Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Marie Burns Bergan

Modelling the Spread of SARS-CoV-2 in a Hospital Setting

Master's thesis in Biotechnology
Supervisor: Eivind Almaas
Co-supervisor: André Voigt
June 2022

NTNU
Norwegian University of
Science and Technology

Marie Burns Bergan

Modelling the Spread of SARS-CoV-2 in a Hospital Setting

Master's thesis in Biotechnology
Supervisor: Eivind Almaas
Co-supervisor: André Voigt
June 2022

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Acknowledgements

This thesis concludes my master's degree in Biotechnology at the Norwegian University of Science and Technology in Trondheim. The past five years have been the most challenging, but also the most memorable years of my life. I am very grateful for all the people that I have met and the experiences I take with me.

I would like to express my gratitude to my supervisor, Eivind Almaas, for the opportunity to be a part of his research group and for his enthusiasm towards the field of systems biology. I would also like to thank my co-supervisor, André Voigt, for his invaluable guidance and advice throughout this project.

I must also thank my family and friends for their moral support and encouragement during these last few months. A special thanks to my mum for reading through the thesis and providing valuable feedback. Finally, a big thanks to Adrian for his love and support during the stressful time of writing a master thesis. Thank you all.

Marie Burns Bergan
Trondheim, June 2022

Abstract

Since the emergence of SARS-CoV-2 in the latter part of 2019 in China, the virus has spread worldwide. The pandemic has so far claimed more than 6 million lives, and approximately 540 million cases are reported globally. Computational models offer useful tools for obtaining an understanding of the epidemiological mechanisms affecting the viral transmission. Such knowledge is essential for mitigating the spread of the virus and limiting its profound consequences.

This master's thesis presents a modelling framework for simulating the spread of SARS-CoV-2 in a hospital setting. The model is agent-based and the behaviour of the agents are governed by either an empirical contact network or a simulated contact network. Simulated networks are generated based on the properties of an empirical close-proximity interaction network obtained from a geriatric hospital ward. The epidemiological model is an extension of the COVID-19 model developed at NTNU during the spring of 2020. The structure and logic of the model is adapted to simulate the spread of SARS-CoV-2 on an inter-individual contact network in a hospital setting. A motivation for the project was to generate a hospital module that can be incorporated into the larger societal modelling framework to increase the granularity.

The results show that the simulated networks capture the most important network properties affecting virus spread, despite the lack of heterogeneity in the node degrees compared to the empirical network. Simulations were run with several different parameter values to evaluate the effect of pathogen characteristics and network structure on spreading dynamics. The epidemiological analysis demonstrates that the computed reproduction number and the responses to intervention measures are highly dependent on the selected input.

Since most of the parameters in the model are tunable, a wide range of epidemiological analysis is possible. This project therefore lays a foundation for further analysis of spreading dynamics and evaluation of non-pharmaceutical interventions in a hospital setting.

Sammendrag

Siden det første utbruddet av SARS-CoV-2 i Kina i slutten 2019 har viruset spredt seg til alle deler av verden. Så langt har omkring 540 millioner tilfeller blitt rapportert og over 6 millioner liv har gått tapt som følge av pandemien. Datamodellering er et nyttig verktøy for å forstå spredningsmekanismene til viruset. Slik kunnskap er viktig for å kunne slå ned smitten og begrense konsekvensene av pandemien.

Denne masteroppgaven presenterer et rammeverk for å modellere spredning av SARS-CoV-2 på sykehus. Modellen er agentbasert og smittespredningen simuleres enten på et empirisk kontaktnettverk eller et konstruert kontaktnettverk. Simulerte nettverk er generert basert på observerte kontaktmønstre i et empirisk nettverk hentet fra en geriatrisk sykehusavdeling. Den epidemiologiske modellen er en utvidelse av COVID-19 modellen utviklet ved NTNU våren 2020. Strukturen og logikken i NTNU modellen har blitt tilpasset for å kunne simulere smittespredning over mellommenneskelige kontaktnettverk på et sykehus. Motivasjonen for prosjektet var å generere en sykehusmodul som kan inkorporeres i en større samfunnsmodell for å øke detaljnivået i smittesimuleringene.

Resultatene viser at de viktigste egenskapene til det empiriske nettverket når det gjelder smittespredning har blitt reprodusert, til tross for at antallet kontkater per node er mer homogent i de simulerte nettverkene. Smitte har blitt modellert med ulike parameterverdier for å evaluere effekten av nettverkstopologi og virusegenskaper på spredning. De epidemiologiske analysene viser at reproduksjonstallet og effekten av å implementere tiltak varier for ulike parameterverdier.

Modellen kan brukes til å utføre en rekke epidemiologiske analyser, ettersom de fleste parmeterne kan justeres. Dette prosjektet legger derfor et grunnlag for videre undersøkelse av spredningsmekanismene til ulike patogener og evaluering av tiltak som kan implementeres for å slå ned smitten.

Table of Contents

List of Tables

List of Figures

Abbreviations

ABM	=	Agent-Based Model
BC	=	Betweenness Centrality
CC	=	Closeness Centrality
CDF	=	Cumulative Distribution Function
CI	=	Confidence Interval
CFR	=	Case Fatality Rate
COVID-19	=	Coronavirus Disease 2019
ER	=	Erdős-Rényi
FHI	=	Folkehelseinsitituttet (Norwegian Public Health Institute)
HSØ	=	Helse Sør-Øst (Southern and Eastern Norwegian Regional Health Authority)
MERS-CoV	=	Middle East Respiratory Syndrome Corona Virus 2
ORF	=	Open Reading Frame
RFID	=	Radio-Frequency IDentification
SARS-CoV	=	Severe Acute Respiratory Syndrome Corona Virus
SARS-CoV-2	=	Severe Acute Respiratory Syndrome Corona Virus 2
SEIR	=	Susceptible - Exposed - Infected - Recovered
SIR	=	Susceptible - Infected - Recovered
SIS	=	Susceptible - Infected - Susceptible
ssRNA	=	Single-Stranded RNA
WHO	=	World Health Organization

ADM	=	Administrative staff
MED	=	Medical doctors
NUR	=	Nurses and nurses' aides
PAT	=	Patients

Chapter 1

Introduction

The COVID-19 pandemic, which has ravaged the world for the past two years, has had devastating consequences [1]. The pandemic has claimed millions of lives and caused disastrous societal and economic damage. Non-pharmaceutical control measures have been implemented worldwide to reduce the transmission of SARS-CoV-2. Examples are lockdowns, social distancing, isolation of infected individuals and closure of schools and workplaces [2, 3]. As of June 2022, a number of COVID-19 vaccines are approved and according to the WHO, almost 12 billion vaccine doses have been administered globally [4]. The increasing immunity in the population and the emergence of less severe variants will arguably cause the pandemic to end or to at least transition into a new phase [1, 5]. The question of how SARS-CoV-2 will affect the world in the future is, however, uncertain.

Infectious diseases have been prevalent in humans for a long time, and in extreme cases they spread to extended geographical areas. Many infectious diseases causing pandemics originate from zoonotic pathogens, meaning they are transmitted between species e.g. from animals to humans [6]. Smallpox, malaria and measles are examples of emerging zoonotic diseases with fatal outcomes [7]. The emergence of infectious diseases with pandemic potential are still considered a major global threat due to the profound societal consequences they result in [8]. Throughout history, several pandemics have afflicted humankind. For instance, infections caused by the flea-born bacteria *Yersinia pestis* has lead to the plague of Justinian and the Black Death. As many as 100 million people are estimated to have been killed by the Justinian plague, while 200 million were killed by the Black Death. Nowadays, the influenza A viruses are responsible for seasonal epidemics all across the world, due to mutations in the structural proteins which are major targets for neutralising antibodies. Influenza claims approximately 500 thousand lives annually [6]. Strains of the influenza A viruses have also caused pandemics in the past, such as the Spanish flu in 1918 and and the Swine flu in 2009, which are both caused by an H1N1 virus [9]. The increased globalisation in recent times has made it more difficult to contain outbreaks of infectious diseases as the world population becomes more interconnected [10]. To prevent the loss of many human lives in future pandemics, non-pharmaceutical interventions (NPIs) must be in place to control the infection in the absence of effective

treatments, such as vaccines [6]. Since NPIs are costly, it is important to identify which ones are the most cost efficient.

Pathogens causing pandemics are often transmitted via droplets during close-proximity interactions, which is the case for coronaviruses and influenza viruses. Other pathogens may be airborne or vector-borne, such as the *Yersinia pestis* which is transmitted through fleas [11]. Since infections are transmitted from one individual to another, networks describing inter-individual contacts are important for understanding the dynamics of human disease spread [12]. Such knowledge can facilitate the development of appropriate mitigation strategies.

Computational modelling can be used to simulate the spread of pathogens on contact networks describing complex systems. Although models are only simplified representations of the real world, they can provide valuable insight into the underlying spreading mechanisms. The spreading dynamics in a hospital are of particular interest because individuals with underlying medical conditions are often the most vulnerable to infectious diseases. In addition, control measures such as social distancing cannot be implemented at a hospital, as close-proximity interactions are a necessity for providing patient treatment. Identifying effective intervention measures suitable for a hospital setting is therefore important to reduce the transmission and prevent deaths.

The aim of this project is to simulate a hospital ward contact network and to adapt and extend the modelling framework developed at NTNU during the spring of 2020 to model the spread of SARS-CoV-2 in a hospital setting.

To achieve this aim, three main objectives have been identified. The first objective is to extract the contact patterns of an empirical hospital ward interaction network. This will be used as the basis for building a network model generating simulated contact networks which mirror the empirical network properties. The second objective is to build an epidemiological model for simulating disease spread in a hospital, based on the existing NTNU agent-based model. The third objective is to assess how well the simulated network performs in reproducing the empirical network topology, and to investigate the effect of different parameter values on the epidemiological simulations.

Chapter 2

Theory

This chapter describes relevant theory used throughout this project, and is divided into four main parts: The first part describes theory related to Coronavirus Disease 2019, focusing on transmission and the course of the disease. The second part provides an introduction to network theory, while the third part presents the three primarily used epidemic models: compartmental models, network models and agent-based models. Lastly, the empirical data set used as a basis for the project is presented.

2.1 Coronavirus Disease 2019

In December 2019, a novel human coronavirus named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was discovered in the city of Wuhan in China. The disease caused by SARS-CoV-2 is called Coronavirus Disease 2019 (COVID-19) [13]. The global spread of the virus led the World Health Organisation (WHO) to declare a pandemic on 11 March 2020 [14]. As of 27 June 2022, 542 188 789 cases and 6 329 275 deaths have been reported globally [4]. Two highly pathogenic human coronaviruses have emerged prior to SARS-CoV-2 within the past two decades; the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV), causing fatal outbreaks in 2003 and 2012, respectively [15].

The described coronaviruses are positive-sense single-stranded RNA viruses (+ssRNA) that belong to the *Betacoronavirus* genus. SARS-CoV-2 shares 79% genome sequence identity with SARS-CoV and 50% with MERS-CoV. The genome comprises 14 open reading frames (ORFs). These encode nonstructural proteins that make up the replicase complex, accessory proteins and the structural proteins; spike (S), envelope (E), membrane (M) and nucleocapsid (N) [16, 17]. The spike protein mediates the entry into host cells by binding to ACE2 (angiotensin-converting enzyme 2) receptors, and are the main target of antibodies [18].

2.1.1 Transmission

The transmission of SARS-CoV-2 occurs mainly trough the respiratory route. The main recognised mode of transmission is through droplets, but aerosol transmission is also believed to be of significant importance [16, 19]. Infected people can transmit the virus during activities such as expiration, coughing and sneezing that forms droplets or aerosols contaminated with viral particles that are released into the air. Aerosols are smaller in size than droplets, and according to the WHO, particles with a diameter less than 5 µare considered as aerosols [20]. Aerosols have the capacity of travelling further and lasting longer in the air than droplets [19]. SARS-CoV-2 RNA has also been detected on surfaces and studies have proven that the virus can persist for a sufficient amount of time for continued transmission. Hence, contaminated surfaces, often found at common areas such as public bathrooms and public transport, provide another route of SARS-CoV-2 transmission [19, 21].

The transmission potential of a pathogen is often indicated by the basic reproduction number, R_0, which is defined as the average number of secondary cases caused by an individual infection when the entire populations is susceptible to infection. If $R_0 < 1$, the outbreak dies out, however if $R_0 > 1$, the virus continues to spread [22]. An accurate R_0 can be very beneficial when implementing intervention measures, as it is an epidemiological metric used to assess the contagiousness of a pathogen. Estimating R_0 during a pandemic can however be challenging, as it describes the transmission potential under no external control, which is usually not the case during a pandemic. Examples of control measures that can affect the estimation of R_0 are lockdowns, social distancing and isolation of infected individuals [23]. R_0 depends primarily on three parameters; how long infected individuals are contagious, the probability of infection per contact between a susceptible and an infectious individual and, lastly, the contact rate. Contact rates are in turn dependent on social organisation and behaviour, which means the basic reproductive number is not exclusively determined by the characteristics of the virus [24]. The general estimates of R_0 for COVID-19 vary. A systemic review and meta-analysis consisting of 23 studies found the mean estimated R_0 to be 3.38, ranging from 1.90 to 6.49. A pooled estimation based on these studies resulted in an R_0 of 3.32, with a 95% confidence interval (CI) ranging from 2.81 to 3.82 [25]. The R_0 is however very dependent on the variants of SARS-CoV-2, which differ greatly in contagiousness.

2.1.2 Course of Disease

The course of COVID-19 can include an incubation period, symptomatic infection and the disease outcome; which can result in recovery, long-time illness or death. The clinical presentation of COVID-19 varies from mild to severe. Elderly and individuals with underlying health conditions, e.g. diabetes, lung disease and heart disease, are considered high risk groups for severe COVID-19 [26]. The main symptoms of COVID-19 are fever, couch, fatigue, expectoration and dyspnea [27].

The incubation period of a virus is defined as the duration from exposure to the virus to the onset of symptoms [28]. The incubation time is a central epidemiological parameter essential for applying appropriate public health measures, such as determining the optimal duration of quarantine [29]. In the beginning of the COVID-19 pandemic the WHO pro-

posed an incubation time ranging from 0 to 14 days [28]. When sufficient data became available, several estimates have been made. A systemic review including 14 studies estimated the pooled mean incubation period of COVID-19 to be 6.5, (95% CI: 5.9-7.1) days [30], while a different meta-analysis of seven studies resulted in an estimated pooled mean incubation period of 5.6 days (95% CI: 5.1–6.1) [29].

When attempting to mitigate the spread of an infectious disease, early recognition of infection and interrupting routes of transmission are key points [31]. However, infected individuals do not always experience symptoms. This is the case for COVID-19 where asymptomatic and presymptomatic infections are believed to contribute to a significant proportion of transmissions [32]. Asymptomatic refers to infectious individuals without symptoms, while presymptomatic individuals are infectious before developing symptoms. Investigating the epidemiological parameter called serial interval gives insight into the role of presymptomatic infection. The serial interval refers to the time span between symptom onset in the primary case and symptom onset in the secondary case [30]. If the serial interval is shorter than the incubation time, the transmission of the virus has occurred during the incubation period of the infector [32]. Asymptomatic and presymptomatic transmission makes it difficult to contain the pandemic, as these individuals are likely to behave as normal and interact with other people.

The duration of symptomatic infection of COVID-19 varies between mild and severe disease courses, and also for different age groups. A study in India from April 2020 performed by Barman et al., found the average recovery time of patients above 60 years old to be 25 (95% CI: 17.22, 32.78) days and younger than 60 years to be 21 (95% CI: 12.82, 29.32) days. The wide confidence intervals show the heterogeneity in recovery time. According to Raveendran et al., recovery from the acute COVID-19 infection occurs within 7-10 days after symptom onset for mild infections, while for severe infections the recovery time is 3-6 weeks. However, recent evidence shows that symptoms can persist for weeks and months after the acute SARS-CoV-2 infection. This condition is commonly referred to as long covid [34, 35].

COVID-19 has proven to be a fatal disease across the world, so a parameter for investigating the severity of the disease is the case fatality rate (CFR). The CFR is the proportion of confirmed cases resulting in deaths. A system review based on international databases including 34 studies in the period from January to august 2020 estimated the CFR to be 2.67 (95% CI: 2.25, 3.13), with variations between the continents [36].

2.2 Network Theory

Many systems can be represented as networks or graphs. Examples include communication systems such as the World Wide Web, the telephone network, transportation infrastructure, biological systems and various social interaction structures [37, 38]. Network analysis is therefore a useful method for analysing the topological structure of real-life systems. The theory presented in this section is based on Network Science by Barabási [39], unless otherwise is specified.

A network consists of a set nodes which are connected by edges (also called links). The number of nodes, denoted N, represent the number of components in the system. The nodes are labeled $i = 1, 2, ..., N$. The total number of edges in the network is denoted

L. The links of a network can either be directed or undirected, depending on whether the interaction has a direction or not. An example of an undirected social network is the friendship network on Facebook. A friendship is only established if both people agree on connecting, thus the resulting network is undirected as the interactions are symmetrical. Other social media platforms that have a 'follow' feature, such as Instagram and Twitter, are examples of undirected networks. A link between two people does not require a mutual agreement, hence the link is undirected [40]. In addition, we differentiate between weighted and unweighted networks. In a weighted network the edges have a unique magnitude, as opposed to an unweighted network where all edges have the same weight. The world-wide airport network is an example of a weighted network where the edge weight represents the number routes present between two airports [37]. Figure 2.2.1 illustrates the differences between undirected, directed and weighted networks.

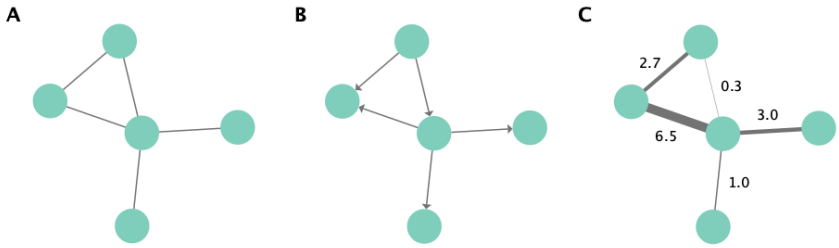


Figure 2.2.1: Three different networks with the same number of nodes ($N = 5$) and edges ($L = 5$). Panel A shows an undirected network where the edges are symmetrical. A directed network is presented in panel B with the arrows depicting the direction of the edges. In the weighted network in panel C, the thickness of the edges are mapped according to their edge weight.

A network can be uniquely described in several ways. An example is storing the nodes connected by an edge in a list: $[(1, 2), (1, 3), (2, 3), (2, 4)]$. This list represents the network in Figure 2.2.2. For mathematical purposes, a network is often represented as a matrix. The edges of an unweighted network are binary; a link between two nodes is either present or not. Such networks can be represented through an NxN adjacency matrix A, where N

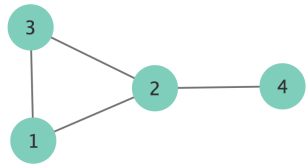


Figure 2.2.2: This network, consisting of four nodes and four edges, can be represented as a list of edges as presented in the text.

represents the number of nodes in the system [38]. Entry A_{ij} in the adjacency matrix describes the interaction between nodes i and j, as shown in Equation 2.2.1. For undirected, unweighted networks, element A_{ij} is 1 if nodes i and j are connected, and 0 if not. For undirected networks, the adjacency matrix is symmetrical around the diagonal because there are two entries for each link, hence $A_{ij} = A_{ji}$. In a directed network on the other hand, each entry in the adjacency matrix represents a unique contact, $A_{ij} \neq A_{ji}$. For weighted networks, the entry is equal to the weight of the interaction between i and j instead of simply 0 or 1, $A_{ij} = w_{ij}$ [38].

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1j} \\ A_{21} & A_{22} & \dots & A_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i1} & A_{i2} & \dots & A_{ij} \end{pmatrix} \tag{2.2.1}$$

2.2.1 Network Measures

This section provides an introduction to basic network measures used for network analysis. The following network measures are described: degree, degree distributions, centrality measures and degree correlations.

Degree

A key property of a node is its degree which describes the number of edges connected to the node. The degree of node i is denoted k_i. The number of links, L, in an undirected network is calculated by summing the node degrees and dividing by two, as each link is counted twice when the links are undirected. In directed networks we distinguish between the incoming degree and the outgoing degree. The incoming degree, k_i^{in}, represents the number of links directed to node i, while the outgoing degree, k_i^{out}, is the number of links directed from i to other nodes. The total degree of a node in a directed network is given by Equation 2.2.2.

$$k_i = k_i^{in} + k_i^{out} \tag{2.2.2}$$

For an undirected network, the node degree can be obtained directly from the adjacency matrix by summing either the rows or the columns of the matrix (Equation 2.2.3), while for directed networks the sum of the rows and the columns represents the incoming and outgoing degrees, respectively (Equation 2.2.4).

$$k_i = \sum_{j=1}^{N} A_{ij} = \sum_{i=1}^{N} A_{ji} \tag{2.2.3}$$

$$k_i^{in} = \sum_{j=1}^{N} A_{ij}, \quad k_i^{out} = \sum_{j=1}^{N} A_{ji} \tag{2.2.4}$$

The equivalent to the node degree in weighted networks is called node strength, denoted s_i. The strength of node i is defined as the total weight of its edges [37], as shown in Equation 2.2.5

$$s_i = \sum_{j=1}^{N} w_{ij} \tag{2.2.5}$$

Degree Distributions

As described previously, the degree of a node is the number of its neighbours in the network. A central network property is therefore the average node degree, $\langle k \rangle$. One of the most common statistics used in the analysis complex networks is the degree distribution [41], which is an important characteristic of a network. The degree distribution, p_k, specifies the probability P of a selected node of having a degree k. The degree distribution is often visualised as a normalised histogram where the number of k-degree nodes is plotted against the degree. For larger networks it is more useful to plot the cumulative degree distribution $P(X > k)$. The shape of the degree distribution determines many network phenomena, thus giving valuable insight into the network structure.

In network science, random network models are applied to reproduce the properties of real networks. An example are the Erdős-Rényi (ER) networks, which consist of N nodes and each node pair is linked with a probability p. The degree distribution of ER networks is binomial and when N is large, it is well approximated by the Poisson distribution given by Equation 2.2.6.

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} = \langle k \rangle \tag{2.2.6}$$

The Poisson distribution is the preferred p_k for ER networks due to its analytical simplicity and the fact that it is only dependent on the average degree $\langle k \rangle$, and not the number of nodes N. In ER networks the degree distribution is narrow around $\langle k \rangle$, meaning that most nodes have a comparable degree to the average. However, this is not true for most real-world networks which portray a more heterogeneous organisation. The degree distribution of real networks often form a straight line on a log-log plot, indicating that is is well fitted by the power law distribution shown in Equation 2.2.7.

$$p_k = k^{-\gamma} \tag{2.2.7}$$

The main difference between the Poisson distribution and the power law distribution is in the region of small and high k. In the power law distribution there is a large number of small degree nodes as well as highly connected nodes, while for the Poisson distribution there is an excess of nodes with their degree approximately equal to the average. Systems following a power law degree distribution are called scale-free as they are not dependent on the networks average degree, such as in random networks. An example of this phenomena in the real world is social friendship networks; all individuals do not have

the same amount of friends, some are more popular and interact with many others while there also exists individuals with only a few friends. The highly connected individuals are called hubs. Hubs are not present in random networks, while in scale-free networks the size of the hubs increases polynomially with network size. A common explanation for the emergence of a power law distribution is the edge formation mechanism called preferential attachment. This mechanism assumes that the probability of a node receiving a new link is proportional to the number of links it already has [42], i.e. in social networks newly introduced individuals are more likely to connect with the already popular individuals. As described, many network properties can be extracted from its degree distribution as it gives valuable information about the underlying topology. In a disease spreading network for instance, hubs may be of great importance as they can serve as super-spreaders if infected with the pathogen.

