Master's thesis

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

Signe Onstad Sævareid

An Agent-Based Modeling Framework for Simulating the Spread of SARS-CoV-2 on a Temporal Contact Network

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Eivind Almaas
June 2020

NTNU
Norwegian University of Science and Technology

Signe Onstad Sævareid

An Agent-Based Modeling Framework for Simulating the Spread of SARS-CoV-2 on a Temporal Contact Network

Master's thesis in Chemical Engineering and Biotechnology
Supervisor: Eivind Almaas
June 2020

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

NTNU
Norwegian University of
Science and Technology

I simply wish that, in a matter which so closely concerns the wellbeing of the human race, no decision shall be made without all the knowledge which a little analysis and calculation can provide.

– Daniel Bernoulli, 1760

Acknowledgments

If you had told me a few years ago that I would spend the spring of 2020 with corona and models, I would have pictured a very different scenario!

This thesis concludes a master's degree in Chemical Engineering and Biotechnology at the Norwegian University of Science and Technology in Trondheim. The last five years have taught me a lot. I take with me valuable experiences and knowledge, both from the lecture halls and laboratories, but also from fellow students and friends.

I would like to express my gratitude to my supervisor Eivind Almaas for introducing me to the field of systems biology and biological networks, and for the inspiration and invaluable support throughout the semester. I would also like to thank my co-supervisor Pål Røynestad for all help and for believing in me and this project. Thank you, Vetle and Aslak, for informative and uninformative discussions and pre-corona coffee breaks. Thank you, Marcus, for all exhausting and mind-cleaning hill and stair runs.

I must also thank my family and Jan Eskil for their unconditional love and support. Special thanks to the cat for expressing her overwhelming enthusiasm over numerous video calls during the last months. And to my mom, and all other healthcare workers who have been standing in the frontline in this pandemic, thank you all.

Signe Onstad Sævareid
June 15, 2020

Abstract

The ongoing coronavirus pandemic has so far claimed more than $430\,000$ human lives worldwide. The crisis has led to a massive global effort in the search for knowledge about the characteristics of the virus and how the spread can be limited. In order to obtain a better understanding of the epidemiological mechanisms affecting the viral transmission among human beings, computational modeling provides a useful tool.

This master's thesis presents a novel modeling framework built for the purpose of investigating how SARS-CoV-2 spreads on a temporal contact network in a virtual hospital. The model is agent-based and simulates both the generation of the inter-individual contacts and the virus transmission based on assigned agent attributes and a set of rules governing their interactions. The temporal contact network on which the virus spreads is based on rules derived from the observed contact patterns in an empirical close-proximity interaction network. The network evolves simultaneously as the simulation and takes staff shifts, patient hospitalization and any quarantine restrictions or isolation into account. The rules governing the epidemiological part of the model are derived from the available literature. Most of these parameter values are tunable, giving the user the possibility to explore a wide range of epidemiological parameters and to evaluate the effect of several key control measures.

Self-consistency tests were conducted to validate the model functionality, yielding model output consistent with the expectations. The model stability was assessed by investigating the variability in output of several simulations run with identical input parameters. Although the infection only spread to a few agents in some of the simulations, a large proportion of the runs resulted in considerable outbreaks. Static and temporal analysis of the generated contact network showed that the simulated network outperformed the randomized reference network in terms of approximating the empirical network properties.

The developed modeling framework is detailed and allows the user to investigate how a range of key epidemiological parameters affect the spread of SARS-CoV-2. The work lays a solid foundation for future epidemiological analyses and evaluation of infection control measures.

Sammendrag

Den pågående koronapandemien har så langt krevd mer enn $430\,000$ menneskeliv på verdensbasis. Krisen har ført til en massiv global innsats i søken etter kunnskap om virusets egenskaper og hvordan spredningen kan begrenses. I arbeidet mot en bedre forståelse av de epidemiologiske egenskapene som påvirker spredningen av viruset blant mennesker, utgjør datamodellering et nyttig verktøy.

Denne masteoppgaven presenterer et nytt modelleringsrammeverk bygget for å undersøke hvordan SARS-CoV-2 spres på et temporalt kontaktnettverk på et virtuelt sykehus. Modellen er agentbasert og simulerer både genereringen av de mellommenneskelige kontaktene og virusspredningen basert på tildelte egenskaper hos agentene og et sett med regler som styrer interaksjonene mellom dem. Det temporale kontaktnettverket som viruset spres på er basert på utledede regler fra de observerte kontaktmønstrene i et empirisk kontaktnettverk. Nettverket utvikles samtidig med simuleringen og tar hensyn til personalskift, pasientopphold og eventuelle karantenerestriksjoner eller isolasjon. Reglene som styrer den epidemiologiske delen av modellen er utledet fra tilgjengelig litteratur. De fleste av disse parameterverdiene er regulerbare, noe som gir brukeren mulighet til å utforske et bredt spekter av epidemiologiske parametere og å evaluere effekten av mange viktige smitteverntiltak.

Selvkonsistenstester ble utført for å validere modellens funksjonalitet, og ga modellresultater som var konsistente med forventningene. Modellens stabilitet ble evaluert ved å undersøke variabiliteten i modellresultatene av flere simuleringer med identiske parameterverdier. Mens infeksjonen ebbet ut i noen av simuleringene, resulterte en stor andel av kjøringene i betydelige utbrudd. Statiske og temporale analyser av det genererte kontaktnettverket viste at det simulerte nettverket approksimerte de empiriske nettverksegenskapene bedre enn et randomisert referansenettverk.

Det utviklede modelleringsrammeverket er detaljert og gir brukeren muligheten til å undersøke hvordan en rekke epidemiologiske parametere påvirker spredningen av SARS-CoV-2. Arbeidet legger et solid grunnlag for fremtidige epidemiologiske analyser og vurdering av smitteverntiltak.

Table of Contents

Abbreviations

ABM	=	Agent-Based Modeling
BC	=	Betweenness Centrality
CC	=	Closeness Centrality
CDF	=	Cumulative Distribution Function
CFR	=	Case Fatality Rate
COVID-19	=	Coronavirus Disease 2019
CPI	=	Close-Proximity Interaction
EEA	=	European Economic Area
EU	=	European Union
GC	=	Giant Component
GEXF	=	Graph Exchange XML Format
GLEaM	=	Global Epidemic and Mobility
HCW	=	HealthCare Worker
HIV	=	Human Immunodeficiency Virus
IBM	=	Individual-Based Model
ICU	=	Intensive Care Unit
IFR	=	Infection Fatality Rate
MERS	=	Middle East Respiratory Syndrome
MSE	=	Mean Squared Error
MRSA	=	Methicillin-Resistant Staphylococcus Aureus
MSSA	=	Methicillin-Sensitive Staphylococcus Aureus
NIPH	=	Norwegian Institute of Public Health
PDF	=	Probability Density Function
RFID	=	Radio-Frequency IDentification
SARS	=	Severe Acute Respiratory Syndrome
SARS-CoV-2	=	Severe Acute Respiratory Syndrome CoronaVirus 2
SIR	=	Susceptible - Infected - Recovered
WHO	=	World Health Organization
ADM	=	Administrative staff
MED	=	Medical doctors
NUR	=	Nurses and nurses' aides
PAT	=	Patients

Chapter 1

Introduction

The ongoing coronavirus pandemic is not the first pandemic to ravage the world [1]. Infectious diseases have existed since the early hunter-gathering days. The pathogens that established in these societies, often consisting of 50 to 100 individuals, were characterized by high transmission rates and by inducing low immunity [2]. The domestication of animals and the increasing population density following the Neolithic Revolution played an important role in the increasing prevalence of communicable diseases [3]. Several of the currently existing infectious diseases are thought to originate from domestic animals. For instance, smallpox is likely to have evolved from cowpox, whereas the virus causing measles is closely related to rinderpest, also known as cattle plague [2]. The denser living conditions facilitated the spread of disease and did also have an impact on hygiene aspects, such as the quality of drinking water [4]. In recent times, globalization has shrunk the time and space by erasing economic, political and technological borders. The increased human interactions across the globe facilitate a more rapid spread of infectious diseases and make it more difficult to contain outbreaks [5]. Although we now dispose a completely different arsenal of weapons concerning medical resources and infection control strategies compared to the hunter-gathering societies, detailed knowledge on the underlying spreading mechanisms is crucial to implement well-founded and effective control measures.

Because human infectious diseases transmit from one individual to another, networks describing human interactions are central in the efforts to understand how these diseases evolve and spread [6]. The routes of transmission may vary [7]: Respiratory transmission involves the spread of pathogens via respiratory droplets, for instance the transmission of the influenza virus in droplets resulting from a sneeze or a cough [8]. Other pathogens transmit via fecal-oral routes, such as several *Salmonella* species transmitting from the feces of an infected individual by contaminating food or water sources [9]. Vector-borne diseases transmit between two organisms via a vector, an agent carrying the pathogen. Examples include the mosquito-borne pathogens causing diseases such as malaria and dengue fever [10] or the oriental rat flea carrying *Yersinia pestis* [11]. The latter is the causative agent for the Black Death, which spread across Europe in the middle of the 14$^{\text{th}}$ century, killing an estimated one-third of the European population [12]. Other pathogens spread

via sexual transmission, such as the bacterial species *Neisseria gonorrhoeae* and *Chlamydia trachomatis* causing gonorrhea and chlamydia, respectively. Thus, when studying the spread of an infectious disease, its route of transmission may affect the choice of a suitable inter-individual network. For instance, a network based on reported sexual contacts is likely to capture the spread of a sexually transmitted infection better than a network created from detected physical proximity would. In addition, the scale of the network may vary [13]: Networks may be large-scale, as the global air traffic network [14] connecting Wuhan Tianhe International Airport in the Hubei province with airports across the globe, national as inter-hospital networks consisting of healthcare institutions connected by patient transfers [15], or local as inter-individual networks connecting physicians, nurses and patients through close-proximity interactions in a hospital ward [16].

Computational modeling is a helpful tool in the work of obtaining a better understanding of how a disease spreads [17]. By creating a digital imitation of the world, a country or a hospital ward, modeling allows for the exploration of control measures and variations in epidemiological variables, circumnavigating the ethical dilemmas these experiments would entail in real-life scenarios [18]. Several modeling frameworks have been developed for this purpose, such as the Global Epidemic and Mobility (GLEaM) model which incorporates spatial population data and travel patterns to simulate worldwide epidemic spread [19]. A branch of the modeling universe called agent-based modeling (ABM) has proven particularly useful for exploring how macro-level patterns evolve from micro-level rules [20]. This modeling approach is built upon interacting agents, all with a set of properties and actions governing their behavior [21]. Compared to more traditional equation-based models, ABMs are able to capture minor differences in agent behavior or characteristics. Hence, the system is not averaged out, but able to capture heterogeneity [22].

The urge to obtain a better understanding of the nature of human pathogens and their spreading potential is everlasting and fundamental. Originally, this thesis aimed to investigate the evolution and spread of antibiotic resistance, an increasingly severe medical challenge that the World Health Organization (WHO) characterizes as "one of the biggest threats to global health" [23]. The rising prevalence of the novel coronavirus over the past months, however, offered an opportunity to shift the focus and take part in the ongoing work towards obtaining a better understanding of the spreading nature of SARS-CoV-2. Therefore, **this thesis aims to develop a novel modeling framework for simulation of the spread of SARS-CoV-2 on a time-evolving contact network, which can be used to examine underlying spreading mechanisms and identify effective control measures.** We divide this aim into three objectives: The first objective is to extract parameter values from observed patterns in an empirical, temporal contact network. The second objective is to build an agent-based model that combines the simulation of a temporal contact network and the spread of SARS-CoV-2 through the evolving inter-individual contacts. The third objective is to demonstrate some of the model features, test the model consistency and stability, and lastly evaluate how well the simulated contact network performs in recreating the empirical contact network properties.

Chapter 2

Theory

This chapter is divided into four main parts: The first part gives an introduction to relevant theory regarding the Coronavirus Disease 2019, with emphasis on epidemiological characteristics and the course of disease. The second part presents theory related to networks and network spreading phenomena, including an introduction to temporal contact networks. The third part discusses epidemic modeling approaches and describes agent-based modeling in more detail. Lastly, the fourth part presents a selection of relevant theory regarding statistics and probability distributions.

Parts of this chapter dealing with network theory and epidemic modeling is based on material presented in my previous project report "Agent-Based Modeling of Evolution and Spread of Antibiotic Resistance on Networks: A Literature Review" [24]. An overview of the relevant sections can be found in Appendix A.1.

2.1 Coronavirus Disease 2019

In late December 2019, the China National Health Commission received reports of several cases of pneumonia of unknown origin [25]. Many of the cases were linked to the Huanan seafood wholesale market in Wuhan, the capital city of the Hubei province in China [26]. Analysis of the isolated virus revealed a novel *Betacoronavirus* strain showing a considerable nucleotide identity with the viral strains causing Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) [27]. The first fatal outcome among confirmed cases was reported on January 11 [28, 29]. Two months later, on March 11, there were reported more than $118\,000$ cases in 114 different countries, and the WHO declared the outbreak a pandemic [30]. At the time of writing (June 15, 2020), there are over $7\,873\,221$ confirmed cases and $432\,173$ deaths reported worldwide [31].

The ongoing pandemic of the coronavirus disease 2019 (COVID-19), is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [32]. In humans, coronaviruses can cause diseases ranging from mild respiratory infections to severe pneumonia. Coronaviruses can also be pathogenic in other mammals and birds [33]. The term 'coronavirus' was firstly used in 1968 by a group of virologists describing a new family of viruses

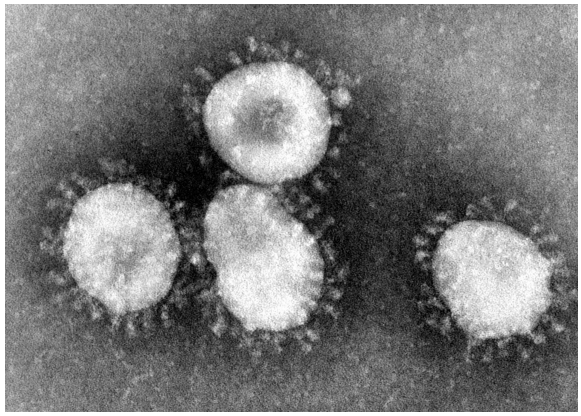

Figure 2.1.1: Transmission electron micrograph of an avian coronavirus belonging to the same viral family as SARS-CoV-2. The surface protein spikes resemble a solar corona and has given the family its name [34, 35]. Photo credit: Centers for Disease Control and Prevention, Dr. Fred Murphy.

[34]. When studying the virus by electron microscopy, the encircling halo of protein spikes on the virus surface gave associations to the solar corona [35], illustrated in Figure 2.1.1.

The current pandemic joins the ranks of two other prominent coronavirus outbreaks during the last twenty years. The outbreak of SARS in 2002-2003 started in the Guangdong Province in the southern part of China. The outbreak resulted in just over 8000 cases and 774 deaths, giving a case fatality rate of approximately $9.6\,\%$ [33]. After mainland China, Hong Kong, Taiwan, Canada and Singapore reported the most number of cases [36]. Several activities in the southeast part of Asia and Toronto, Canada were shut down to prevent spreading. It is estimated that the SARS outbreak resulted in a loss of economic activities worth nearly $40 billion US dollars [33]. In comparison, the Asian Development Bank has estimated that the economic impact of COVID-19 could reach $8.8 trillion US dollars [37]. A new coronavirus started spreading in the Middle East approximately ten years after SARS emerged [33]. The novel virus caused several cases of severe respiratory tract infections. The disease, known as MERS, killed over one-third of the infected individuals, 858 out of approximately 2500 confirmed cases [38]. The case fatality rate of $34.4\,\%$ is hence significantly higher than for SARS. There have been several cases of resurgence in the years following the initial outbreak in 2012, for instance in South Korea in 2015 where a total of 168 cases were confirmed in the months following the return of an infected 68-year-old man who had been visiting several Middle East countries on a business trip [39]. Both SARS-CoV and MERS-CoV have shown to originate from bats: The former spread from bats to humans via Asian palm civets, whereas the latter spread from bats to humans via camels [40, 41].

The global death toll due to COVID-19 has long since passed the number of deaths caused by SARS and MERS combined, even though COVID-19 so far has shown to have a lower case fatality rate than its previous relatives [38]. The three following sections give an introduction to relevant theory related to the transmission of SARS-CoV-2, the course of disease and lastly control measures implemented in order to slow the spread.

2.1.1 Transmission

At an early stage, there was no evidence of inter-individual transmission of SARS-CoV-2 in the scientific literature [28]. Many of the early cases were directly linked to the Huanan seafood wholesale market and pointed towards a possible animal-to-human transmission [42]. Studies do suggest that this coronavirus, like its predecessors, originates from bats, and that an animal sold at the seafood market may have served as an intermediate animal host enabling transmission to humans [43–45]. One of the earliest indications of human-to-human transmission was presented by Chan *et al.* who analyzed five cases of pneumonia in a family who had returned to Shenzhen after visiting Wuhan. An additional family member, who had not visited Wuhan, was infected with the virus after several days of contact with the infected family members [26]. Li *et al.* points out that the rapid spread of SARS-CoV-2 must imply a high degree of human-to-human transmission and not by spillover events from animals to humans alone [45].

After it became clear that SARS-CoV-2 spreads among humans, efforts have been made to determine *how* the virus spreads [47]. The question of whether the virus can transmit through the air has been of particular interest. Whereas some pathogens transmit through droplets generated from coughing or sneezing, other pathogens can spread through exhaled aerosols [47]. The size of the particle carrying the pathogen affects both how long and the particle can linger in the air before it settles on a surface, and how far down the respiratory tract of a susceptible individual the virus can reach [46]. Figure 2.1.2 illustrates how the continuous specter of particle sizes can be classified based on the aerodynamic diameter, d_A: The largest droplets, $d_A > 100\,\mu\mathrm{m}$, will settle on a surface within seconds after being expelled. Pathogens carried by large droplets can thus only be transmitted through direct or indirect contact, for instance through fomites [46] such as door handles or payment terminals. Droplets smaller than $100\,\mu\mathrm{m}$, on the other hand, are inhalable and can enter the body through the nose or mouth. The smallest particles are referred to as respiratory droplet nuclei. These settle slowly, travel further than larger droplets and can penetrate to the alveolar region of the lungs [46].

Studies have established that SARS-CoV-2 transmits from one individual to another

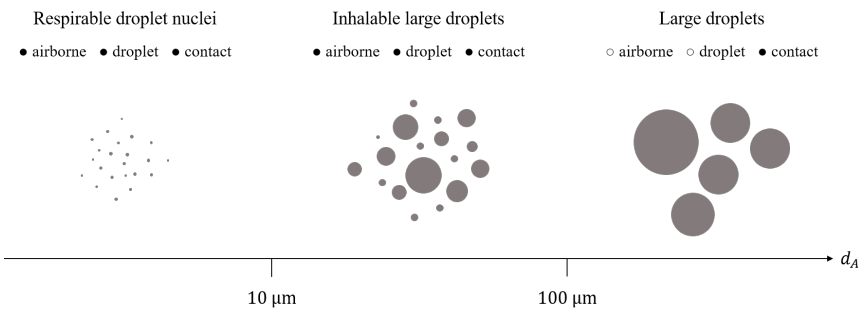

Figure 2.1.2: Respiratory droplets can be classified based on their aerodynamic diameter: Respirable droplet nuclei and inhalable large droplets contribute to airborne, droplet and contact transmission, whereas large droplets settle on surfaces within seconds only can spread via contact transmission. Based on [46].

via respiratory droplets in close contact settings[1] or indirectly via contaminated fomites [29, 42, 50]. According to the WHO, there is not sufficient evidence for claiming that the novel coronavirus is airborne [51]. However, a recent study measuring the level of SARS-CoV-2 in aerosols in two Wuhan hospitals shows elevated concentrations of viral RNA in patients' toilets and public areas. The analysis does not confirm the infectivity of the detected viruses but suggests a potential airborne route of transmission [52].

The Basic Reproduction Number

There is also uncertainty related to the contagiousness of SARS-CoV-2, i.e. how easily it spreads from one individual to another [53]. A frequently used quantity describing the transmission potential of a pathogen in a population is the basic reproduction number, R_0. This quantity is defined as the number of secondary cases one single infective individual can cause in a fully susceptible population, i.e. where no one has immunity towards the given pathogen [54]. The basic reproduction number is not solely determined by the characteristics of the pathogen: The number depends on how long the infected individuals are infectious, the number of contacts they make with susceptible individuals in this period and lastly the probability of transmission during each of these contacts [55]. Hence, R_0 for the same disease can vary from one place to another, for instance depending on social behavior. As long as the value of R_0 is kept below one, the disease will eventually become extinct as each infectious individual infects less than one new individual. If R_0 exceeds one, on the other hand, the initial cases may result in a large outbreak affecting nearly all individuals in the population [55].

There exists several estimates of R_0 for SARS-CoV-2. Some of the estimates are presented in Table 2.1.1, ranging from below 2 to almost 6. In February, WHO estimated the value of R_0 to be between 2.0 and 2.5 [56]. In comparison, R_0 is estimated to 12-18 for measles and approximately 1.3 for the seasonal influenza [57, 58].

Table 2.1.1: Estimated values for the basic reproduction number R_0 for SARS-CoV-2 with 95 % confidence intervals. Most results are based on numbers from Wuhan and/or mainland China, except the results presented by Zhang *et al.* which are based on data from the outbreak on the Diamond Princess cruise ship. *The interval is given as a 90 % high density interval. †Assumes an 8-fold increase in reporting rate. ‡Assumes a 2-fold increase in reporting rate.

Authors	R_0 **estimate**	**95% CI**	**Citation**
Wu *et al.*	1.94	(1.83 - 2.06)	[59]
Ferretti *et al.*	2.0	(1.7 – 2.5)	[60]
Riou & Althaus	2.2	(1.4 - 3.8)*	[61]
Li *et al.*	2.2	(1.4 - 3.9)	[50]
Zhao *et al.*	2.24†	(1.96 - 2.55)	[62]
Zhang *et al*	2.28	(2.06 - 2.52)	[63]
Zhao *et al.*	3.58‡	(2.89 - 4.39)	[62]
Sanche *et al.*	5.7	(3.8 - 8.9)	[64]

[1]The World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) define close contact as being within a range of one and two meters, respectively [48, 49].

