*, 5(1):1–31, 2017. 82

M. Roberts and H. Heesterbeek. Bluff your way in epidemic models. *Trends in Microbiology*, 1(9):343–348, 1993. ISSN 0966-842X. doi: https://doi.org/10.1016/0966-842X(93)90075-3. URL https://www.sciencedirect.com/science/article/pii/0966842X93900753. 4

M. Roberts, V. Andreasen, A. Lloyd, and L. Pellis. Nine challenges for deterministic epidemic models. *Epidemics*, 10:49–53, 2015. ISSN 18780067. doi: 10.1016/j.epidem.2014.09.006. URL http://dx.doi.org/10.1016/j.epidem.2014.09.006. 64

M. Roy and M. Pascual. On representing network heterogeneities in the incidence rate of simple epidemic models. *Ecological Complexity*, 3(1):80 – 90, 2006. ISSN 1476-945X. doi: https://doi.org/10.1016/j.ecocom.2005.09.001. URL http://www.sciencedirect.com/science/article/pii/S1476945X05000814. 45

A. R. Ryan and K. A. Nesreen. The network data repository with interactive graph

analytics and visualization. In *AAAI*, 2015. URL http://networkrepository.com. 58

T. Sellke. On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability*, pages 390–394, 1983. 35

C. I. Siettos and L. Russo. Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4):295–306, 2013. ISSN 21505608. doi: 10.4161/viru.24041. 8

H. Silk, G. Demirel, M. Homer, and T. Gross. Exploring the adaptive voter model dynamics with a mathematical triple jump. *New Journal of Physics*, 16(9):093051, 2014. 64

P. L. Simon and I. Z. Kiss. From exact stochastic to mean-field ODE models: a new approach to prove convergence results. *IMA Journal of Applied Mathematics*, 78 (5):945–964, 2013. 41

P. L. Simon, M. Taylor, and I. Z. Kiss. Exact epidemic models on graphs using graph-automorphism driven lumping. *Journal of Mathematical Biology*, 62(4): 479–508, 2011. 41

J. Stack, S. Bansal, V. S. Anil Kumar, and B. Grenfell. Inferring population-level contact heterogeneity from common epidemic data. *Journal of The Royal Society Interface*, 10(78):20120578, 2013. doi: 10.1098/rsif.2012.0578. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2012.0578. 38

D. Stadlbauer, J. Tan, K. Jiang, M. Hernandez, S. Fabre, F. Amanat, C. Teo, G. A. Arunkumar, M. McMahon, J. Jhang, et al. Seroconversion of a city: Longitudinal monitoring of sars-cov-2 seroprevalence in new york city. *medRxiv*, 2020. 122

C. Stegehuis, R. Van Der Hofstad, and J. S. Van Leeuwaarden. Epidemic spreading on complex networks with community structures. *Scientific reports*, 6(1):1–7, 2016. 25

A. Szabó-Solticzky, L. Berthouze, I. Z. Kiss, and P. L. Simon. Oscillating epidemics in a dynamic network model: stochastic and mean-field analysis. *Journal of Mathematical Biology*, pages 1–24, 2015. ISSN 0303-6812. 11

J. Teixeira da Silva, P. Tsigaris, and A. Erfanmanesh. Publishing volumes in major databases related to covid-19. *Scientometrics*, 126:831–842, 01 2021. doi: 10.1007/s11192-020-03675-3. 4, 15

R. N. Thompson, C. A. Gilligan, and N. J. Cunniffe. Control fast or control smart: When should invading pathogens be controlled? *PLoS computational biology*, 14 (2):e1006014, 2018. 92

L. Thunstrom et al. The benefits and costs of flattening the curve for covid-19. *Available at SSRN 3561934*, 2020. 157, 159

Tom Whipple. Professor neil ferguson: People don't agree with lockdown and try to undermine the scientists, 2020. https://www.thetimes.co.uk/article/people-don-t-agree-with-lockdown-and-try-to-undermine-the-scientists-gnms7mp98, Last accessed on 2020-08-06. 180

F. Trabelsi and N. Naouara. Boundary classification and simulation of one-dimensional diffusion processes. *International Journal of Mathematics in Operational Research*, 11:107–138, 08 2017. doi: 10.1504/IJMOR.2017.10006270. 68

R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, and N. M. Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 20(6):669–677, Jun 2020. ISSN 1473-3099. doi: 10.1016/S1473-3099(20)30243-7. URL https://doi.org/10.1016/S1473-3099(20)30243-7. 145

E. M. Volz, J. C. Miller, A. Galvani, and L. A. Meyers. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput Biol*, 7 (6):e1002042, 06 2011. 129, 139

C. J. Wang, C. Y. Ng, and R. H. Brook. Response to covid-19 in taiwan: big data

analytics, new technology, and proactive testing. *Jama*, 323(14):1341–1342, 2020. 91

Y. Wang, J. Cao, X. Li, and A. Alsaedi. Edge-based epidemic dynamics with multiple routes of transmission on random networks. *Nonlinear Dynamics*, 91 (1):403–420, jan 2018. ISSN 1573269X. doi: 10.1007/s11071-017-3877-3. URL https://link.springer.com/article/10.1007/s11071-017-3877-3. 11, 38

H. Ward, C. J. Atchison, M. Whitaker, K. E. Ainslie, J. Elliot, L. C. Okell, R. Redd, D. Ashby, C. A. Donnelly, W. Barclay, et al. Antibody prevalence for sars-cov-2 in england following first peak of the pandemic: React2 study in 100,000 adults. *MedRxiv*, 2020. 114

W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, and V. J. Lee. Presymptomatic transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. *Morbidity and Mortality Weekly Report*, 69(14):411–415, apr 2020. ISSN 1545861X. doi: 10.15585/MMWR.MM6914E1. URL https://www.cdc.gov/mmwr/volumes/69/wr/mm6914e1.htm. 131

J. Y. Wong, P. Wu, H. Nishiura, E. Goldstein, E. H. Lau, L. Yang, S. Chuang, T. Tsang, J. M. Peiris, J. T. Wu, et al. Infection fatality risk of the pandemic a (h1n1) 2009 virus in hong kong. *American journal of epidemiology*, 177(8): 834–840, 2013. 114

World Health Organisation. Media statement: Knowing the risks for covid-19, 2020. https://www.who.int/indonesia/news/detail/08-03-2020-knowing-the-risk-for-covid-19, Last accessed on 2020-08-06. 114, 168

Y.-C. Wu, C.-S. Chen, and Y.-J. Chan. The outbreak of covid-19: An overview. *Journal of the Chinese Medical Association*, 83(3):217, 2020. 91

L. e. a. Xu. Full spectrum of COVID-19 severity still being depicted. *Lancet. 2020*, 2020. 92

S. Xu and Y. Li. Beware of the second wave of covid-19. *The Lancet*, 395(10233): 1321–1322, 2020. 156

C. You, Y. Deng, W. Hu, J. Sun, Q. Lin, F. Zhou, C. H. Pang, Y. Zhang, Z. Chen, and X.-H. Zhou. Estimation of the time-varying reproduction number of COVID-19 outbreak in china. *medRxiv*, 2020. doi: 10.1101/2020.02.08.20021253. URL https://www.medrxiv.org/content/early/2020/02/17/2020.02.08.20021253. 91