Centrality Measures

Identifying central nodes in a network is essential in terms of e.g. disease spread. Since the definition of centrality depends on its application, several measures have been proposed [43]. A simple measure already described is the degree centrality which is defined by its node degree. Hubs are central in terms of having many connections, but they may be placed at the periphery of a network without any significant impact on the whole network. Thus, the degree centrality measures the local influence of a node and is considered a local centrality measure [43, 44].

Another measure of local centrality is the shortest path; the number of edges between nodes i and j in the shortest path connecting them. [43]. The distance between i and j is denoted d_{ij}, and for undirected networks this distance is equal both ways, i.e. $L_{ij} = L_{ji}$, while for directed networks $L_{ij} \neq L_{ji}$. The average shortest path length of a node, $\langle L_i \rangle$, is defined by the average shortest path length to all neighbours, given by Equation 2.2.8.

$$\langle L_i \rangle = \frac{1}{N} \sum_{j=1}^{N} L_{ij} \tag{2.2.8}$$

A node with a short average path length can be considered central as it is close to all other nodes in the network [43]. This is captured by the global centrality measure called closeness centrality, which is the distance between a node and all other nodes reachable from it [45]. The closeness centrality of a node, CC_i, is defined in Equation 2.2.9.

$$CC_i = \frac{N}{\sum_{j=1}^{N} L_{ij}} \tag{2.2.9}$$

Another global centrality measure dependent on shortest paths is the betweenness centrality. The betweenness of a node i is defined as the fraction of shortest paths between all nodes in a network that passes through i. The betweenness centrality is therefore a measure of the influence a node has on the information flow in a network. [45].

Degree Correlations

The clustering coefficient is a node property quantifying how interconnected the neighbours of a node are. The local clustering coefficient, C_i, defined in Equation 2.2.10, is between 0 and 1 and represents the probability of two neighbours of node i forming a link.

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \tag{2.2.10}$$

The overall tendency of nodes to form clusters in a network is characterised by the average clustering coefficient, $\langle C \rangle$. In many real networks, the clustering coefficient decreases with node degree. This indicates a hierarchical structure, meaning hubs have low clustering and small-degree nodes tend to form clusters [46].

The correlation between degree and the degree of neighbours can also be quantified by measuring the average degree of the neighbours, as described in Equation 2.2.11.

$$k_{nn}(k_i) = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j \tag{2.2.11}$$

If hubs tend to link to one another and small-degree nodes tend to link to one another, the network exerts assortativity, a positive degree correlation. An example of this in the real world is the tendency of celebrities, which are hubs in social networks, to date each other. If the average neighboured degree decreases with increasing degree, the degree correlations are disassortative (negative) [47]. This is often found in metabolic networks where hubs prefer connecting to small-degree nodes. Networks with no correlation between their degree and the average degree of their neighbours, are called neutral networks.

2.3 Epidemic Modelling

Epidemic modelling uses mathematical and computational tools to study the underlying mechanisms of the spread of infectious diseases throughout a population [48]. Epidemic models can be divided into three main approaches: compartmental modelling, network modelling and agent-based modelling. The following sections describes the applications of these models, as well as their limitations.

2.3.1 Compartmental Modelling

Compartmental models are simplified epidemic models that can give valuable insight into the macroscopic aspects of epidemic diffusion, such as epidemic threshold and size. The advantage of compartmental models is that they do not require much computing power or high fidelity data [49].

Compartmental modelling relies on two key assumptions: compartmentalisation and homogeneous mixing. Compartmentalisation means that individuals are classified based on their health state, and all individuals in one compartment are treated equally. With the

second assumption, homogeneous mixing of the population, all individuals have the same probability of interacting with all other individuals and thus infecting one another [39].

Due to variance in epidemic progress in different diseases, individuals may have a variety of health states [49]. One of the most common compartmental models is the SIR model, where the individuals are either susceptible (S), infectious (I) or recovered (R). Susceptible individuals are healthy individuals who can be infected by a pathogen. When infected with an infection rate, β, their state becomes infectious. Next, the infected individual reaches the recovery state with a recovery rate, γ. The individual has now acquired immunity and cannot be infected again [39, 49]. Compartmental models may be comprised of several other health states depending on the characteristics of the disease and the purpose of the model [50]. In other variants of the SIR model, such as the SIS model, the infected individuals return to the susceptible state instead of recovering after infection. In the SEIR model the individuals additionally experience an exposed state before reaching the infectious state, where the individuals are infected but not sick. [49].

In compartmental models with SIR dynamics, the transition between states is represented with differential equations. The flux of individuals from one state to another per time unit is controlled by the infection rate, β, and the recovery rate, γ[49]. Examples of equations describing these transitions are shown in Equations 2.3.1 – 2.3.3.

$$\frac{dS}{dt} = -\beta IS \tag{2.3.1}$$

$$\frac{dI}{dt} = \beta IS - \gamma I \tag{2.3.2}$$

$$\frac{dR}{dt} = \gamma I \tag{2.3.3}$$

The simplifications and assumptions in compartmental models enables analysis of macroscopic aspects of epidemics, but also limits the capability of simulating the realistic spread of infections. The assumption of homogeneous populations does not capture the differences in individual attributes and behaviour or the contact patterns of social networks, which impacts epidemic spread [51]. Additionally, compartmental models use only a few variables that are parameterised with average quantities. Thus, they do not account for the time scale in disease progress and the variation in infectivity [49].

2.3.2 Network Modelling

Due to the limitations of compartmental models, new approaches for modelling epidemics have emerged. The main difference between compartmental models and computational approaches, including complex network models and agent-based models, is that the latter models epidemics on an individual level. Although it requires large amounts of data and greater computational power, computational approaches are currently widely used to study epidemic spread [49]. Network models are considered an extension of simple compartmental models, but not as complex as agent-based models. Individuals are categorised based on their health state as in compartmental models, but network theory that describes the underlying social structure is also incorporated to capture the heterogeneity across individuals [49, 52].

The primary limitation of compartmental models is the assumption of homogeneous mixing, causing all individuals to have the same probability of coming into contact with one another and infecting one another. As described in Section 2.2, the node degrees in real-world contact networks are highly heterogeneous, i.e. individuals are not equally likely to be infected. To account for this, the degree block approximation adds compartments such that nodes are distinguished based on their degree. This method considers the degree of each node as a variable, as opposed to only considering the common transmission rate, β, and recovery rate γ. In addition to acquiring a more realistic representation of populations, network models can also capture the impact of network topology during a disease outbreak by incorporating other network properties [39].

2.3.3 Agent-Based Modelling

Agent-based models (ABMs), also known as individual-based models (IBMs), are complex models aiming to simulate the dynamics of real-life systems [53]. ABMs consist of agents that interact based on a set of rules within a defined environment [54]. Since agent-based modelling is a bottom up approach where the micro-behaviour of individuals produces system-level outcomes, ABMs are capable of capturing the stochastic and heterogeneous nature of epidemic spread [53]. They provide a more accurate representation of the reality than traditional compartmental models and network models and are therefore a popular choice in epidemic modelling [49].

Agent-based modelling is a relatively recent modelling approach that was originally developed for economics [55]. ABMs can, however, be applied to a wide range of fields and have in the recent years been widely used in social sciences, ecology and epidemiology, among others. An example, in terms of social sciences, is an ABM that models the development of friendship links within Facebook [56]. In epidemic modelling, the agents represent humans. The agents have a unique set of attributes that can include age, gender, location and degree, and the attributes of an agent at a certain point in time is referred to as the agent's state. These characteristics can affect the probability of being infected by the disease directly, or by influencing the decisions made by the agent, hence which agents it encounters. Because the model allows the agents to make their own decisions based on set rules, they have realistic heterogeneous contact patterns [57]. ABMs are particularly appropriate for epidemic modelling as they can incorporate properties from social network analysis and individual behaviour, which may have great impact on the spread of infectious diseases [57, 58].

An important aspect of ABMs are the stochastic elements. Individual attributes and behaviour, such as infectivity rate and contact rate, are often modelled as stochastic processes where random variables are drawn from a probability distribution [59]. This random sampling results in fluctuations in the model output across model runs. To account for this, several simulations are usually run and the average model output from the different simulations is extracted. Here, there is a trade-off between computational time and the number of simulations needed to sufficiently mimic the real system. This stochastic variability means that model simulations with the exact same input parameters can lead to a variation in outputs [54].

There are obvious advantages of ABMs compared to traditional epidemic models regarding its capability of capturing the heterogeneous and stochastic nature of epidemics.

However, ABMs require a large amount of empirical contact data to be able to understand the social behaviour of humans. Additionally, they are more computationally demanding than the previously described models. Despite these challenges, the popularity of ABMs in epidemic modelling continues to grow [49].

2.4 Empirical Data set

This section presents the empirical data used in this project. The data is retrieved from `www.sociopatterns.org`, which is a collaboration between several researchers and developers. The study, Estimating potential infection transmission routes in hospital wards using wearable proximity sensors, performed by Vanhems et al. is the basis for this section.

SocioPatterns have developed a method for measuring face-to-face proximity contacts with a high temporal resolution. The system is based on Radio-Frequency Identification (RFID) devices embedded in wearable badges. These devices exchange ultra-low-power radio packets when they are located within 1 - 1,5 meters from one another. The individuals participating in the study wore these devices on their chest to ensure that packet exchange between two RFID sensors only occurred during face-to-face contacts, as the body acts as a radio-frequency shield. This replicates close-range encounters where disease infections may be transmitted, for example, by coughing, sneezing or hand contact. Radio receivers installed throughout the hospital unit receive and register the information detected by the wearable sensors.

If at least one packet is exchanged between the two sensors during a 20-second interval, a contact is registered, as the probability of detecting a contact over a time period of 20 seconds is larger than 99%. By definition a contact is therefore symmetric, and all contact pairs are considered if several individuals are involved in a contact. As long as the RFID devices continue to exchange at least one packet for every subsequent 20-second interval, the contact is considered ongoing. With this method, Vanhems et al. have quantified human proximity face-to-face contacts with a high temporal resolution in environments such as social gatherings, schools and hospitals.

The data presented by Vanhems et al. consists of an empirical contact network from a hospital ward in a university hospital in Lyon, France. More specifically, the data was collected from Monday, December 6, 2010 at 1:00 pm to Friday, December 10, 2010 at 2:00 pm at a short stay geriatric unit with 19 beds. During these four days, 32 424 contacts were registered among the 75 individuals who participated, consisting 46 staff members and 29 patients. The participation rates were 92% and 94% for staff members and patients, respectively.

The individuals participating in the study were categorised into four classes based on their status in the ward: ADM, MED, NUR and PAT. The administrative staff, denoted ADM, consisted of 8 participating individuals. Medical doctors, including physicians and interns, are represented by the group MED. There were 11 individuals belonging to this group in the study. NUR represents paramedical staff, including nurses and nurses' aides. In addition, staff who worked without a fixed schedule or on demand, such as physiotherapists, nutritionists and social counsellors, were considered as nurses. 27 participants in the study belonged to the group NUR. MED and NUR professionals together form a group named health care workers (HCWs). In addition to the staff members 29 patients,

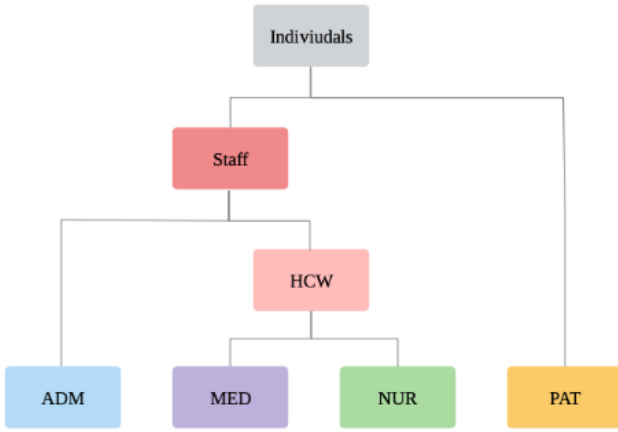


Figure 2.4.1: An overview of the categorisation of the individuals participating in the study, which comprises staff members and patients (PAT). The staff consists of the administration (ADM) and the health care workers (HCWs) who are either medical doctors (MED) or nurses (NUR).

denoted PAT, participated in the study. An overview of the categorisation of individuals is presented in Figure 2.4.1.

The working hours for different shifts are provided for nurses and nurses' aides who are present around the clock and medical doctors who are present during the day time. An overview of the working shifts of the HCWs is provided in Table 2.4.1. The article provides no information about patient admissions and discharges or working hours for administrative staff.

Table 2.4.1: Overview of shift hours and the number of employees present at different times. The shifts for nurses and nurses' aids (NUR) are divided into three, and in addition there is an extra shift covering the day time. Medical doctors (MED) are only present during the day, while no information was provided about shift hours for the administration (ADM).

Status	Shift	Time period	Staffing
ADM	-	-	-
MED	-	08:00 - 17:00	4
NUR	Morning	07:00 - 13:30	5
	Afternoon	13:30 - 20:00	5
	Night	20:00 - 07:00	2
	Day	09:00 - 17:00	2

The empirical data is available in the supplementary materials of the original article [60]. The data file contains a tab-separated list representing the active contacts during 20-second intervals of data collection. Each line represents a contact and consists of the

IDs of the two people in contact, i and j, their status in the hospital unit, S_i and S_j, and the interval during which the contact occurred [t – 20s, t]. Ongoing contacts over several 20-second intervals are represented with a line for each interval the contact was active. An excerpt from the empirical raw data are presented in Table 2.4.2.

Table 2.4.2: The structure of the raw data obtained from [60]. t represents the time interval the contact was active, while (i, j) and (S_i, S_j) represents the IDs and the status of the individuals in contact, respectively.

t	i	j	S_i	S_j
140	13	4	MED	ADM
160	13	16	MED	MED
500	13	14	MED	MED
520	13	14	MED	MED
560	14	16	MED	MED
⋮	⋮	⋮	⋮	⋮
347 640	41	45	NUR	PAT

The objective of the study was to obtain high-resolution contact data describing interactions between individuals in a healthcare setting. The number and duration of contacts displayed large variation across the classes, but also across individuals belonging to the same class, emphasising a heterogeneous contact pattern. According to Barabási, the detailed contact data provided in the study can be utilised to accurately model the spread of infection in a confined hospital setting.

Chapter 3

Methods and Software

This chapter describes the methodology of the work carried out in this project, where the aim was to construct a contact network based on empirical data and build an epidemiological modelling framework for simulating the spread of SARS-CoV-2.

The chapter is divided into four sections, starting with an introduction to the software used in the project. Secondly, the network analysis performed on the empirical contact data is described followed by the methods used for generating a simulated contact network. In the third section, the mechanisms of the epidemiological model are presented, and the final section outlines the complete modelling framework.

3.1 Software

3.1.1 Python

The data analysis and modelling work in this project is performed in Python [61], version 3.8.5, which is an open-source programming language available for most operating systems. Python is an interpreted, high level programming language, meaning that the code is expressive and readable. Given that the original model was written in Python and also the usability and clear syntax of Python made it a good choice for this project.

In addition to a library of standard modules, several supplementary modules can be imported to handle a wide range of operations. Imported modules from the Python Standard Library and installed modules are presented in Appendix A.1. Python can be downloaded from `www.python.org`.

3.1.2 Cytoscape

The network visualisations in this thesis are produced in Cytoscape [62], version 3.9.0. Cytoscape is an open-source software platform for visualisation and analysis of complex networks, originally developed for biological research. Additional features, such as new

layouts and generation of random networks, are available as apps. Cytoscape can be downloaded from `www.cytoscape.org`.

3.2 Generating a Contact Network

One of the objectives of this thesis is to recreate a geriatric hospital ward contact network that can be utilised to model the spread of SARS-CoV-2. To mirror the interactions between individuals, a set of rules governing their contact patterns must be derived. The goal is to capture individual variations that may affect the pathogen transmission. Contact patterns from an empirical contact network are extracted and translated into model parameters and agent attributes. When developing a model there is always a trade-off between complexity and computational running time. Therefore, only the most central network properties are selected as a basis for the model to prevent the simulation from being too computationally demanding.

The empirical contact data presented by Vanhems et al. serves as a basis for the generation of a simulated contact network. The data is collected from a short stay geriatric hospital unit in Lyon, France. 75 individuals participated in the study, 46 staff members and 29 patients, and more than 32 thousand contacts were registered over a time period of four days. Face-to-face proximity interactions within 1.5 meters were detected using wearable RFID sensors. The data is described in more detail in Section 2.4.

The empirical contact network has a temporal resolution of 20 seconds, but generating a synthetic network with the same resolution would be very time consuming. A possibility would be to decrease the temporal resolution by dividing the data into hours, shifts or even days. However, to accomplish the goals of recreating an empirical inter-individual network and developing an epidemiological model to simulate pathogen spread, it was decided to aggregate the empirical contact data to a static network. This simplifies the task of generating a contact network and can also be justified as the time span of the empirical data is only four days and the main contact patterns are still captured.

3.2.1 Network Analysis

The initial raw data file consisted of a table of five columns; time point in which the contact started t, IDs (i, j) and status (S_i, S_j) of the individuals in contact. The individuals can either have status as administrative staff (ADM), medical doctor (MED), nurse and nurses' aide (NUR) or patient (PAT). Each line in the file represents an active contact during a 20 second interval. To prevent the network analysis from being too complex, the empirical dynamic contact network from four consecutive days were aggregated to a static network. The data was processed to create a new data file where each line corresponds to a contact between two unique individuals. The temporal resolution was excluded, so the time stamp was replaced with an edge weight, w_{ij}, represented as the number of contacts within the contact pair during the entire study period. An excerpt of the processed empirical data is presented in Table 3.2.1.

With the data extracted from this file, a weighted adjacency matrix was created. An entry in the matrix represents the number of contacts between nodes i and j. All the contacts are undirected, which results in a symmetric weighted adjacency matrix. Since the node

Table 3.2.1: An excerpt of the processed empirical data where each line represents a unique contact. The file contains the IDs of the individuals in contact (i, j), their status at the hospital (S_i, S_j) and the number of contacts between them, represented as an edge weight (w_{ij})

i	j	S_i	S_j	w_{ij}
24	36	NUR	NUR	1059
41	44	NUR	NUR	576
41	45	NUR	NUR	571
35	36	NUR	NUR	563
24	35	NUR	NUR	553
⋮	⋮	⋮	⋮	⋮
8	68	MED	PAT	1

IDs are sorted by group, the array can be viewed as an grid of 16 blocks corresponding to all the combinations of the four groups at the hospital, as shown in Figure 3.2.1.

The contacts array was used to create a heatmap to visualise the contact patterns within the hospital ward. A heatmap is a graphical representation of data where the values are depicted as colours. In this case, the pixel at entry i, j reflects the number of contacts between these two nodes. A heatmap is useful for identifying correlations in the data, rather than just looking at the numerical values. It can give an impression of the contact patterns within and between the different groups, which is important for developing a model to recreate the empirical contact network.

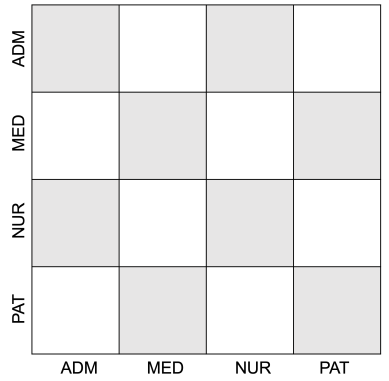


Figure 3.2.1: An illustration of how the interactions between individuals belonging to different groups are arranged in the adjacency matrix. Each combination of interactions between the groups is referred to as a block.

The individuals have different properties depending on their role in the hospital ward and the groups have different interaction patterns within the group and with other groups. To capture these patterns, the hospital ward network was divided into subnetworks for all combinations of groups. This resulted in 10 subnetworks subject to topological analysis.

Four of the networks are internal interaction networks for the groups at the hospital; the administrative staff (ADM), the medical doctors (MED), the nurses and nurses' aides (NUR) and the patients (PAT), describing the contacts within the group. The remaining six networks are bipartite networks between two of the groups. Bipartite refers to a network with two sets of nodes, U and V. Nodes belonging to the U-set are only connected to nodes in the V-set, and vice versa [39]. An example of a bipartite subnetwork is the interactions between the administration (ADM) and the doctors (MED). Since these interactions are symmetrical, the network is represented by two blocks in the weighted adjacency matrix.

The degree distribution is one of the most important network characteristics associated with epidemics because it naturally captures the heterogeneous potential of individuals to become infected and further transmit the infection [63]. Intuitively, individuals with many contacts are at higher risk of being in contact with infected individuals and consequently more likely to cause further infection. To uncover the contact patterns of the different groups, the degree distribution was plotted for all subnetworks. The bipartite subnetworks are described by two degree distributions, one from each of the groups point of view. The ADM-MED degree distribution shows the number of interactions each ADM-individual has with MED-individuals, while the MED-ADM degree distribution displays how many contacts each MED-individual has with ADM-individuals. The resulting 16 degree distributions served as a basis for recreating the empirical contact network. Due to the small system containing a total of only 75 nodes, it was more useful to plot the cumulative degree distributions. The degree distribution was plotted for the entire network as well as for all the subnetworks.

3.2.2 Network Models

Erdős-Rényi Model

In the process of attempting to reproduce the empirical contact pattern, an Erdős-Rényi model for generating random graphs was developed as a starting point. The interactions between all the combinations of the four groups at the hospital were treated as independent subnetworks. A network was constructed for each of the subnetworks by connecting a predefined number of nodes randomly. Each edge was included in the network with a probability p, independently from every other edge. Specifically, the model creates a Poisson distributed variable for all the node pairs in the given subnetwork, which represents the number of contacts between the two individuals. This was repeated every hour for 24 hours to produce a contact array with the entries representing the edge weight added up for every hour. The probability of contact for the different subnetworks was adjusted according to the discovered patterns of the empirical network. The resulting network was too mixed compared to the empirical network and was not able to capture its underlying structures. This was expected as real-life interaction networks are usually characterised by a heterogeneous degree distribution, i.e. hubs and small-degree nodes are prominent. In contrast, an Erdős-Rényi random network model produces nodes with a degree close to the networks average degree. To increase the complexity of the model, a different approach for generating random networks, called a configuration model, was applied. The advantage of a configuration model is that it connects nodes based on a given degree sequence.