2.1.2 Course of the Disease

The previous section gave an introduction to how SARS-CoV-2 transmits and presented estimates of the basic reproduction number describing its transmission potential in a population. The following section introduces key aspects of the course of disease of COVID-19, including the duration of the incubation period, common symptoms, the rate of asymptomatic cases, presymptomatic contagiousness and the observed case fatality rate.

Incubation Period and Main Symptoms

The delay from an individual is invaded by a pathogen until symptoms appear is referred to as the incubation period [65]. In order to determine the incubation period for SARS-CoV-2, one must therefore estimate the time of exposure, for instance by mapping travel patterns from Wuhan or by identifying contact periods with infected individuals, in addition to the time of symptom onset [66].

In February 2020, the WHO published a mean incubation period of approximately 5 to 6 days, with a range spanning 1 to 14 days [56]. This estimate is supported by several analyzes: Zhang *et al.* analyzed data from over 8500 cases in 30 Chinese provinces and estimated the mean incubation period to 5.2 days, with a $95\,\%$ confidence interval spanning 1.8 to 12.4 days [67]. Linton *et al.* estimated the incubation period based on data from 158 individuals traveling in Wuhan and found a mean incubation time of 5.6 days, with a range spanning from 2 to 14 days with $95\,\%$ confidence [68]. Lauer *et al.* analyzed the incubation time among 181 confirmed cases in China where both exposure and symptom onset were identifiable. They estimated the median incubation time to 5.1 days. Based on their analysis, they concluded that $97.5\,\%$ of all infected who develop symptoms, do so within 11.5 days after exposure. This analysis also pointed out that approximately $1\,\%$ of infected individuals develop symptoms *after* 14 days [66]. Backer *et al.* studied 88 confirmed cases by linking their travel history to or from Wuhan together with symptom onset. They estimated the mean incubation time to be 6.4 days, ranging from 2.1 to 11.1 days [69]. As these estimates show, the incubation period of COVID-19 can vary greatly among individuals.

Also the clinical picture of COVID-19 has shown to manifest very differently from person to person [70]. In February, the WHO presented a list of the most common symptoms based on almost $56\,000$ confirmed cases [56]. The top five most common symptoms and their associated frequencies are presented in Table 2.1.2.

Table 2.1.2: The five most frequent symptoms of COVID-19, based on approximately $56\,000$ confirmed cases. The numbers were presented in the middle of February 2020 [56].

Symptom	Frequency (%)
Fever	87.9
Dry cough	67.7
Fatigue	38.1
Sputum production	33.4
Shortness of breath	18.6

Probability of Asymptomatic Infection

While the incubation period describes the duration from infection to symptom onset, not all individuals infected with SARS-CoV-2 experience symptoms [71]. Asymptomatic courses of disease have been observed in all age groups [71], but children and younger adults seem to have a greater probability of not developing symptoms [72, 73].

As per May 2020, there is limited literature on the rate of asymptomatic infections. An important contribution comes from the COVID-19 outbreak on the Diamond Princess cruise ship, where over 3700 people were put in quarantine after a former passenger had tested positive for SARS-CoV-2 [74]. The ship arrived in Yokohama, Japan on February 5, and by February 21, 634 people had tested positive for the virus. Based on the results from the 3063 conducted tests, Mizumoto *et al.* estimated the asymptomatic proportion to be $17.9\,\%$ with a $95\,\%$ credible interval ranging from 15.5 to $20.2\,\%$. Approximately three-quarters of those who tested positive were aged 60 years or older [74]. In Vo', a municipality in the Italian region Veneto, the entire population was put under lockdown for 14 days after a resident, the first in Italy, died from a COVID-19 infection on February 21, 2020 [75]. During this two-week period, the entire population was tested twice for SARS-CoV-2, using nasopharyngeal swabs. Results from the two surveys showed that $43.2\,\%$ of the confirmed cases were asymptomatic [75]. Additionally, according to results from voluntary testing conducted on Iceland, approximately $50\,\%$ of the individuals who tested positive reported no symptoms of the disease [76]. These results underpin that the rate of asymptomatic COVID-19 cases could be high.

Infectiousness in Absence of Symptoms

When discussing asymptomatic carriers of SARS-CoV-2, another question arises: To which extent do infected individuals without symptoms contribute to the overall transmission? Bai *et al.* were one of the first to report evidence of transmission from an asymptomatic individual. They analyzed the case history of a family of five individuals in China, who were all infected with SARS-CoV-2 after having contact with an asymptomatic family member returning from Wuhan. Based on the sequence of events, it is likely that the family was infected by the asymptomatic individual [77]. Analysis of samples taken from the upper respiratory tract of patients with confirmed COVID-19 infections has shown similar levels of viral load in asymptomatic individuals as in symptomatic individuals [75, 78]. These results suggest that also asymptomatic carriers can shed the virus and hence infect others. The studies comparing levels of viral load or presenting examples of probable cases of asymptomatic transmission suggest that individuals lacking symptoms may contribute to the spread of the novel coronavirus but does not quantify the contribution to the overall spread. The actual role of asymptomatic individuals in the spread of SARS-CoV-2 remains to be determined. [79].

The question of transmission in the absence of symptoms applies to presymptomatic carriers as much as asymptomatic carriers. One way to evaluate the role of presymptomatic transmission is to investigate the serial interval, also known as the generation time [65, 80]. This term refers to the time span between two analogous phases in the course of a disease, for instance the symptom onset, in a chain of successive infection cases [65]. An example of two successive cases is illustrated in Figure 2.1.3. If the duration of the serial interval

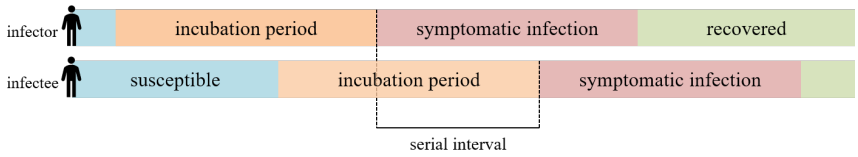

Figure 2.1.3: The serial interval, also called the generation time, is defined as the time span between two analogous phases, for instance the symptom onset, in a chain of successive cases [65]. The figure illustrates two successive cases, an infector and an infectee, where the serial interval is shorter than the incubation period. The different phases of the disease are color-coded and named in the respective boxes. Based on theory presented in [80].

is shorter than the incubation time, the virus must necessarily have been transmitted while the infector was still in the incubation period [80].

Several studies present numbers supporting presymptomatic transmission of SARS-CoV-2: Tindale *et al.* investigated data from 93 confirmed cases in Singapore and 135 confirmed cases in Tianjin, China. In both data sets, the mean serial interval was found to be shorter than the estimated mean incubation time, see Table 2.1.3. Hence, they inferred that the virus on average was transmitted 2.55 and 2.89 days before symptom onset, indicating presymptomatic transmission [80]. Another, larger study comprising 468 cases from mainland China estimated the average serial interval to be 3.96 days [81], thus even shorter than 4.56 and 4.22 days. In the same study, Du *et al.* reported that in 12.6 % of these cases, the infectee had symptoms earlier than the infector, hence yielding a negative serial interval. These results also support presymptomatic transmission [81]. He *et al.* estimated the serial interval of 77 transmission pairs and found that the infected individuals were infectious from 2.3 days before symptom onset. They also estimated 44 % of the infectees were infected before symptom onset of the infector [82].

Table 2.1.3: Estimated durations of incubation time and serial intervals for two data sets obtained from confirmed cases in Singapore and Tianjin, respectively. The time spans are given as the mean number of days, with a 95 % confidence interval in brackets [80].

Data set	Incubation time	Serial interval
Singapore	7.11 (6.13, 8.25)	4.56 (2.69, 6.42)
Tianjin	9.02 (7.92, 10.2)	4.22 (3.43, 5.01)

Other studies suggest that the serial interval exceeds the incubation time and do therefore speak against presymptomatic transmission. Ki *et al.* estimated the incubation period and serial interval based on 28 cases in Korea to 3.9 and 6.6 days, respectively [83]. Based on the first 425 confirmed cases in Wuhan, Li *et al.* estimated the average incubation time to 5.2 days and the serial interval to 7.5 days [50]. The results presented by Ki *et al.* and Li *et al.* do hence suggests that SARS-CoV-2 transmits after symptoms have appeared.

Duration of Infection and Infectiousness

The duration of a COVID-19 infection has shown to depend on the severity of disease. In case of a mild infection, the symptoms most often pass within one to two weeks, whereas in case of a more severe disease course, the symptoms may last for three to six weeks [73]. Thus, there is a great variety in the infection duration among individuals. There exist several estimates of the duration COVID-19 infections, several of them investigating the time span from symptom onset to death. Wang *et al.* analyzed data from the first seventeen reported deaths in China. They reported a median of 14 days from the first symptoms to death, with a range spanning 6 to 41 days [25]. Two other estimates are given as lognormal distributions[2]: Jung *et. al* deduced a lognormal distribution with location and scale parameters of 2.84 and 0.52, respectively, yielding a median value of approximately 17 days from symptom onset to death [84]. Linton *et al.* report a median value of 13.2 days, and that $99\,\%$ of deaths occurred within 36.0 days from symptom onset [68].

The time span of which an infected individual is infectious has a big impact on the spreading potential of a disease [55]. The existing literature on the infectiousness of COVID-19 suggest that the infectiousness varies throughout the course of the disease [78, 82, 85, 86]: In one of these studies, He *et al.* investigated the temporal transmissibility dynamics of COVID-19 by combining analysis of viral load throughout the disease course with estimated serial intervals among infector-infected pairs. They observed that the viral load was highest at the time of symptom onset, followed by a gradual decrease towards towards the limit of detection around three weeks. They also observed a substantial transmission potential before symptom onset [82]. Another study, which showed a similar decline of viral load in the sputum, also observed that the shedding of viral RNA outlived the presence of symptoms [85].

Case Fatality Rate

As previously introduced, COVID-19 has shown to be less deadly per infection case than its coronavirus relatives SARS and MERS [38]. An important measure of the severity of a disease is the case fatality rate (CFR). This term describes the fraction of fatalities among confirmed cases [87, 88], see Equation (2.1.1). Due to limited testing capacity and the perpetual challenge of identifying infected individuals with asymptomatic or mild disease courses, the number of confirmed cases in the denominator is often lower than the actual number of infection cases. Hence, the calculated CFR can be an overestimated representation of the reality [88]. The term 'infection fatality rate' (IFR) is sometimes used to capture fatal outcomes among all infected individuals [89].

$$\text{CFR} = \frac{\text{number of deaths}}{\text{number of confirmed cases}} \tag{2.1.1}$$

Several risk factors are affecting the severity and mortality of a COVID-19 infection [90]. One of them is the presence of underlying health conditions, such as chronic respiratory disease, cardiovascular disease, cancer, hypertension or diabetes. There has also been reported an increased risk in men than in women, an observation that may be influenced by higher smoking rates and related comorbidities. Another key risk factor is related to

[2] See Section 2.4.2 for details on probability distributions.

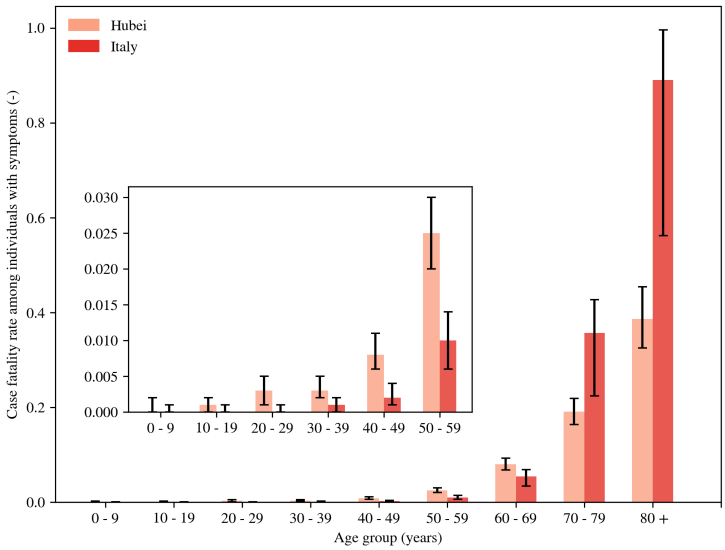

Figure 2.1.4: Estimated age-specific case fatality rates among symptomatic individuals from Hubei and Italy. The error bars indicates a 95 % credible interval [91]. The inset figure gives a closer look at the age groups with lower mortality rates. Based on data presented in [91].

older age [90]. Hauser *et al.* presented estimated age-specific case fatality rates based on data from the Hubei province in China and the northern part of Italy [91]. Their results show that the mortality increases with age and are thus in compliance with previous analyses [90]. Figure 2.1.4 presents their estimated age-specific case fatality rate among individuals with symptoms.

Secondary infections may also affect the mortality of COVID-19 [92]. Opportunistic pathogens take advantage of the absence of a normal host resistance, for instance in individuals where the pulmonary tissue is weakened after a respiratory infection or in patients requiring mechanical ventilation. In the latter case, the tube that is passed down the windpipe offers an artificial surface ideal for bacterial biofilm formation [93–95]. Increased morbidity and mortality associated with bacterial secondary infections have, for example, been shown for both seasonal and pandemic influenza [95, 96]. In a study investigating demographic, clinical and laboratory data from 191 COVID-19 patients admitted to two hospitals in Wuhan, Zhou *et al.* observed that half of the patients who died from COVID-19 had contracted a secondary infection, compared to 1 % of the survivors [32]. Almost one-third of the patients who received invasive mechanical ventilation experienced ventilator-associated pneumonia [32].

Moreover, 95 % of the hospitalized patients were treated with antibiotics [32], which leads to the question of whether the prevalence of antibiotic resistance may affect the mortality of the ongoing pandemic. Through varying mechanisms of action, resistant bacteria

can withstand antibiotic treatments. Either, the bacteria are able to reduce the concentration of drugs reaching its cellular target, for instance by inactivating the drug, decreasing the uptake or by producing efflux pumps that transport the drug out of the cell. Or, the bacteria circumnavigate the metabolic target of the drug, for instance by using an alternative metabolic pathway or by altering the structure of the enzyme of attack so that it is no longer recognizable [97]. Thus, resistant bacteria do not respond to antibiotic treatment as susceptible bacteria would, resulting in increased morbidity and mortality [98]. A central opportunistic bacteria associated with an increasing antimicrobial resistance is *Staphylococcus Aureus* [99]. This bacterial species can cause diseases ranging from skin infections to respiratory tract infections and sepsis [100]. Table 2.1.4 shows how the prevalence of methicillin-resistant *S. Aureus* (MRSA) has evolved in some European countries from 2015 through 2018 [101]. The population-weighted mean of the countries belonging to the European Union (EU) or the European Economic Area (EEA) is included for reference. There exist several opportunistic bacteria that potentially can contribute to an increased COVID-19 mortality. To narrow the scope of this work, MRSA is used for demonstration.

Table 2.1.4: Overview of the determined percentage of methicillin-resistant bacteria among invasive isolates of *S. aureus* (MRSA) in a selection of countries belonging to EU or EEA. The population-weighted mean is marked in boldface. Obtained from [101].

Country	**2015**	**2016**	**2017**	**2018**
Norway	1.2	1.2	1.0	0.9
Sweden	0.8	2.3	1.2	1.9
United Kingdom	10.8	6.7	6.9	7.3
EU/EEA	**19.0**	**17.7**	**16.8**	**16.4**
Spain	25.3	25.8	25.1	24.2
Italy	34.1	33.6	33.9	34.0

There exist several studies comparing the mortality of infections caused by methicillin-resistant and methicillin-sensitive *S. aureus* (MSSA) strains: Hanberger *et al.* investigated data from 13 796 patients admitted to intensive care units (ICUs) from 75 countries. Approximately 1000 of these patients contracted an *S. aureus* infection, roughly equally many susceptible as resistant towards methicillin. They observed a statistically significant difference in mortality rates between patients infected with MRSA compared to MSSA [102],

Table 2.1.5: Comparison of some observed case fatality rates of *S. aureus* infections caused by methicillin-resistant and -susceptible bacteria, denoted MRSA and MSSA, respectively.

Authors	**MRSA**	**MSSA**	**Ratio**	**Comment**	**Citation**
Hanberger *et al.*	36.4 %	27.0 %	1.35	Hospital	[102]
Hanberger *et al.*	29.1 %	20.5 %	1.42	ICU	[102]
Gastmeier *et al.*	16.9 %	7.0 %	2.41	-	[103]
Blot *et al.*	63.8 %	23.7 %	2.69	In-hospital mortality	[104]
Blot *et al.*	53.2 %	18.4 %	2.89	30-day mortality	[104]

see Table 2.1.5. In another study, Gastmeier *et al.* investigated the mortality among patients with *S. aureus* pneumonia. The case fatality rate among patients with a resistant strain was more than double as high as for patients with the susceptible strain [103]. Blot *et al.* found even larger ratios when investigating the mortality among 85 patients with *S. aureus* bacteremia, a condition where the bacteria is present in the blood, with ratios equal to 2.69 and 2.89 for in-hospital mortality and 30-day mortality, respectively [104].

Immunity and Risk of Reinfection

As per May 2020, there exists no approved vaccine providing immunity towards SARS-CoV-2 [105]. Immunity can hence only be acquired by being exposed to the virus, resulting in a production of antibodies [65]. Reports of patients who have tested positive after recovery, however, have raised questions about whether convalescing patients could be reinfected [106, 107].

Two recent studies suggest that COVID-19 infections do results in the production of antibodies: Bao *et al.* infected four adult rhesus macaque monkeys with SARS-CoV-2 and monitored clinical signs and immune response in the following weeks [107]. 28 days after the initial infection, all monkeys tested positive for the specific antibodies. Then, two of the monkeys were rechallenged with the same viral dose to investigate a potential reinfection. Analyses of viral loads in nasopharyngeal and anal swabs five days after re-exposure showed no signs of infection recurrence, hence indicating that the primary infection protected the monkeys from the secondary exposure [107]. Fafi-Kremer *et al.* tested 162 recovered hospital staff members from Strasbourg University Hospitals for anti-SARS-CoV-2 antibodies [108]. All participants had recovered from a mild COVID-19 infection. They found that antibodies against SARS-CoV-2 were detectable in nearly all hospital staff thirteen days after symptom onset. Their analysis also showed that the proportion of individuals with detectable antibodies increased with the number of weeks after symptom onset [108].

2.1.3 Control Measures

Several infection control measures have been implemented in order to slow the spread of SARS-CoV-2. This section briefly presents some of the key interventions and the current guidelines on testing, isolation, quarantine and contact tracing given by the Norwegian Institute of Public Health (NIPH).

- **Social distancing**: By increasing the physical distance between individuals, social distancing aims to reduce the mixing of infectious and susceptible individuals [109].
- **Testing**: A test can either be conducted to confirm or deny a current infection, or to determine whether an individual has antibodies from a previous infection [110]. Only the former test is relevant here. The limited testing capacity has necessitated a prioritizing of suspected COVID-19 cases based on their clinical picture, profession within healthcare services or close contact with confirmed cases of COVID-19 [111]. As of April 1, 2020, NIPH recommends testing of all people with "[...] acute respiratory tract infection with fever, cough or breathing difficulties." The following prioritized list is given (directly obtained from [111]):

1. Patients in need of hospital admission
2. Patients / residents in health institutions
3. Employees in the healthcare service with work that puts them in the vicinity of patients. (This group may be considered for testing even for milder acute respiratory symptoms with no other probable cause that have lasted for more than 2 days.)
4. People over the age of 65 who have underlying chronic diseases, e.g.:
   - cardiovascular disease
   - diabetes
   - chronic lung disease
   - cancer
   - high blood pressure.
5. People who have been in close contact with a confirmed case of COVID-19.

- **Isolation**: Isolation of confirmed cases is implemented in order to prevent the virus from spreading to susceptible individuals. As per April 1, 2020, individuals with a confirmed COVID-19 infection should be isolated until $SI7$ days after symptoms are gone [112].
- **Quarantine**: Individuals who have had close contact with a confirmed COVID-19 case should enter a quarantine lasting 14 days[3] after the last contact with the infected individual [113].
- **Contact Tracing**: A positive COVID-19 test initiates tracing of contacts with potential infectees. These include all individuals who have had close contact with the confirmed case from 24 hours[4] before symptom onset until the infected individual comes out of isolation [114]. Indoors, NIPH defines 'close contact' as being closer than two meters for more than 15 minutes continuously.

This section has introduced relevant theory and quite a few numbers describing the transmission of SARS-CoV-2 and the course of disease of COVID-19. The next main section discusses theory related to networks and network spreading phenomena, which is useful for describing how SARS-CoV-2 spreads in a human population.

[3] On May 8, 2020 the duration of quarantine was changed to 10 days after last contact with a confirmed case.
[4] On May 8, 2020 the definition was changed to include close contacts from 48 hours before symptom onset.

2.2 Networks and Spreading Phenomena

A network consists of nodes connected by links [115]: A social network may consist of users connected by friendships, or an air traffic network of airports connected by flights, illustrated in Figure 2.2.1. The links of a network may be directed or undirected, depending on whether the interactions have directionality or not. In the social network example, Facebook is an example of an undirected network where a friendship is a two-way interaction between two users. Twitter, on the other hand, is an example of a directed network where one user may follow the other one without being followed back. The links of a network may also be weighted or unweighted. The former case is exemplified through an air traffic network where the frequency of flights between two destinations matter; a route flown several times a day will have a greater weight than a route flown once a week.

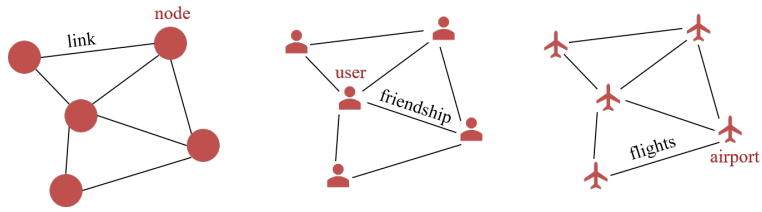

Figure 2.2.1: Illustration of networks with nodes connected by links. A social network with users connected by friendships and an air traffic network with airports connected by flights.

Networks are useful for describing several spreading phenomena [115]: Biological spreading phenomena comprise the transmission of pathogens on networks interconnecting human beings. An example of such a network is the airline network and its role in the 2003 outbreak of SARS [116]. Another is the networks of sexual partners and their significance in the transmission of the human immunodeficiency viruses (HIV) [117]. Digital spreading phenomena may include the spread of digital viruses, malicious software which for instance can spread through an e-mail network [118]. Social networks can capture the spread of knowledge, behavior or rumors. Twitter has proven useful for studying such phenomena as all messages by default are public [119]. When investigating spreading phenomena on networks, the spreading ability of different nodes is of interest. Locating nodes central to the spreading process in a network may, for instance, be an important step in the process of designing efficient measures to prevent epidemic spreading [120]. Examples include targeted vaccination in a population [121] or targeted allocation of infection control resources in an inter-hospital network [122].

The following section will give an introduction to network theory and relevant tools for describing network topology and identifying central nodes. The first part addresses static networks, whereas the second part discusses networks that evolve with time. The third part introduces empirical contact networks and how these can be generated based on close-proximity interactions. This part also presents a hospital ward contact network in more detail. The theory presented in this section presents only a small excerpt of an extensive field of study. The reader is referred to Barabàsi [115] or Newman [123] for a more comprehensive introduction. Unless otherwise specified, the presented network theory is based on these references.

2.2.1 Static Networks

A network can be represented in several ways. For instance, all nodes connected by a link might be stored in a list: {(1, 2); (2, 3); (1, 3)}. This list represents a network where three nodes are linked in a triangle. For mathematical network analysis, a network is often represented by its adjacency matrix A, where entry ij represents the relationship between nodes i and j, as shown in Equation (2.2.1): In an undirected, unweighted network, an entry equal to 1 represents an existing link between nodes i and j, whereas an entry equal to 0 represents the absence of a link. The adjacency matrix of an undirected network is symmetric, giving that if there exists a link between nodes ij, the same link exists between ji. For a directed network, on the other hand, the value of entry ij only tells whether a link from i to j exists, not if there is a link from node j to i. For weighted networks, the entries of the adjacency matrix also give information about the weight of the links.

$$A = \begin{matrix} A_{1,1} & A_{1,2} & ... & A_{1,j} \\ A_{2,1} & A_{2,2} & ... & A_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i,1} & A_{i,2} & A_{i,3} & A_{i,j} \end{matrix} \tag{2.2.1}$$

The number of nodes in a network is denoted N and equals the number of rows or columns in the square adjacency matrix. The total number of links is denoted M. For an undirected network, the number of links is calculated by summarizing the number of all non-zero entries in A and divide this number by two. Due to undirected links and symmetry in the adjacency matrix, all links are counted twice. As the following sections will demonstrate, the adjacency matrix is useful for determining important network characteristics, both for specific nodes and the network as a whole.

Local Centrality Measures

One of the most central node properties in a network is the node degree, describing the number of neighbors the node is connected to. The degree for node i is denoted k_i. For an undirected and unweighted network, k_i is calculated by summarizing the entries of its corresponding column in the adjacency matrix, as shown to the left in Equation (2.2.2). For directed networks, one distinguishes between incoming and outgoing degrees from a node. For weighted networks, the equivalent of the node degree is called the node strength, denoted s_i [124]. The node strength takes the link weights into account and is calculated by summarizing the entries of the corresponding column for node i in the weighted adjacency matrix w, as shown to the right in Equation (2.2.2).

$$k_i = \sum_{j=1}^{N} A_{i,j} \qquad s_i = \sum_{j=1}^{N} w_{i,j} \tag{2.2.2}$$

Figure 2.2.2 illustrates three static networks. The network in panel A is both undirected and unweighted, whereas the networks in panel B and C are directed and weighted, respectively. The degree of the red-colored node in panel A simply equals the number of nearest neighbors, yielding $k_i = 4$. In panel B, the in-degree equals $k_i^{in} = 3$, whereas the

out-degree equals $k_i^{out} = 2$. The strength of the red-colored node in panel C takes the link weights into account, yielding $s_i = 11.25$.

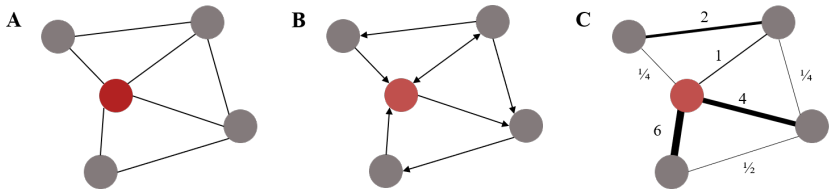

Figure 2.2.2: Local centrality measures in three static networks: Panel A shows an undirected, unweighted network where the red-colored node has a degree equal to 4. In the directed network in panel B, the red-colored node has an in-degree and out-degree of 3 and 2, respectively. Panel C shows a weighted network where the red-colored node has a strength of 11.25.

The node degrees in a network can be used to say something about the network as a whole. The average degree represents the average number of links per node, and can, for instance, be used to characterize how dense or sparse the network is. Another central network characteristic is the degree distribution, i.e. the probability P of a randomly chosen node to have a degree k. The shape of the degree distribution can have a great impact on how easily something spreads across the network. If a spreading agent reaches a node with a high degree, this "super-spreader" can pass it on to many neighbors at once.

Path and Distances

In light of spreading phenomena, several important network measures relate to the path between nodes, i.e. the number of steps required to move from one node to another. Figure 2.2.3 illustrates the four steps required to move from node i to j. More specifically, these steps constitute the shortest path between the two nodes. Additionally, one could reach node j by either of the two other paths depicted with dashed links. These paths, however, include one extra step compared to the shortest path. The shortest path between node i and j is denoted $L(i, j)$ and is defined as the lowest amount of steps required to get one node to the other. The average shortest path for a node, $\langle L_i \rangle$, is calculated by summarizing the shortest paths to all other reachable nodes and divide by the total number of nodes N, as shown in Equation (2.2.3).

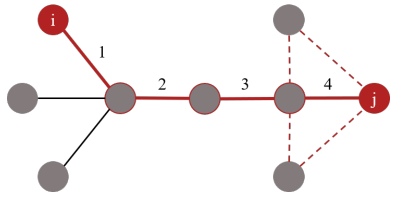

Figure 2.2.3: The shortest path from node i to node j, enumerated and depicted with red-colored links, consists of four steps. Additionally, one could reach node j by replacing step four with the routes depicted by dashed links.

$$\langle L_i \rangle = \frac{1}{N} \sum_{j=1}^{N} L(i,j) \tag{2.2.3}$$

In undirected networks, all nodes are reachable from one another as long as they belong to the same connected component. Figure 2.2.4 shows an example of a network consisting of two disconnected components. The three nodes in the smallest component are not reachable from any of the nodes in the larger component. The largest connected component in a network is often referred to as the giant component (GC). In directed networks, however, the directionality of the links may provide a path from node i to j without offering a returning path. However, if a path exists from node i to j and from node j to k, this implies that a path from i to k exists.

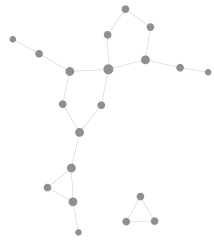

Figure 2.2.4: A network consisting of two disconnected components. The three nodes in the smallest component are not reachable from any of the nodes belonging to the larger component. Once you have seen the football player in this figure, there is no way to unsee it.

Global Centrality Measures

The node degree and node strength are local centrality measures for a node but does not take the node's position in the network as a whole into account. Two important centrality measures that do are the closeness centrality (CC) and betweenness centrality (BC). The closer to all other nodes a node is, the higher its CC is. The closeness centrality is defined as the inverse of the average shortest path, as shown to the left in Equation (2.2.4). Consequently, the shorter path node i has to all other nodes in the network, the higher CC_i is. The betweenness centrality is also based on shortest paths, but instead of looking at the average shortest paths from node i, the BC is a measure of how many of the other nodes' shortest paths are going through node i. This centrality measure is therefore suitable for identifying important "bridges" in the network. Removing a node with a high BC from the network would result in a drastic increase in the average shortest path in the network, or split the network into separate, disconnected components. By denoting the number of shortest paths between node j and k as $\rho(j,k)$, and the number of these shortest paths going through node i as $\rho_i(j,k)$, the BC for node i is calculated as the sum of the ratio between the two, as shown to the right in Equation (2.2.4).

$$CC_i = \langle L_i \rangle^{-1} = \left(\frac{1}{N-1} \sum_{j=1}^{N} L(i,j) \right)^{-1} \qquad BC_i = \sum_{j,k=1; j \neq k \neq i}^{N} \frac{\rho_i(j,k)}{\rho(j,k)} \tag{2.2.4}$$

The two global centrality measures CC and BC may rate the centrality of the same node in a network differently. Figure 2.2.5 shows an illustration of calculated CCs and BCs of nodes in the same network. The panel to the left shows that the red-colored node in the very middle has the highest CC of all nodes. This node reaches all other nodes with an average of 1.75 steps. The inverse of 1.75 is approximately 0.57. The panel to the right shows that the most central node according to BC no longer is the node in the middle, but the red-colored node to its left. Some of the nodes in this network have a betweenness centrality of zero, implying that no shortest paths go through these nodes.

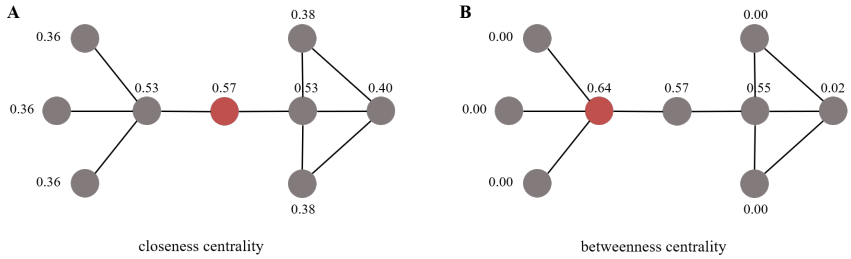

Figure 2.2.5: Illustration of closeness centrality and betweenness centrality of nodes in a simple network. The red-colored node represents the most central node. The figure illustrates how different centrality measures can range the same node in a network differently. The values are found by using the software Cytoscape, see Section 3.1.

Another approach for identifying nodes that are central in a network is by peeling off less central nodes until an interconnected core remains [125]. For unweighted networks, the k-core is calculated based on the node degree, where k denotes the minimum degree of all nodes in the remaining subset of the network [126]. Thus, to find, for instance, the 2-core of a network, all nodes with a degree less than two is recursively removed until all remaining nodes have a degree equal to or greater than two. The 1-core of a network refers to the connected component where no isolated nodes in the network remain [126]. Figure 2.2.6 illustrates different k-cores of a simple network. For weighted networks, the equivalent is called the s-core, peeling of nodes having a strength s_i less than s [125].

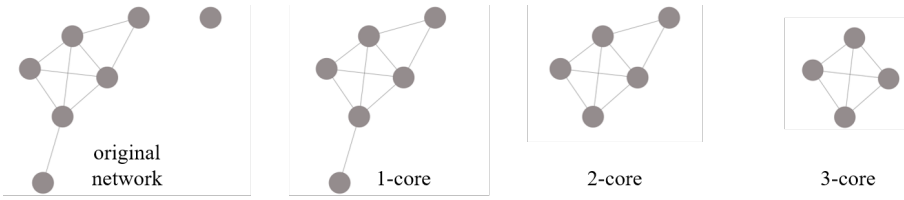

Figure 2.2.6: Illustration of a simple network and its k-cores. Nodes with degree less than k are recursively peeled off, leaving a remaining core where all nodes have degree k or more. The given network has no 4-core.

2.2.2 Temporal Networks

So far, the discussion of networks has revolved around static networks, where the adjacency matrix A describes a set of non-dynamic interactions between nodes. In many real-life cases, however, the interaction between nodes is not continuous. In cases where the network changes with time, it is referred to as a temporal network [115]. These networks can be described using a time-evolving adjacency matrix, $A_{ij}(t)$, where entry i, j is equal to 1 if nodes i and j are connected at time t. Temporal networks are especially important in terms of describing spreading phenomena where the spreading and network evolve approximately at the same time scales [127]. An illustrative example is the spread of sexually transmitted diseases. In sexual contact networks, the transmission of the infection is directly correlated to the evolution of the contact network itself [128].

In the same way as the network topology influences the spreading dynamics on a network, the temporal structure can have a great impact on the transmission process [127]. Panel A of Figure 2.2.7 illustrates how a temporal network can be visualized, denoting the time-points of when the links are active indicated on each link. Panel B shows the corresponding static aggregation. Panel C shows how the links of the temporal network evolve with time. This network is rather simple, but illustrates an important concept related to spreading phenomena on temporal networks: If the temporal network was projected to a static network, a spreading agent would reach all nodes in the network in the course of maximum two steps, regardless of which node it started at. In the temporal network, on the other hand, the temporal structure sets limits for how large parts of the network a spreading agent can reach and how fast it spreads [127]. If node A carries a spreading agent, and this agent spreads immediately upon link-activation, node B is infected at time-step 6, whereas node C is infected at time-step 8. Node D will not become infected. If node D is the initially infected node, however, all nodes in the network are infected at time-step 6.

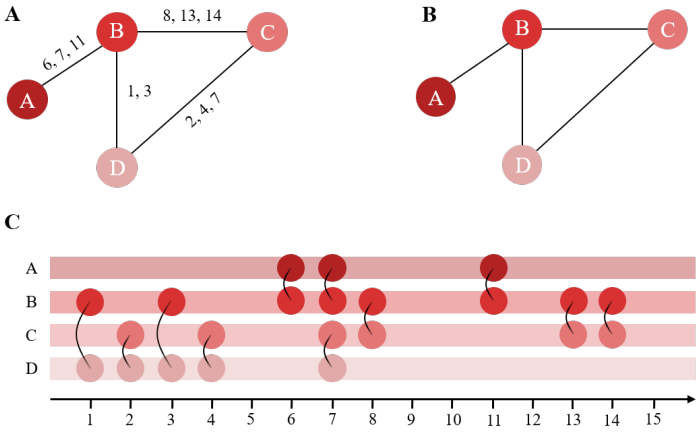

Figure 2.2.7: Panel A shows a temporal network with time-evolving links. The time-points of activation are indicated on each link. Panel B shows a static projection of the temporal network. Panel C shows explicitly how the links evolve through a series of time-steps. Adapted from [127].

As the example above points to, projecting a temporal network to a static network, may result in significant loss of information about the system. In the following, a selection of measures of temporal-topological structures are presented. Where not specified otherwise, the following theory is based on Holme & Saramaki [127].

Time-respecting paths

As discussed in Section 2.2.1, a path refers to the sequence of steps required to get from one node to another. In temporal networks, these paths are necessarily constrained by the order of link activation, as the example presented above underlines. Within a certain observation window $t \in [t_0, T]$, a time-respecting path defines which nodes are reachable from which other nodes in the temporal network. In the example above, for example, there exists no time-respecting path from node A to node D. Similarly to static, directed networks, an existing path from node i to j does not guarantee an existing path from j to i. However, as opposed to directed networks, where an existing path from node i to j and from node j to k implies that there is a path from i to k, the paths in a temporal network are not transitive. The paths of a temporal networks are also temporal.

Two key measures follow the definition of time-respecting paths. Firstly, the set of influence is a node specific property representing the set of nodes which can be reached by time-respecting paths starting at node i. Thus, the set of influence for node A in the example above includes nodes B and C, but not D. Secondly, the reachability ratio is a network specific property based on the sets of influence for all nodes in the network. The fraction of nodes in the set of influence is calculated for each node, contributing to the average fraction of reachable nodes in the network as a whole. Hence, a reachability ratio of 1.0 implies that all nodes can reach all other nodes through time-respecting paths in the given time window.

In some spreading phenomena, one could imagine that a clock starts ticking when a node becomes infected. The spreading agent could for instance be a rumor that is passed

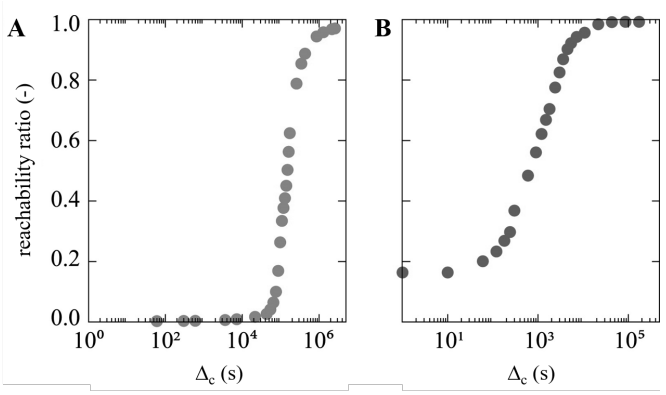

Figure 2.2.8: Reachability ratio as function of maximum allowed waiting time for two real-life temporal networks. Panel A shows a network of mobile telephone calls, whereas panel B shows an airline network of passenger flights. Obtained from [127], based on data from [129].

on from person to person or an infectious virus transmitting upon contact. When the given time has passed, the node recovers and can no longer pass the spreading agent on to its neighbors. Thus, in order for the spread to progress on the network, the node must be able to infect its neighbors while still being infectious. The upper limit of this time span from infection to recovery is called the maximum allowed waiting time, Δ_c, and affects the reachability ratio of the network. If a node never recovers once infected, $\Delta_c \to \infty$, a significantly larger proportion of the network can become infected compared to a case where the the nodes recover almost immediately. How the reachability ratio varies with maximum allowed waiting time gives important information about the temporal characteristics of a given network. Figure 2.2.8 shows two examples: Panel A presents a mobile telephone call network consisting of over $150\,000$ calls between almost 2000 users [129]. The reachability ratio increases rapidly when the maximum allowed waiting time passes two days. Hence, if the interest of conveying an interesting rumor languishes within two days, the story remains within a rather low proportion of the network. In panel B, which presents an airline network consisting of approximately $180\,000$ flights connecting 279 US airports, the corresponding increase is observed around 30 minutes, which is consistent with the minimum allowed transfer time between two connected flights [127, 129].

Temporal Centrality Measures

Many of the tools presented in Section 2.2.1, useful for describing static networks and identifying nodes central to the spreading process, can be transferred to temporal networks. For instance, the time-dependent degree of a node i can be defined as the number of links connecting the node to its neighbors in a given time window. For calculating global centrality measures depending on paths, the path in static networks can be replaced by the time-respecting path in temporal networks. The shortest path in a temporal network, however, can be defined in two different ways: The *distance* refers to the number of links required to get from one node to another, whereas the *latency* refers to the duration. The average latency can be used to characterize the "velocity" of the temporal network and describes how fast something can spread.

Regardless of choice of shortest path, either distance or latency, the temporal closeness centrality and betweenness centrality can be expressed as presented in Equation (2.2.5). In addition to deciding whether to describe the shortest path in terms of distance of latency, one also must define the time interval t.

$$CC_{i,t} = \langle \lambda_{i,t} \rangle^{-1} = \left(\frac{1}{N-1} \sum_{j \neq i}^{N} \lambda_t(i,j) \right)^{-1} \qquad BC_{i,t} = \sum_{i \neq j \neq k}^{N} \frac{\rho_{i,t}(j,k)}{\rho_t(j,k)} \tag{2.2.5}$$

2.2.3 Empirical Contact Networks

The two examples of temporal networks presented in Section 2.2.2 illustrate how modern inventions such as mobile telephones and air travels facilitate communication and interaction between human beings. Despite the technological development, face-to-face interactions continue to be an important part of human behavior [130]. The social networks

formed by these physical proximity interactions play an important role in the dissemination of information, opinions and not least infectious diseases. Therefore, building temporal networks based on accurate empirical contact data may provide valuable insight into epidemic spreading and the design of efficient control measures [130].

Traditionally, empirical contact networks have been generated based on self-reporting systems such as contact diaries and surveys [131]. During the period of study the participants note down information about the contacts they make, such as time, duration and the name of the contact partner. This method of recording contacts has, however, shown to have limited correlation with the actual contact patterns, largely influenced by underreporting of contacts [131]. The same contact data could also be registered by third party observers monitoring the contacts between individuals [132]. A key limitation associated with these approaches is the bias coming from the observation itself, as the presence of observers may give rise to behavioral modifications [16].

A third approach is offered by the use of electronic proximity sensors [132]. Radio-Frequency IDentification (RFID) technology registers close-proximity interactions (CPIs) between all participants wearing badge-size electronic devices [133]. These wearable devices exchange small ultra-low-power radio packets when being within a certain range, for instance 1.5 meters. The registered contact results in a link between the two individuals in the forming temporal network, as illustrated in Figure 2.2.9. The human body functions as a radio frequency shield. Thus, by wearing the badges on the chest, only face-to-face contacts are registered. By detecting contacts every 20 seconds, for instance, the use of RFID tags offer a high resolution empirical contact network [16].

One RFID system used to generate several empirical inter-individual contact networks is developed by SocioPatterns, a collaboration between several researchers and developers. Many of the datasets obtained through this platform is made publicly available. These include contact networks from a primary school [134], a high-school [135], an office building [136], households in Kenya [137], two conferences and a museum exhibition [138, 139]. A network generated from registered contacts in a hospital ward [16] serves as basis for the model developed in this thesis, and is presented in more detail below.

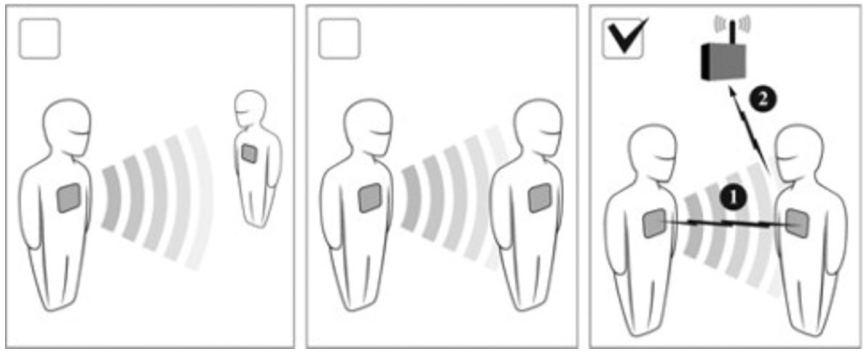

Figure 2.2.9: Contacts among participants wearing RFID devices are only detected when two individuals are located within a certain distance and facing each other. The signal is registered by an RFID reader. Obtained from [133].