Simple Configuration Model

A configuration model builds a network of nodes with predefined node degrees. Based on these node degrees, the network is assembled randomly. The algorithm consists of two main steps. First, half-links called stubs are assigned to all the nodes in the network. The degree sequence is either generated analytically from a pre-selected power-law distribution or an adjacency matrix of a real network. The number of stubs must be even to avoid leaving half-links unpaired. In the second step, two stubs are selected and paired repeatedly until there are no remaining stubs. This process may result in different networks depending on the order the stubs are selected [39].

As a starting point, a simple configuration model was developed. In the model, the number of stubs assigned to each node in the system is generated from a uniform distribution. A uniform distribution means that all outcomes are equally likely within certain limits, so the node degrees are generated within the span of the empirical node degrees in the given subnetwork. For example, the number of stubs assigned to ADM-nodes in the ADM-MED subnetwork was limited by the minimum and maximum node degrees of ADM-individuals in the empirical ADM-MED network. For the bipartite networks, a random stub was selected from both of the groups and paired repeatedly until one of the groups ran out of stubs. For the internal group networks, two random stubs were selected and paired. This resulted in simulated networks for all the combinations of groups.

Preferential Attachment Configuration Model

The cumulative degree distribution of the simulated networks from the simple configuration model deviated significantly from the empirical degree distribution. The edges were more evenly dispersed without the same tendency of forming hubs as in the empirical network. To increase the heterogeneity of the node degrees, a preferential attachment mechanism was implemented in the model. Preferential attachment means that the more connected a node is, the more likely it is to receive new links. In this case, node pairs that are already connected are modelled to have an increased probability of linking to each other. When selecting stubs to be paired, the selection is therefore weighted according to the number of links they already have. This increases the tendency of highly connected nodes to link to other nodes.

The configuration model has so far been based on a uniform degree distribution, however, this does not fit with the empirical data for most of the subnetworks. On a cumulative degree distribution plot with linear axes, a uniform distribution forms a straight line. This pattern is only true for the internal MED-subnetwork. To improve the fit of the simulated degree distributions, the stubs must be generated from more compatible distributions. This shows the advantage of a configuration model which can generate random networks with arbitrary degree distributions. By distinguishing between the subnetworks it is easier to capture the variations between the groups.

Apart from the MED-subnetwork, the cumulative distribution functions (CDFs) of the 16 blocks were found to be approximately linear on either a log-log plot or a semi-log plot, meaning that the scale of either the x-axis or the y-axis is logarithmic. The blocks were therefore classified based on the scaling of the axes resulting in a CDF following a straight line. The four classifications are lin-lin, lin-log, log-lin and log-log. For example,

blocks classified as log-lin display an approximately linear cumulative degree distribution on a plot with logarithmic x-axis and linear y-axis. Table 3.2.2 provides an overview of the classifications:

Table 3.2.2: An array presenting the classification of the cumulative degree distribution of the different blocks. The degree distributions are classified based on the scaling of the x-axis and the y-axis.

	ADM	**MED**	**NUR**	**PAT**
ADM	Log-Lin	Log-Log	Log-Lin	Lin-Log
MED	Log-Log	Lin-Lin	Lin-Log	Lin-Log
NUR	Lin-Log	Lin-Log	Log-Lin	Lin-Log
PAT	Log-Lin	Lin-Log	Lin-Log	Lin-Log

To account for the scaling of the axes, the uniformly distributed variable used to generate stubs for each node was transformed. A function for transforming the variable was found for each of the four classifications. When the transformation functions are descending, the cumulative degree distribution, $P(y > Y)$, can be deduced from the following equations:

$$P(x < X) = X \tag{3.2.1}$$
$$P(y > f(X)) = X \tag{3.2.2}$$
$$P(y > Y) = X \tag{3.2.3}$$
$$P(y > Y) = f^{-1}(Y) \tag{3.2.4}$$

When appropriate functions for generating stubs were defined, the parameters were adjusted to fit the empirical degree distributions as closely as possible.

The simulated networks generated by the configuration model had a tendency of generating extreme values for the number of stubs in one of the groups in some of the bipartite subnetworks. This resulted in fewer stubs being assigned to the individuals in the other group of the bipartite network, causing an uneven distribution of stubs. When pairing stubs from two different groups, it is desirable that the number of stubs match to be able reproduce empirical contact patterns. Extreme values were generated at the tail of the distributions because of the high level of stochasticity in the model. A mechanism to correct for outliers was therefore implemented. This was done by removing stubs from the node with the highest degree until it reached a level of stubs equal to the node with the second highest degree. This was repeated until the sum of degrees in each degree distribution was satisfactory.

3.3 Building an Epidemiological Model

The agent-based model used in this project is based on the NTNU COVID-19 model developed by Voigt et al. in the spring of 2020. A description of the modelling framework

and additional documentation about the NTNU COVID-19 Modelling Taskforce can be found at `www.ntnu.edu/biotechnology/ntnu-covid-19`. The following section describes the mechanisms of the computational modelling framework. Most of the model structure is reused from the NTNU model, so this section is mainly based on the article; Containing pandemics through targeted testing of households, by Voigt et al..

3.3.1 Network Structure

The model is agent-based, also known as individual based, which is widely used in epidemiological modelling. Individual agents are assigned a behaviour pattern which is meant to simulate the society. In the original model developed by Voigt et al., nodes (agents) are generated from population demographics and placed into k-cliques, meaning that all members in the clique are in contact with each other. Nodes are assigned attributes such as age, domicile, layer membership and disease state. The model consists of 9 layers: Household, Daycare, Primary school, Secondary school, High school, Workplace, Nursing home, Hospital and Generic contact network. All nodes are present in the household layer and the generic contact network, as well as one of the other layers.

The model in this project comprises the hospital layer, and more specifically a hospital ward. The layer structure of the original model is therefore not included in the model and the hospital ward is considered as an isolated system. The agents are nodes in an inter-individual contact network of a hospital ward. The behaviour of the nodes are governed by a simulated contact network generated as described in Section 3.2. In the ABM the nodes are initialised with an ID, an age and a status at the hospital. The hospital status of a nodes is either ADM, MED, NUR or ADM. Other attributes are also assigned, such as if the node is present at the hospital, if it's sick and if it's a health-care worker. The configuration model generates a network with the same number of nodes as in the empirical network. IDs are therefore assigned according to the distribution of individuals in the different groups in the empirical data. The age of each node is generated as a random number within an appropriate range considering their status as either a staff member or a patient. For instance, the age of the patients is determined to be between 70 and 100 because the hospital unit modelled is a geriatric unit. In addition to these immutable attributes, individuals are assigned a disease state that changes with time. Apart from the seed node of the epidemic, all nodes are initialised in the susceptible state. In order to enable an outbreak to take hold in the population, seed nodes are initialised in the asymptomatic or presymptomatic state because symptomatic nodes are instantly removed from the system until they recover. This is grounded upon the assumption that symptomatically infectious individuals are quarantined without the ability of infecting others. The personal attributes affect the potential of infection and further transmission, i.e. the model accounts for individual behaviour when modelling the spread of COVID-19.

3.3.2 Epidemiological Dynamics

The spread of SARS-CoV-2 in the ABM follows SEIR-type dynamics. Each individual is in a given disease state ranging from healthy to recovered. The states used in this model are Susceptible (S), Exposed (E), Infected asymptomatic (Ia), Infected pre-symptomatic (Ip), Infected symptomatic (Is), Hospitalised (H), Intensive care (ICU), Recovered (R),

and Dead (D). The disease states and the possible transitions between them are illustrated in Figure 3.3.1.

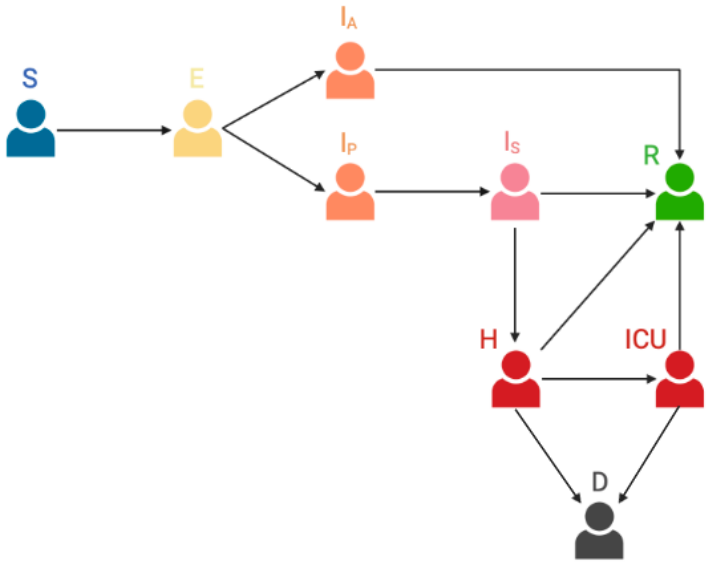


Figure 3.3.1: A flowchart describing the SEIR-dynamics of the ABM showing the possible transition between the disease states.

A susceptible individual enters an exposed state followed by an incubation period leading to infection. The infection may be asymptomatic or symptomatic. In this model, individuals with asymptomatic infection are assumed to recover, while symptomatic individuals may be hospitalised and subsequently enter the ICU, recover or die. Since, the transitions are adapted from a model which incorporates several layers rather than just a hospital layer, the hospitalisation state is considered in a different way in this model. When staff members are hospitalised, they are removed from the hospital unit, which is equivalent to individuals being removed from their workplace layer in the original model. This is governed by the attribute 'present', which is set to 'False' when individuals have a symptomatic infection. Patients are, however, already hospitalised and therefore assumed to be isolated from the rest of the system when they enter the hospitalisation state in the model. In practice, they are not present at the hospital unit and thus not able to cause further disease spread. This is based on the infection control measures carried out by health care workers which are assumed to completely prevent further transmission. The control measures can include isolating the patient in a single room, as well as infection control equipment which successfully protects the health care workers from infection. In addition to the transitions displayed in Figure 3.3.1, individuals may transition directly from susceptible to recovered through vaccination. However, this aspect is not included in the ABM.

The probability of a susceptible individual with an infected neighbour entering the ex-

posed state is governed by an infection probability, β, dependent on the type of contact. The infectivity of asymptomatic and presymptomatic individuals relative to symptomatic individuals is set to 0.3 and 3.0, respectively. Once infected, an individual's disease progression follows SEIR-dynamics. The duration of each state is determined by a stochastic Poisson process with appropriate transition rates based on empirical data for COVID-19. The parameter estimations are performed by Voigt et al.. As shown in Figure 3.3.1, several routes of infection are possible so the probability of particular state transitions are defined. Examples with probabilities equal for all individuals are the chance of developing symptoms during infection (P_I), and the chance of needing the ICU when hospitalised (P_{HI}). The transition rates, λ, and the transition probabilities, p, used in the model are presented in Table 3.3.1. The probability of entering certain states of infection is also age dependent. For instance, the probability of being hospitalised or dying from an infection increases with age, as shown in 3.3.2. In conclusion, the epidemiological spread is dependent on network dynamics, state transition probabilities and the distribution of the duration of each state.

Table 3.3.1: The parameters used for the COVID-19 disease dynamics in the agent based model. The symbol, value and unit of each parameter is provided, as well as the distribution it is drawn from.

Parameter	**Symbol**	**Value**	**Unit**	**Function**	**Source**
Probability of infection	β	0.05	-	Network effect	-
Incubation time	λ_E	5	Days	Poisson	FHI
Asymptomatic time before recovery	λ_{IaR}	8	Days	Poisson	HSØ
Pre-symptomatic duration	λ_{IpS}	2	Days	Poisson	FHI
Symptomatic time before recovery	λ_{IsR}	5	Days	Poisson	FHI
Symptomatic time before hospitalisation	λ_{IsH}	6	Days	Poisson	HSØ
Hospital time before recovery	λ_{HR}	8	Days	Poisson	HSØ
Hospital time before ICU	λ_{HI}	4	Days	Poisson	HSØ
ICU time before recovery	λ_{IR}	12	Days	Poisson	HSØ
ICU time before death	λ_{ID}	12	Days	Poisson	HSØ
Exposed developing symptoms	P_I	50	%	Bernoulli	FHI
Hospitalised needing ICU	P_{HI}	30	%	Bernoulli	FHI
Not developing immunity	P_{RS}	0	%	Bernoulli	FHI

For each day of simulation, the nodes store information about their state progression. Specifically, the current state, the next state, the date of the last state change and the date

Table 3.3.2: Overview of the age dependent transition probabilities. (P_{IsH}) represents the probability of a symptomatic individual becoming hospitalised, while (P_{HD}) describes the probability of a hospitalised individual of dying.

Age group (years)	P_{IsH} (%)	P_{HD} (%)
0 – 9	0	1.61 e-3
10 – 19	0.048	6.95 e-3
20 – 29	1.04	3.09 e-2
30 – 39	3.43	8.44 e-2
40 – 49	4.25	0.161
50 – 59	8.16	0.595
60 – 69	11.8	1.93
70 – 79	16.6	4.28
80 +	18.4	7.8

of the following state change is tracked. The subsequent state is determined by the SEIR-dynamics and the transition probabilities. Once the next state is selected, the duration of the current state is generated with a Poisson-distributed variable according to the state's typical duration, plus one day. One day is added to prevent the ability of changing states twice in one day. For example, the waiting time before a presymptomatic individual becomes symptomatic ($I_p - I_s$) is determined by the Poisson draw of $p(1 + \lambda_{IpS})$. The date for entering the next state is set accordingly.

3.3.3 Reproduction number

There are several ways to calculate the basic reproduction number, R_0, of an epidemic outbreak. R_0 can be defined as the average number of secondary cases caused by an individual infection when the entire population is susceptible. In the ABM, all individuals are susceptible at the beginning of the simulation, except the seed node. R_0 is therefore calculated by running a simulation until the seed node recovers and counting the number of individuals infected by the seed node. The resulting R_0 may vary due to stochasticity, but it also depends on the type of node initiated as seed. Highly connected nodes can potentially cause more secondary infections than nodes with fewer connections. For this reason, 100 simulations are run, seeding individuals from each of the four hospital groups; ADM, MED, NUR and PAT. For each group, the R_0 is calculated as the average number of secondary infections caused by the seed node in the 100 simulations.

In the original model developed by Voigt et al., from which the modelling framework is based on, R is calculated using a different approach. Initially, 20 randomly selected individuals are set as seed nodes and infection is allowed to spread without limitations until R stabilises. Subsequently, the model runs for 40 days with a specific set of parameters, including the fraction of the population vaccinated and the amount of random contact. An individual reproduction number is calculated for the individuals that recovered before the 40-day mark. A daily R is expressed as the average individual reproduction number for

all infected individuals on that given day. The average of the daily R-values from day 12 through 17 is reported as the final R. The first period is excluded to pass the transient phase, while the last period is neglected to ensure that all infected individuals have the time to recover [65].

The approach used by Voigt et al. is suitable for larger systems simulated for longer periods of time. The contact network in this project contains only 75 nodes, meaning it would saturate very quickly if the model was initialised with 20 seed cases. This would lead to inaccurate computations of the reproduction number. Considering only the seed nodes is more appropriate for such a small system, because the rest of the population is susceptible at the beginning of the simulation. This method is however computationally demanding and wold be unfeasible in a national model. In a smaller setting such as a hospital ward, the cost is acceptable.

3.3.4 Intervention Measures

During the pandemic, several intervention measures have been implemented by governments to mitigate the spread of COVID-19. Examples are social distancing, lockdowns and various testing regimes. Although intervention measures have not been the main focus of the model, testing regimes with subsequent isolation of infected individuals are implemented in the model.

The goal of regular pooled testing is to detect asymptomatic and presymptomatic infections, which are the sources of transmission in the model as symptomatic individuals are instantly removed from the system. Individuals testing positive are isolated to limit the outbreak. In the model, the duration of isolation was determined according to the most recent isolation policy in Norway before the restrictions were lifted, which was 4 days, i.e. nodes with detected infection are removed from the system for 4 days or until they recover. Although the accuracy of COVID-19 tests vary for different test methods, the sensitivity and specificity of the test in the model is set to 80% and 100%, respectively. Sensitivity refers to a test's ability to correctly identify patients with the disease, while specificity refers to a test's ability to correctly identify patients without the disease. This means that infected individuals without symptoms have a 20% chance of receiving a false negative test result, while the possibility of false positive test results are neglected.

Two test strategies were used in the model; daily and weekly testing of all the individuals present at the hospital ward. Pooled testing is assumed to be feasible due to the small size of the network and the resources available in a hospital. There are two main reasons for implementing test strategies in the simulation: First, to investigate if testing affects the overall progression of the epidemic. Second, to investigate how regular testing affects the absence of health care workers. Strict testing and isolation regimes may lead to a deficiency in available health care workers, causing the gain of testing to be outweighed by the loss. Simulating with two different strategies can give information about the testing frequency required for testing to be beneficial.

3.4 The Modelling Framework

This section provides an overview of the modelling framework of the project, including a description of the main algorithm. The Python code used in this project is available in the following GitHub repository: `https://github.com/mariebergan/Hospital`.

3.4.1 Algorithm

The model can be divided into four main simulation steps, as shown in the flowchart in Figure 3.4.1.

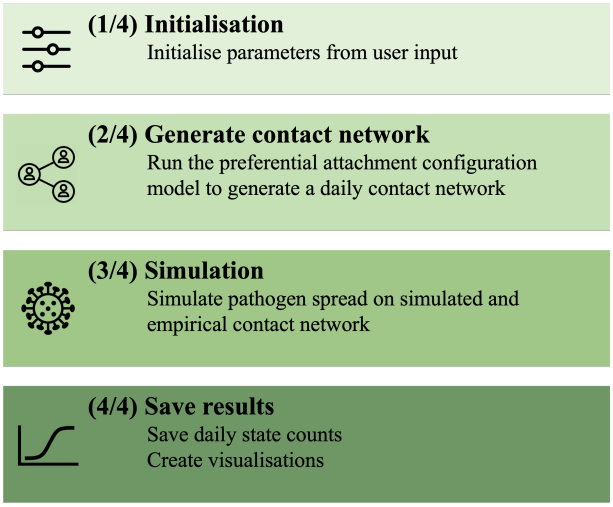


Figure 3.4.1: A overview of the main steps of the simulation algorithm.

In the first step, parameters are initialised from user input and default values. Second, the preferential attachment configuration model is run to generate a daily contact network. Next, the epidemiological model is simulated on both the empirical and the generated contact network, before the results are saved and summarised.

3.4.2 Code Structure

The code for the network analysis, the generation of simulated contact networks and the epidemiological model are separated into different files. The functions for the initial network analysis of the empirical network, and later the simulated network, are stored in the file `networkAnalysis.py`. The `empContacts.py` file creates an aggregated contact array from the empirical contact data, while `configurationModel.py` file contains the code for generating a simulated aggregated contact network for four days, as well as daily contact networks. The daily contact networks are obtained by dividing the number of contacts for each node pair by four, since the empirical network contains contact data from four consecutive days. The file includes functions for generating stubs, connecting stubs and removal of outliers.

The epidemiological model is divided into three main files. `modelFuncs.py` contains all the main functions for simulating epidemiological spread, while the actual simulation is run from `run.py` with model parameters imported from `parameters.py`. Prior to simulation, the input parameters set by the user and constant parameters are defined. A contact network network is also fed into the model, either an empirical or a simulated contact network. In the first step of simulation, the agents are initialised followed by the main simulation. Analysis of the simulations and visualisations are also generated in the `run.py` file. An overview of the main files in the model is outlined in Table 3.4.1.

Table 3.4.1: An overview of the files in the modelling framework with a short description.

Filename	Description
`networkAnalysis.py`	Network analysis of empirical and simulated network
`empContacts.py`	Generates a contact array for the empirical data
`configurationModel.py`	Generates simulated contact networks
`modelFuncs.py`	Main functions, including setup and running functions
`parameter.py`	Disease dynamics and important model parameters
`run.py`	Runs simulation and creates output

3.4.3 Model Parameters

When running a simulation of the epidemiological model, several user input parameters and constant parameters are fed into the model. The main model parameters used in the simulations are shown in Table 3.4.2. Additionally, either the empirical contact network or a simulated contact network is fed into the model to account for the individual behaviour. Since the empirical data only contains contact data from four days, the contact arrays for these four days are reused throughout the simulation period, which is usually set to 60 days. Although the empirical data is divided into days when running the epidemiological simulation, the model for generating a simulated contact network is based on the aggregated empirical network. Therefore, only one simulated contact network is generated and reused throughout the simulation. Since the simulated network is generated based on empirical contact data from four days, all the entries in the contacts array are divided by four do generate daily arrays of contact. This way of simulating spread on the constructed network is referred to as Scenario 1 in this section. However, other scenarios for simulating spread are also possible, as described below:

1. Generate one aggregated contact network which is divided by four and reused every day.
2. Generate four aggregated contact networks that are divided by four. The four daily contact arrays are reused throughout the simulation, similar to the empirical daily contact networks.
3. Generate a new aggregated network that is divided by four for each simulation day.

Table 3.4.2: An overview of the main parameters of the epidemiological model. The data type and a short description of the parameters is provided, as well as the default values.

Parameter	Type	Description	Default
contactNetwork	array	Empirical or simulated contact network	None
infP	float	Probability of infection	0.01
runDays	int	Number of simulation days	60
nSeedNodes	int	Number of seed nodes	1
seedState	str	Disease state of seed node	I_P
testing	bool	If a testing regime is active or not	None
testFrequency	int	Number of days between every pooled testing	7

3.4.4 Model Output

The main output of the model can be divided into two parts: text summaries and visualisations. The main output is the state log which saves the number of individuals in each disease state for each day simulated. Other extracted outputs used are described in Table 3.4.3 The output data has been used to create several visualisations. A typical illustration of a simulation is a SEIR-plot, where the number of individuals in each of the SEIR-states are plotted for each day, as shown in Figure 3.4.2.

Table 3.4.3: An overview of the model output used for visualisations

Output	Type	Description
stateLog	2D dict	Daily count of the number of individuals in each state
infLog	list	Daily count of the number of infected individuals
absenceLog	list	Daily count of the number of absent HCWs
infIDs	list	ID of individuals infected during the simulation period
R_0	int	Number of individuals infected by the seed node

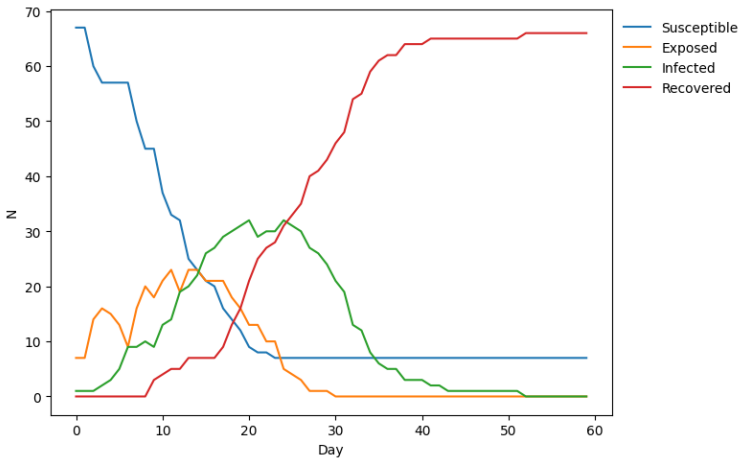


Figure 3.4.2: An example of a visualisation of the model output. The plot shows the daily state count for the susceptible, exposed, infected and recovered state at each simulation day.

Chapter 4

Results and Analysis

The previous chapter introduced the structure and logic of the modelling framework governing the network generation and spreading dynamics.

This chapter presents the results and analysis of the configuration model and the epidemiological model. First, the network analysis performed on the empirical contact network is described, followed by the network analysis of the simulated contact network. Network statistics for empirical and simulated contact networks are compared. Subsequently, the results from the epidemiological simulations on the empirical and simulated contact networks are outlined. This part includes disease spreading dynamics, calculations of the reproduction number and testing of intervention measures. Lastly, the error analysis of the model is presented, which consists of an assessment of the configuration model and an analysis of the contact network period used in the epidemiological model.

4.1 Analysis of Empirical Data Set

4.1.1 Heatmap

To get an initial impression of the contact patterns in the hospital ward, a heatmap was produced, shown in Figure 4.1.1. The colour of the pixel at entry i, j represents the number of contacts between nodes i and j. Dark colours correspond to few contacts, while brighter colours correspond to many contacts. The number of contacts between two nodes vary from 0 to over 1000, consequently the scale of the heatmap is log-transformed to improve the visualisation. White grid lines are added to the plot to distinguish the subnetworks from each other. The heatmap is symmetrical due to the undirected edges. This results in the interactions between two different groups, i.e. the blocks on the off-diagonal, being represented twice. The diagonal blocks display the interactions within the groups of ADM, MED, NUR and PAT. Each of these are viewed as subnetworks of the hospital, either bipartite networks or internal networks.

The brightest colours are found in the MED-block and NUR-block, indicating that the highest-degree nodes are medical doctors or nurses. The MED group can almost be

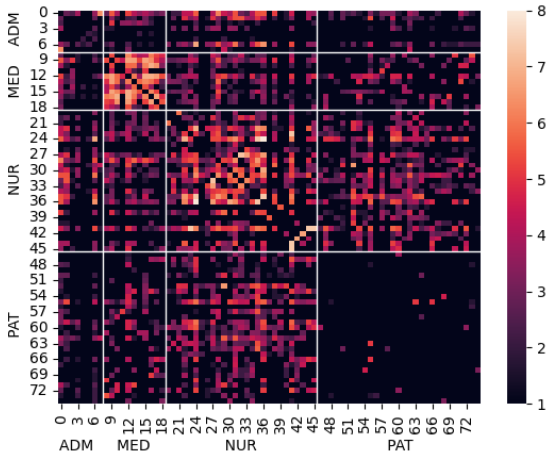


Figure 4.1.1: A heatmap visualising the contact network from the entire empirical data. The colour at pixel i, j represents the number of contacts detected between nodes i and j for the entire study period. The white grid lines separate the four groups in the hospital.

described as a clique, as most of the nodes are in direct contact with each other. The number of contacts seems to be quite homogeneous within this group, but also relatively high compared to the rest of the network. These observations indicate that most of the doctors at the hospital unit work closely together. In the NUR block, the contacts are not as evenly dispersed. There seems to be a distinction between social and antisocial nodes, meaning that nodes with many contacts with some individuals also have many contacts with most of the nurses. In contrast, there are a few individuals with almost no contact with other nurses. This is represented by the column or row for the given node being almost completely black. This pattern is true for the interactions between the MED and NUR groups as well. The nurses with few contacts with other nurses are also less connected to the doctors. Except for these antisocial nurses, the interaction pattern between the NUR and MED groups is mixed. However, MED individuals have fewer contacts with each of the NUR individuals than with the individuals in the same group, and vice versa. A reason for this may be that the doctors have their own office area and that the nurses share a break room, whereas the contact between them may be mainly during patient visits.

The ADM subnetwork exhibits a one-to-all pattern with the rest of the nodes having few contacts with each other. This can be consistent with leaders being in contact with most of the administration, while other ADM staff work separately from each other. An assumption is therefore that the administration consists of loose nodes and a few central nodes. It is however difficult to extract rules governing contact patterns for the administration of a hospital unit from a subnetwork containing only 8 nodes. In contact with other groups, the hubs of the ADM-network seem to be the most active. They are in contact with most doctors and nurses, and also some patients. A few of the individuals in ADM with little internal contact also serve as external contacts points.

There are very few contacts between the patients, and the pattern seems to be random without any complex underlying structure. Presumably these contacts can be neglected or reproduced with a simple Erdős-Rényi model with a low probability of contact. A lack of regular contact between patients can be a result of them having individual rooms or that they, for the most part, are lying in bed without face-to-face proximity contact with each other, which is required for the RFID-sensors to detect a contact. Out of all the other groups, the nurses interact the most with the patients. The contact pattern seems homogeneous without a clear structure. This is also the case for the bipartite network consisting of the MED-group and the PAT-group. Overall, there is large variety in the number of contacts across the blocks. These differences need to be captured by the reconstructed network to be able to realistically simulate disease spread.

4.1.2 Degree Distribution

The degree distribution of a network is an important statistic for describing its contact patterns. The cumulative degree distribution of the static empirical network with contact data aggregated for the entire study period of four days is illustrated in Figure 4.1.2. Since the distribution is plotted cumulatively, the y-axis refers to the probability of a node X of having a degree equal to or larger than the degree k on the x-axis, $P(X \geq k)$. With a linear x-axis and a logarithmic y-axis, the degree distribution is approximately linear. The y-axis is logarithmic to better illustrate the highly connected nodes. Typical scale-free interaction networks follow a power law distribution, indicated by a straight line on a log-log plot. Scale-free networks are characterised by the presence of nodes with a degree that greatly exceeds the average, known as hubs. The CDF of the empirical degree distribution is however approximately straight on a semi-log plot, not a log-log plot. In light of that, there seems to be more nodes with comparable degrees to the average and less distinct hubs in the empirical contact network compared to the typical scale-free networks.

The focus of this project has been to study the interactions between groups and within

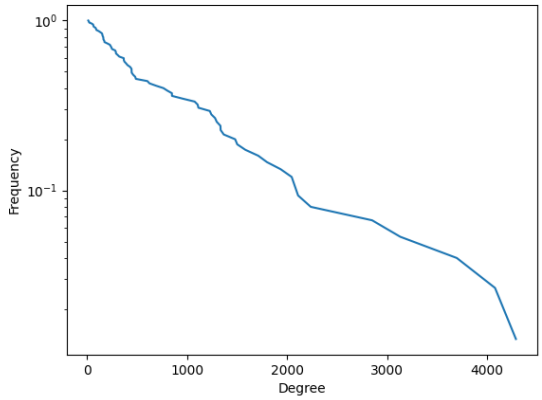


Figure 4.1.2: The cumulative degree distribution of the aggregated empirical contact network.

groups as separate subnetworks to be able to capture individual differences in contact patterns dependent on their role at the hospital. As described in Section 3.2.1, the interactions between two different groups can be expressed from the perspective of both of the groups. Consequently, two degree distributions are generated for the bipartite subnetworks. The internal networks are, on the other hand, represented by only one degree distribution. This results in a total of 16 cumulative degree distributions for all the subnetworks of the empirical contact network, shown in Figure 4.1.3. The rows and the columns of the subplot correspond to the groups from which the degree distributions are calculated. Specifically, the row indicates the perspective of the group from which the degree distribution is plotted from. The scaling of the axes varies between the blocks so that the CDF follows an approximately straight line. However, this is not accomplished for all of the blocks, such as the NUR-MED degree distribution which distinctly deviates from linearity. Nevertheless, by assuming that the degree distributions are linear, the main contact patterns of blocks classified with the same scaling of axes can be extracted.

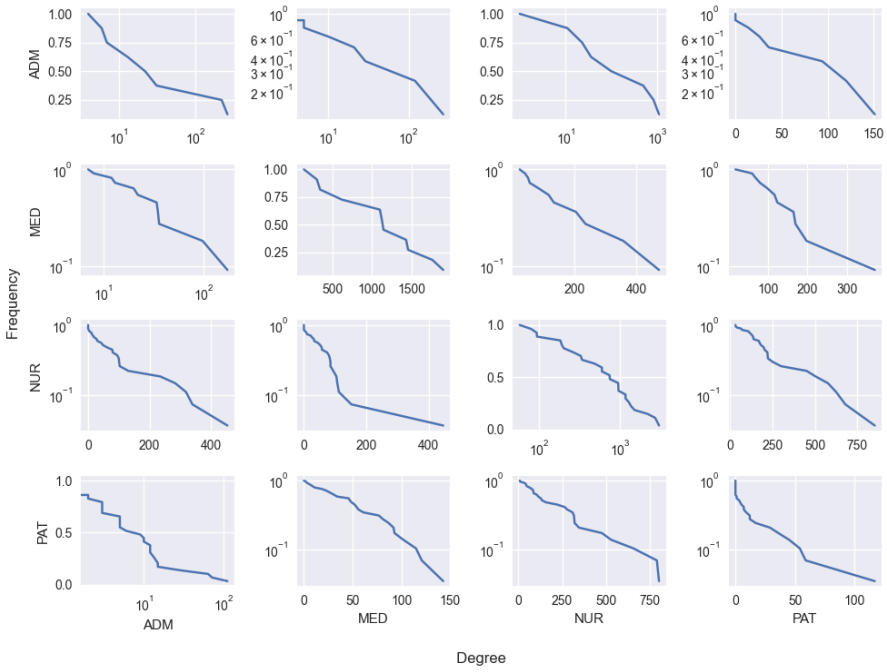


Figure 4.1.3: The cumulative degree distributions for each of the blocks in the empirical contact network.

As described above, CDFs that are linear on the log-log scale follow a power law distribution, indicating the presence of many small-degree nodes as well as a few hubs. This is the case for the ADM-MED and the MED-ADM block. Blocks with log-lin axes, which comprises ADM-ADM, ADM-NUR, NUR-NUR and PAT-ADM, have smaller hubs and more average degree nodes compared to log-log blocks. A straight CDF on a lin-lin plot on the other hand indicates a uniform distribution, meaning that a node has an

equal probability of obtaining any node degree. This pattern is only found for the MED-MED block. The degree distribution of the majority of the blocks, including ADM-PAT, MED-NUR, MED-PAT, NUR-ADM, NUR-MED, NUR-PAT, PAT-MED, PAT-NUR and PAT-PAT, have CDFs that are approximately linear on a lin-log plot. An overview of the classification of the blocks can be found in Table 3.2.2.

4.2 Simulated Contact Network

To assess the configuration model's ability to reproduce the interaction patterns of the empirical contact network, several network statistics were analysed for the simulated network. The main network analysis performed on the empirical contact network produces a heatmap of the weighted adjacency matrix and the degree distribution for the entire network as well as for each of the blocks. The same network properties are therefor characterised for the simulated network. The following section presents a comparison of the network characteristics of the empirical network and simulated networks.

4.2.1 Heatmap

Similarly as for the empirical network, a heatmap was generated for the simulated contact networks. A heatmap for a network generated from a single simulation is presented in Figure 4.2.1. As with the empirical contact network, the scale of the heatmap is log-transformed to improve the visualisation due to the great variation in the number of contacts. The scaling of the colour bar is equal to the empirical heatmap, allowing for a comparison between the two. Generally, the distribution of contacts in the empirical heatmap is well approximated in the simulated heatmap for all the subnetworks. For instance, the

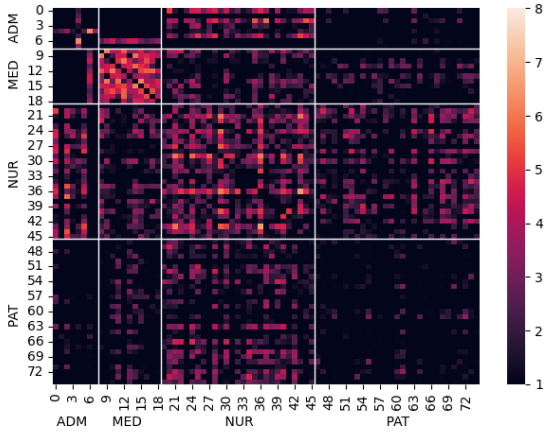


Figure 4.2.1: A heatmap visualising a simulated contact network generated with the preferential attachment configuration model. The colour at pixel i, j represents the number of contacts detected between nodes i and j for the entire study period. The white grid lines separate the four groups in the hospital.

strongest connections, referring to the number of contacts between two nodes, are found in the MED and NUR subnetworks. Similar to the empirical network, the MED subnetwork is a homogeneous group where most nodes are in direct contact with each other. The contacts in the NUR subnetwork are however slightly more mixed than in the empirical network, without antisocial nodes with almost no contacts. The one-to-all pattern in the empirical ADM subnetworks is also present in the simulated network. There are however no interactions between the remaining administrative staff. There are few contacts between the patients, but still slightly more than in the empirical PAT subnetwork. The distribution of contacts in the bipartite subnetworks are similar to that found in the empirical heatmap.

It is however clear that the brightest, almost white, pixels are not present in the simulated heatmap and the colours are generally darker compared to those in the empirical heatmap. Hence, the connected node pairs are not as highly connected as in the empirical network. Overall, the distribution of contacts in the simulated network is comparable to the empirical network, but the strength of the contacts is significantly weaker. This implies that the pairs of individuals that are connected have fewer or shorter encounters, which can decrease the transmission potential of an infectious disease.

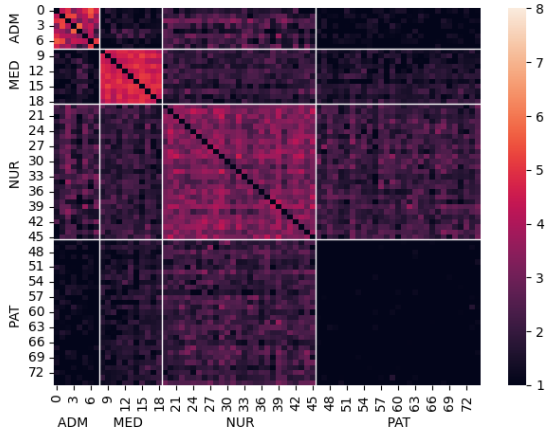


Figure 4.2.2: Heatmap produced from 20 simulations of the preferential attachment configuration model. The pixel at location i, j represents the average number of contacts between nodes i and j in 20 simulated contact networks.

Since the heatmap in Figure 4.2.1 is a result of only a single network simulation, it is difficult to draw conclusions about how well it approximates the empirical heatmap. The stochastic element in the configuration model leads to variation in the generated contact networks, i.e. the heatmap of one simulated network is not necessarily representative for the contact patterns generated with the model. To account for the variation in the simulated networks, several simulated contact networks were generated. The heatmap in Figure 4.2.2 is produced from 20 simulations of the configuration model, with the number of contacts between nodes i and j representing the average number of contacts from the 20 generated networks. The resulting heatmap is very homogeneous within the subnetworks compared to the empirical heatmap and the heatmap from a single model simulation. Calculating in

average from several independent simulations results in mixed contact patterns, which are difficult to analyse.

4.2.2 Degree distribution

The development of the configuration model is mainly based on the characteristics of the degree distribution of the blocks in the empirical data. In other words, the goal was to reproduce these degree distributions in the simulated networks. As described in Section 3.2.2, a simulated network is produced by generating a set of stubs for each node and subsequently connecting them. The distribution of stubs is determined by a transformed uniform distribution. The cumulative degree distribution, $P(y > Y)$, and the transformation function for each of the four classifications are presented in 4.2.1.

Table 4.2.1: The table presents the function for the cumulative degree distribution as well as the function for transforming a uniformly distributed variable, which is the inverse of the cumulative degree distribution function, for the four classifications of degree distributions in the empirical network.

Scale of axes	$\mathbf{P(y > Y)}$	**Transformation function**
Lin-Lin	$X = \frac{1-Y}{a}$	$Y = a(1-x)$
Lin-Log	$X = 10^{-\frac{Y}{a}}$	$Y = -a \cdot log_{10}(X)$
Log-Lin	$X = 1 - b \cdot log_{10}\left(\frac{Y}{a}\right)$	$Y = a \cdot 10^{\frac{1-X}{b}}$
Log-Log	$X = a \cdot 10^{-b \cdot log_{10}(Y)}$	$Y = a \cdot x^{-b}$

The parameters, a and b, of the transformation functions are adjusted according to the empirical degree distributions. The parameters for each block used to generate simulated contact networks in the preferential attachment configuration model are presented in Table 4.2.2.

Due to the stochasticity in the configuration model, several simulated networks are generated in comparison to the empirical network. The cumulative degree distributions for all blocks in both the empirical network and the simulated networks are displayed in Figure 4.4.1. The blue curve represents the empirical degree distributions while the light green curves represent the degree distribution of 20 simulated networks.

For all blocks, the degree distributions of the simulated networks are able to replicate the shape of the empirical degree distributions fairly well. However, there is a significant variability in the degree distributions of the simulated networks for a few of the blocks. For instance, in the ADM-MED block and NUR-ADM the stochasticity is large. In addition, for the NUR-MED block the simulated networks do not capture the extreme values at the tail of the empirical distribution. Hence, the simulated networks do not generate hubs of the same size as in the empirical NUR-MED subnetwork. The reason for this deviation may be due to the fact that the empirical degree distribution of the NUR-MED block is considered to be linear on the lin-log scale in the model, which is not completely true. Despite this simplification, the main characteristics of the node degrees are captured. Overall,

Table 4.2.2: An overview of the parameters used in the transformation function for the different blocks. For a given block, Group 1 refers to the nodes in which the degree distribution is calculated for in contact with individuals in Group 2.

Group 1	Group 2	*a*	*b*
ADM	ADM	1.5	0.5
ADM	MED	3	2.5
ADM	NUR	1	0.3
ADM	PAT	150	–
MED	ADM	15	1.1
MED	MED	131	1908
MED	NUR	430	–
MED	PAT	340	–
NUR	ADM	325	–
NUR	MED	200	–
NUR	NUR	60	0.6
NUR	PAT	0	610
PAT	ADM	0.8	0.5
PAT	MED	115	–
PAT	NUR	670	–
PAT	PAT	55	–

the shape of the simulated degree distributions fit the empirical degree distributions well. Since, the distribution of degrees is the most important characteristic determining the contact patterns in an interaction network, the preferential attachment configuration model seems to serve as a good basis for generating networks that can be utilised in disease spread simulations.

4.3 Epidemiological Model

This section describes the results from the epidemiological simulations run on both the empirical and the simulated networks. The simulations are mainly run with three different probabilities of infection; 0.01, 0.005, and 0.001. The rationale for selecting these probabilities was to generate a sufficient amount of cases to cause an outbreak without the epidemic growth rate being unreasonably high. A relatively high infection probability is required as the hospital ward is considered as an isolated system, meaning no new cases are imported during the simulation. All simulations are run for a period of 60 days, and as described in Section 3.4.3, Scenario 1 describes the contact network period used for the simulated networks. That means that a single contact network is generated by the configuration model which is reused for all 60 simulation days. The elements in the generated contact array is divided by four to replicate the contacts from one day, instead of

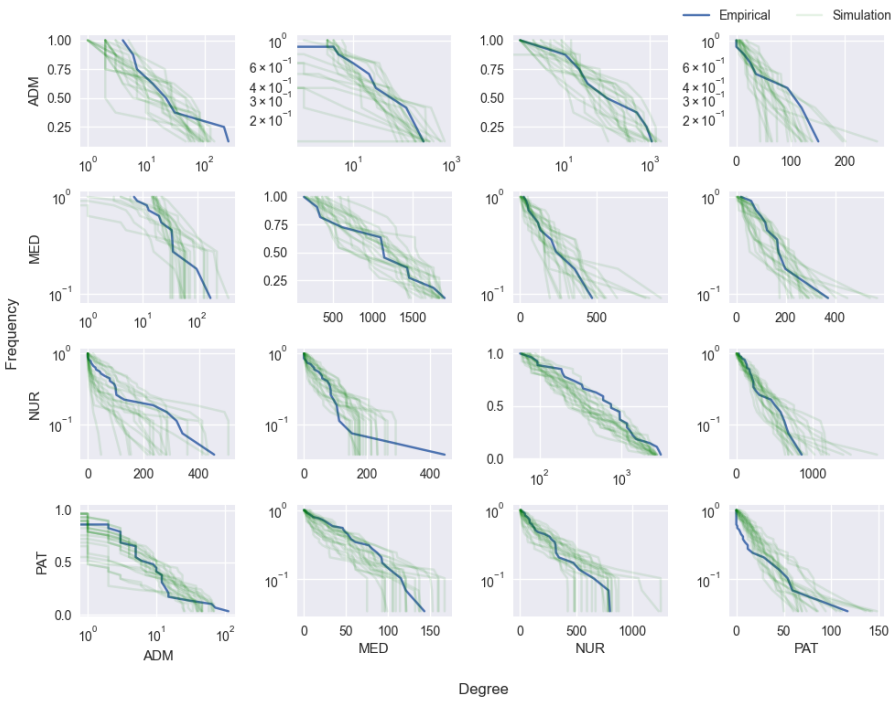


Figure 4.2.3: The cumulative degree distributions of all blocks in the empirical network are shown in blue, while the cumulative degree distribution of 20 simulated networks are visualised in green.

the aggregated network of four days.

4.3.1 Disease Spread

In the model, disease spread is simulated on both the empirical network and the simulated networks generated with the configuration model. This allows for a comparison of the results between the two simulation methods and further an assessment of the configuration model. Similar results imply that the configuration model succeeds in reproducing the contact patterns important for disease spread in a hospital ward setting.

Disease Dynamics

Figure 4.3.1 presents the disease dynamics of simulations run with an infection probability of 0.01, for both empirical and simulated networks. The simulations were run for 60 days, and the number of infected end recovered individuals are calculated for each day. Each run is initiated with one seed node in the presymptomatic state, meaning the individual is infectious without currently manifesting symptoms. The results are obtained from 20 simulations and an average is visualised as a solid line in the plots.

The area under the red curves, representing infected individuals, looks similar for both

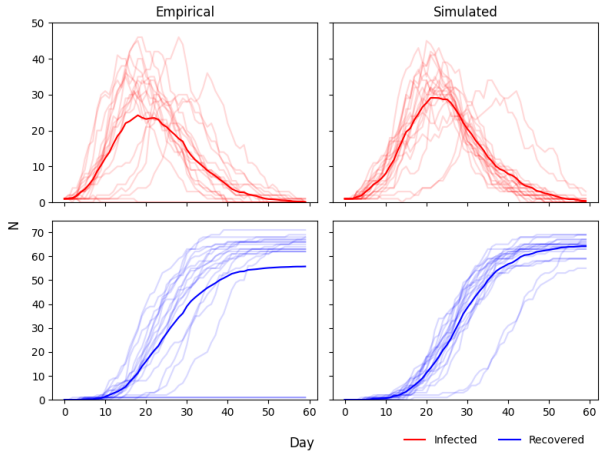


Figure 4.3.1: The figure presents the number of infected (red) and recovered (blue) individuals on each day of the simulation, for both the empirical (left panels) and simulated networks (right panels). The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.01. The transparent curves represents the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

the empirical and the simulated networks. The recovered curve (blue) is, however, slightly lower for the empirical network, reflecting a higher total number of infected individuals during the simulation period for the simulated networks. In the beginning of the simulation, the outbreak grows somewhat faster for the empirical network which reaches its peak of infected individuals immediately before 20 days. The peak for the simulated network is, on the other hand, just after 20 days. The outbreaks die out towards the end of the simulation period for both contact networks, but slightly earlier for the simulated networks. A few of the empirical simulations result in practically no disease spread, indicating that there are impasses in the contact routes preventing transmission. The reason for this may be that a single aggregated simulated network is reused throughout the simulation, while data from four separate days are utilised for the empirical simulations. Consequently, the contact patterns are more mixed in the constructed networks leading to more consistent simulations. The overall trend of the disease dynamics are however similar for both contact networks, indicating that the simulated networks mirror the most important topological features impacting disease spread.