A Hospital Ward Contact Network

The data constituting the empirical contact network presented by Vanhems *et al.* were collected in a short stay geriatric unit with 19 beds in a hospital in Lyon, France in 2010. The data were collected in the course of four days, lasting from Monday 6 December at 13:00 to Friday 10 December at 14:00. In the course of these 97 hours, over $14\,000$ contacts were registered among the 46 staff members and 29 patients who participated, 75 individuals in total. The participation rate was $92\,\%$ and $94\,\%$, for staff and patients respectively. The material presented in this section is based on the work presented by Vanhems *et al.* [16].

The individuals were categorized into four different classes based on their role in the ward. An overview of the classification is presented in Figure 2.2.10. One of the roles belonging to the healthcare worker category, denoted NUR, consisted of paramedical staff, including nurses and nurses' aides. This category also encompassed several staff members working on-demand or without any fixed schedule, such as a physiotherapist, a nutritionists, a counselor and a physical therapist. There were 27 individuals belonging to this category participating in the study. The second healthcare worker role, MED, consisted of medical doctors, including physicians and interns. There were eleven individuals belonging to this role in the study. In addition to healthcare workers, the workforce also consisted of eight individuals belonging to the administrative staff, denoted ADM. The fourth role, PAT, consisted of 29 patients. The article provides information about working shifts for nurses and nurses' aids, as well as medical doctors. The given information is summarized in Table 2.2.1. No information regarding working hours for the administrative staff, nor patient admissions or discharges are provided.

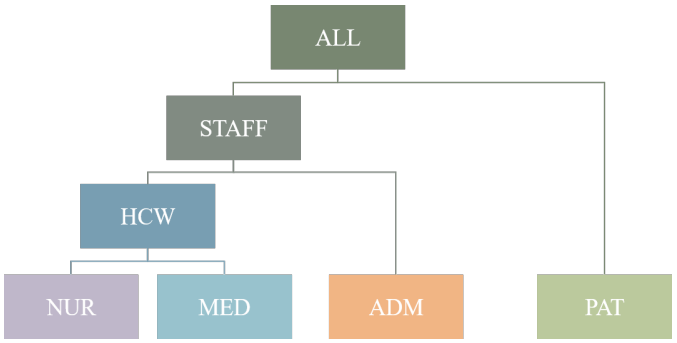

Figure 2.2.10: Overview of the four different role categories in the ward; NUR, MED, ADM and PAT. The two former roles constitute the healthcare workers, which together with the administrative staff constitute the staff in the ward. Based on information presented in [16].

The empirical network is available as supplementary material of the original article and can be accessed here. The network is stored in Graph Exchange XML Format (GEXF), divided into five different files representing each day Monday through Friday. The dataset has a temporal resolution of 20 seconds, and does not provide any spatial data. The combined network consists of 75 nodes, each representing a unique RFID tag, and 1139 links corresponding to registered contacts between them. Each node has one attribute describing

Table 2.2.1: Overview of the given shifts for all three roles belonging to the ward staff. In the course of 24 hours, a total of fourteen nurses and nurses' aids (NUR) work in the ward, divided into four shifts. Four medical doctors (MED) work during daytime. No information is provided regarding the administrative staff (ADM). Based on information presented in [16].

Role	Shift	Duration	Staffing
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
MED	-	08:00 - 17:00	4
ADM	-	-	-

the role. Each link has three attributes, namely the total number of contacts, the cumulative duration of them and lastly a list of all the time intervals of registered contacts. The node and link attributes are summarized in Table 2.2.2. An example of two RFID tags and the information contained in the link interconnecting them is provided in Figure 2.2.11. Links to other nodes are omitted for simplicity.

Table 2.2.2: Overview of nodes and links and their corresponding attributes in the empirical close-range interaction network presented by Vanhems *et al.* [16]

	Corresponds to	Attributes
nodes	RFID tag	role
links	contact	Ncontacts
		cumulativeduration
		list_contacts

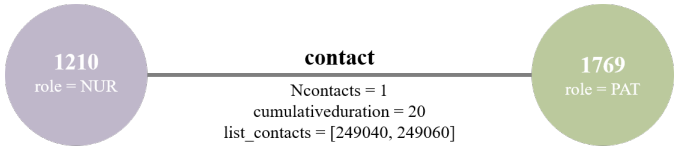

Figure 2.2.11: Example of two nodes corresponding to two different RFID tags carried by a nurse and a patient, respectively. The node attribute is given below the tag id, whereas the link attributes are given below the link. This pair of individuals had only one contact in the course of the observation period, which lasted for 20 seconds. All other links are omitted for simplicity.

The empirical network presented by Vanhems *et al.* has been analyzed and used by several others. Examples include analysis of temporal characteristics [140] and the simulation of several spreading phenomena on the temporal network [141] or its corresponding static aggregation [142]. To my knowledge, the network has not been used to derive agent-based model parameters used for subsequent network simulation.

2.3 Epidemic Modeling

It is clear that the spread of SARS-CoV-2 is a complex process affected by a wide range of epidemiological parameters, and that empirical contact networks may serve as a basis for investigating the inter-individual transmission. In order to gain a solid understanding of the interplay between these elements, computational modeling provides a valuable tool. Using mathematical models to investigate epidemiology and infectious diseases is not a new phenomenon. In 1760, the Swiss scientist Daniel Bernoulli presented a mathematical analysis investigating the mortality of smallpox and the potential advantages of universal vaccination [143]. In 1927, the British scientists Kermack and McKendrick presented their work on epidemiological modeling in "A Contribution to the Mathematical Theory of Epidemics" [144], laying an important foundation for the modern mathematical epidemiology.

A model can be defined as an abstract and simplified representation of an existing or planned reality and is often designed to explain observed or predicting future phenomena [145]. Computational models have become helpful tools for public health decision-making and evaluation of potential control measures [146]. The tree of computational modeling splits into two main branches: Compartmental models aggregate the population into different categories and use differential equations to describe the transitions from one category to another. Agent-based models shift the focus on to each individual and interactions between them [147]. The following section will give a brief introduction to both approaches, including main principles, capabilities and limitations. Although computational modeling can be used to model any kind of population, only individuals of human populations are used in the following.

2.3.1 Compartmental Modeling

In compartmental models, individuals are classified into different compartments based on given states. For instance, an individual can be categorized as susceptible to, infected by or recovered from a disease. The flux of individuals from one state to another is described by differential equations. The models are often denoted with the acronym of the flow pattern between the states, where the SIR model (susceptible, infected, recovered) is one of the most central [148]. Compartmental models are based on two key assumptions [115]:

1. **Compartmentalization**: All individuals are divided into a set of compartments describing their states. All individuals within one compartment are treated equally.

2. **Homogeneous mixing**: All individuals are homogeneously mixed, meaning that all individuals have the same chance of interacting with all other individuals.

The SIR model is useful for capturing the development of immunity against a pathogen after recovering from an infection [115]. In this model, S denotes healthy, non-infected individuals who can become infected with a pathogen. When infected, the individual changes state from S to I. The individual will eventually recover from the disease and return to a healthy state. When recovered, R, the person is assumed to have developed immunity to the pathogen, and cannot be infected again [148]. The three states are presented in Figure 2.3.1. In order to model the transition dynamics between the states, differential

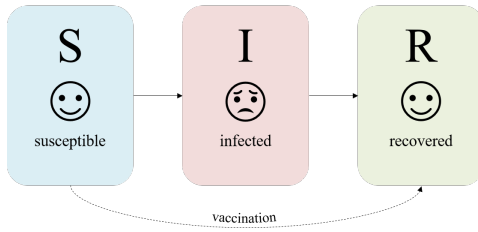

Figure 2.3.1: Illustration of the three epidemiological states in the SIR model, susceptible, infected and recovered. The arrows indicate possible transitions. The dashed line from susceptible to recovered state exemplifies how control measures such as vaccination can be incorporated in the model through introduction of additional transitions. Adapted from [115].

equations are used to describe the change in the number of individuals in the states S, I and R per time unit. Examples of equations describing these transitions are shown in Equation (2.3.1). β denotes a transmission coefficient, which combines both the contact rate and the transmission rate in case of contact between a susceptible and an infected individual. $1/\gamma$ describes the average infectious period. The larger value of γ, the shorter it takes for an individual to recover [149, 150].

$$\begin{aligned} \dot{S} &= -\beta SI \\ \dot{I} &= \beta SI - \gamma I \\ \dot{R} &= \gamma I \end{aligned} \tag{2.3.1}$$

In order to investigate the impact of control measures, compartmental models can be modified in two ways. One can either quantitatively change the parameters, such as transmissibility or length of infectious periods, or one can introduce new transition routes between states. The introduction of a vaccination program can, for instance, be modeled by introducing a transition directly between the S and R states, as vaccination provides immunity without first being infected [151].

Although compartmental models allow for a relatively accessible mathematical approach for modeling epidemics, one primary limitation is the assumption of homogeneity within each compartment [152]. Individual differences affecting factors such as dissemination and disease progression are not captured in the common transmission parameter β and recovery rate γ. Compartmental models can, however, be extended to take heterogeneity into account. For instance, age-related differences in hospital admission and treatment rates in a model describing the transmission of MRSA have been incorporated by subdividing each compartment into smaller age classes, letting parameters describing admission and discharge, contact and treatment rates vary with age [153].

Networks can also be incorporated into compartmental models by a degree block approximation [115]. Similar to the subdivision into age classes, this approach introduces a new level of compartments. Nodes are separated into blocks based on their degree. Figure 2.3.2 illustrates an example of a SIR model on a network with four different degrees. In total, this model has $4 \cdot 3$ different compartments. Nodes with equal degrees are assumed to be statistically equivalent [115].

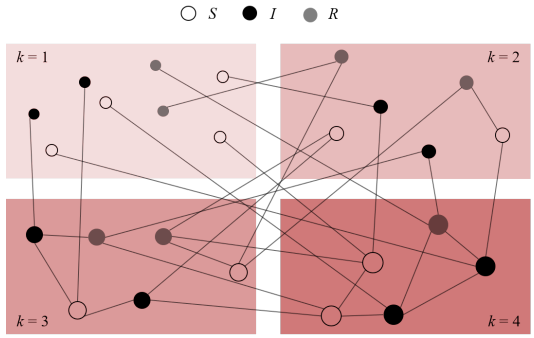

Figure 2.3.2: Illustration of a network incorporated into a compartmental SIR model. The nodes have degrees between $k = 1$ and $k = 4$, and can take one of the following epidemiological states: Susceptible (S), infected (I) or recovered (R). Adapted from [115].

The SIR model has been compared to the ideal gas and simple harmonic oscillator within the field of physics, as it represents an elegant and tractable simplification of the real world. Like the other mentioned simplifications, however, the SIR model is not fully capable of capturing the full complexity of real-world settings [151].

2.3.2 Agent-Based Modeling

Agent-based models are capable of capturing heterogeneity. Instead of treating all individuals within a (sub)compartment equally, each agent has its own set of characteristics and abilities governing its behavior [20]. For instance, an individual may be assigned with properties that are expected to affect its susceptibility of being infected or to have an effect on the course of disease. Such properties may be age, gender, occupation, degree of skepticism towards vaccines or immune response. Agent-based models are also called individual-based models (IBM) [154]. Compared to compartmental models, ABMs do not "average out" the system [22]. ABMs are particularly suitable for capturing emerging population phenomena based on differences on microscale level [119].

Agent-based modeling was originally developed for the field of economics, where economists formulated equations mirroring the behavior of different economic entities [155]. Within microeconomics, a field dealing with the behavior of individual consumers and producers, entities would typically be individuals, households and firms. Within macroeconomics, a field dealing with the total economic behavior of a nation, entities could be all households and all firms. With increasing computational abilities, economists were able to use ABM to combine simulations on micro- and macroeconomic scale [155].

One of the earliest agent-based models was presented by the American economist Thomas Schelling in 1971. In his article "Dynamic Models of Segregation", he presented several models investigating the dynamics of discriminatory individual behavior, conscious or unconscious [156]. Schelling modeled how black and white people moved in an abstract neighborhood, aiming to obtain neighbors with a higher proportion of individuals with the same skin color as themselves. These agents were able to change their position based on a set of rules depending on their environment. The simulations were

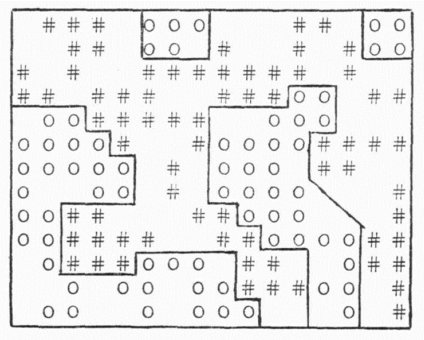

Figure 2.3.3: Emerging segregation patterns between black and white individuals in a two-dimensional abstract neighborhood. Obtained from [156].

performed by hand and eye with rather simple instruments such as dice and graph paper [156]. Schelling demonstrated how patterns of segregated communities emerged as a result of individual behavior, as shown in Figure 2.3.3.

Even though Schelling simulated emerging patterns based on agent behavior using dice and hand drawings, this is a rather tedious approach. Agent-based simulations are computationally demanding [157]. ABMs are time discrete, meaning that for each time point, the agents behave accordingly to their current set of characteristics and the environment [158]. Thus, large amounts of data are generated in the course of a simulation. If investigating the position of each cell in a bacterial population of 100 million cells, each time point would generate over 1 GB of raw data. It is worth mentioning that this number of cells is modest in bacterial context [22]. In addition to being computationally demanding, agent-based models require high-quality data to specify realistic agent behavior [155, 158]. Uncertainty in the input parameters contributes to variability in the model output. Besides, one often makes unverifiable assumptions about one or more key processes in the model. There is also a level of stochastic variability due to run-to-run fluctuations [158]. Overall, agent-based models are prone to uncertainty in input and model assumptions, in addition to variability brought about by stochasticity.

Despite computational demand, uncertainty and variability, agent-based models are useful for bridging the gap between microscale rules and emerging macroscale behavior [20, 22, 119]. ABMs do also allow for capturing interdependent behavior: In the short run, individuals tend to respond to their environment, whereas in the long run, the accumulation of individual choices tends to shape the environment [158]. Agents do also have the ability to remember and adapt to past experiences. Mathematical analysis does typically have a limited ability to capture such interactions [20].

A central part of agent-based models is the element of stochastic processes [159]. The agents' actions in each time step are based on certain probabilities. Hence, running a simulation several times with identical input parameters will produce different results. This variability can to some extent be balanced by using a sufficiently large population and running the simulation multiple times. The next section provides an introduction to some statistical tools useful for processing the output of such simulations.

2.4 Statistics and Statistical Distributions

As mentioned in the previous section, stochastic processes are central to agent-based modeling [159]. Identical input parameters can produce a myriad of different results. In order to collect, display and analyze these data, the field of statistics comes in handy [160]. The following section gives a brief introduction to relevant theory regarding descriptive statistics, including measures for location and variability and statistical inference, including discrete and continuous probability distributions. The theory presented in the following section is mainly based on Walpole [161].

2.4.1 Descriptive Statistics

An important part of statistics deals with organizing and summarizing the data for the purpose of interpretation and to facilitate further analysis: The descriptive statistics deals with characterizing the nature of the data [161, 162]. Two important measures in that regard is the location, often the center, and the spread of the data. One key measure of the center is the sample mean, which is the numerical average of n observations, see Equation (2.4.1).

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n} \tag{2.4.1}$$

Whereas the sample mean is susceptible to extreme values, the sample median is not affected by ouliers. This centrality measure describes the middle value when the data are sorted in ascending order, as presented in Equation (2.4.2).

$$\tilde{x} = \begin{cases} x_{(n+1)/2}, & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & \text{if } n \text{ is even} \end{cases} \tag{2.4.2}$$

A third location measure is the sample mode, i.e. the value appearing most frequently in the data set. There also exist several measures of variability in a sample. One example is the sample range, simply describing the difference between the largest and smallest value in the sample. The most widely used measure of spread, however, is the sample standard deviation, which is given by the positive square root of the sample variance, as presented in Equation (2.4.3). As the expression shows, the variance is based on the summarized distances of each sample measurement from the sample mean.

$$s = \sqrt{s^2} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} \tag{2.4.3}$$

To summarize, the above-mentioned expressions can be used to describe the nature of a sample, more specifically the location and variability, also known as the scale. These tools can, for instance, be used to describe the average and variability in the height of 100 individuals randomly drawn from the population of Trondheim. The field of inferential statistics is helpful in order to draw conclusions about the population of which the sample is drawn from, for example the height distribution of the entire population of Trondheim.

The following section gives a brief introduction to discrete and continuous probability distributions.

2.4.2 Probability Distributions

When conducting a statistical experiment, the set of all potential outcomes is referred to as the sample space. If the number of outcomes are finite and countable, such as the six possible outcomes of throwing a die when playing Ludo, the sample space is discrete. Returning to the example of the population height in Trondheim, on the other hand, one could measure a countless numbers of different heights, depending on the degree of accuracy. In such cases, where the sample space constitutes an infinite number of possible outcomes, the sample space is referred to as continuous. The probability distribution, giving the probability of a given outcome in the sample space, can hence also be discrete or continuous. A selection is presented below.

Discrete Probability Distributions

Many phenomena are binary, meaning they can take one out of two possible outcomes. For instance, a coin flip may result in heads or tails, Schrödinger's cat may be dead or alive[5], or a randomly chosen person may or may not be infected with a given disease. If an experiment consists of independent, repeated trials where the probability of success p remains constant, each trial of the experiment is referred to as a Bernoulli trial. The probability of failure is hence $q = 1 - p$. The probability of x successes in n Bernoulli trials is described by the binomial distribution, see Equation (2.4.4).

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, ..., n \tag{2.4.4}$$

In the special case where only a single Bernoulli trial is conducted, $n = 1$, the Binomial distribution is referred to as a Bernoulli distribution. In such cases, Equation (2.4.4) is simplified to the following expression [163]:

$$f(x; p) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases} \tag{2.4.5}$$

As an example, p could for instance denote the chance of becoming infected with a pathogen. If $p = 0.01$, the Bernoulli distribution would return 'yes' for infection in 1 in 100 cases on average. Figure 2.4.1 shows three Bernoulli distributions with probabilities of success equal to $p = 0.2$, $p = 0.4$ and $p = 0.7$, respectively.

In cases where a trial has more than two possible outcomes, the experiment is referred to as multinomial. The corresponding multinomial distribution gives the probabilities $p_1, p_2, ..., p_k$ for the k possible outcomes. In a complete deck of cards, the probabilities of drawing a card belonging to one of the four suits can be described as a multinomial experiment, given that the card is replaced. There exists several other discrete probability distributions, including the hypergeometric distribution, geometric distribution andPoisson distribution.

[5] Requires that an observer opens the box to check its state, off course.

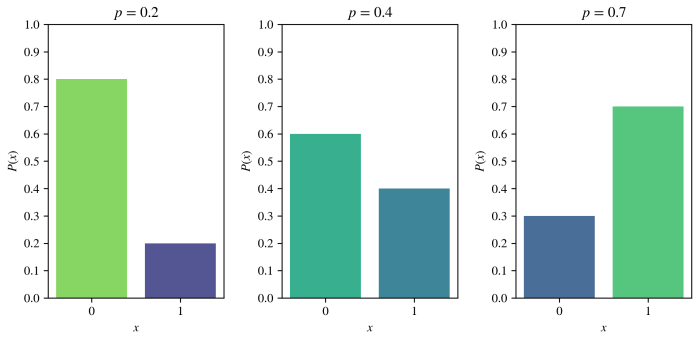

Figure 2.4.1: Three different Bernoulli distributions with probability of success ($x = 1$) equal to $p = 0.2$, $p = 0.4$ and $p = 0.7$, respectively. The probability of failure ($x = 0$) equals $q = 1 - p$.

Continuous Probability Distributions

The most central continuous probability distribution is the normal distribution. The bell-shaped graph of a normal distribution is fully described by the mean, μ, and variance, σ^2, as shown in Equation (2.4.6).

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad x \in (-\infty, \infty) \tag{2.4.6}$$

Figure 2.4.2 shows the probability density function for a variable $x \sim \mathcal{N}(\mu, \sigma^2)$, with a mean value $\mu = 2$ and variance $\sigma^2 = 1$. This figure illustrates several key aspects of the normal distribution. The curve is symmetric around the mean μ, and both the median and mode coincides with this vertical line. In both directions from the mean, the curve approaches the horizontal axis asymptotically. The total area under the curve still equals one. In addition, a variable drawn form a normal distribution can take both positive and negative values.

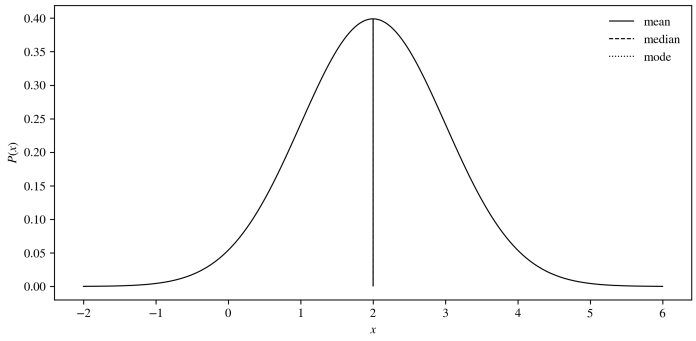

Figure 2.4.2: The probability density curve for a normal distributed variable x, with mean $\mu = 2$ and variance $\sigma^2 = 1$. As the coinciding lines illustrate, the mean, median and mode are equal.

In some cases, however, it is desirable to use a normal distribution to describe a variable which cannot take negative values [164]. For instance, given that a laboratory weight is correctly calibrated, the measured weight of a substance should only yield non-negative values. In order to preserve important characteristics of the normal distribution, while avoiding negative or another range of extreme values, one could use a truncated normal distribution [164]. This distribution is derived from the normal distribution, and is bounded by a lower limit a and/or a upper limit b. The left panel of Figure 2.4.3 shows the truncated form of the normal distribution presented above, where the lower limit a is set to zero. This distribution has no upper limit, meaning that $b \to \infty$.

Another distribution ensuring positive variable values is the folded normal distribution [165]. This distribution is based on the absolute value of the normal distribution, as illustrated in the right panel of Figure 2.4.3: The dotted line shows the probability density function of the normal distribution folded around the vertical line in $x = 0$. The blue curve makes up the sum of the normal probability density function of both $-x$ and x, [165].