The corresponding results from simulations with infection probabilities of 0.005 and 0.001 can be found in Appendix B.1.1.

Test of Seed Node

As shown in the previous figure, most individuals are infected during a simulation running for 60 days with an infection probability of 0.01. The reason that some individuals do not get infected may be due to the stochasticity of the model, or dead ends in the contact network physically preventing transmission. This may be dependent on the individual ini-

tiated as the seed case. To test how the selected seed node affects the disease spread, 100 simulations were run with each of the 75 nodes of the system set as seed node. The nodes catching the infection during each simulation was tracked, generating a 75x75 matrix with entry i, j representing how many times node i was infected out of the 100 simulations with node j as the seed node. The test was run for presymptomatic and asymptomatic seed nodes with the three infection probabilities used for simulations, for both the empirical and the simulated networks. The generated arrays were translated into heatmaps for visualisation. The colours of the heatmaps range from black to white according to the infection incidents ranging from 0 to 100. Figure 4.3.2 displays the heatmaps generated with presymptomatic seed nodes. The corresponding figure generated from simulations with asymptomatic seed nodes can be found in Appendix B.1.1.

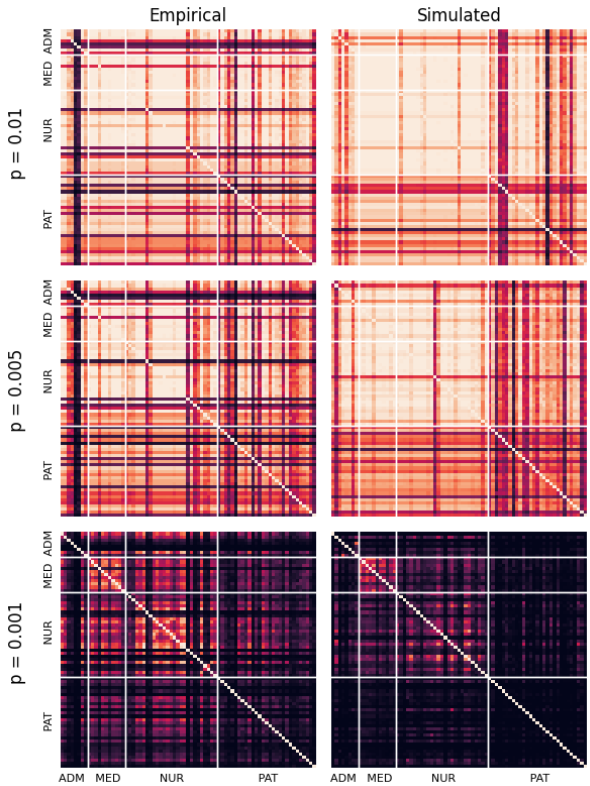


Figure 4.3.2: Heatmaps visualising the number of times node i gets infected out of 100 simulations with node j as the seed node, which is initiated in the presymptomatic state. The subplot contains heatmaps for simulations with both the empirical network (left panels) and a simulated network (right panels) for three infection probabilities; 0.01, 0.005, and 0.001.

Generally, the patterns of the heatmaps are similar for all probabilities. However, the empirical heatmaps are darker than the simulated heatmaps for $p = 0.01$ and $p = 0.005$, while the simulated heatmap is darker for $p = 0.001$. Hence, the nodes in the

empirical network have a lower probability of becoming infected for high and intermediate infection probabilities. Conversely, at a lower transmission rate the empirical nodes have a higher probability of catching the infection. In other words, the variation is greater for the empirical heatmaps and the most significant deviations are found at the extreme infection probabilities.

The rows and columns that are all black, regardless of the infection probability, reflect impasses in the contact network. The topology of the network physically prevents certain nodes from becoming infected when particular nodes are initialised as the seed. This is more prominent in the empirical simulations, implying that the contact patterns are more heterogeneous than in the constructed ones, which is consistent with previous discoveries. Overall the results of the simulated networks are close enough to the results in the empirical network, even though all topological complexities are not captured.

4.3.2 Reproduction Number

The basic reproduction number, R_0, quantifies the transmission potential of a pathogen. R_0 can be defined as the average number of secondary cases caused by an individual infection when the entire populations is susceptible to infection. In the model, R_0 is therefore calculated as the number of individuals infected by the initial seed node. Due to the stochastic variability in the model, the basic reproduction number is calculated as an average from 100 simulations. To investigate if seeding individuals from certain groups cause larger outbreaks, R_0 is quantified for simulations with seed nodes from all of the hospital groups; ADM, MED, NUR and PAT. Figure 4.3.3 presents an overview of pie charts and the computed R_0 for both the empirical and simulated networks with seed nodes initiated in the asymptomatic or the presymptomatic state. The segments of a pie chart reflect the proportion of simulations that resulted in a given number of secondary infections caused by the seed node, while the colours represent the number of individuals infected by the seed node, shown as one to six or more than six. Each of the pie-charts are obtained from 100 simulations with an infection probability of 0.005. Corresponding figures simulated with $p = 0.01$ and $p = 0.001$ can be found in Appendix B.1.2.

For all scenarios, the R_0 is significantly higher for simulations run with a doctor (MED) or a nurse (NUR) as the seed node. This is expected as the doctors and the nurses have more contacts than administrative workers and patients in both the empirical and simulated networks. For example, simulations with presymptomatic MED seed nodes generate basic reproduction numbers that are more than twice as high compared to presymptomatic ADM seed nodes. The R_0 of the simulated network scenarios are generally very similar to the empirical network scenarios. For instance, an asymptomatic ADM seed node generates an R_0 of 1.20 and 1.16 for the empirical and simulated network, respectively. The largest deviation is found in the scenario where the seed node is a presymptomatic nurse. The empirical R_0 is 7.84 while the simulated R_0 is 4.95. Apart from this case, the average values of R_0 are comparable even though the proportions of the pie charts vary between empirical and simulated. The calculated R_0 from simulations with a presymptomatic seed node is significantly higher than for asymptomatic seed node cases. This is according to expectations, as the the relative infectivity of an presymptomatic interaction is distinctly higher than an asymptomatic interaction. Overall, the simulated R_0-values correspond well with the empirical R_0-values, and the difference between the groups is

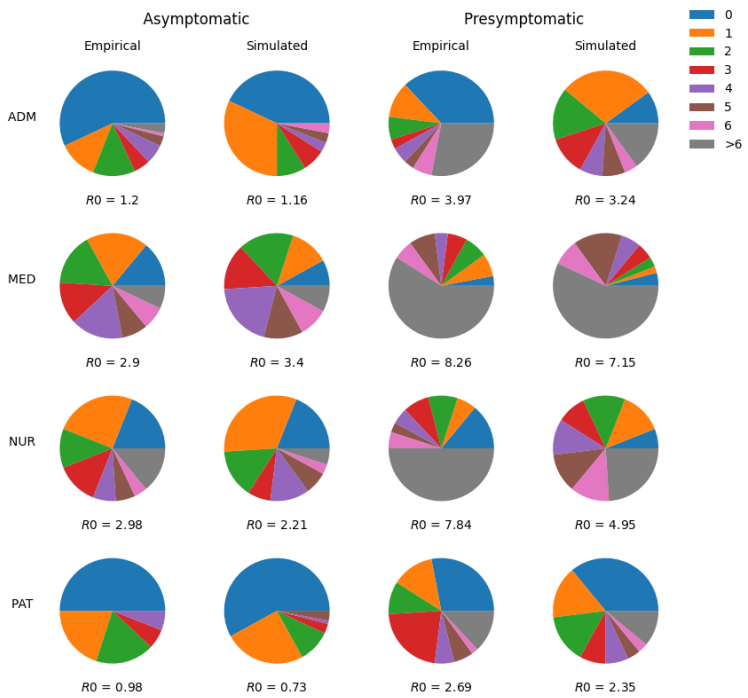


Figure 4.3.3: An overview of pie charts describing the proportion of 100 simulations with $p = 0.005$ resulting in a given R_0. Pie charts are generated for simulations on the empirical and constructed networks with an asymptomatic or presymptomatic seed node. The pie charts are also distinguished by the hospital status of the seed node. The average R_0 for each scenario is denoted the below pie chart.

according to expectations.

The relationship between the infection probability, p, and the basic reproduction number, R_0, is presented in Figure 4.3.4. It is expected that R_0 will be saturated at some point with an increasing p because the contact networks limit the spreading potential. This is however not shown for the selected infection probabilities. There is an approximately linear relationship between p and R_0 for all scenarios, but R_0 is not necessarily proportional to p. The results based on the simulated network fit the empirical network well, and the trends are similar for simulations with asymptomatic and presymptomatic seed nodes, although the R_0 is generally higher for the latter. Additionally, it is clear that MED and NUR seed nodes cause significantly larger outbreaks than ADM and PAT seed nodes, which is consistent with the contact patterns discovered in the network analysis. Individual plots, with 95 % confidence intervals, for all the curves shown in Figure 4.3.4 can be found in Appendix B.1.2.

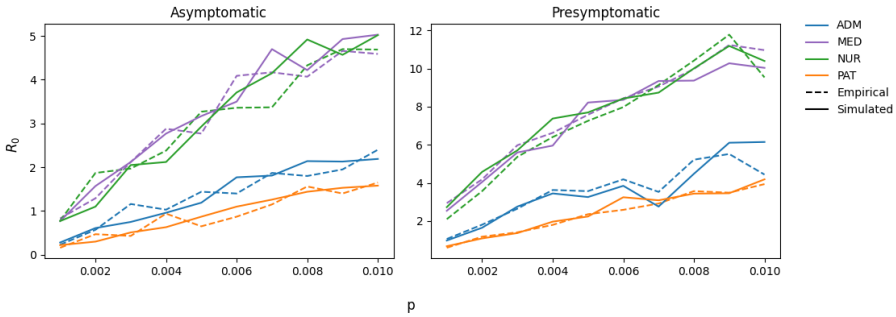


Figure 4.3.4: A representation of the relationship between the infection probability, p, and the basic reproduction number, R_0. R_0 is calculated for ten infection probabilities ranging from 0.001 to 0.01. The colours represent the hospital status of the seed node, while the dotted and solid lines represent the empirical and simulated network, respectively. The results in the left panel are generated with asymptotic seed nodes, while the right panel is generated from presymptomatic seed nodes.

4.3.3 Pooled Testing

An application for an agent-based model simulating the spread of SARS-CoV-2 is to test the effect of intervention measures. A testing regime was implemented in the model to investigate how regular testing affects the epidemic spread as well as the absence of health care workers. Two test strategies were simulated; weekly and daily testing of all individuals present at the hospital ward. Individuals with positive test results are quarantined for four days. The goal of testing symptom free individuals is to detect the asymptomatic and presymptomatic cases to isolate the individuals, preventing them from causing further transmission. It is expected that frequent pooled testing will reduce the number of confirmed cases, but it will also lead to a higher absenteeism of health care workers. The cost of isolating HCWs capable of working may outweigh the benefit of reducing the spread. For different variants of SARS-CoV-2 with different contagiousness, the cost of isolation versus the benefit of reducing the infection rate is not the same. For highly contagious variants, such as the Omicron variant, eradicating the virus is extremely demanding. Therefore, in order to maintain the operations of the hospital, pooled testing of the entire unit is not beneficial. For less contagious variants, regular testing can be an efficient method to mitigate the spread and still maintain activity.

To investigate the effect of pooled testing, the daily count of absent health care workers and the total infected individuals in all groups are compared for simulations run with and without testing. The group of HCWs is comprised of only the medical doctors and the nurses. Administrative staff are not included as their absence is believed to have less impact on the ability of a hospital unit to stay open. Illustrations of the results from 20 simulations run with $p = 0.01$ and $p = 0.005$ are shown in Figure 4.3.5 and Figure 4.3.6, respectively. In addition, Figure 4.3.7 provides a comparison of absent HCWs with and without a daily testing regime with $p = 0.01$. The left panels represent the empirical results, while the right panels describe the results from simulations on networks generated with the configuration model. The daily count of absent HCWs with no testing regime implemented is shown in blue in the upper panels, while the results from weekly testing are

shown in the middle panels represented by orange curves. The transparent lines represent the results from each of the 20 runs, while the solid lines are the average values. In the lower panels the average number of absent HCWs are visualised in the same plot for better comparison. In addition, the daily count of infected individuals across all groups are included, represented by dotted lines. This allows for an instant analysis of the effect of testing on both the absenteeism and daily infections.

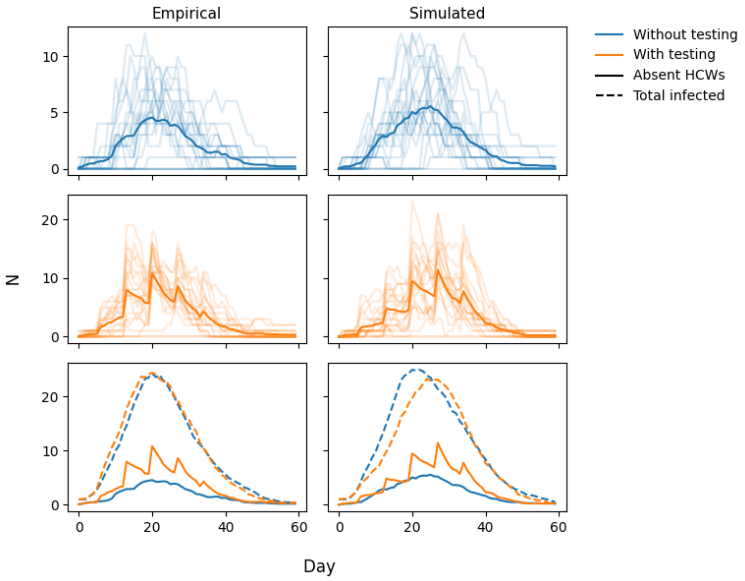


Figure 4.3.5: A comparison of the number of absent HCWs and the total number of infected individuals in all groups for simulations run with weekly testing (orange) and without testing (blue). The results are obtained from 20 simulations with an infection probability of 0.01, using both the empirical network and simulated networks. The transparent lines represent the number of absent HCWs at each simulation, while the solid lines are an average of the 20 simulations. The dotted lines represent an average of the total number of infected individuals, not only HCWs.

As shown in Figure 4.3.5, the impact of weekly pooled testing is similar for the empirical and simulated model runs with $p = 0.01$. The average number of daily infections are not affected by a weekly pooled testing regime. However, the absence of HCWs is higher when weekly testing is implemented. The spikes in the absence rate depicts the length of the imposed quarantine, which is four days, before HCWs can return to work. Since the resources spent on testing only result in higher absenteeism of HCWs without reducing the number of infections, weekly testing is not beneficial at such a high infectivity level. This is according to expectations, as it is difficult to mitigate the spread of highly contagious virus variants. Testing every seven days of the simulation period is therefore not frequent enough to reduce the number of cases, and only leads to fewer available HCWs.

For model simulations run with an infection probability of 0.005, as shown in Figure 4.3.6, the results tell a different story. The absence rate varies when weekly testing is implemented, nevertheless it is closer to the absence rate generated without testing than

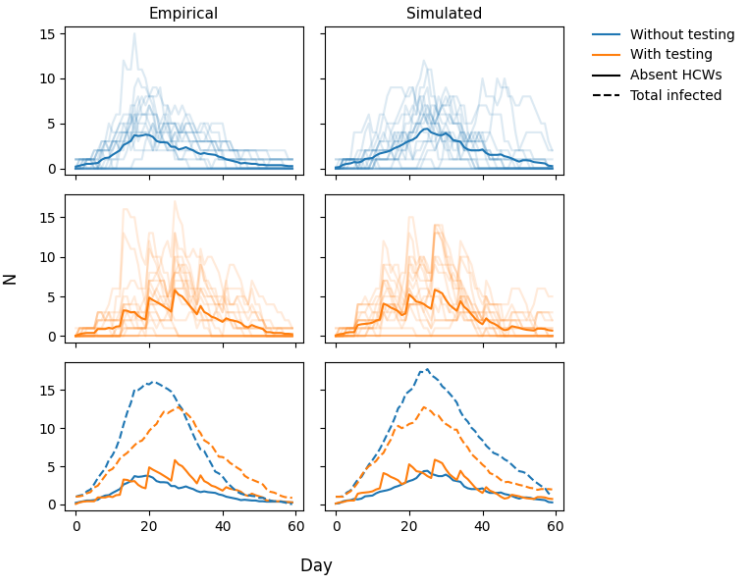


Figure 4.3.6: A comparison of the number of absent HCWs and the total number of infected individuals in all groups for simulations run with weekly testing (orange) and without testing (blue). The results are obtained from 20 simulations with an infection probability of 0.005, using both the empirical network and simulated networks. The transparent lines represent the number of absent HCWs at each simulation, while the solid lines are an average of the 20 simulations. The dotted lines represent an average of the total number of infected individuals, not only HCWs.

for simulations with $p = 0.01$. In addition, weekly testing causes a significant reduction in the number of daily infections for both the empirical and simulated networks. Weekly pooled testing therefore seems to be an effective intervention measure to reduce the spread of SARS-CoV-2 in a geriatric hospital unit, without causing a high absenteeism of HCWs. Again, it is expected that the effect of testing will be more prominent at lower infectivity levels. In conclusion, weekly pooled testing is only beneficial when the probability of infection transmission for each contact is below 0.01. The results obtained from weekly testing with $p = 0.001$ can be found in Appendix B.1.3.

To be able to mitigate the spread of a virus with an infection probability of 0.01, the frequency of pooled testing must be increased. At a hospital, it is assumed that the resources required to perform daily pooled testing of a unit consisting 75 people are available. Figure 4.3.7 presents the results obtained from simulations with daily pooled testing. In the lower panels, it is clear that both the daily number of absent HCWs and infected individuals is reduced to approximately zero when daily testing is implemented. This is true for both the empirical and simulated contact networks. Hence, testing every day is sufficient to completely eradicate the virus and prevent absence of HCWs. In conclusion, a strategy of pooled testing can also be beneficial at high infectivity levels if it is performed frequently enough. In the model, the only test strategies that were used were weekly and daily testing. However, based on the results from daily testing, less frequent testing than daily testing

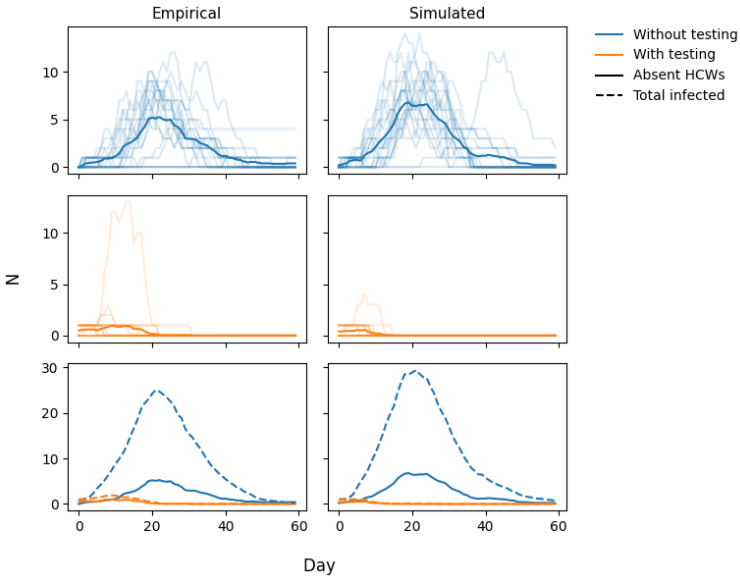


Figure 4.3.7: A comparison of the number of absent HCWs and the total number of infected individuals in all groups for simulations run with daily testing (orange) and without testing (blue). The results are obtained from 20 simulations with an infection probability of 0.01, using both the empirical network and simulated networks. The transparent lines represent the number of absent HCWs at each simulation, while the solid lines are an average of the 20 simulations. The dotted lines represent an average of the total number of infected individuals, not only HCWs.

may also be sufficient to mitigate the spread. The same trends are found for daily testing with infectivity levels of 0.005 and 0.001, as illustrated in Appendix B.1.3. The modelled hospital is assumed to be an isolated system, meaning no new cases of COVID-19 are imported. Due to this, the probability of infection must be sufficiently high for an outbreak to persist. As a result, the selected transmission rates are possibly artificially high, influencing the observed effect of the simulated testing strategies.

4.4 Error Analysis

The following section describes the error analysis performed on the model and possible explanations for model deviations are proposed. First, an assessment of the preferential attachment configuration model is presented. Network statistics which are not accounted for in the model are calculated to provide insight about the deviations in the simulated networks which may affect the epidemic modelling. Subsequently, an analysis of the contact network periods in the epidemiological model is presented.

4.4.1 Assessment of Configuration Model

Although the preferential attachment configuration model has succeeded in generating networks with similar degree distributions within the blocks, other empirical network properties may not be captured in the model. To discover possible explanations for deviations in simulated networks in the epidemiological modelling, several network properties were characterised. In other words, this section investigates if reproducing the degree distributions of the empirical subnetworks was sufficient to capture other underlying structures of the network. First, the cumulative degree distribution of aggregated simulated networks is compared to the empirical degree distribution. Additionally, the assortativity of the complete networks as well as the subgroups are presented for both the simulated and empirical network.

Total degree distribution

In the preferential attachment configuration model, the number of contacts assigned to each node is determined independently for all of the blocks. In consequence, the degree distribution of the entire simulated network may not fit the empirical network even though the degree distribution within the blocks matches. Hence, an explanatory factor for deviations in the model may be that the contact patterns across the blocks are not sufficiently reconstructed. To investigate this possibility, the total cumulative degree distribution for 20 simulated networks were plotted, as shown in Figure 4.4.1. Apart from a few outliers, the simulated networks tend to generate fewer large hubs than in the empirical network. Consequently, a larger proportion of the nodes in the simulated networks have comparable degrees, i.e. the networks are more homogeneous. This characteristic can possibly affect the transmission of an infectious disease in the population.

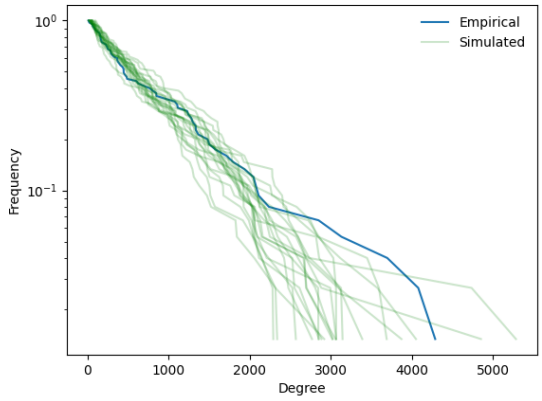


Figure 4.4.1: The cumulative degree distribution of the aggregated empirical contact network is shown in blue, while the cumulative degree distributions of 20 simulated networks are depicted in green.

Assortativity

Another possible reason for discrepancies in the simulated networks is that there is a degree of assortativity in the empirical network that is not captured by the configuration model. The degree correlation of a node refers to the relationship between its degree and the average degree of its neighbours. The degree correlations for the nodes of the aggregated empirical network are presented in Figure 4.4.2. The dots of the scatter plot are coloured based on the hospital group each node belongs to. For assortative networks, the average neighboured degree increases with k, meaning that hubs tend to connect to each other and small-degree nodes connect to each other. As shown in Table 4.4.1, the degree correlation coefficient of the total empirical network is calculated to be 0.25. The positive coefficient and the slightly increasing trend of the dots in Figure 4.4.3 proves that there is a degree of assortativity in the empirical network. Since it is difficult to visualise degree correlations for several networks in a single plot, only the calculated degree correlation coefficients are presented for the simulated networks, as shown in Table 4.4.1. The coefficients are an average from 100 simulated networks with a 95% confidence interval. The total degree of assortativity for the simulated networks is calculated to 0.51 (95% CI: 0.49, 0.53). This value deviates significantly from the empirical network, proving that this network property has not been reproduced in the simulated networks.

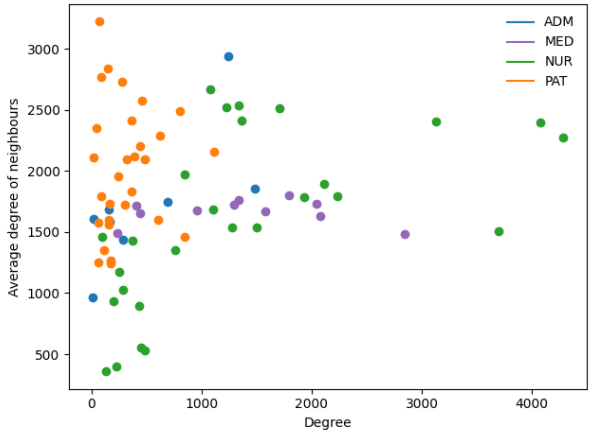


Figure 4.4.2: A scatter plot which presents the correlation between degree and average degree of neighbours for all nodes in the empirical network. The nodes are distinguished by colour based on the hospital group the belong to.

Although the overall network is assortative, it does not mean that the subgroups of the network are. For example, the degree correlation of a doctor or a nurse may differ from an administrative worker or a patient. The calculated degree correlation coefficients for the subgroups, consisting of ADM, MED, NUR and PAT, for both the empirical and the simulated networks are presented in Table 4.4.1. In addition, the degree correlation for each group in the empirical network are visualised in Figure 4.4.3. For the ADM and the NUR group there is a clear tendency of increasing average neighboured degree with increasing

k. The degree correlation is 0.76 and 0.68 for the ADM and the NUR group, respectively, meaning there is a high degree of assortativity. The values for the simulated networks are, however, significantly lower; 0.36 (95% CI: 0.29, 0.43) for the administration and 0.40 (95% CI: 0.36, 0.45) for the nurses. In contrast, the degree of assortativity for the patients is remarkably higher for the simulated network than the empirical network, with a correlation coefficient of 0.40 (95% CI: 0.37, 0.44) compared to 0.10 for the empirical network. The empirical degree correlation coefficient for the medical doctors is 0.01, indicating a neutral network. The degree of assortativity for the simulated networks is closest to the empirical network for this group with a value of 0.09 (95% CI: 0.02, 0.16). The empirical coefficient does however not lie within the 95% confidence interval. On the whole, all of the groups in both the empirical and the simulated networks admittedly have positive degree correlation coefficients, meaning that the networks are assortative. However, the deviations in the simulated networks are clear and the analysis proves that the the configuration model is not able to capture the degree of assortativity found in the empirical network.

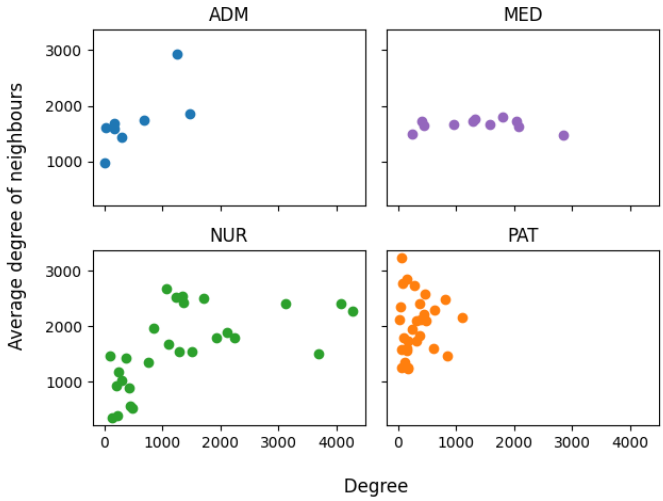


Figure 4.4.3: The correlation between degree and average degree of neighbours plotted for all nodes in their respective groups. The upper panels present the degree correlations of the administration and the medical doctors, while the lower panels display the degree correlation of nurses and patients.

4.4.2 Contact Network Period

The empirical contact network contains data from four consecutive days, but for modelling purposes the data was aggregated into one static network. When reproducing the empirical contact network, topological properties were extracted from the aggregated network resulting in a single aggregated simulated network. The epidemiological model is built to simulate state progression and count the number of individuals in each state every day, meaning that daily contact networks are used in the model. The empirical contact data was

Table 4.4.1: The calculated degree correlation coefficients for the four groups and the total degree correlation for the entire network of both the empirical and simulated network. The simulated degree correlation coefficients are an average of 100 simulated networks with a 95 % confidence interval.

	Empirical	**Simulated**
ADM	0.76	0.36 [0.29, 0.43]
MED	0.01	0.09 [0.02, 0.16]
NUR	0.68	0.40 [0.37, 0.44]
PAT	0.10	0.40 [0.36, 0.45]
Total	0.25	0.51 [0.49, 0.53]

therefore divided into four days which were reused alternately throughout the simulation. In this way the variations between the days and the heterogeneity of the empirical contact network is captured in the model. Conversely, for the simulated contact network, a single contact network is generated for each epidemiological simulation. Since the network is generated based on empirical contact data from four days, all the entries in the array are divided by four to generate a daily contact array which is reused every day. This method is referred to as Scenario 1. In Section 3.4.3, two other possible scenarios for simulating spread were proposed. In Scenario 2, four aggregated contact networks are generated and then divided by four. The four daily contact arrays are reused throughout the simulation, similar to the empirical daily contact networks. Lastly, in Scenario 3 a new aggregated contact network, divided by four, is generated for each simulation day.

To analyse how the selected contact network period impacted the epidemiological simulations, all three scenarios are compared in Figure 4.4.4. The presented results are obtained from simulations with an infection probability of 0.01. Corresponding figures generated with infection probabilities of 0.005 and 0.001 can be found in Appendix B.2.1.

In theory, Scenario 2 is the most similar to the empirical approach, as there are four daily contact networks that are used repeatedly throughout the simulation. However, the number of cases is significantly higher than in Scenario 1, where only one daily contact network is utilised. In general, Scenarios 2 and 3 produce more infections than Scenario 1. It is expected that increasing the amount of generated contact arrays will lead to more cases since the daily contact networks are generated independently of each other, producing a more mixed population. In the empirical network, the daily contact networks emerge from four consecutive days of data collection. The contact patterns of the four days are therefore more dependent on each other, assuming that doctors and nurses work somewhat similar shifts over a certain period of time and that patients are admitted for several consecutive days. Although the contagiousness of Scenario 3 is too high compared to the empirical network, it provides a more realistic basis for predictive simulations. However, for the purpose of comparing the empirical and simulated contact networks, selecting Scenario 1 appears to be acceptable because reusing a single contact network prevents the number of infections from exceeding the outcome of empirical simulations. This is substantiated by the results described in Section 4.3, which prove that the transmission dynamics are similar for the empirical network and the simulated network from Scenario 1.

Similar to how the simulations were run with the constructed network, the empirical

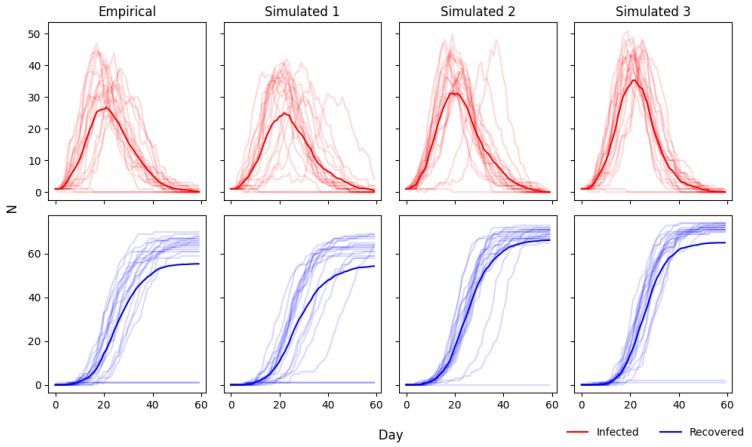


Figure 4.4.4: The number of infected (red) and recovered (blue) individuals on each day of the simulation for the empirical contact network and three different contact network periods of the simulated networks. Simulated 1, 2 and 3 refers to the three scenarios described above. The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.01. The transparent curves represent the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

simulations could have been run by reusing the contact data from only one day. Figure 4.4.5 presents the dynamics of the number of infected and recovered individuals for five different empirical contact network periods obtained from simulations with an infection probability of 0.01. Corresponding figures generated with infection probabilities of 0.005 and 0.001 can be found in Appendix B.2.1. In the figure, total refers to the method used in the epidemiological model, where the contact networks from four consecutive days alternate throughout the simulation. Day 1 to Day 4 refers to runs with the contact network from a single day reused for the entire simulation period. As shown in the figure, there are noteworthy variations in the results between Total and the four individual days. In addition, the infection levels are significantly higher for the simulations using all of the contact data. The empirical data contains contact data from only four days, which is essentially scarce. Utilising only one day of the contact data available therefore reduces the robustness of the model since the results are biased towards the contact patterns of a single day, which may not be representative for the general contact patterns of the hospital ward. In conclusion, the most realistic simulations are obtained by alternately using the contact data from all four days.

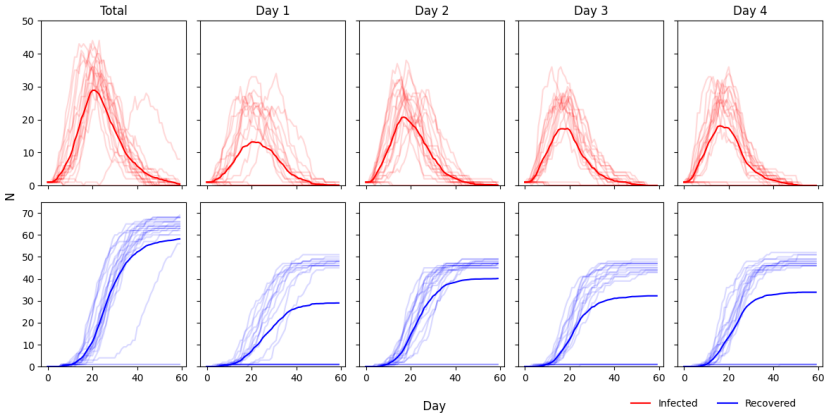


Figure 4.4.5: The number of infected (red) and recovered (blue) individuals at each day of the simulation for different contact network periods of the empirical network. Total refers to simulating with the four days alternately, while Day 1 to Day 4 refers to repeatedly simulating with one daily contact network. The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.01. The transparent curves represent the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

Chapter 5

Discussion

Models are simplified representations of real life systems constructed to provide understanding of observed phenomena or to predict future phenomena [66]. This is also the case for the model developed in this thesis. As described in Chapter 3, several simplifications and assumptions have been made for the simulations to be feasible and effective. The following section discusses the limitations of the model, the main assumptions and the challenges they imply, as well as key observations.

In agent-based modelling (ABM), complex systems are viewed as sets of interacting agents. ABMs can be used to investigate how macro-level properties of a system are dependent on micro-level rules [66]. Uncertainties in the behaviour of agents will consequently affect the model output. High-quality data is therefore required to develop a realistic representation of a real-life system. Hence, the most obvious limitation of the modelling framework in this project is the size of the data it is based on.

The simulated network is constructed based on a single empirical data set containing contact data collected over the course of four days. 75 individuals at a geriatric hospital ward participated in the study. They are further divided into four groups based on their status at the hospital, resulting in group sizes ranging from 8 to 27 individuals. Essentially, the data serving as a basis for this project is scarce, which gives rise to three main issues. Firstly, the size of empirical contact network and its subnetworks are small. Rules governing individual behaviour are therefore derived from few data points, causing a very stochastic system. For example, drawing conclusions about the behaviour of administrative workers based on 8 individuals in the empirical data is challenging. Hence, the small system can lead to uncertainty in the contact patterns generated in the simulated networks. Secondly, the data is collected from a limited period of time. Specifically, the contact patterns at the hospital ward during four days may not be representative for the general activity. Due to this, a synthetic network with the most central network properties of the empirical network was generated to model the epidemic over a longer period of time. However, the model parameters are derived from uncertain data which can greatly impact the output. For instance, if a special occasion influenced the contact patterns of the staff members in one of the days in the study, it creates a bias in the generated contact

network. Thirdly, the contact patterns of a single hospital unit may not representative for other hospital settings. The structure of shifts and the type of patients treated at the given unit presumably impacts the contact patterns. For example, HCWs interact more closely with patients in an intensive care unit than in an observation unit. For this reason, several empirical interaction networks should be utilised to extract general contact patterns. However, this was not feasible due to time constraints and the need to prevent the empirical analysis from being too complex. Overall, the model is sensitive to the behaviour of few individuals over the course of few days, causing uncertainty in the results. Additionally, it is difficult to draw conclusions about the transferability of the findings to other hospital settings.

In order to generate a simulated contact network, the characteristics of the empirical network were simplified. Arguably, the degree distribution is one of the most central network properties associated with epidemics because it naturally captures the heterogeneous potential of individuals to become infected and to further transmit the infection [63]. Based on this, the configuration model assumes that reproducing the degree distribution of the subnetwork blocks is sufficient to model realistic spreading phenomena. The results described in Section 4.3 substantiates this hypothesis as the disease dynamics are similar for the empirical and the simulated network. The error analysis proved that several network statistics, such as the total degree distribution and the the degree of assortativity, deviated significantly from the empirical network. This implies that the configuration model fails in capturing several underlying structures of the empirical network. However, these discrepancies have seemingly limited effect on the outcome of the epidemiological simulations.

To summarise, recreating the degree distributions of the empirical blocks seems to be sufficient to model realistic disease spreading phenomena. Drawing general conclusions from the model is however challenging due to the scarce data. Nevertheless, it is believed that the modelling framework can provide valuable insight into the spreading dynamics of SARS-CoV-2 in a hospital setting.

5.1 Key Assumptions

In addition to the simplifications outlined above, several key assumptions have been made in the model. The most important assumptions and the challenges they bring about are discussed in the following section:

- **Hospital ward as an isolated system:** A considerable simplification in the model is that the hospital ward is assumed to be an isolated system. The only source of infection is therefore the initially infected seed node. Since no new cases are imported into the system, the probability of infection must be sufficiently high to prevent the outbreak from dying out. The transmission parameters are therefore possibly artificially high, which would affect the spreading dynamics. Consequently, an exaggerated transmission rate may cause unrealistic responses to implemented intervention measures. Certain mutants of SARS-CoV-2 can potentially generate a local reproduction number of this magnitude, however, based on existing data it is difficult to come to that conclusion. In real life, new cases can be introduced to the hospi-

tal through admitting infected patients or by staff members catching the infection from other sources. An interesting extension of the model would be to account for external import of new cases. This could be achieved by defining a probability of introducing new cases based on empirical data.

- **Transmission through close proximity contacts:** The interactions between agents of the model are based on empirical contact data collected with wearable RFID-sensors. The sensors are worn on the chest, and the threshold for detecting a contact is 1.5 meters. First of all, the model therefore assumes that the inter-individual transmission is captured by face-to-face proximity contact. Secondly, the model neglects other possible routes of SARS-CoV-2 transmission. Throughout the pandemic the FHI's definition of a close proximity contact has varied between 1 and 2 meters. The reach of the RFID-sensors therefore seem reasonable for capturing the direct contacts, which presumably account for the majority of transmissions. However, several other routes of transmission are proposed in reality. For instance, aerosol transmission and indirect transmission through contaminated surfaces are believed to be of importance [16, 19]. Restricting the spread of SARS-CoV-2 in the model to close proximity interactions is therefore a simplification which may reduce the transmission potential compared to a real life scenario. Increased transmissibility would increase the requirements of the restrictions to limit the spread, and the response to control measures in the model would be affected. For instance, infected patients are assumed to have no transmission potential in the hospital due to isolation and infection control equipment worn by the HCWs. However, if other routes of transmission are included, such as aerosol transmission, the implemented restrictions may not be sufficient to prevent patients from infecting others. Consequently, isolated patients could be a source of new infections in the hospital making it more challenging to mitigate an outbreak. Implementing a spatial component in the model would be interesting to account for indirect transmission. Individuals residing in the same room can have a probability of becoming infected despite not having a close proximity encounter. Examples are staff members having lunch break in the same room, or HCWs treating the same patient without being in direct contact with one another. Spatial information about the participants in the empirical study is, however, not provided.

- **Compliance to intervention measures:** Strategies for pooled testing with varying frequency are implemented in the model. Individuals with positive test results are required to quarantine for four days. In the model, all individuals are assumed to comply to these control measures. In addition, symptomatic individuals are instantly removed from the system, meaning individuals are assumed to isolate themselves instantly at symptom manifestation. Presumably, some individuals will not adhere to these intervention measures. People can possibly shorten their quarantine or refrain from isolating themselves when exhibiting symptoms. It would be interesting to include this phenomena in the model to investigate if it notably affects the spread. Since an agent-based model can account for individual behaviour, it is possible to model agents that ignore the implemented intervention measures.

- **Low temporal resolution:** In the epidemiological model, disease progression and

state transitions are carried out at each simulation step, i.e. once a day. It assumes that a temporal resolution of 24 hours is sufficient to realistically simulate spread. Modelling spread day by day makes it difficult to draw conclusions from the responses to daily testing. In reality, individuals can infect one another during the 24 hour interval, causing a delay from infection to having a positive test result. In the model however, individuals are tested immediately after infection if the daily testing regime is implemented. The temporal resolution of the original empirical contact data is 20 seconds. This high-resolution contact data is aggregated to 24 hours to fit the epidemiological modelling framework. In reality, one could think that an individual working the early shift could indirectly infect an individual working the late shift. This issue can however be ignored due to the incubation period of SARS-CoV-2, which is longer than a day [28, 29]. Additionally, in the empirical data only direct, symmetrical contacts are registered, meaning that the time component can be ignored. Nevertheless, initial analysis of the empirical data demonstrated great variance in the contact patterns between the early, late and night shifts, caused by the differing working hours of groups of staff. A weakness of the model is that the contact patterns across shifts are not captured. Due to time constraints, the shift structure was neglected to significantly simplify the process of recreating the contact patterns of the empirical network in the configuration model. Presumably, the overall contact patterns affecting the disease dynamics are still captured. However, implementing a shift structure gives rise to opportunities such as investigating the effect of control measures at varying levels of activity.

5.2 Key Observations

The key observations from the results presented in Section 4 are discussed in the following section. Possible explanations for deviations in the configuration model and the epidemiological model are outlined.

- **Lack of heterogeneity in simulated networks:** Compared to the empirical network, the simulated networks lack heterogeneity in the node degrees. This can be seen in the block degree distributions shown in Figure 4.2.3. For most of the blocks, the simulated degree distributions replicate the empirical degree distribution sufficiently. However, there is a prominent stochastic element in the model which causes varying results. Specifically, the ADM-MED and the NUR-ADM block vary greatly. Since the administrative staff generally have few contacts, the number of interactions with doctors and nurses can significantly impact their potential of catching the virus. For several blocks, such as the ADM and the NUR-MED blocks, the simulated networks do not reproduce the heavy tail of the empirical network. Hence, the simulated networks do not generate hubs of the same size as in the empirical network. This characteristic is also observed in the total degree distribution, displayed in Figure 4.4.1. Hubs are potential super spreaders due to their large number of contacts. Fewer large hubs would presumably reduce the number of infections, however the epidemiological simulations run with a transmission rate of 0.01 show otherwise. As can be seen in Figure 4.3.1, the average number of infected individu-

als is actually slightly higher for the simulated contact networks. With that said, the results fluctuate due to the stochastic variability in the model. The spreading dynamics also heavily depend on the transmission rate and the behaviour of the seed node. Figure 4.3.2 shows that the probability of infection is higher in the simulated network for high and intermediate levels of p. At a lower transmission rate ($p = 0.001$) the dynamics shift; the probability of infection is higher in the empirical network. A possible explanation is that the infection level is saturated at higher transmission rates, i.e. certain nodes avoid infection because of physical impasses in the contact network. Since the simulated network is a more homogeneous system, the majority of the nodes are reached by the virus when the transmission rate is sufficiently high. The empirical network, on the other hand, contains more small-degree nodes physically prevented from infection. In contrast, the large hubs of the empirical network generate more infections compared to the simulated network at lower probabilities of infection. This is expected, as the analysis of the simulated network proves that it does not capture the heterogeneity of the empirical network. Since individual behaviour is mainly based on their hospital status, assigning other personal attributes could improve the granularity of the model. An example of this is distinguishing between social and anti-social nodes.

- **R_0 is dependent on the seed node:** As shown in Figure 4.3.3, the reproduction number is highly dependent on the type of node selected as the seed. Doctors and nurses generate significantly higher R_0 compared to the administrative staff and patients, which is as expected. The transmission potential is higher for MED and NUR individuals because they are more connected, which is clear in the interaction matrices visualised in Figure 4.1.1 and 4.2.1. This pattern is consistent for varying levels of infection (Figure 4.3.4). The increased values of R_0 resulting from presymptomatic seed cases reflects the relative infectivity of presymptomaitc contacts, which is 3.0, compared to 0.3 for the asymptomatic contacts. Since the doctors and the nurses seem to be the super spreaders of the hospital, targeting control measures towards them could be effective for mitigating spread. Their increased level of contact is presumably inevitable, as they must interact with each other and the patients to provide the best treatment. However, reducing the purely social contacts of MED and NUR individuals and imposing the use of infection control equipment at all close proximity interactions could be effective. It would be interesting to test such assumptions in the model, however it is very difficult to determine the level of protection that, for example, the use of face masks provide.