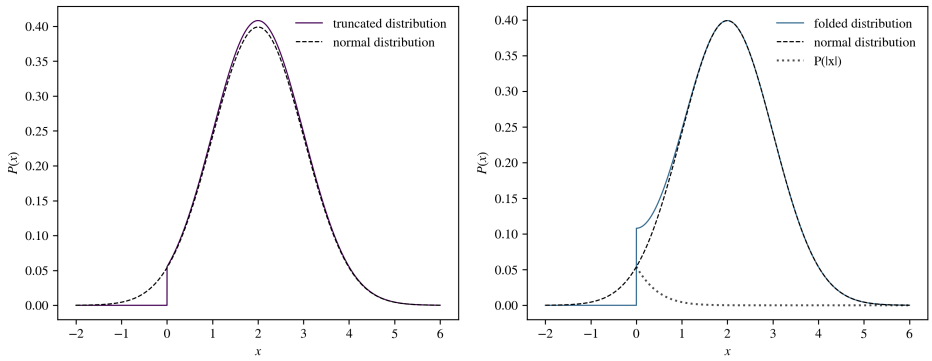

Figure 2.4.3: Two distributions derived from the normal distribution: The truncated and the folded normal distribution. The left panel shows a purple-colored truncated distribution with lower limit $a = 0$. The right panel shows a blue-colored folded distribution. The dotted line shows the mirror image of the normal distribution taking negative values. In both panels, the density function of the normal distribution with mean $\mu = 2$ and variance $\sigma^2 = 1$, is plotted in a dashed curve.

The normal distribution and its derivatives, however, are not suitable for describing all variables. Some data are best described using an asymmetric distribution, being skewed to the left or to the right. One of these distributions having a positive skew is the lognormal distribution. This distribution describes variables where the natural logarithm is normally distributed. The probability density function is presented in Equation (2.4.7).

$$f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2.4.7}$$

As the probability density function shows, a variable drawn from a lognormal distribution will always yield non-negative values. The left panel of Figure 2.4.4 shows a lognormal distribution with $\mu = 2$ and $\sigma^2 = 1$. As the figure shows, the mean, median and mode

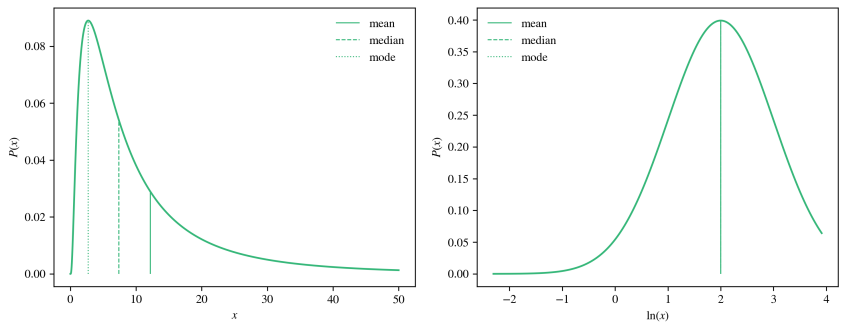

Figure 2.4.4: The left panel shows a lognormal probability density function with a mean $\mu = 2$ and variability $\sigma^2 = 1$. The mean, median and mode are depicted as vertical solid, dashed and dotted lines, respectively. The right panel shows a normal distribution natural logarithm of x.

values do not coincide as they do in a normal distribution: The mean is greater than the median which is greater than the mode. The right panel shows the the normal distribution of the natural logarithm of the corresponding x values.

2.4.3 Correlation Analysis and Mean Squared Error

So far, this section has discussed descriptive statistic tools and probability distributions of the one variable x. In cases where the dependence between two variables x and y comes of interest, correlation analysis is useful for describing the relationship between them. The Pearson correlation coefficient, see Equation (2.4.8), is commonly used as a measure of linear correlation between two variables. The coefficient takes values in the range $[-1, 1]$, where $\rho = -1$ indicates a perfect negative linear correlation and $\rho = 1$ indicates a perfect positive linear correlation. If the coefficient equals zero, there is an absence of linear relationship between the two variables [166], although other non-linear relationships may occur.

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.4.8}$$

Given two sets of values $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$, one can use the mean squared error (MSE) to determine the overall deviation between them. The values of x and y could for instance be estimated and actual values of a given variable. As the name implies, the mean squared error is calculated by determining the mean value of all squared errors $e_i = x_i - y_i$ [167]. The expression is given in Equation (2.4.9).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{2.4.9}$$

Chapter 3

Methods and Software

The overall aim of this thesis is to build an agent-based model for exploring the mechanisms governing the spread of COVID-19 on a temporal contact network. This chapter gives a detailed introduction to the underlying data processing and algorithms. After a brief presentation of the relevant software, the first part of the chapter describes how the model simulates a time-evolving contact network among staff members and patients in a virtual hospital. The second part of the chapter gives an introduction to the epidemiological part of the model, including a description of the different epidemiological states and the transitions between them. This part also includes the implementation of several control measures. The chapter is concluded by a summary of the values constituting the baseline model parameter set.

3.1 Software

3.1.1 Python

Most of the data analysis and modeling work in this thesis is performed in Python [168], version 3.7.3. Python is an open source, high-level general-purpose programming language, available for several operating systems. In addition to a library of standard modules, several additional modules can be imported to handle a wide range of operations. Imported modules from the Python Standard Library, in addition to installed modules, are presented in Table B.1.1. Python can be downloaded from `www.python.org`.

3.1.2 Cytoscape

All network visualizations are produced using Cytoscape [169], version 3.8.0. Cytoscape is a software platform for analysis and visualization of complex networks, originally developed for biological research. Additional features, such as random network generation or additional layout algorithms, can be downloaded as Apps. Cytoscape is open source and can be downloaded from `www.cytoscape.org`.

3.2 Generating a Temporal Contact Network

A key sub-goal is to develop a temporal contact network mirroring the inter-individual contacts in a geriatric hospital ward. More specifically, the aim is not to construct *one* single network, but to derive a set of rules governing the evolution of a contact network building simultaneously as a simulation runs. In this way, the network rules could be set to adapt to changes in the system, for example to take staff shifts into account, replace regular staff members with temporary staff in case of quarantine or sick leave, or prolonging patient hospitalizations in case of illness.

The contact network should also be able to capture temporal and individual variations. In the model, this is done by extracting patterns from an empirical contact network and translate the information into model parameters and agent attributes. Figure 3.2.1 illustrates the two main steps in this process. As the empirical contact network contains large amounts of information, the step of pattern extraction requires that some network properties are selected at the expense of others. Focusing on a few patterns is also useful in regards to the simulation step, where a trade-off between complexity and computational running time must be made.

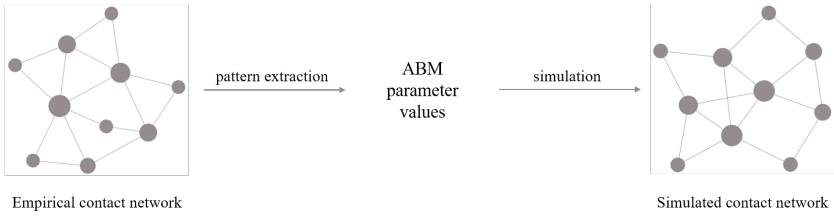

Figure 3.2.1: Illustration of the process of simulating a temporal network capturing inter-individual contacts based on an empirical contact network: The extracted patterns are translated into parameter values incorporated into an agent-based model. The simulated contact network is then built stepwise and simultaneously as the model simulation proceeds.

The empirical contact network presented by Vanhems *et al.* [16] serves as a basis for the inter-individual contact network in this model. See Section 2.2.3 for a detailed introduction. This empirical network contains data on over $14\,000$ registered contacts between 75 individuals in a short stay geriatric ward of a hospital in Lyon, France. As earlier described, the contacts in this network refer to detected close-range interactions, where wearable sensors carried by the participants signalize proximity if the individuals are located face-to-face within 1.5 meters from each other. The network has a high temporal resolution and provides potential real-life transmission routes for the spread of a pathogen. However, the short time span of only four days makes it challenging to model realistic epidemiological spreading phenomena on the network directly. In order to investigate pathogen transmission dynamics over a longer period of time, a synthetic network based on the empirical one is simulated to recreate the most central network properties.

The empirical network has a temporal resolution of 20 seconds. Proceeding with this resolution would generate 4320 time steps in the course of 24 simulation hours and would be computationally demanding when running a simulation over several days or weeks.

On the opposite side, one could aggregate the network to a static network. This solution greatly reduces running complexity but results in loss of all temporal information. The question of choice of temporal resolution does not only concern running time, but also the statistical foundation in the process of extracting patterns: Examining the contact data using time slots of 20 seconds would yield considerably fewer data points per time slot than if choosing periods of minutes, hours or days. In the original article, the network data is aggregated on an hourly basis. This resolution seems reasonable for ensuring an adequate amount of data per time slot and would also reduce the complexity in terms of running time.

The following sections give an introduction to how the temporal contact network is built. As the rules governing the construction of this network are based on patterns extracted from an empirical network, some of the material presented in these sections comprises results from the processing and analysis of the given empirical data. They are included in this chapter for the sake of coherence and because the choice of model behavior and parameters are derived directly from these intermediate results. The choices of which contact properties to focus on are based on the key findings presented by Vanhems *et al*. They involve a repeating daily contact pattern and large variations in the observed number of contacts both between and within the four ward roles.

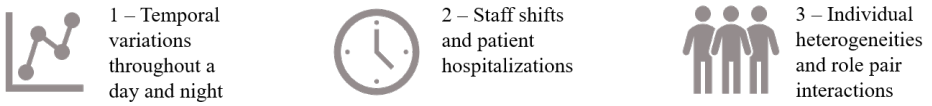

Figure 3.2.2: The pattern extraction from the empirical contact network involves three main steps.

The section is divided into four main parts, see the illustration of the three first in Figure 3.2.2: The first part describes how the temporal contact variations throughout a day and night are replicated, the second part discusses how staff shifts and patient hospitalizations are determined, and the third part presents how individual heterogeneities in terms of relative contact rates and interactions between the role pairs are identified. Lastly, the fourth part summarizes the section by showing how the contact network is implemented into the model.

3.2.1 Introducing Temporal Variations Throughout a Day and Night

In a most simplistic approach, one could simulate the temporal contact network based on the average number of registered contacts per hour. Given $14\,037$ detected contacts in the course of the 97 hours, the average number of contacts per hour is approximately 145. Panel A of Figure 3.2.3 shows a network consisting of 75 nodes and all possible links between them. Each node represents an individual and each link a possible inter-individual interaction. Panel B through D shows three realizations where 145 contacts are randomly drawn among the links in panel A. The node size increases with the node degree and the link thickness increases with the number of contacts in the given hour.

Given a simulation over $97\,\mathrm{hours}$, this temporal network would generate approximately the same number of unique contacts as the empirical network, but without taking

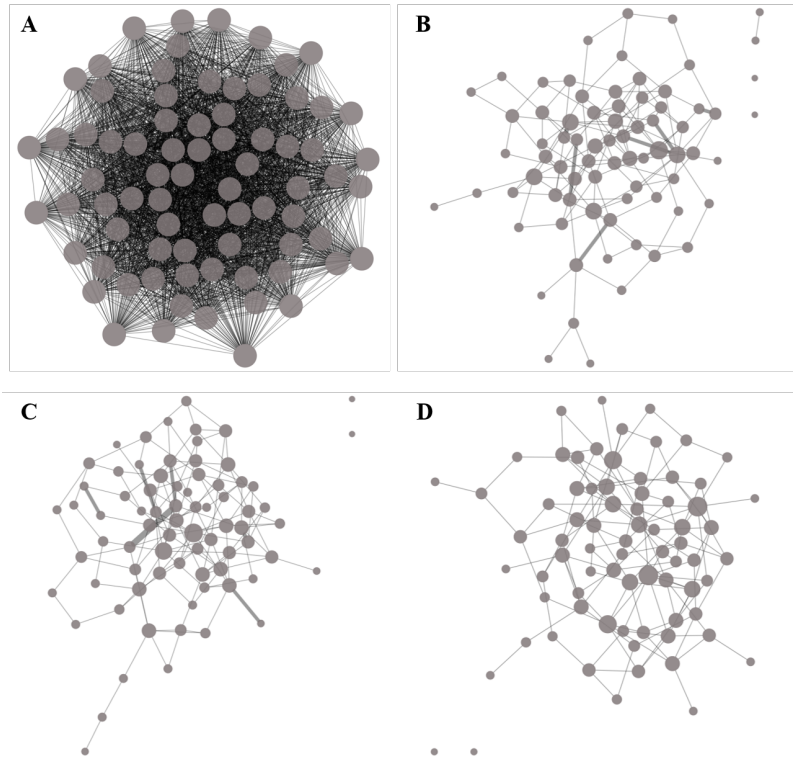

Figure 3.2.3: Panel A shows a network consisting of 75 nodes and all 2775 possible links between them. Panel B through D shows three time step examples, where 145 contacts have been randomly drawn among the possible node pairs. The node size increases with the node degree and the link thickness increases with the number of contacts in the given hour.

any temporal fluctuations or variations between roles or individuals into account. In addition, this network assumes that all individuals are present in the ward at any hour in the simulation, and is hence neglecting both patient admissions and discharges, staff shifts and potential sick leaves or quarantines.

Daily Contact Number Variations

Vanhems *et al.* observed large variations in the number of contacts registered throughout a day and night, but a similar evolution from day to day [16]. The number of registered contacts for each hour in the study period is presented in Appendix B.2.3. A visual representation of these data is presented in Figure 3.2.4: The upper panel presents the total number of contacts registered for each of the 97 hours in the study period, whereas the lower panel shows the same data split into each of the five days Monday through Friday. Corresponding plots where the contacts are filtered based on role or role pairs are presented in Figures B.2.2 to B.2.15. These figures correspond to Figure 2 in the original article and are generated from the available raw data.

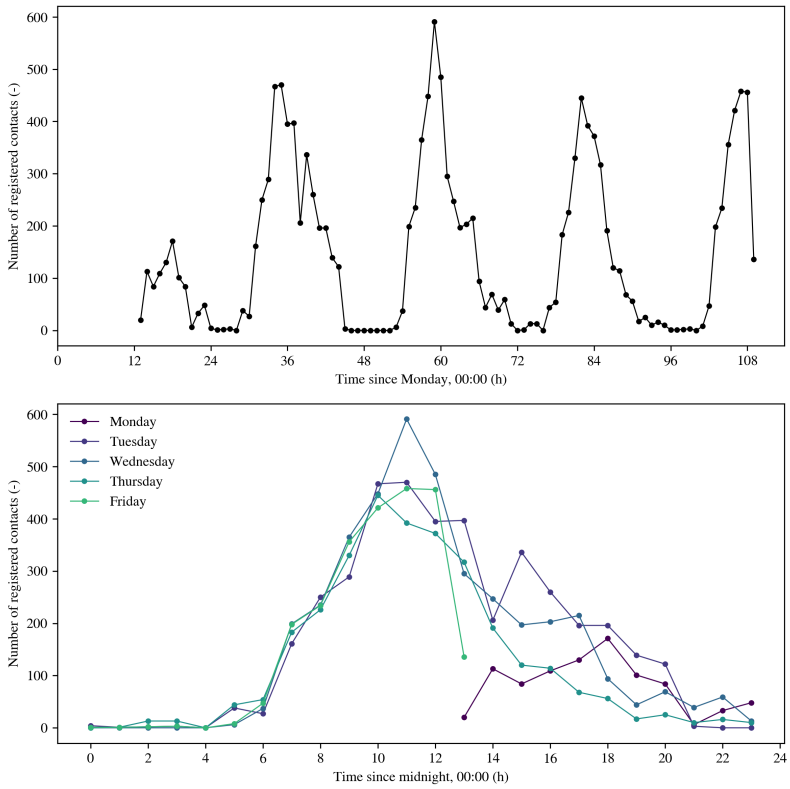

Figure 3.2.4: Upper panel: Number of registered contacts per hour, starting from Monday 13:00. Lower panel: Number of registered contacts per hour for each day Monday through Friday.

As the figure shows, the hourly contact number is relatively similar from day to day. Thus, instead of using an average of 145 contacts per hour, the hourly contact numbers could be based on the observed daily temporal contact pattern, for instance by drawing the number of contacts for each hour from distributions based on the sample mean and standard deviation. This approach would for instance yield few contacts throughout the night and several hundred contacts in the hours around noon. Due to the size of the data set, with four (or five) data points per hour, the process of determining a suitable distribution becomes rather arbitrary. Assuming that the hourly number of contacts could vary in both directions from a mean value in case of more quiet or more busy days, a normal distribution could be appropriate for describing this parameter. Thus, for each of the 24 hours, the number of contacts is drawn from a normal distribution with mean μ equal to the sample mean $\bar{x}$, and variance σ^2 equal to the sample variance s^2 for that given hour. For the purpose of allowing for up-scaling of number of wards and/or beds, the number of contacts per hour is calculated with the basis of contacts per bed. Figure 3.2.5 presents the average value of contacts per bed for each hour, with error bars corresponding to the sample standard deviation.

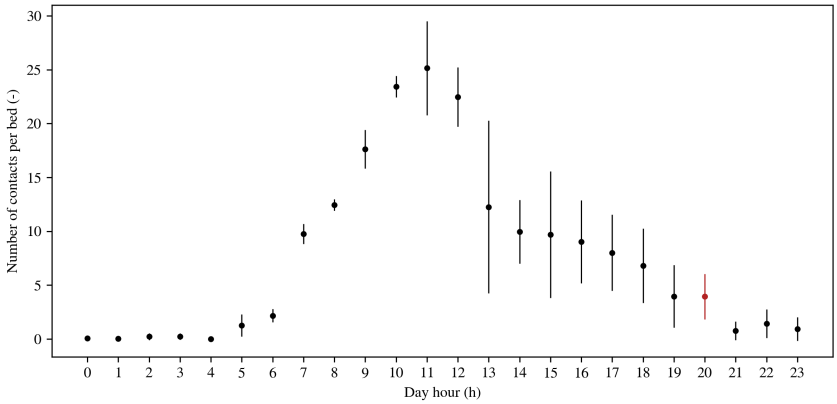

Figure 3.2.5: Number of contacts per bed for each hour throughout the day and night. Each point represents the sample mean calculated from the empirical data, whereas the error bars represent the sample standard deviation. The data for hour 20:00 is marked in red and is used as an example in the following.

The red colored data point in Figure 3.2.5 corresponds to the sample mean and standard deviation for the contacts registered in the hour starting from 20:00 in the empirical network. The corresponding normal distribution, $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 3.9474$ and $\sigma = 2.1110$, is plotted in Figure 3.2.6. The mean value is plotted using a black, dashed line, whereas the interval spanning $\mu \pm \sigma$ is marked with dotted lines. The red colored histogram presents $10\,000$ values drawn from the normal distribution.

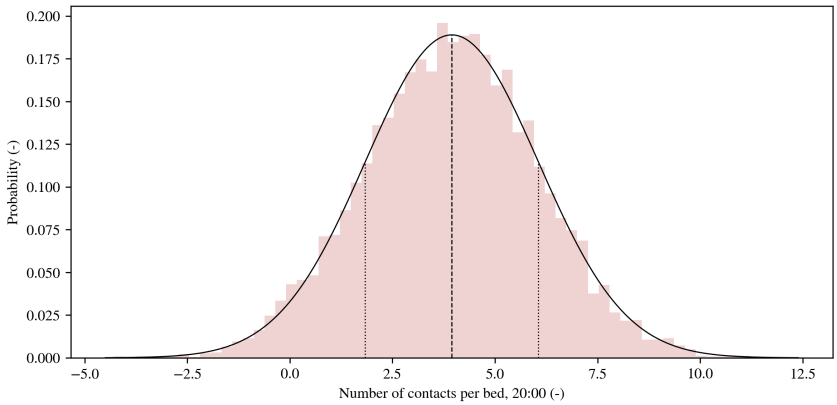

Figure 3.2.6: The solid curve represents the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 3.9474$ and $\sigma = 2.1110$ from which the number of contacts per bed for the hour 20:00 is drawn. The mean value μ and the interval spanning $\mu \pm \sigma$ are plotted using black dashed and dotted lines, respectively. The red colored histogram represents $10\,000$ values drawn from the distribution.

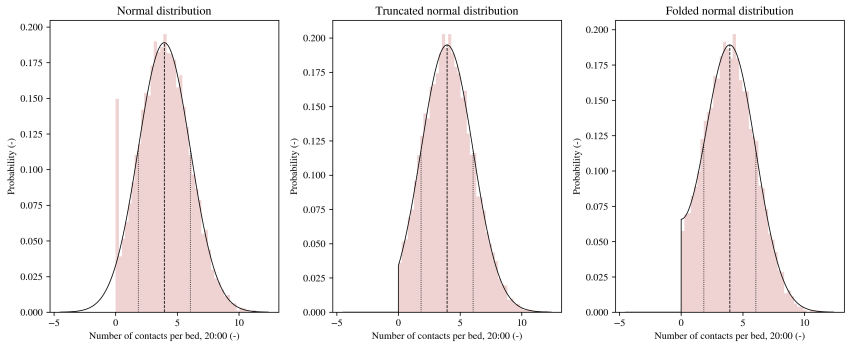

Figure 3.2.7: Three different distributions (solid lines) illustrating the number of contacts per bed for the hour 20:00. The dashed lines refers to the average value, whereas the dotted lines encloses $\mu \pm \sigma$. Normal, truncated normal and folded normal distribution.

As Figure 3.2.6 shows, a considerable proportion of the drawn contact numbers are negative. This is also the case for several of the other time points. A relatively simple solution is to set all negative values to zero. Other alternatives include using a truncated normal distribution with lower bound set to zero or a folded normal distribution. The probability density function and a corresponding histogram presenting $10\,000$ drawn values from each other three distributions are presented in Figure 3.2.7. As for the distribution presented in Figure 3.2.6, the dashed and dotted lines present the mean μ and the interval spanning $\mu \pm \sigma$, respectively. In order to determine which of the three alternatives to use, the mean of $1\,000\,000$ drawn values for each of the 24 hours were calculated and compared to the empirical means. The difference between the simulated and empirical means were calculated using the mean squared error (MSE), see details in Appendix B.2.2. The lowest MSE was found when drawing the contact numbers from the normal distribution with negative values set to zero. Therefore, this distribution is used to determine the hourly contact numbers.

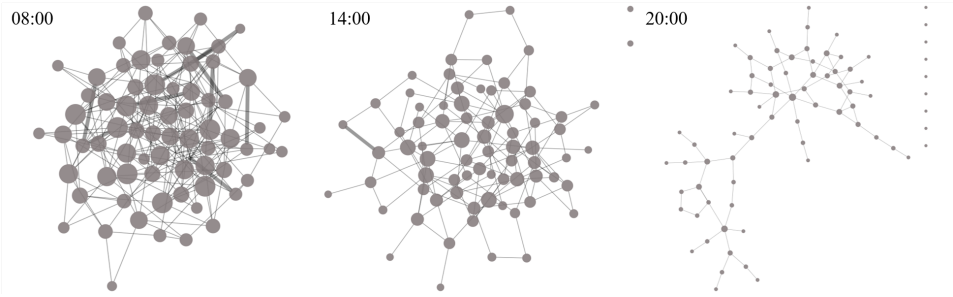

Figure 3.2.8: The panels show three different randomly generated networks representing contact pattern in the hours 08:00, 14:00 and 20:00. The numbers of contacts are drawn from normal distributions based on the registered number of contacts for the given hours. The node size increases with the node degree and the link thickness increases with the number of contacts in the given hour.

By drawing the hourly number of contacts from a normal distributions with parameters derived from the empirical contact network, the randomly generated networks presented in Figure 3.2.3 with an average of 145 contacts per hour can be updated to mirror the temporal evolution throughout a day and night. Figure 3.2.8 shows three different snippets from a temporal network taking the discussed temporal variations into account. The node pairs are still chosen randomly. The implementation of these temporal variations result in a time-evolving network where the number of hourly contacts pulse in pace with the pattern observed in the empirical network. Still, however, the network assumes all individuals to be present in the ward and available for contacts at all times. In a real life scenario, however, it is natural to assume that the staff members are present in the ward during shifts and absent when off work, and that patients only are present in the ward in the time span between admission and discharge. The following section discusses how staff shifts and patient admissions and discharges are estimated based on the data in the empirical network.

3.2.2 Estimating Staff Shifts and Patient Hospitalization

As presented in Section 2.2.3, the original article provides some information about the duration of shifts and how many individuals are required to cover them. However, initial analysis of the registered contacts suggests that there exists some deviation between the given information and the recorded data. In addition, no information about the administrative staff shifts or the duration of patient hospitalizations are provided. Since a recorded contact necessarily implies that the given individual is present in the ward, the staff shifts and patient hospitalization can be estimated from the detected contacts.

The original data set consists of five different files storing the network information for each of the days Monday through Friday, see details in Section 2.2.3. After merging the information into one file, the data contained in the network nodes and links were extracted and stored in a data structure, see a snippet in Table 3.2.1. The id and role were retrieved directly from the node, whereas the total number of contacts, the cumulative duration and the listed intervals of contacts were found by combining the information contained in each of the links connecting the node to its neighbors. Figure 3.2.9 was made by stripping the

Table 3.2.1: The data structure contains the summarized contact network information for each of the 75 individuals; the RFID id number and role, as well as the total number of contacts, the cumulative duration and a list containing all contact time intervals of registered contacts.

	id	**role**	**contacts**	**duration**	**listed**
0	1152	MED	79	4720	[(8300, 8320), (8480, 8500), ...]
1	1157	MED	1195	56 960	[(8300, 8320), (8480, 8500), ...]
2	1159	MED	710	31 540	[(13 840, 13 860), (14 080, 14 100), ...]
3	1164	NUR	1053	42 140	[(71 300, 71 320), (75 760, 75 780), ...]
...	...	...	...	...	...
73	1784	PAT	35	1220	[(330 700, 330 740), (334 760, 334 780), ...]
74	1469	PAT	87	3280	[(255 620, 255 660), (255 700, 255 720), ...]

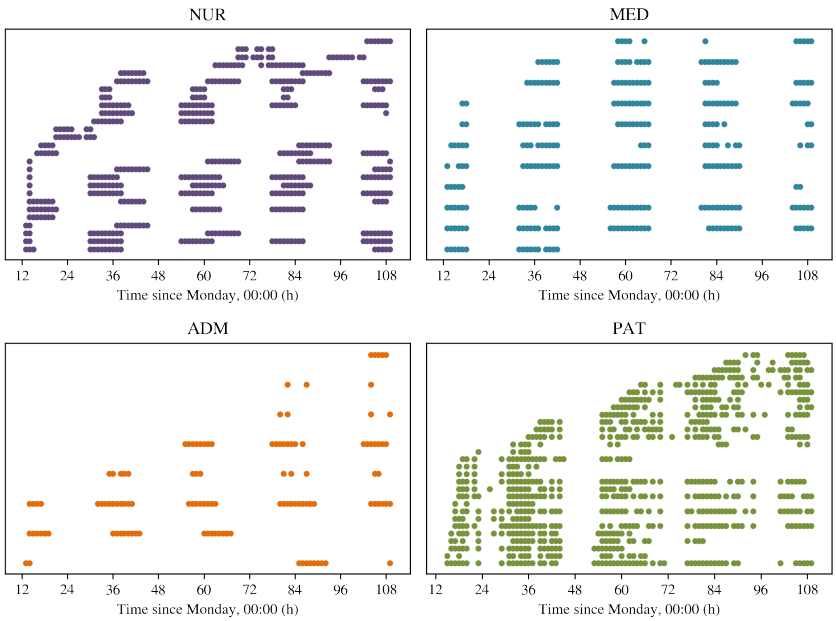

Figure 3.2.9: The four panels present hours of detected contacts for individuals belonging to each of the four roles NUR, MED, ADM and PAT. The x-axis corresponds to the time in the study period, counting from Monday at 00:00, whereas each row corresponds to one unique individual.

listed contact intervals down to a set of unique hours of detected contacts: The colored points in each of the four panels, each representing one role, denote hours of which one or more contacts has been registered. Each row corresponds to one individual belonging to the given role.

There are two main challenges related to the process of estimating shifts and hospitalizations from the given contact data. Firstly, although a registered contact implies presence in the ward, the lack thereof does not necessarily imply that the individual is absent. For instance, a consecutive sequence of hours with registered contacts may contain enclosed vacancies, as observed several places in Figure 3.2.9 and illustrated by arrow number one in the clipart presented in Figure 3.2.10. This figure presents hours of registered contact for three different individuals belonging to the MED role at a given day. In the hour where no contacts have been registered, the individual could for instance have had contact with another individual *not* wearing an RFID sensor ($8\,\%$ of staff members and $6\,\%$ of patients). In case of silent hours enclosed by hours of contact supporting the assumption of presence, one could, within reasonable limits, assume that the individual is present in the ward. It is more challenging to determine whether leading and trailing vacancies, as illustrated by arrow number two in Figure 3.2.10, are a result of the individual not being present in the ward or not. Examining the hours of contact registered for the patients in the lower right panel of Figure 3.2.9, for instance: Is the first hour of contact equivalent with the patient admission or could the patient have been hospitalized from before the study period started?

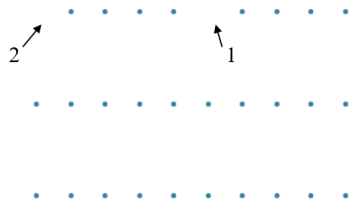

Figure 3.2.10: A clipart showing the registered hours of contact for three individuals belonging to the MED role on one of the five week days. Arrow number one points at an enclosed vacancy where no contacts have been registered. The second arrow points at a leading vacancy.

Secondly, whereas the individuals of the MED and seemingly the ADM roles only work during day time, the NUR individuals work in four different, partially overlapping shifts. In addition, there seems to be large variations in the shift patterns among individuals in all three staff roles. In order to reduce the stochasticity in the model, it is desirable to derive fixed rules governing the shift schedule for the three roles; the duration and staffing per shift. The three following sections present algorithms and analysis behind the derivation of these rules and describe the reasoning behind choices of simplifications. The fourth section deals with patient admissions and discharges.

Nurses or Nurses' Aides

According to Vanhems *et al.*, the nurses and nurses' aids work in four different shifts, see an overview of the shift durations and staffing in Table 2.2.1. As this table shows, the morning and afternoon shifts overlap at 13:30. Due to the temporal resolution of one hour, both shifts are extended in order to capture the whole hour: The model morning shift ends at 14:00 and the afternoon shift begins at 13.00. In addition to the nurses and nurses' aids, the NUR role includes several staff members working on-demand or without a fixed schedule. As they are included as NUR individuals in the empirical network and contribute to the overall contacts, the model attempts to recreate the observed patterns of presence among all individuals registered as NURs, not only the nurses and nurses' aids shifts provided in the staffing information.

The hours of detected contacts presented in Figure 3.2.11 serve as basis in this analysis. The first aim is to identify all independent shifts, i.e. tell hours of absence apart from hours of presence without contacts. The second aim is to match the identified shifts up to the provided shift hours by comparing the relative overlaps. The NUR shifts are established by the following algorithm. The contact data for the individual with RIFD id number 1485 is used as an example.

1. The **starting time** of listed contact intervals presented in Table 3.2.1 are extracted. [202 180, 202 180, 202 240, 202 240, 202 300, 202 300, ..., 323 460]

2. Each list entry is converted to its **corresponding hour**, starting from Monday 00:00. [69, 69, 69, 69, 69, 69, ..., 102]

3. All **unique hours** are extracted, resulting in a set of all hours of detected contacts. [69, 70, 71, 73, 74, 75, 77, 78, 93, 94, 95, 96, 97, 98, 99, 101, 102]
4. The set of unique hours is divided into groups of **consecutive hours**. [[69, 70, 71], [73, 74, 75], [77, 78], [93, 94, 95, 96, 97, 98, 99], [101, 102]]
5. Groups where the gap between them is smaller than a certain **cutoff value** are spliced. The cutoff value is set to 5 hours to avoid having a shift spanning 23 hours. [[69, 70, 71, 73, 74, 75, 77, 78], [93, 94, 95, 96, 97, 98, 99, 101, 102]]
6. The hours in each spliced list are assumed to **belong to the same shift**. The enclosed vacant hours are filled in. [[69, 70, 71, 72, 73, 74, 75, 76, 77, 78], [93, 94, 95, 96, 97, 98, 99, 100, 101, 102]]
7. Each spliced list of hour is **checked up against all possible shifts** in order to find the greatest overlap.
   - The list of hours $[69, 78]$ has a $91\,\%$ overlap with the 3^{rd} night shift $[68, 78]$.
   - The list of hours $[93, 102]$ has a $91\,\%$ overlap with the 4^{th} night shift $[92, 102]$.

By following this algorithm, NUR 1485 is found to have worked two shifts in total, the third and the fourth night shift, respectively. In cases where several shifts have an equally high overlap score, the shift type is assigned according the the following hierarchy: Day $>$ night $>$ afternoon $>$ morning. Such cases were most frequently occurring in the very beginning and end of the study period, where data points were few. In addition, an exception

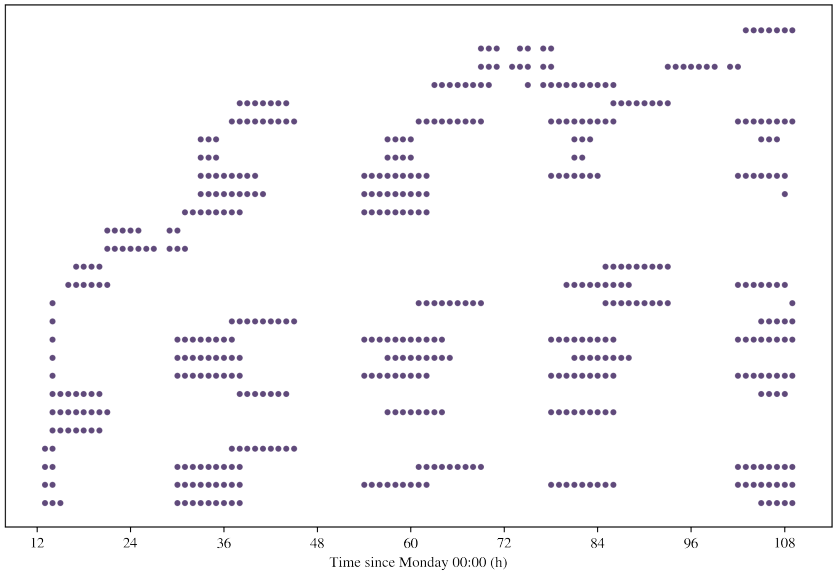

Figure 3.2.11: Each point marks an hours where one or more contacts has been registered for the individuals belonging to the NUR role. The x-axis corresponds to the time in the study period, counting from Monday at 00:00, whereas each row corresponds to one unique individual.

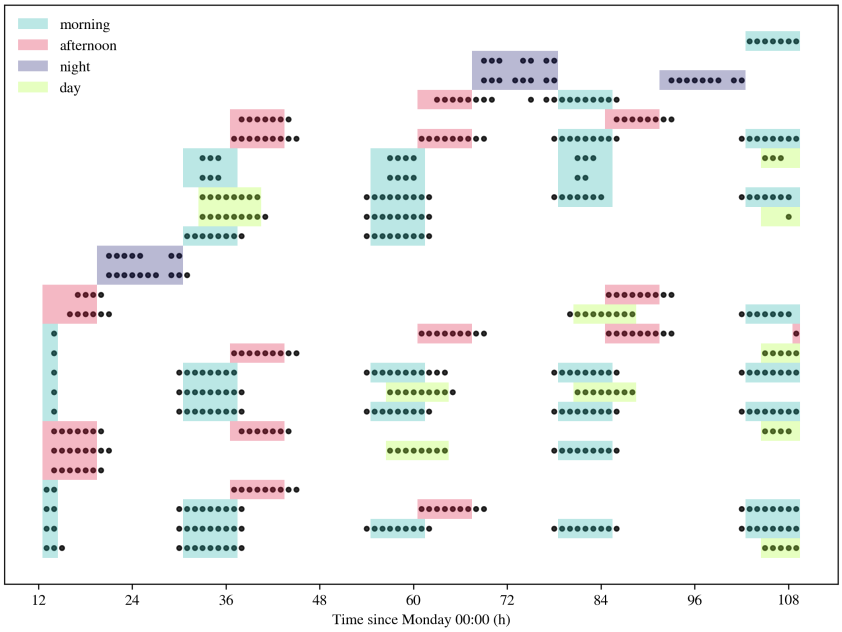

Figure 3.2.12: The figure shows hours of detected contacts (black dots) and the corresponding shifts (colored patches) having the highest overlap for all individuals belonging to the NUR category. Each individual has its own row, and are sorted based on their order of first detected contact. The patches are color coded after the type of shifts.

from the provided shift hours were made to capture five single hours of contact detected at hour 14: The very first morning shift was extended to capture these data points, based on the assumption that both the day and afternoon shifts probably would have resulted in contacts in the hours beyond 14:00.

As Figure 3.2.12 shows, the predicted shifts cover parts of the registered contact hours quite nicely. The number of staff members covering each shift tends to vary from day to day, as exemplified by the night shifts which are covered by two, zero, two and one nurse in the course of the week. Again, the lack of data may be a result from the participation rate not being $100\,\%$. As the figure shows, the predicted shifts fail to capture some of the hours of detected contact and covers other hours where no contacts are detected. A Venn diagram representing the overlap between hours of detected contact (contact hours) and predicted shift hours (shift hours) is presented in Figure 3.2.13. In total, 436 out of 504 shift hours ($87\,\%$) are supported by detected contacts, whereas 436 out of 519 contact hours ($84\,\%$) are captured by the estimated shifts. It is worth mentioning that the 68 shift hours not supported by contact hours include the enclosed vacancy hours. A considerable proportion of the 83 hours of contact *not* captured by the shifts seems to come from leading and/or trailing hours of a shift. If an individual has a shift from 09:00 to 17:00, but has detected contacts at 08:59 and 17:01, for instance, both hours from 08:00 to 09:00 and 17:00 to 18:00 are registered.

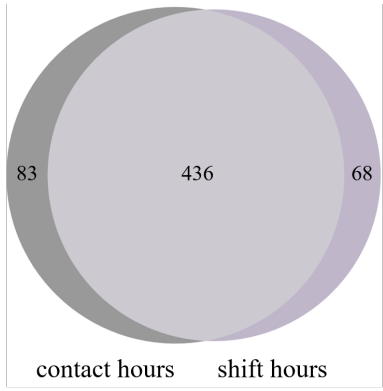

Figure 3.2.13: The Venn diagram represents the overlap between hours containing registered contacts ('contact hours') and hours covered by the estimated shifts ('shift hours') for NURs.

Since the estimated shifts appear to cover the observed contact hours relatively well, duration of the shifts in the model are directly retrieved from the provided shift information, with the exception of the discussed adjustment of the morning and afternoon shifts due to the temporal resolution. The number of staff members required to cover each shift is adjusted accordingly to the observations made in Figure 3.2.12. The day, afternoon and night shifts are covered by 2, 5 and 2 NUR individuals respectively, as provided in the original article, whereas the staffing of the morning shift is increased from 5 to 9 individuals. This information is summarized in Table 3.2.4.

When a new shift is to begin, the model can only choose from NUR agents that are available: An agent cannot begin a new shift it is already preoccupied with another, or if it is subject to restrictions such as quarantine or isolation. The shift can be assigned by randomly drawing among the available agents. In the model, this shift assignment approach is called 'random'. Two additional elements can included in order to prioritize the shift assignment:

1. Percentage of employment: The agents can be assigned with percentages of employment based on the observed distribution of number of shifts in the study period, see the left panel of Figure 3.2.14. After determining the pool of available NURs, the required number of agents are drawn with probabilities depending on their relative working percentages. Thus, a person having a work percentage of 1.0 has a greater chance of being chosen for a shift than a person having a percentage of 0.4. This shift assignment is referred to as 'percentage' in the model.

2. Resting hours after shift: The model can take resting hours after an ended shift into account. As the right panel of Figure 3.2.14 shows, the number of hours between two consecutive shifts may vary from 12 to almost 70 hours. Then median number of resting hours, 18, is used as baseline value in the model. When taking resting hours into account, the NUR agents with the lowest number of remaining resting hours are prioritized.

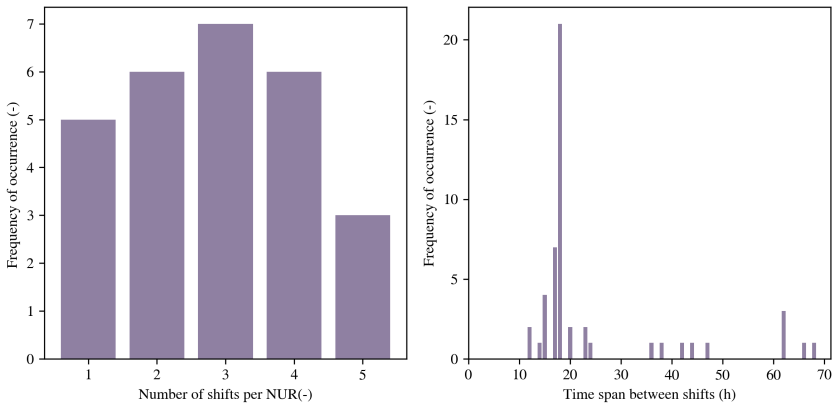

Figure 3.2.14: The left panel shows a bar plot representing the estimated number of shifts in the course of the study period for each individual belonging NUR role. The bar plot in the right panel represents the number of hours between the estimated shifts.

Medical Doctors

The process of replicating the shift pattern for the 11 medical doctors is somewhat easier than replicating the shift pattern for the nurses since the individuals belonging to the MED role only work during daytime. According to the original article, there were two physicians and two interns present between 08:00 and 17:00 every day. Figure 3.2.15 shows all hours of registered contacts for each of the medical doctors, indicating that more than four individuals from this category were present each day. In addition, several of the individuals have contact hours indicating presence exceeding the given nine-hour shift. The number of detected MED contacts for each hour is plotted as bars in Figure 3.2.16. The

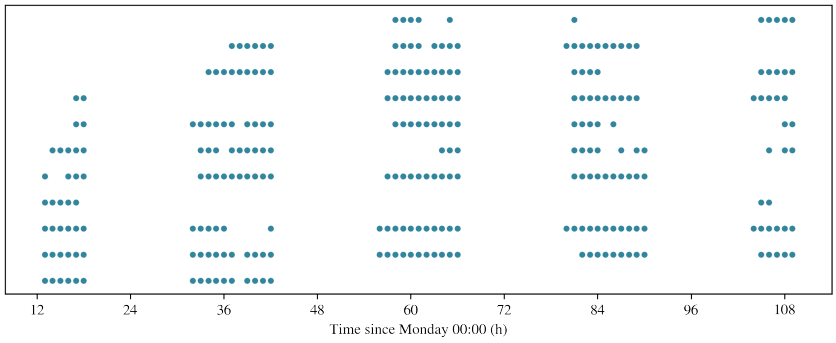

Figure 3.2.15: Each point marks an hours where one or more contacts has been registered for the individuals belonging to the MED role. The x-axis corresponds to the time in the study period, counting from Monday at 00:00, whereas each row corresponds to one unique individual.

given working hours from 08:00 to 17:00 are indicated with lighter blue shades, cropped at the beginning and end of the study period. As the this figure shows, a relatively large proportion of the contacts are registered after the original working hours. The relative fraction of contacts registered for each hour is presented in the left panel of Figure 3.2.17. These numbers do for instance show that a very low proportion of the contacts are made in the hour from 08:00 to 09:00.

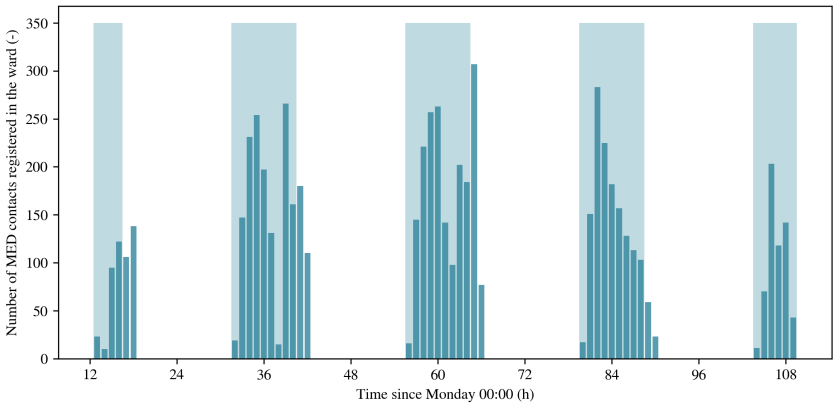

Figure 3.2.16: The bars represent the number of contacts registered for MEDs for each hour. The lighter blue shaded areas represent the given working hours, 08:00 to 17:00.