- **Effect of pooled testing:** In the model, weekly pooled testing when the transmission rate is 0.01 is not beneficial, as can be seen in Figure 4.3.5. The absenteeism of HCWs is increased, while the level of infection is equally high despite testing. This is expected, as mitigating the spread of a highly contagious virus variants is very challenging. During the pandemic in Norway, most intervention measures were lifted when the the Omicron variant of SARS-CoV-2 dominated, because eradicating it by imposing social distancing and the use of face masks would be impossible due to its transmissibility. This was justified by the fact that the Omicron variant was reported to cause less severe disease, relative to previous variants such as

Delta [67]. For evolving variants of SARS-CoV-2, there is believed to be a trade-off between transmissability and virulence [68]. Hence, the strategy for handling the pandemic is adapted accordingly. For less infective variants with greater disease potential, mitigating the spread is more important. As shown in Figure 4.3.6, weekly pooled testing had a moderate effect on the infection levels for simulations run with a transmission rate of 0.005. However for further reducing the spread, the frequency of testing must be increased. In the model, daily pooled testing (Figure 4.3.7) completely eradicates the virus and consequently prevents HCWs from being quarantined for all the infection probabilities tested. However, it might not be worth spending the resources required for daily pooled testing if the end goal is not to eradicate the virus. For less severe variants, preventing HCWs capable of working from being quarantined is more important. It would be interesting to investigate the effect of pooled testing with other frequencies to discover the optimal frequency for reducing spread and minimising the absenteeism of HCWs.

Chapter 6

Conclusion and Outlook

The COVID-19 pandemic has turned life upside down for the majority of the world population over the past two years. As of now, the vaccine coverage in many countries is high and control measures have been lifted, allowing for people to return to their normal everyday lives. Although the pandemic seems to be coming to an end, modelling the spread of SARS-CoV-2 can provide valuable insight into the underlying spreading mechanisms. Such knowledge is useful for understanding the current situation, as well as preparing for pandemics caused by pathogens utilising similar transmission routes, which will arise in the future.

In this project, an agent-based modelling framework is developed to simulate the spread of SARS-CoV-2 in a hospital setting. The motivation was to generate a hospital module that can be implemented in a larger societal model. The behaviour of the agents are governed by either an empirical contact network or a simulated contact network. The simulated networks are generated with a configuration model based on the degree distribution of the blocks in the empirical network. A block refers to a portion of the interactions in the weighted adjacency matrix of the network. The diagonal blocks describe the internal contacts within a group, while the blocks on the off-diagonal describe the interactions of a bipartite network, i.e. between all combinations of groups. The structure and logic of the epidemiological model is reused from the original COVID-19 model developed at NTNU. The model has been tested with different parameter values, and the user is free to tune input parameters such as the infection probability and the selected test strategy.

In general, the simulated contact networks lack the heterogeneity in the node degrees found in the empirical contact network. However, the epidemiological dynamics for model runs on the empirical network and the simulated networks are very similar. So, despite the deviations in the network topology, the configuration model succeeds in reproducing the most important network properties of the empirical networks associated with epidemics. Hence, modelling spread on the simulated networks can presumably provide realistic predictions for disease dynamics of the particular hospital ward that the contact data is derived from. However, the results are probably not generalisable to other hospitals because of the scarce data the model is based on.

The effect of pooled testing varied for the different infection probabilities and testing frequencies that were simulated. Weekly testing was not beneficial at a high transmission rate, while daily testing was sufficient to eradicate the virus at all modelled infection probabilities. The simulations illustrate the importance of identifying the optimal frequency of testing, balancing the costs detecting asymptomatic and presymptomatic cases versus the benefits of mitigating spread. The intervention measures are, however, tested on a small system which is assumed to be isolated from the rest of society. Hence, the observed responses may not be applicable to other hospital systems.

There exist endless possibilities for improving the model, some of which are mentioned in Chapter 5. Further work should focus on gathering more empirical contact data from hospital environments to model individual behaviour relevant for other systems. An interesting extension would be to utilise the temporal resolution of the data to simulate spread with an evolving contact network that can adapt to changes in the system. Modelling such detailed behaviour of agents is computationally demanding, but would lead to more realistic spreading dynamics. Despite the limitations of the model, further work should also include testing the model within the framework of the original societal model. It is important to evaluate how the added granularity of the hospital module would affect the overall spreading dynamics.

Further, the model is flexible and can potentially be utilised to simulate disease spread in other comparable systems, such as schools and workplaces. This can be achieved by modifying the degree distribution functions in the configuration model. Another possibility is to adjust the epidemiological parameters to capture the spreading dynamics of other human pathogens. In conclusion, the modelling framework presented in this thesis is flexible and can be easily modified to simulate pathogen spread in multiple settings.

Bibliography

[1] Michela Biancolella, Vito Luigi Colona, Ruty Mehrian-Shai, Jessica Lee Watt, Lucio Luzzatto, Giuseppe Novelli, and Juergen KV Reichardt. Covid-19 2022 update: transition of the pandemic to the endemic phase. *Human Genomics*, 16(1):1–12, 2022. ISSN 1479-7364.

[2] Orla McBride, Jamie Murphy, Mark Shevlin, Jilly Gibson-Miller, Todd K Hartman, Philip Hyland, Liat Levita, Liam Mason, Anton P Martinez, and Ryan McKay. Monitoring the psychological, social, and economic impact of the covid-19 pandemic in the population: Context, design and conduct of the longitudinal covid-19 psychological research consortium (c19prc) study. *International journal of methods in psychiatric research*, 30(1):e1861, 2021. ISSN 1049-8931.

[3] Imen Ayouni, Jihen Maatoug, Wafa Dhouib, Nawel Zammit, Sihem Ben Fredj, Rim Ghammam, and Hassen Ghannem. Effective public health measures to mitigate the spread of covid-19: a systematic review. *BMC public health*, 21(1):1–14, 2021. ISSN 1471-2458.

[4] 2022. URL `https://covid19.who.int`.

[5] John PA Ioannidis. The end of the covid-19 pandemic. *European Journal of Clinical Investigation*, page e13782, 2022. ISSN 0014-2972.

[6] Jocelyne Piret and Guy Boivin. Pandemics throughout history. *Frontiers in microbiology*, 11:631736, 2021. ISSN 1664-302X.

[7] David M Morens and Anthony S Fauci. Emerging pandemic diseases: how we got to covid-19. *Cell*, 182(5):1077–1092, 2020. ISSN 0092-8674.

[8] Brian McCloskey, Osman Dar, Alimuddin Zumla, and David L Heymann. Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. *The Lancet infectious diseases*, 14(10):1001–1010, 2014. ISSN 1473-3099.

[9] Aitor Nogales, Luis Martinez-Sobrido, Kevin Chiem, David J Topham, and Marta L DeDiego. Functional evolution of the 2009 pandemic h1n1 influenza virus ns1 and pa in humans. *Journal of Virology*, 92(19):e01206–18, 2018. ISSN 0022-538X.

[10] Pol Antràs, Stephen J Redding, and Esteban Rossi-Hansberg. Globalization and pandemics. Report, National Bureau of Economic Research, 2020.

[11] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51): 22020–22025, 2010. ISSN 0027-8424.

[12] Jonathan M Read, Ken TD Eames, and W John Edmunds. Dynamic social networks and the implications for the spread of infectious disease. *Journal of The Royal Society Interface*, 5(26):1001–1007, 2008. ISSN 1742-5689.

[13] Marco Ciotti, Massimo Ciccozzi, Alessandro Terrinoni, Wen-Can Jiang, Cheng-Bin Wang, and Sergio Bernardini. The covid-19 pandemic. *Critical reviews in clinical laboratory sciences*, 57(6):365–388, 2020. ISSN 1040-8363.

[14] Domenico Cucinotta and Maurizio Vanelli. Who declares covid-19 a pandemic. *Acta Bio Medica: Atenei Parmensis*, 91(1):157, 2020.

[15] Zhixing Zhu, Xihua Lian, Xiaoshan Su, Weijing Wu, Giuseppe A Marraro, and Yiming Zeng. From sars and mers to covid-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respiratory research*, 21(1):1–14, 2020. ISSN 1465-993X.

[16] Andrew G Harrison, Tao Lin, and Penghua Wang. Mechanisms of sars-cov-2 transmission and pathogenesis. *Trends in immunology*, 41(12):1100–1115, 2020. ISSN 1471-4906.

[17] Ben Hu, Hua Guo, Peng Zhou, and Zheng-Li Shi. Characteristics of sars-cov-2 and covid-19. *Nature Reviews Microbiology*, 19(3):141–154, 2021. ISSN 1740-1534.

[18] Alexandra C Walls, Young-Jun Park, M Alejandra Tortorici, Abigail Wall, Andrew T McGuire, and David Veesler. Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. *Cell*, 181(2):281–292. e6, 2020. ISSN 0092-8674.

[19] Ali A Rabaan, Shamsah H Al-Ahmed, Maysaa Al-Malkey, Roua A Alsubki, Sayeh Ezzikouri, Fadel Hassan Al-Hababi, Ranjit Sah, Abbas Al Mutair, Saad Alhumaid, and Jaffar A Al-Tawfiq. Airborne transmission of sars-cov-2 is the dominant route of transmission: droplets and aerosols. *Infez Med*, 29(1):10–19, 2021.

[20] World Health Organization. *Infection prevention and control of epidemic-and pandemic-prone acute respiratory infections in health care*. World Health Organization, 2014. ISBN 9241507136.

[21] Dinh-Toi Chu, Vijai Singh, Suong-Mai Vu Ngoc, Thanh-Lam Nguyen, and Damià Barceló. Transmission of sars-cov-2 infections and exposure in surfaces, points and wastewaters: A global one health perspective. *Case Studies in Chemical and Environmental Engineering*, page 100184, 2022. ISSN 2666-0164.

[22] Farnaz Sheikhi, Negar Yousefian, Pardis Tehranipoor, and Zahra Kowsari. Estimation of the basic reproduction number of alpha and delta variants of covid-19 pandemic in iran. *PloS one*, 17(5):e0265489, 2022. ISSN 1932-6203.

[23] Arni SR Srinivasa Rao, Steven G Krantz, Michael B Bonsall, Thomas Kurien, Siddappa N Byrareddy, David A Swanson, Ramesh Bhat, and Kurapati Sudhakar. How relevant is the basic reproductive number computed during the coronavirus disease 2019 (covid-19) pandemic, especially during lockdowns? *Infection Control & Hospital Epidemiology*, 43(1):125–127, 2022. ISSN 0899-823X.

[24] Paul L Delamater, Erica J Street, Timothy F Leslie, Y Tony Yang, and Kathryn H Jacobsen. Complexity of the basic reproduction number (r0). *Emerging infectious diseases*, 25(1):1, 2019.

[25] Yousef Alimohamadi, Maryam Taghdir, and Mojtaba Sepandi. Estimate of the basic reproduction number for covid-19: a systematic review and meta-analysis. *Journal of Preventive Medicine and Public Health*, 53(3):151, 2020.

[26] Dominik Wolff, Sarah Nee, Natalie Sandy Hickey, and Michael Marschollek. Risk factors for covid-19 severity and fatality: a structured literature review. *Infection*, 49 (1):15–28, 2021. ISSN 1439-0973.

[27] Rodrigo da Rosa Mesquita, Luiz Carlos Francelino Silva Junior, Fernanda Mayara Santos Santana, Tatiana Farias de Oliveira, Rafaela Campos Alcântara, Gabriel Monteiro Arnozo, Etvaldo Rodrigues da Silva Filho, Aisla Graciele Galdino dos Santos, Euclides José Oliveira da Cunha, and Saulo Henrique Salgueiro de Aquino. Clinical manifestations of covid-19 in the general population: systematic review. *Wiener klinische Wochenschrift*, 133(7):377–382, 2021. ISSN 1613-7671.

[28] Nazar Zaki and Elfadil A Mohamed. The estimations of the covid-19 incubation period: A scoping reviews of the literature. *Journal of infection and public health*, 14(5):638–646, 2021. ISSN 1876-0341.

[29] JA Quesada, A López-Pineda, VF Gil-Guillén, JM Arriero-Marín, F Gutiérrez, and C Carratala-Munuera. Incubation period of covid-19: A systematic review and meta-analysis. *Revista Clínica Española (English Edition)*, 221(2):109–117, 2021. ISSN 2254-8874.

[30] Muluneh Alene, Leltework Yismaw, Moges Agazhe Assemie, Daniel Bekele Ketema, Wodaje Gietaneh, and Tilahun Yemanu Birhan. Serial interval and incubation period of covid-19: a systematic review and meta-analysis. *BMC Infectious Diseases*, 21(1):1–9, 2021. ISSN 1471-2334.

[31] Zhiru Gao, Yinghui Xu, Chao Sun, Xu Wang, Ye Guo, Shi Qiu, and Kewei Ma. A systematic review of asymptomatic infections with covid-19. *Journal of Microbiology, Immunology and Infection*, 54(1):12–16, 2021. ISSN 1684-1182.

[32] Muluneh Alene, Leltework Yismaw, Moges Agazhe Assemie, Daniel Bekele Ketema, Belayneh Mengist, Bekalu Kassie, and Tilahun Yemanu Birhan. Magnitude

of asymptomatic covid-19 cases throughout the course of infection: A systematic review and meta-analysis. *PloS one*, 16(3):e0249090, 2021. ISSN 1932-6203.

[33] Manash Pratim Barman, Tousifur Rahman, Krishnarjun Bora, and Chandan Borgohain. Covid-19 pandemic and its recovery time of patients in india: A pilot study. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5):1205–1211, 2020. ISSN 1871-4021.

[34] AV Raveendran, Rajeev Jayadevan, and S Sashidharan. Long covid: an overview. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 15(3):869–875, 2021. ISSN 1871-4021.

[35] Harry Crook, Sanara Raza, Joseph Nowell, Megan Young, and Paul Edison. Long covid—mechanisms, risk factors, and management. *bmj*, 374, 2021. ISSN 1756-1833.

[36] Tanvir Ahammed, Aniqua Anjum, Mohammad Meshbahur Rahman, Najmul Haider, Richard Kock, and Md Jamal Uddin. Estimation of novel coronavirus (covid-19) reproduction number and case fatality rate: A systematic review and meta-analysis. *Health science reports*, 4(2):e274, 2021. ISSN 2398-8835.

[37] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, 101(11):3747–3752, 2004. ISSN 0027-8424.

[38] Mark EJ Newman. Analysis of weighted networks. *Physical review E*, 70(5):056131, 2004.

[39] Albert-László Barabási. Network science book. *Network Science*, 625, 2014.

[40] Daniel Schall. Link prediction in directed social networks. *Social Network Analysis and Mining*, 4(1):1–14, 2014. ISSN 1869-5469.

[41] Michael PH Stumpf and Carsten Wiuf. Sampling properties of random graphs: the degree distribution. *Physical Review E*, 72(3):036118, 2005.

[42] Andrew T Stephen and Olivier Toubia. Explaining the power-law degree distribution in a social commerce network. *Social Networks*, 31(4):262–270, 2009. ISSN 0378-8733.

[43] Francisco Aparecido Rodrigues. *Network centrality: an introduction*, pages 177–196. Springer, 2019.

[44] Naveen Gupta, Anurag Singh, and Hocine Cherifi. Centrality measures for networks with community structure. *Physica A: Statistical Mechanics and its Applications*, 452:46–59, 2016. ISSN 0378-4371.

[45] Mark EJ Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005. ISSN 0378-8733.

[46] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113, 2004. ISSN 1471-0064.

[47] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5):056104, 2003.

[48] Tom Britton and David Lindenstrand. Epidemic modelling: aspects where stochasticity matters. *Mathematical biosciences*, 222(2):109–116, 2009. ISSN 0025-5564.

[49] Wei Duan, Zongchen Fan, Peng Zhang, Gang Guo, and Xiaogang Qiu. Mathematical and computational approaches to epidemic modeling: a comprehensive review. *Frontiers of Computer Science*, 9(5):806–826, 2015. ISSN 2095-2236.

[50] Nicholas C Grassly and Christophe Fraser. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, 6(6):477–487, 2008. ISSN 1740-1534.

[51] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4): 599–653, 2000. ISSN 0036-1445.

[52] Marcelo N Kuperman. Invited review: Epidemics on social networks. *arXiv preprint arXiv:1312.3838*, 2013.

[53] Scott De Marchi and Scott E Page. Agent-based models. *Annual Review of political science*, 17:1–20, 2014. ISSN 1094-2939.

[54] Elizabeth Bruch and Jon Atwell. Agent-based models in empirical social research. *Sociological methods & research*, 44(2):186–221, 2015. ISSN 0049-1241.

[55] Leigh Tesfatsion. Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82, 2002. ISSN 1064-5462.

[56] Syed Muhammad Ali Abbas. An agent-based model of the development of friendship links within facebook. *Computational and Mathematical Organization Theory*, 19 (2):232–252, 2013. ISSN 1572-9346.

[57] Elizabeth Hunter, Brian Mac Namee, and John D Kelleher. A comparison of agent-based models and equation based models for infectious disease epidemiology. In *AICS*, pages 33–44, 2018.

[58] Abdulrahman M El-Sayed, Peter Scarborough, Lars Seemann, and Sandro Galea. Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations*, 9(1):1–9, 2012. ISSN 1742-5573.

[59] Jang Won Bae, Euihyun Paik, Kiho Kim, Karandeep Singh, and Mazhar Sajjad. Combining microsimulation and agent-based model for micro-level population dynamics. *Procedia Computer Science*, 80:507–517, 2016. ISSN 1877-0509.

[60] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS one*, 8(9):e73970, 2013. ISSN 1932-6203.

[61] Python software foundation. *Python, version 3.8.5*. URL `www.python.org/`.

[62] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003. ISSN 1088-9051. URL `www.cytoscape.org`.

[63] Leon Danon, Ashley P Ford, Thomas House, Chris P Jewell, Matt J Keeling, Gareth O Roberts, Joshua V Ross, and Matthew C Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011. ISSN 1687-708X.

[64] André Voigt, Nikolay Martyushenko, Emil Karlsen, Martina Hall, Kristen Nyhamar, Stig William Omholt, and Eivind Almaas. Containing pandemics through targeted testing of households. *BMC infectious diseases*, 21(1):1–10, 2021. ISSN 1471-2334.

[65] André Voigt, Stig Omholt, and Eivind Almaas. Comparing the impact of vaccination strategies on the spread of covid-19, including a novel household-targeted vaccination strategy. *Plos one*, 17(2):e0263155, 2022. ISSN 1932-6203.

[66] Sameera Abar, Georgios K Theodoropoulos, Pierre Lemarinier, and Gregory MP O'Hare. Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review*, 24:13–33, 2017. ISSN 1574-0137.

[67] Alex Sigal, Ron Milo, and Waasila Jassat. Estimating disease severity of omicron and delta sars-cov-2 infections. *Nature Reviews Immunology*, 22(5):267–269, 2022. ISSN 1474-1741.

[68] Samuel Alizon and Mircea T Sofonea. Sars-cov-2 virulence evolution: avirulence theory, immunity and trade-offs. *Journal of Evolutionary Biology*, 34(12):1867–1877, 2021. ISSN 1010-061X.

Appendix

Appendix A

Methods Supplementary

A.1 Python Modules

The code in this project is written in Python version 3.8.5. The modules utilised, including modules imported from the Python Standard Library and installed modules, are presented in Table A.1.1.

Table A.1.1: The imported modules from the Python Standard Library and the installed modules used in the project.

Module	Description	Version
os	Operating system interfaces	-
math	Mathematical functions	-
random	Pseudo-random numbers	-
numpy	Scientific computing	1.20.1
matblotlib	Visualisations	3.5.1
seaborn	Visualisations	0.11.1
scipy	Statistics	1.6.2
pandas	Data analysis	1.2.4

Appendix B

Results Supplementary

B.1 Epidemiological Model

The supplementary results and figures from Section 4.3 are presented in the following section.

B.1.1 Disease Spread

The figures provided in this section correspond to the results described in Section 4.3.1, only with different parameter values.

Disease Dynamics

The disease dynamics of model runs with an infection probability of 0.01 are shown in Figure 4.3.1. The same figures visualising the the number of infected and recovered individuals for simulations run with infection probabilities of 0.005 and 0.001 are presented in Figure B.1.1 and Figure B.1.2, respectively.

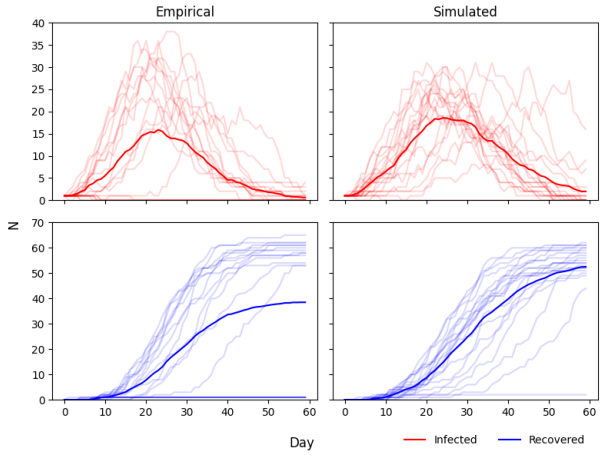


Figure B.1.1: The figure presents the number of infected (red) and recovered (blue) individuals at each day of the simulation, for both the empirical (left panels) and simulated networks (right panels). The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.005. The transparent curves represents the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

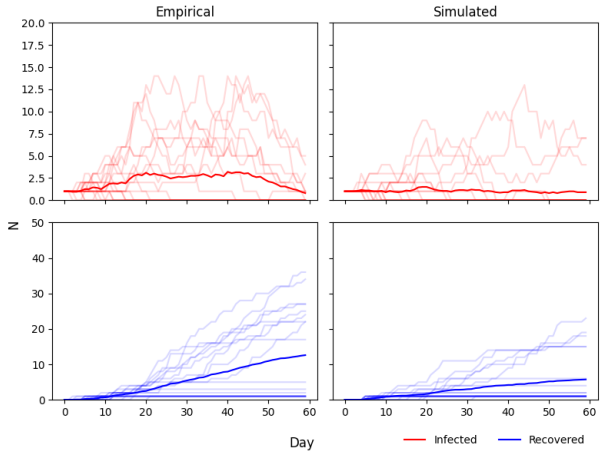


Figure B.1.2: The figure presents the number of infected (red) and recovered (blue) individuals at each day of the simulation, for both the empirical (left panels) and simulated networks (right panels). The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.005. The transparent curves represents the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

Test of Seed Node

The resulting heatmaps from running 100 simulations with each of the nodes initiated in the asymptomatic state are presented in Figure B.1.3. The figure corresponds to Figure 4.3.2 which is generated with presymptomatic seed nodes.

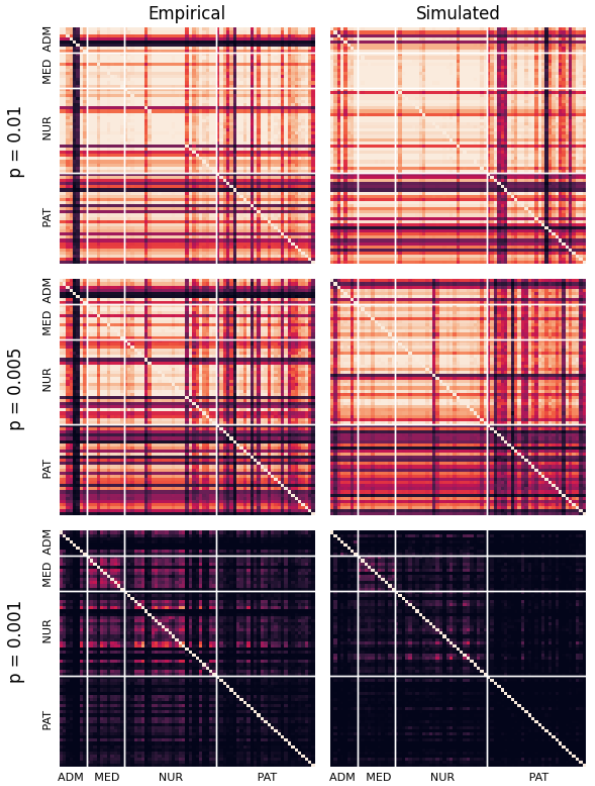


Figure B.1.3: Heatmaps visualising the number of times node i gets infected out of 100 simulations with node j as the seed node, which is initiated in the asymptomatic state. The subplot contains heatmaps for simulations with both the empirical network (left panels) and a simulated network (right panels) for three infection probabilities; 0.01, 0.005, and 0.001.