The data presented in the two other panels of Figure 3.2.17 is based on the assumption that one or more hours of registered contacts between 08:00 and 19:00 corresponds to a

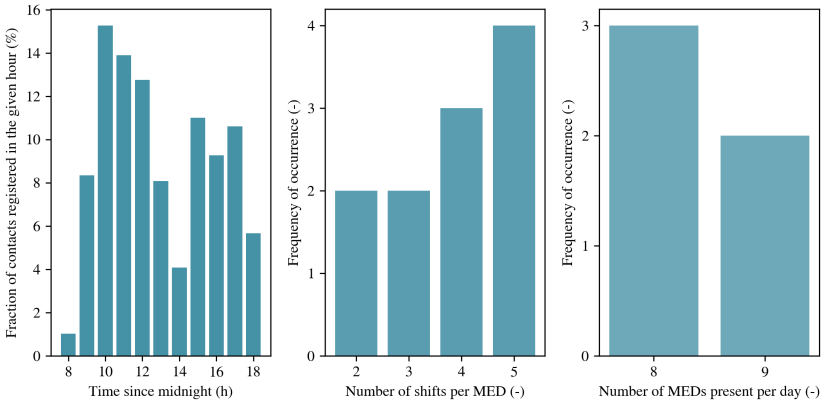

Figure 3.2.17: Left panel: The fraction of contacts registered for each hour summarized over the entire study period. Middle panel: Distribution showing how many shifts the MED individuals have during the study period. Right panel: Distribution showing the number of registered MED shifts per day in the ward.

shift: The middle panel of Figure 3.2.17 shows the distribution of how many shifts each individual worked during the study period. The third panel shows the number of MED shifts that were registered per day. Based on the given contact information and the above-mentioned assumption, this number varies between 8 and 9 individuals per day.

Based on the observed hours of contact in the above-mentioned assumptions, the following rules are set for the model MED shifts: The working hours are changed from 08:00 - 17:00 to 09:00 - 19:00 to cover a larger proportion of the observed contacts. In addition, the number of agents set to cover each shift is changed from 4 to 8, see Table 3.2.4. Also for this role, the user can set the shift assignment to be random or prioritized based on given percentages of employment.

Administrative Staff

Unlike the two other staff categories, no information about the administrative staff shifts are provided in the original article. Therefore, the shift assignment for this role is solely based on the empirical contact data and a range of assumptions. The hourly registered contacts presented in Figure 3.2.18 serves as basis.

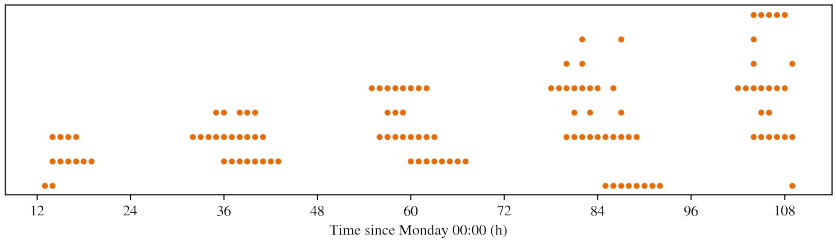

Figure 3.2.18: Each point marks an hours where one or more contacts has been registered for the individuals belonging to the MED role. The x-axis corresponds to the time in the study period, counting from Monday at 00:00, whereas each row corresponds to one unique individual.

As Figure 3.2.18 shows, the ADM individuals seem to work during daytime. The hours of contacts vary greatly from day to day, and also among the individuals working in the ward. The number of contacts registered through the study period is presented as bars in Figure 3.2.19. The orange shaded areas denotes the time span of fifteen hours of registered contacts, going from 06:00 to 21:00. The longest observed shift in this category spans 10 hours. Therefore, the ADM shifts in the model are determined by identifying the time span of 10 hours covering the largest possible number of contacts. The leftmost panel of Figure 3.2.20 shows the fraction of ADM contacts registered in each of the fifteen hours. Setting the ADM shift from 08:00 to 18:00 results in the largest proportion of contacts covered by a ten-hour shift.

Also for the ADM role, one or more registered contacts in the course of a day is set as identification criterion for a shift. This assumption yields the distribution of number of shifts worked in the course of the study period (middle panel) and the number of individuals in the ward per day (right panel) for the ADM role, see Figure 3.2.20. The number of ADM agents drawn for each shift is set to 4, equivalent to the median value of the number

of ADMs present in the ward per day, see Table 3.2.4. Also for the ADM role, the shift assignment can be done randomly of according to percentages of employment.

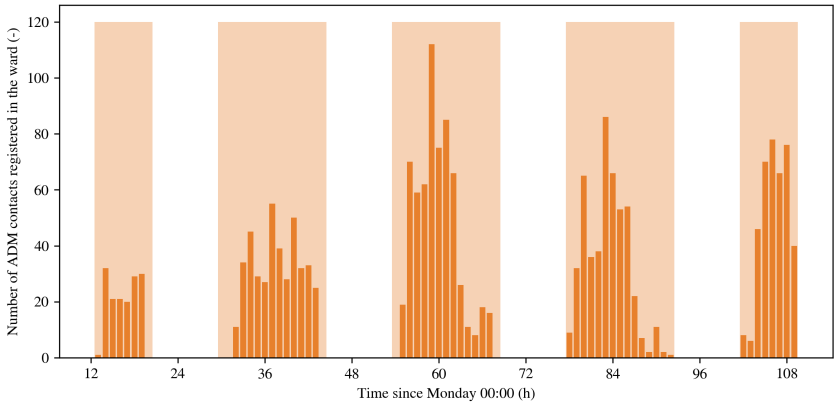

Figure 3.2.19: The bars represent the number of contacts registered for ADMs for each hour hour. The lighter orange, shaded areas represent the span of working hours, from 06:00 to 21:00.

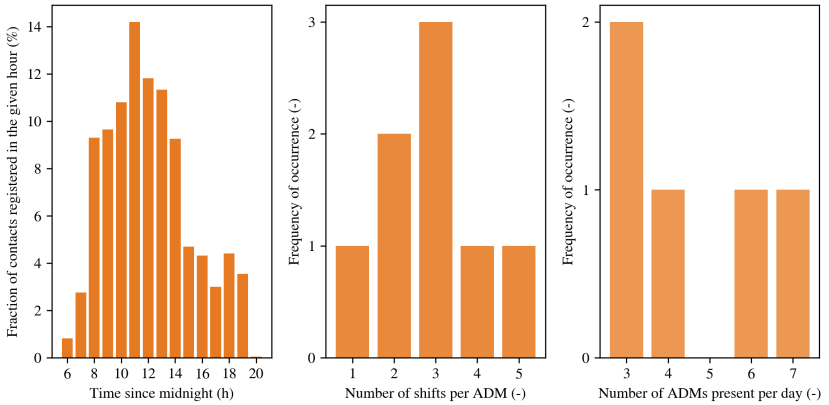

Figure 3.2.20: Left panel: The fraction of contacts registered for each hour summarized over the entire study period. Middle panel: Distribution showing how many shifts the ADM individuals have during the study period. Right panel: Distribution showing the number of registered ADM shifts per day in the ward.

Patients and Hospitalization

In the course of the study period, there are registered contacts for 29 unique patients in the hospital ward. In Figure 3.2.21 each green point represents an hours in which one or more contact has been registered, each row representing one unique patient. Ideally, the duration of hospitalization would be derived directly from the empirical data, for instance by fitting

the observed lengths of hospitalizations with a suitable distribution. Due to the short time span, however, these durations are hard to identify: It is difficult to determine whether the first hour of contact corresponds to the time point of admission, or if it is simply the first hour of registered contact in the time window short corresponding to the study period.

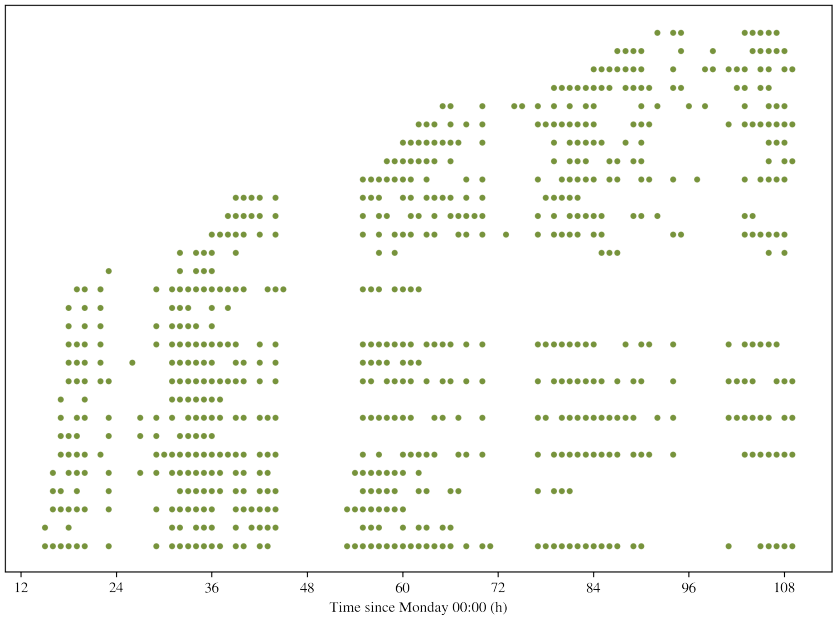

Figure 3.2.21: Each point marks an hours where one or more contacts has been registered for the individuals belonging to the PAT role. The x-axis corresponds to the time in the study period, counting from Monday at 00:00, whereas each row corresponds to one unique individual.

An alternative approach is to base the hospitalization length on the patient turnover rate. The number of admissions and discharges in the course of the 97 hours can be determined based on the following assumptions: Initially, the time points of admission and discharge are set to the first and last hour of registered contacts, respectively. However, if a patient has registered contacts within the first twelve hours, the patient is assumed to have an admission time point preceding the study period. Similarly, if the patient has registered contacts during the last twelve hours, the patient is assumed to have a discharge succeeding the study period. For the individuals to which these assumptions apply, no admission and/or discharge is registered. Figure 3.2.22 presents a visualization of the described assumptions. The light green patches correspond to the time span between first and last registered hour of contact, whereas the darker green patches correspond to hours where the patient is *assumed* to be present in the ward even though no contacts are registered.

Based on these assumptions, there are 13 admissions and 12 discharges in the ward in the course of the 97 hours, giving an average of 12.5 patient replacements. As shown in Equation (3.2.1), the average turnover rate is approximately 6.78×10^{-3} replacements per bed per hour.

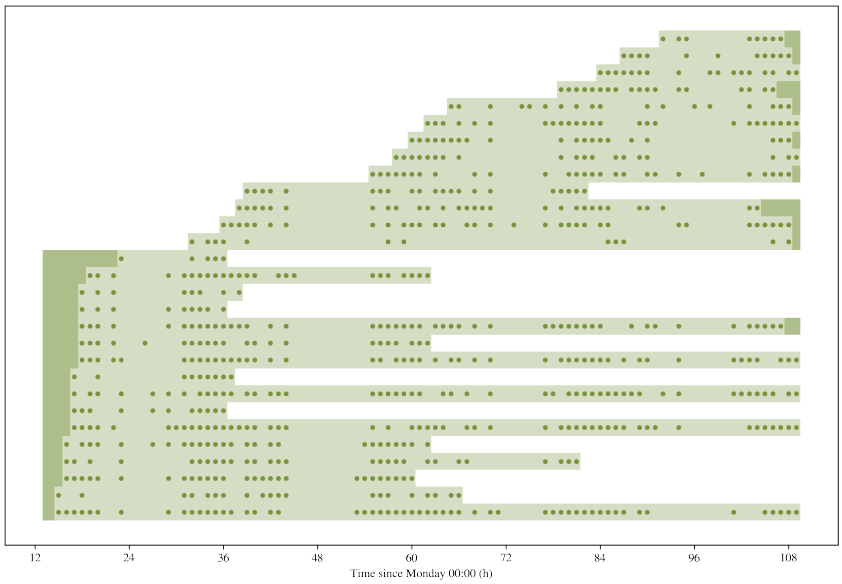

Figure 3.2.22: Each point marks an hours where one or more contacts has been registered for the individuals belonging to the PAT role. The light green shaded areas mark the time span between the first and last detected hour of contact. The darker green shaded areas present additional hours of where the patient is assumed to be present in the ward.

$$r_{\text{turnover}} = 12.5 \cdot \frac{1}{19 \text{ beds}} \cdot \frac{1}{97 \text{ hours}} \approx 6.78 \times 10^{-3} \frac{1}{\text{beds} \cdot \text{hours}} \tag{3.2.1}$$

The average number of bed hours, i.e. the average duration of hospitalization, is given by the inverse of the turnover rate, as shown in Equation (3.2.2).

$$t_{\text{hospitalization}} = {r_{\text{turnover}}}^{-1} \approx 147.44 \text{ bed hours} \tag{3.2.2}$$

The average duration of hospitalization equal to 147.44 hours corresponds to just over six days. Although some variance in the duration of hospitalization is excepted, it is difficult to determine the magnitude of it. Under the above-mentioned assumptions, only one patient is admitted and discharged in the course of the study period. This patient is present in the ward for 44 hours. Therefore, the following assumption is made: The hospitalization duration for each patient is drawn from a normal distribution with expectation value $\mu = 147.44$ days and a standard deviation σ so that the observed hospitalization of $x = 44$ hours falls within the $95\,\%$ confidence interval. Since $95\,\%$ of the area under the normal distribution curve lies within approximately 1.96 standard deviations away from the mean [161], the standard deviation σ can be calculated as shown in Equation (3.2.3). The resulting normal distribution is presented in Figure 3.2.23.

$$1.96 \cdot \sigma = \mu - 44 \qquad \Rightarrow \qquad \sigma = \frac{147.44 - 44}{1.96} \approx 52.78 \tag{3.2.3}$$

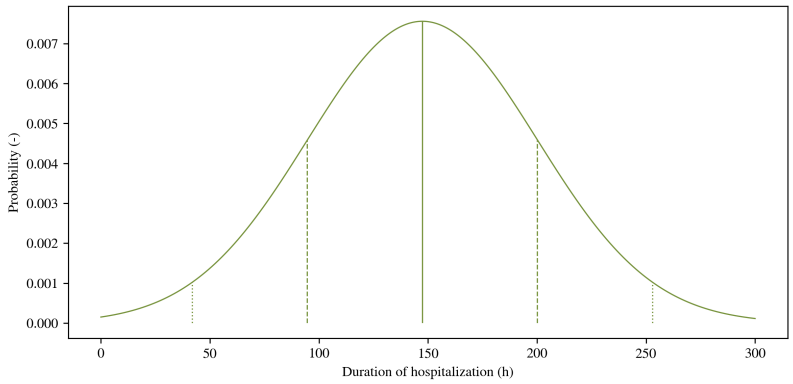

Figure 3.2.23: The length of hospitalization, in hours, for all patients is drawn from a normal distribution with mean value $\mu = 147.44$ and standard deviation $\sigma = 52.78$. The solid curve represents the probability density function, whereas the solid, vertical line represents the mean value. The dashed and dotted lines indicate one and two standard deviations away from the mean, respectively.

When a new PAT agent is admitted into the model hospital ward, the length of hospitalization is thus drawn from a normal distribution $\mathcal{N}(\mu, \sigma^2)$ where $\mu = 147.44$ hours and $\sigma = 52.78$. In order to avoid very short time spans, a minimum limit is set to 8 hours, meaning that all drawn values below this value is set to 8. In order to avoid all initially admitted patients from starting their hospitalization when the simulation begins, their remaining hours of hospitalization is drawn randomly among all hours in the range spanning the minimum value and the drawn hospitalization length.

Some simplifications are made in terms of bed occupancy and the window of admissions and discharges: Figure B.2.16 shows a plot of how the bed occupancy in the ward varies with time, based on the discussed assumptions of presence in the ward given registered contacts within the twelve first or last hours in the study period. Both the mean and median bed coverage in the ward equal 16.0 out of nineteen beds, yielding a coverage of approximately $84.21\,\%$. In the model, the bed coverage is assumed to remain stable at this percentage. Hence, a patient discharge results in an immediate admission in the following hour. In the empirical data, all admissions and discharges occur between 07:00 and 20:00 during the day. In the model, this window is not taken into account, and patients may be admitted and discharged at any hour.

3.2.3 Individual Heterogenities

The somewhat cumbersome work of identifying staff shifts and patient hospitalizations serves another role than just replicating a resembling work schedule and patterns of patient admission and discharge. These estimates can also be used to determine the relative contact numbers for each individual, adjusted for the estimated hours of presence in the ward. Although the original article provides information about the total number of contacts for each participant in the study, it is worth investigating whether the total contact number correlates with the hours of presence, or if some individuals are more extroverted

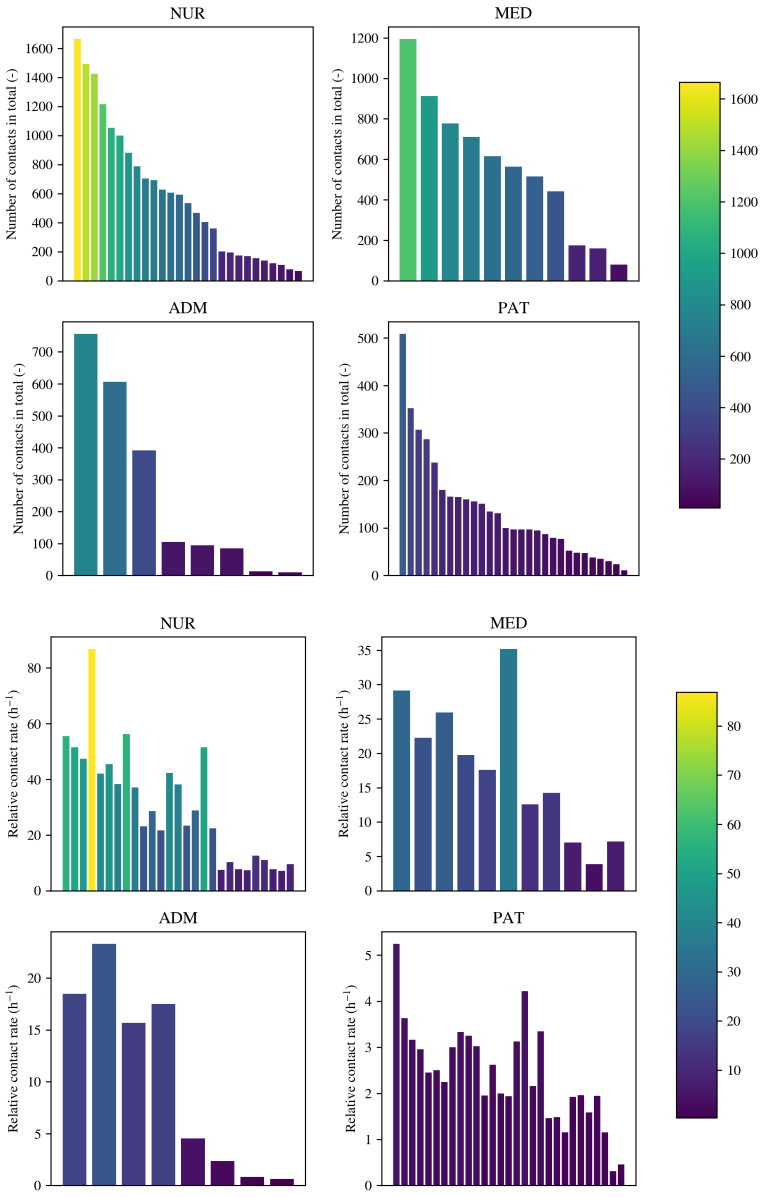

Figure 3.2.24: The upper four bar plots present the total number of registered contacts for each individual, each panel corresponding to one role. A color gradient is included to make it easier to compare the contact numbers across the panels. The lower four bar plots present the relative contact number, given as the number of registered contacts per estimated hour of presence in the ward.

than others. The objective is to assign all agents with a degree of extroversion which is drawn from a distribution mirroring the relative contacts rates for each of the four roles.

The relative contact rates are simply calculated from the total number of contacts registered throughout the study period and the estimated number of hours of presence in the ward. The upper four panels of Section 3.2.3 illustrate the total number of contacts registered for each individual, each panel corresponding to one role. The y-axes, representing the total number of contacts, span quite different numerical ranges in the four different panels. The color gradient is included to make it easier to compare the numbers across the panels. A stacked area plot showing how the accumulated number of contacts evolves with time is presented in Figure B.2.17.

The estimated hours of presence in the ward are based on the registered contacts and the assumptions made regarding staff shifts and patient hospitalizations as described above. The relative contact rate is simply found by dividing the total number of registered contacts with the estimated number of present hours. Table B.2.4 provides a complete overview of the role, ID number and the numbers providing the basis for the calculated contact rate per hour. A graphical illustration of the relative contact rate is presented as the lower four panels in Figure Section 3.2.3, where the order of the individuals from the figure above is preserved for comparison.