B.1.2 Reproduction Number

The reproduction number of modelled epidemics can be visualised in several ways. In the following section the distribution of R_0 from 100 model simulations are presented for various infection probabilities, as well as a representation of the relationship between the reproduction number and infection probabilities.

Pie Charts

Results from computations of R_0 from model runs with infection probability of 0.005 are described in Section 4.3.2. The corresponding results emerging from simulations with infection probabilities of 0.01 and 0.001 are presented in Figure B.1.4 and Figure B.1.5, respectively.

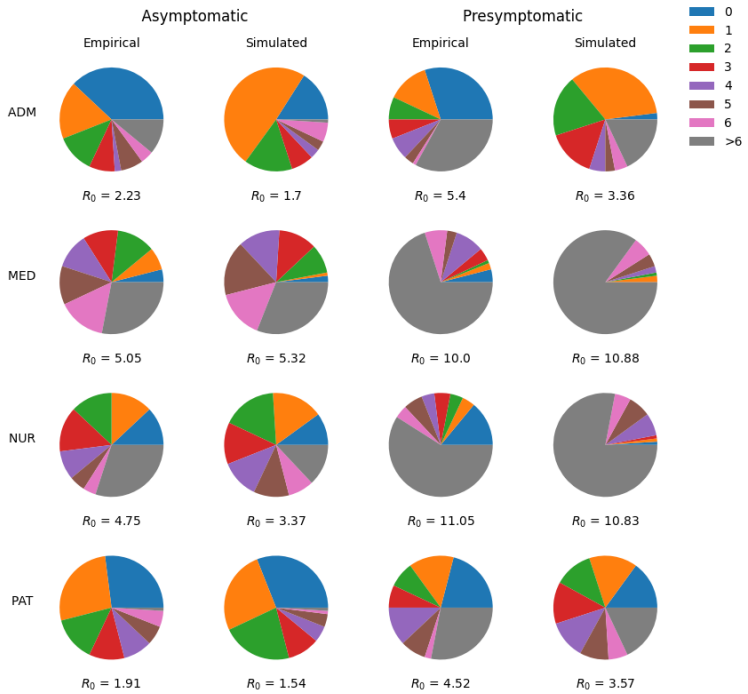


Figure B.1.4: An overview of pie charts describing the proportion of 100 simulations with $p = 0.01$ resulting in a given R_0. Pie charts are generated for simulations on the empirical and constructed networks with an asymptomatic or presymptomatic seed node. The pie charts are also distinguished by the hospital status of the seed node. The average R_0 for each scenario is denoted the below pie chart.

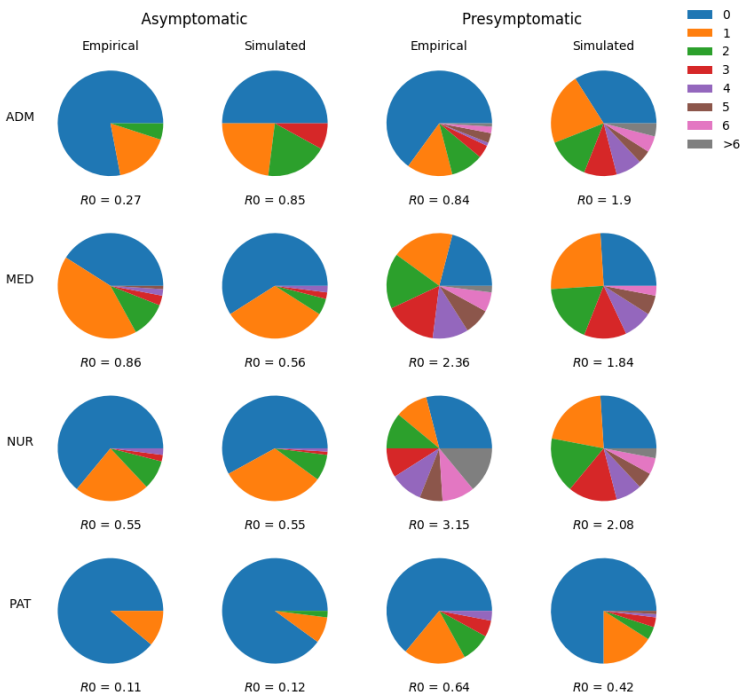


Figure B.1.5: An overview of pie charts describing the proportion of 100 simulations with $p = 0.001$ resulting in a given R_0. Pie charts are generated for simulations on the empirical and constructed networks with an asymptomatic or presymptomatic seed node. The pie charts are also distinguished by the hospital status of the seed node. The average R_0 for each scenario is denoted the below pie chart.

Infection Probability and Reproduction Number

A supplementary visualisation of the relationship between the infection probability, p, and the basic reproduction number, R_0, is presented in Figure B.1.6. The plot consists of individual subplots describing the relationship for all the groups in both the empirical and the simulated network, generated with asymptomatic and presymptomatic seed cases. 95 % confidence intervals are also provided.

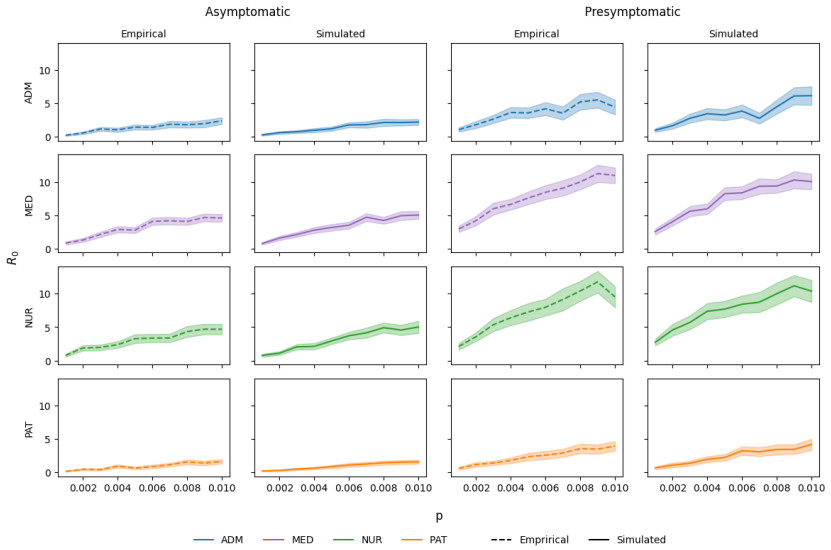


Figure B.1.6: A representation of the relationship between the infection probability, p, and the basic reproduction number, R_0. R_0 is calculated for ten infection probabilities ranging from 0.001 to 0.01. The colours represent the hospital status of the seed node, while the dotted and solid lines represent the empirical and simulated network, respectively. The results in the left panels are generated with asymptotic seed nodes, while the right panels are generated from presymptomatic seed nodes.

B.1.3 Pooled Testing

The results from analysing the effect of pooled testing on the number of infections and absent HCWs are presented in Section 4.3.3. Supplementary results emerging from different infection probabilities and test strategies are provided in the following section. Figure B.1.7 presents the results from weekly pooled testing with $p = 0.001$. The effects of daily pooled testing with $p = 0.005$ and $p = 0.001$ are visualised in Figure B.1.8 and Figure B.1.9, respectively.

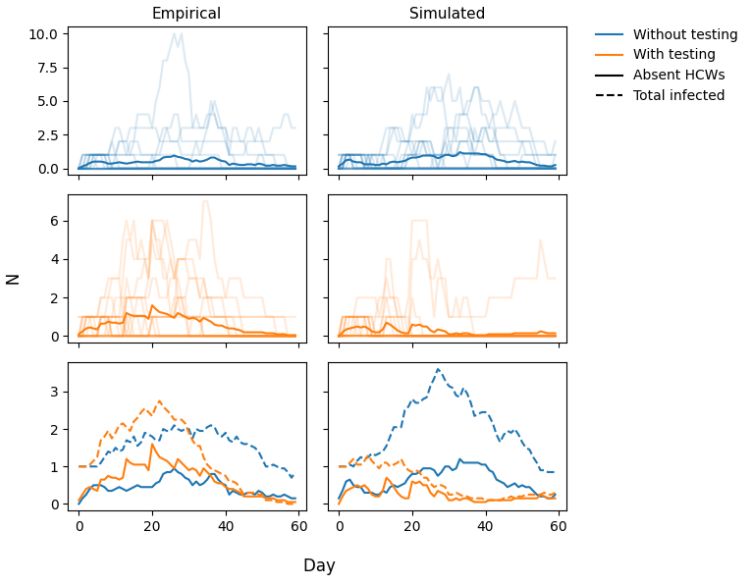


Figure B.1.7: A comparison of the number of absent HCWs and the total number of infected individuals in all groups for simulations run with weekly testing (orange) and without testing (blue). The results are obtained from 20 simulations with an infection probability of 0.001, using both the empirical network and simulated networks. The transparent lines represent the number of absent HCWs at each simulation, while the solid lines are an average of the 20 simulations. The dotted lines represent an average of the total number of infected individuals, not only HCWs.

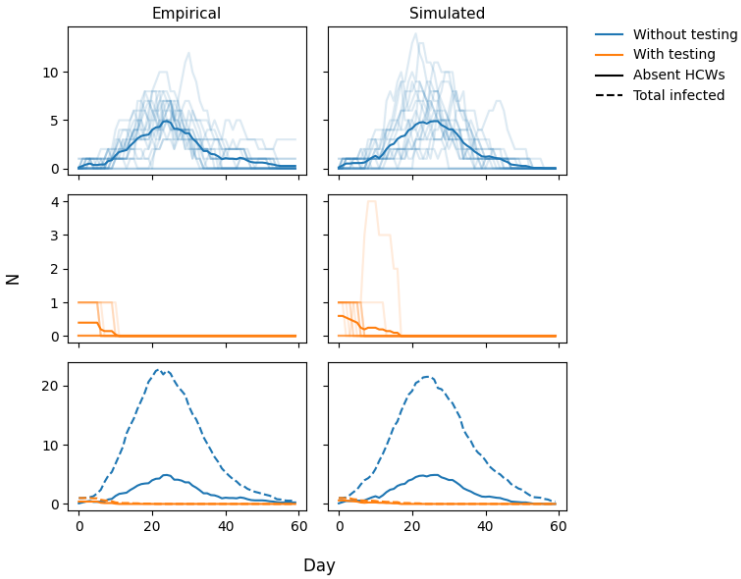


Figure B.1.8: A comparison of the number of absent HCWs and the total number of infected individuals in all groups for simulations run with daily testing (orange) and without testing (blue), with an infection probability of 0.005.

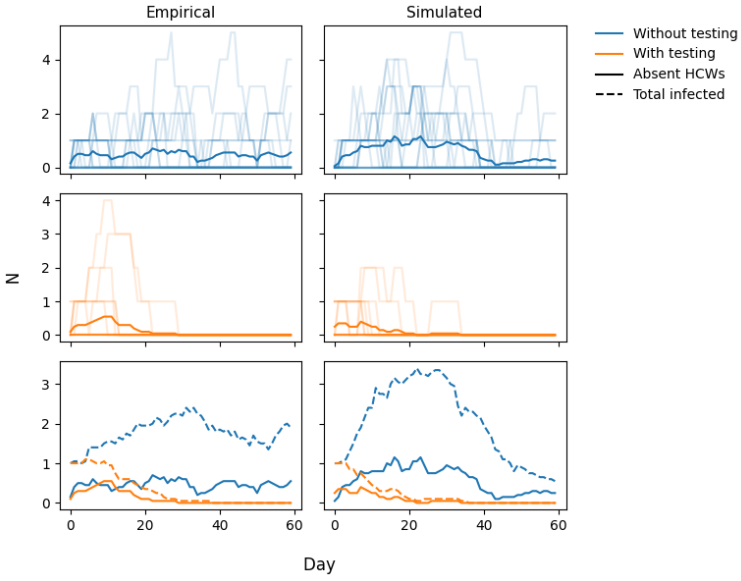


Figure B.1.9: A comparison of the number of absent HCWs and the total number of infected individuals in all groups for simulations run with daily testing (orange) and without testing (blue), with an infection probability of 0.001.

B.2 Error Analysis

The supplementary results and figures error analysis described in Section 4.4 are presented in the following section.

B.2.1 Contact Network Period

Section 4.4.2 presents the analysis of the impact of various contact network periods that can be used for simulating disease spread. Supplementary figures with different infection probabilities are presented in the following section. Figure B.2.1 and Figure B.2.2 compare simulated contact network periods for $p = 0.005$ and $p = 0.001$, respectively. Simulated 1, 2 and 3 in the figure refers to the scenarios described earlier. The contact network periods of the empirical network are compared for $p = 0.005$ and $p = 0.001$ in Figure B.2.1 and Figure B.2.4, respectively.

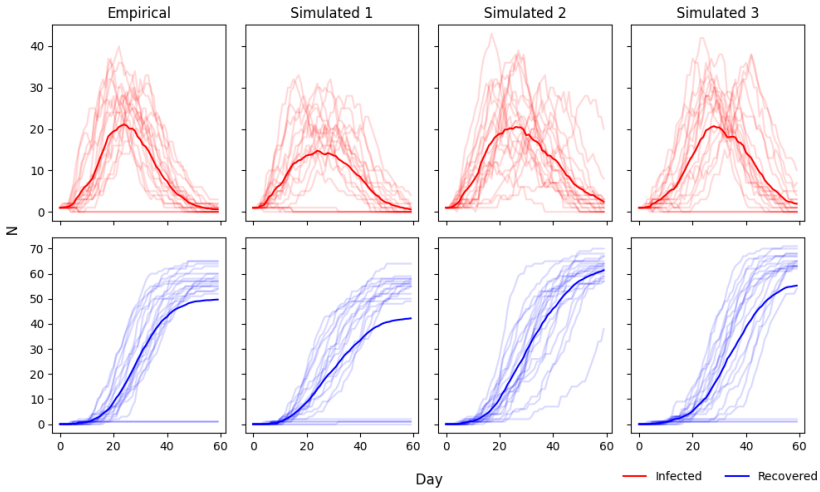


Figure B.2.1: The number of infected (red) and recovered (blue) individuals on each day of the simulation for the empirical contact network and three different contact network periods of the simulated networks. The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.005. The transparent curves represent the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

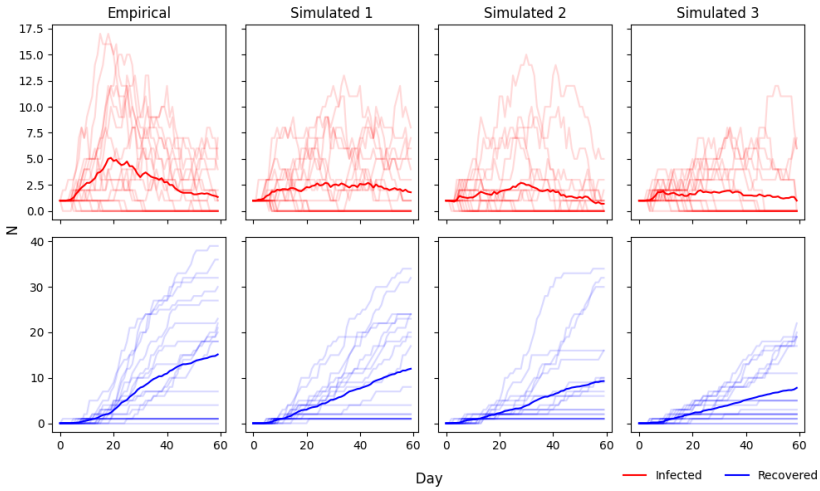


Figure B.2.2: The number of infected (red) and recovered (blue) individuals on each day of the simulation for the empirical contact network and three different contact network periods of the simulated networks. The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.001. The transparent curves represent the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

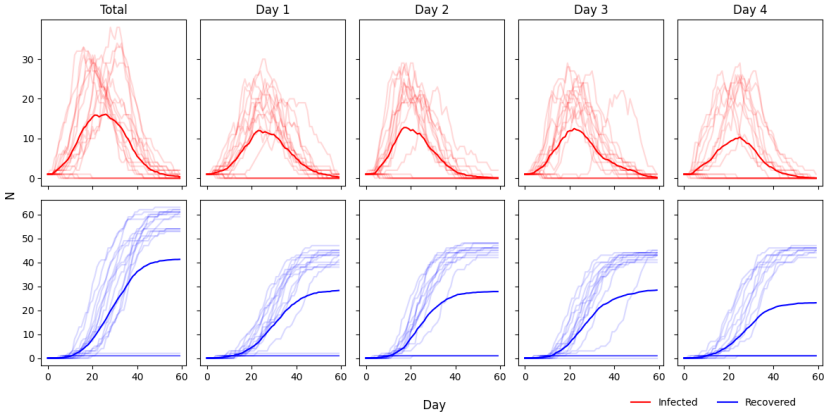


Figure B.2.3: The number of infected (red) and recovered (blue) individuals at each day of the simulation for different contact network periods of the empirical network. Total refers to simulating with the four days alternately, while Day 1 to Day 4 refers to repeatedly simulating with one daily contact network. The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.005. The transparent curves represent the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.

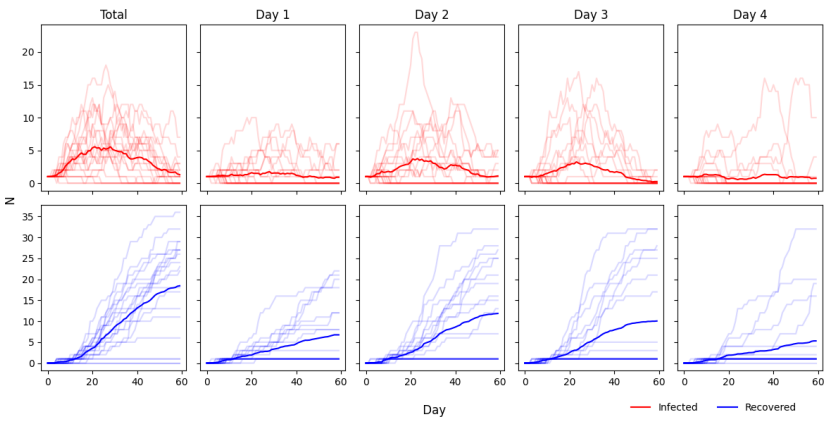
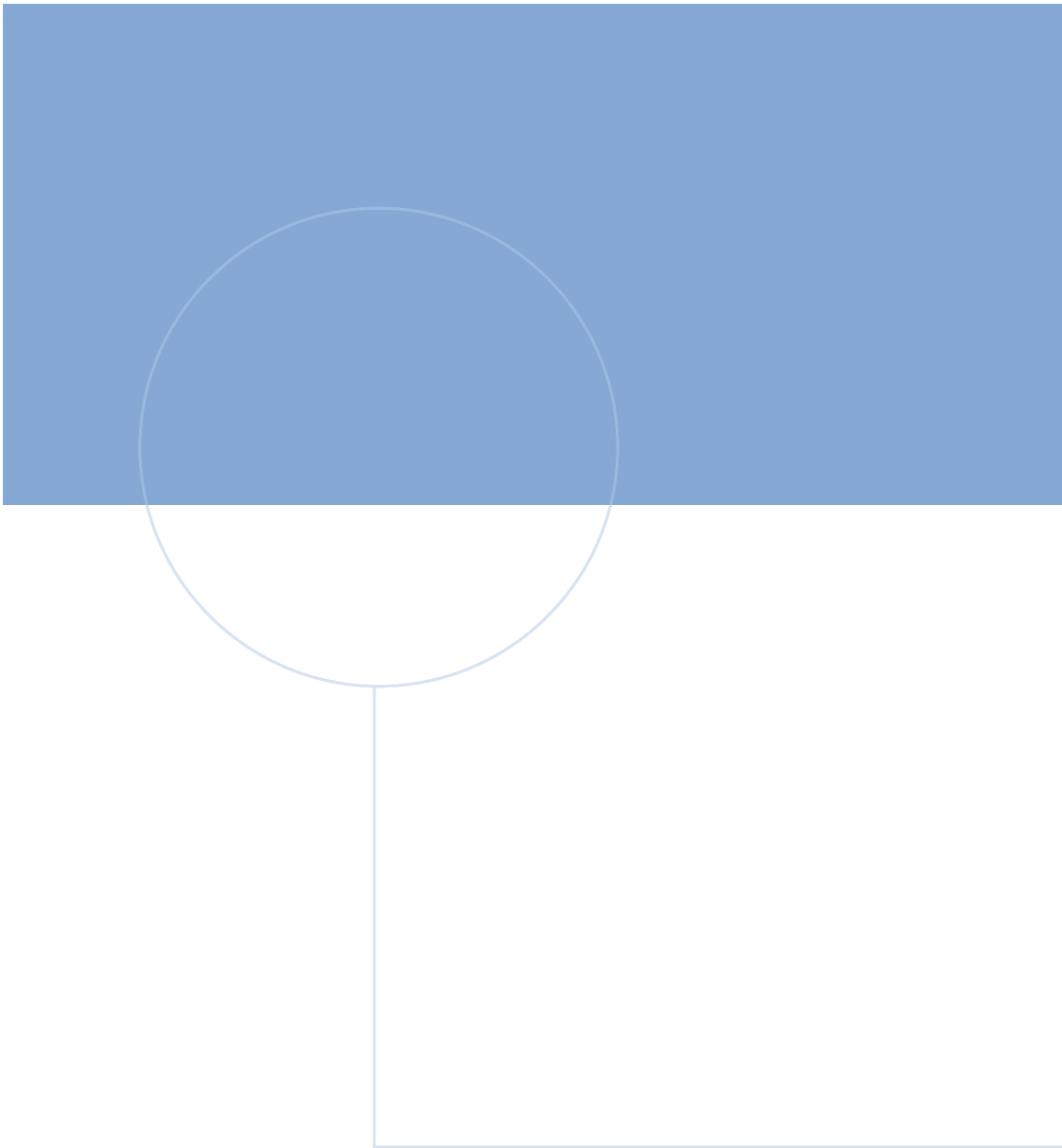


Figure B.2.4: The number of infected (red) and recovered (blue) individuals at each day of the simulation for different contact network periods of the empirical network. Total refers to simulating with the four days alternately, while Day 1 to Day 4 refers to repeatedly simulating with one daily contact network. The results are obtained from 20 simulations over a period of 60 days, with an infection probability of 0.001. The transparent curves represent the dynamics of each simulation, while the solid curves illustrate an average of the 20 simulations.



NTNU

Norwegian University of
Science and Technology