In order to recreate a similar pattern of relative contact rates in the model, the data representing the number of contacts per hour are plotted as histograms and fitted by lognormal curves, see the left panels of each role in Figure 3.2.25. Again, the size of the dataset makes it challenging to make good reasoning for the choice of a distribution fitting the data well, however, the lognormal distribution makes it possible to capture individuals

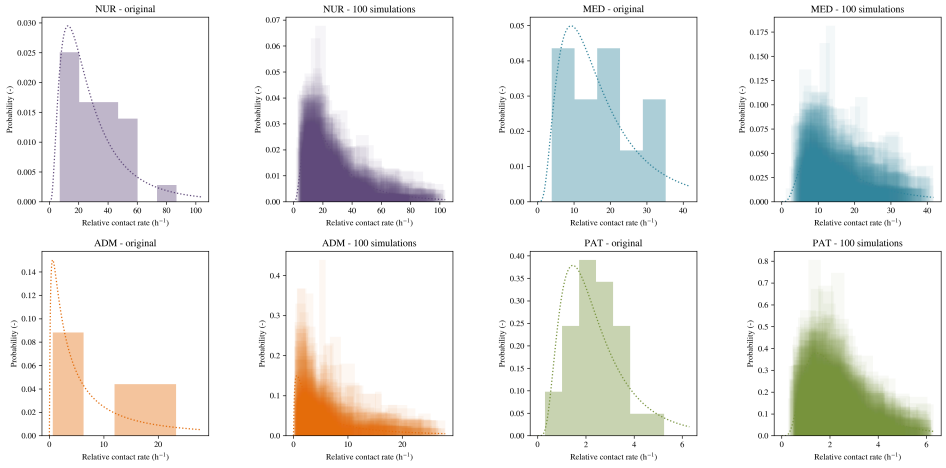

Figure 3.2.25: Out of the eight panels, two and two represents one role. Of these, the panel to the left represents a histogram corresponding to the calculated relative contact rate and a lognormal distribution based on the mean and standard deviation of these data. The panel to the left presents the same distribution and 100 simulations with drawn values from this distribution. The number of drawn values in each simulation equals the number of individuals belonging to each role.

contributing with a particularly high relative contact rate. In order to prevent the lognormal distribution from creating individuals with superficially high relative contact rates compared to the original data, a limit of 1.20 is set relative to the highest observed relative contact rate: If a value drawn from the distribution exceeds 120 % of the highest observed relative contact rate, the drawing is repeated. The resulting distributions with histograms from 100 simulations are presented in the right panels for each role in Figure 3.2.25. A corresponding figure where no upper limit is set, is presented in Figure B.2.19. The parameters determining the lognormal distributions, as well as the upper limit, are presented in Table 3.2.2.

Table 3.2.2: The table presents the location and scale parameters for the lognormal distribution governing the relative contact rates drawn for each agent in the model, in addition to the highest observed number of contacts per hour and a maximum limit set to 1.20 times this value.

Role	**Location, μ**	**Scale, σ**	**Highest**	**Upper limit**
NUR	3.1568	0.7804	86.8571	104.2285
MED	2.6911	0.6861	35.1875	42.2250
ADM	1.6713	1.4697	23.3077	27.9692
PAT	0.7410	0.6019	5.2474	6.2969

In addition to taking individual heterogenities in terms of extroversion into account, the contact network in the model attempts to mirror the observed interactions between the pair of roles. Through a series of contact matrices showing the number of contacts between the four different roles, Vanhems *et al.* observe that some role pair interactions constitute a larger proportion of the total number of contacts than other. Figure 3.2.26 presents a graphical illustration of the relative probability of interaction with an individual of the four different roles. For instance, if one of the individuals in a contact pair belongs to the NUR category, there is a 50 % chance that the individual making up the other half does also belong to the NUR role.

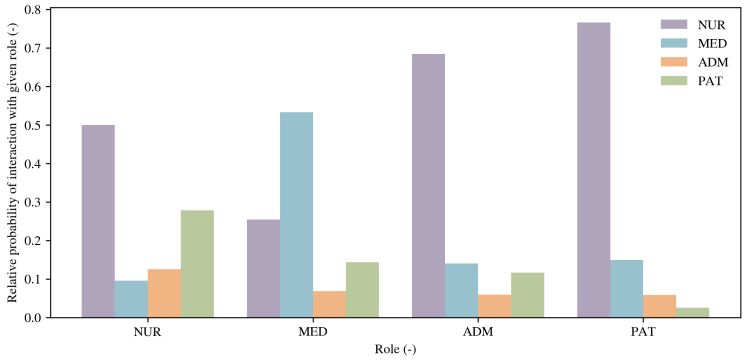

Figure 3.2.26: The bars represent the relative contact rates among roles in the ward for each of the four roles in the ward. The bars for each role summarize to 1.0.

3.2.4 Implementation of Temporal Network Into the Model

This section has given an introduction to how data extracted from the empirical contact network can be used to mirror the daily pulses of contacts in the ward, how staff shifts and patient hospitalizations can be replicated based on hours of registered contacts and lastly how individual variations within and among each of the four roles can be reproduced by fitting the data with lognormal distributions. The last section dealing with the generation of the temporal contact network summarizes how these three approaches are implemented in the model. The first part deals with agent initialization, whereas the second part covers the algorithm of the simulation steps.

Agent Initialization

After the model parameter are set, the model agents are initialized. In this step, the desired number of agents belonging to each of the four roles NUR, MED, ADM and PAT are created and assigned with attributes such as an id number and which role they belong to. After an agent is initialized, it is added to a list keeping track of all agents in the simulation. In addition, a node representing the agent is added to the temporal network.

The data giving rise to the empirical contact network presented by Vanhems *et al.* was obtained from a geriatric hospital ward with 19 beds. This model allows for up-scaling of the number of wards and the number of beds per ward. Hence, the initial number of staff members is calculated based on the staffing per bed in the empirical data. For instance, 8 individuals in the administrative staff per 19 beds in the ward, results in approximately 0.42 administrative staff per bed. Thus, for each ward in the model, the number of initialized agents with an ADM role is calculated using

$$a_{\text{ADM}} = \left\lfloor \frac{8}{19} \cdot n_{\text{beds}} \right\rfloor ,$$

with equivalent expressions for the initial number of staff belonging to the other roles. The same strategy applies for the number of staff members covering each of the given shifts: If two nurses cover the night shift in the original ward, $3 \cdot 2$ NURs have to work during night time if the number of beds in the ward has tripled. This information is summarized in Table 3.2.4. Figure 3.2.27 presents an illustration of how the staff agents are initialized.

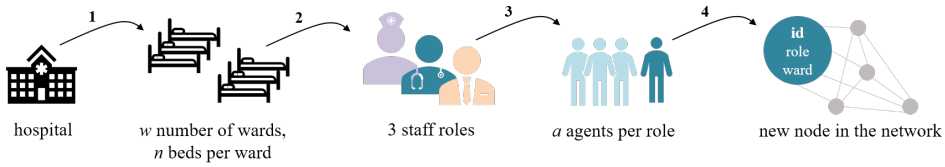

Figure 3.2.27: The initialization of agents belonging to one of the three staff classes can be divided into several steps: The user can create w number of wards with n number of beds in each ward. For each of the three roles NUR, MED and ADM, a agents are initialized in each ward, based on a potential up-scaling of the number of individuals given in the empirical data. A new staff agent results in a new node in the temporal network, storing the role and ward as node attributes.

The staff agents that are initialized in the very beginning of the simulation, are classified as regular staff. That is, if none of the staff members have to stay home due to quarantine or sick leave during the simulation period, these are the only staff agents who are initialized in the course of the simulation. If there are no available staff members to cover a shift, however, temporary staff members are initialized, see the discussion below. Regardless of whether the agent is a part of the regular or temporary staff, all staff agents are assigned with a given ward number. The agent can only cover shifts belonging this ward. The staff agents are also assigned with a percentage of employment. If the user sets the shift assignment to be 'random', all agents receive a percentage of 1.0. If the assignment is set to 'percentage', on the other hand, the percentage is drawn from a multinomial distribution derived from on the observed number of shifts, see Table 3.2.3.

Table 3.2.3: A staff agent is assigned a percentage of employment when being initialized. This table shows the percentages, x, the total number of individuals in the original data working with the given percentage, N, and the corresponding relative probability $P(x)$ for each of the staff roles.

Percentage	**Frequency**			**Probability**		
	NUR	MED	ADM	NUR	MED	ADM
0.2	5	0	1	0.19	0.00	0.13
0.4	6	2	2	0.22	0.18	0.25
0.6	7	2	3	0.26	0.18	0.38
0.8	6	3	1	0.22	0.27	0.13
1.0	3	4	1	0.11	0.36	0.13
Sum	27	11	8	1.00	1.00	1.00

The number of initial patients is calculated by multiplying the number of wards with the number of beds per ward and the average bed occupancy set to $84.21\,\%$. Each of these patients are initialized and assigned an id number and ward, similarly as for staff agents. The remaining number of available beds are updated simultaneously. In addition, the patient agents are assigned with the hour of admission and the number of remaining hours before discharge: The duration of hospitalization, $t_{\text{hospitalization}}$ is drawn from a normal distribution as described above. For the patients initialized in the very beginning of the simulation however, the remaining number of hours in the ward is drawn randomly from the range spanning $[t_{\text{min}}, t_{\text{hospitalization}}]$. In a scenario without a spreading infection, all patients are discharged when their drawn length of hospitalization has come to an end. If a patient is put in a quarantine or isolation due to suspected or confirmed disease, however, the hospitalization is extended to contain this period. Hence, the average duration of hospitalization may increase in case of an extensive disease outbreak.

All agents are assigned with a degree of extroversion when being initialized. As described above, this number is drawn from a distribution corresponding to their given role, and is set to not exceed 1.20 time the highest observed relative contact rate among the participants. All agent attributes that are set during initialization are presented in Figure 3.2.28.

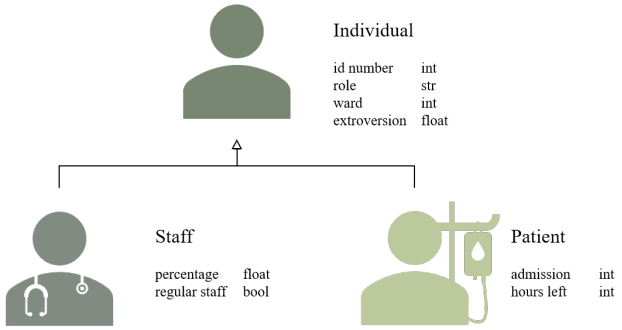

Figure 3.2.28: All agents are assigned with id number, role, ward and degree of extroversion when being initialized. In addition, staff agents are given a percentage of employment and confirmed regular staff if being initialized in the very beginning of the simulation. Patients are assigned with the hour of admission and the number of hours left before discharge. The data types denoted to the right are briefly explained in Table B.1.2.

The Simulation Step

The agent initialization step is followed by the the actual simulation. In this step, the model iterates through each of the H simulation hours. For each of these hours, denoted h, the model goes through five steps, presented in Figure 3.2.29. The three first steps deals with presence in the ward and contact generation and are described below. Step four and five deals with the epidemiological spread and summarizing the spreading state in the system, and are therefore described in the next main section.

1. **Shifts and patient hospitalization**: If the simulation hour h corresponds to the beginning of a new shift, the model shift manager ensures that the shifts are covered according to the model parameter values, see Table 3.2.4. A list of available staff members is made based on the following criteria: The agent is not currently present in the ward, neither restricted to come due to quarantine, for example. In addition, the agent must be part of the regular staff and the ward number and role must com-

Table 3.2.4: Overview of the number of agents belonging to the three different staff roles initialized per ward, in addition to information about staffing, duration and start time of the shifts.

Role	Regular staff	Shift type	Staffing	Duration (h)	Start
NUR	$27 \cdot n/19$	morning	$9 \cdot n/19$	7	07:00
		day	$2 \cdot n/19$	8	09:00
		afternoon	$5 \cdot n/19$	7	13:00
		night	$2 \cdot n/19$	11	20:00
MED	$11 \cdot n/19$	-	$8 \cdot n/19$	10	09:00
ADM	$8 \cdot n/19$	-	$4 \cdot n/19$	10	08:00

ply with the requirements. If the list of available staff is longer than the required number of agents needed to cover the shift, a selection is chosen based on employment percentages and remaining resting time, see next paragraph. If there are not available agents among the regular staff, the same screening process is repeated for any existing temporary staff members. If there are still not enough agents to cover the shift, new staff members are initialized as described above, with the exception of that the regular staff-attribute is set to false. To summarize, the model is built to 1) choose regular staff over temporary staff, and to 2) choose already existing temporary workers over generating new ones.

If there are more available staff members than shift vacancies to cover, the required number is drawn based on the percentages of employment. For instance, if there are three available nurses A, B and C with percentages 0.6, 0.6 and 0.8, and only two of them are required for the night shift, the percentages are first normalized in order to summarize to one (0.3, 0.3 and 0.4). These weights affect the probability of each agent being chosen, implying that the latter NUR agent has a higher probability of

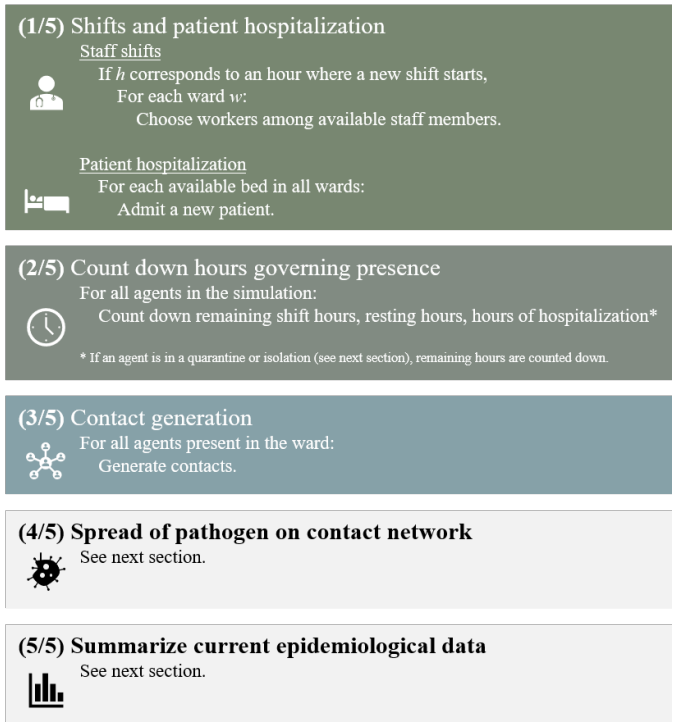

Figure 3.2.29: Each simulation step is divided into five main parts: The first step manages new shifts and admits patients if there are available beds. The second step counts down remaining hours of shifts, resting hours and hospitalization hours. The third step generates contacts among the agents present in the ward. Steps four and five govern the spreading and the summary of the epidemiological status in the ward.

working the given shift. The user can determine if resting hours for the NUR agents are taken into account. If they are, each NUR starts a countdown at 18 hours of rest when ending a shift, and the agents with the lowest remaining hours when a new shift is to start are chosen in front of other agents, even if the agent has a lower working percentage. Returning to the example above, if the NUR agents A and B are fully rested, whereas agent C is not, the two former are chosen to cover the shift.

After any new shift has been covered, potential available beds are filled up by admitting new patients. These are initialized as described above, except from that their duration of hospitalization is drawn directly from the distribution and not shorted down. The remaining hours in the ward is extended if the patient is quarantined or isolated due to illness, as will be discussed in the next main section.

2. **Count down of hours governing presence**: In this step, the simulation runs through all agents in the simulation, not only those present in the ward. If the agent is currently present in the ward, the remaining hours, either of the shift or the hospitalization is reduced with one hour. When reaching zero, the staff agents leave work whereas the patient agents are discharged, freeing up a bed. For agents of the NUR role who are currently not present in the ward, their remaining hours of rest are counted down. For agents in quarantine or isolation, the remaining restriction hours are reduced by one. When reaching zero, the agent returns to a state of no restrictions.

3. **Contact generation**: The contacts for the given hour are generated among the agents who are present in the ward. Firstly, the number of contacts per bed is drawn from a normal distribution with parameters corresponding to the given hour h. If the number of contacts per bed is larger than zero, it is multiplied with the number of wards and number of beds per ward. The total number of contacts for the given hour, here denoted C, can be reduced to mimic social distancing, see Section 3.3.6.

   The C number of contacts must consist of pairs of agents. These are referred to as contact pairs and consist of a person A and a person B. Firstly, persons A of all contact pairs are drawn from a multinomial distribution (with replacement). The probability of being drawn corresponds to their degree of extroversion, again normalized in order to summarize to one. Thus, an agent with a high degree of extroversion may be drawn multiple times, see Figure 3.2.30.

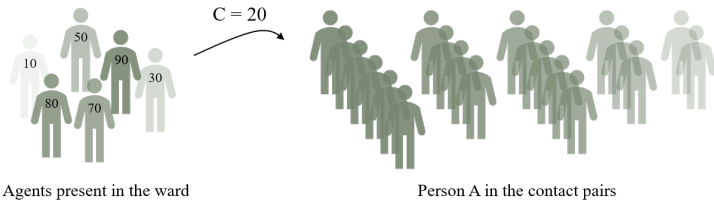

Figure 3.2.30: The person A of all contact pairs is drawn among the present agents in the wards and depend on their degree of extroversion, here denoted by numbers on each of the six example agents. When drawing without replacement, the same agent may serve as person A in several contact pairs.

After person A of all contact pairs have been drawn, the model iterates through the list unique agents. If a given agent was draw 10 times, 10 persons B must be drawn from the remaining agents in the ward. However, the probabilities governing these drawings depend not only on the agent's degree of extroversion, but their role relative to the role of person A and their assigned ward relative to the assigned ward of person A. The probability for each potential person B is calculated by multiplying three values:

(a) *The degree of extroversion:* Agent attribute.

(b) *The probability of interaction between the role pairs*: If person A for instance belongs to the NUR role, a potential person B candidate belonging to the NUR role will have a greater probability of being drawn compared to an agent belonging to the MED role, see Table 3.2.5.

Table 3.2.5: Relative contact rates for each of the four roles in the ward.

	NUR	**MED**	**ADM**	**PAT**
NUR	0.5000	0.2547	0.6843	0.7662
MED	0.0962	0.5332	0.1401	0.1497
ADM	0.1256	0.0681	0.0590	0.0589
PAT	0.2782	0.1440	0.1165	0.0252

(c) *The degree of collaboration across the wards*: In case of simulating multiple wards, the user can determine the degree of collaboration across the wards by tuning the value of a ward isolation parameter. Patients can only interact with patients belonging to the same ward, but all staff members can have contact with agents belonging to other wards. Thus, if this parameter is set to 1.0, the wards are totally isolated and no intra-ward contacts occur. On the other hand, if the parameter is set to 0.0, a staff member is just as likely to have contact with a agent from another ward as an agent belonging to the same ward as itself.

From the available agents the contact partners for person A are drawn using a multinomial distribution with probabilities corresponding to the product of the extroversion, the role pair interactions and the ward isolation parameter. For each of the resulting contact pairs, the number of contacts in this given hour h is registered in both of the agents' lists managing contacts. In addition, their corresponding link in the building contact network is either added or updated.

The user can control the randomness in the described contact generation by tuning a contact randomization parameter. When this is set to 0.0, the contact generation is just as described above. By increasing the parameter value, potentially up to 1.0, an increasingly larger proportion of the C contact pairs are drawn randomly, without taking the degree of extroversion or role pair interactions into account. The ward isolation is not affected.

3.3 Building an Epidemiological Model

The previous section introduced how a temporal contact network is generated based on patterns extracted from an empirical hospital ward network. These inter-individual contacts form the basis for transmission of SARS-CoV-2. The model is agent-based and built upon transitions between several epidemiological states. The following section introduces these states, the transitions between them and attempts to quantify the relevant parameters. The section is concluded by describing how the epidemiological part of the model is implemented, including some key infection control measures.

Epidemiologial States and Transitions

The choice of epidemiological states and associated transitions are based on the current knowledge regarding COVID-19, as presented in Section 2.1. Upon transmission, a susceptible individual (S) enters an exposed state (E). The incubation period is followed by an infection, which could be either symptomatic (I) or asymptomatic (A). In this model, all asymptomatic individuals are assumed to recover (R), whereas individuals with symptoms may experience a fatal outcome (D). In addition, the model includes the possibility of vaccination through a direct transition from susceptible to recovered state. The debated chance of not obtaining immunity after an infection is reflected in the transition from recovered to susceptible state. All six states and the possible transitions are illustrated in Figure 3.3.1.

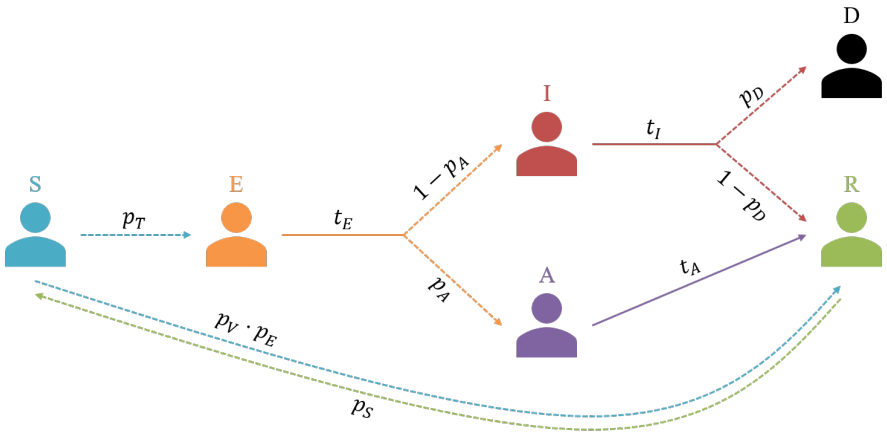

Figure 3.3.1: Flow chart representing the epidemiological states and the transitions between them in the model describing the spread of SARS-CoV-2. Dashed lines indicate transitions determined by probabilities, p, whereas solid lines indicate transitions determined by time spans, t. The states and transitions are introduced in the text.

As the figure shows, the transitions between the states depend on several parameters. The dashed arrows represent probabilities, p, whereas the solid arrows represent time spans, t. These parameters are shortly described in Table 3.3.1 and further discussed in the following.

Table 3.3.1: Overview of the parameters governing the epidemiological model describing the spread of SARS-CoV-2. The parameters are either transition probabilities, p, or time spans, t.

Parameter	Description
p_T	transmission in contact with infectious agent
p_A	asymptomatic course of disease
p_D	fatal outcome given symptomatic course of disease
p_V	being vaccinated towards the virus
p_E	full protection given vaccination, i.e. efficacy
p_S	not gaining immunity after infection
t_E	incubation period
t_I	infection period
t_A	asymptomatic period

Sections 3.3.1 to 3.3.5 give an introduction to five of the epidemiological states, including the characteristics of the agents they comprise and the possible transitions to other states. As the following material will demonstrate, some of the listed parameters depend on several others. The state comprising diseased agents, state D, is not described in any further detail as there are no transitions from this state.

3.3.1 Susceptible state

The susceptible state comprises all agents who can become infected with the virus upon exposure. These agents do not carry the virus and are hence not infectious. Nor do they not have immunity towards it. All agents in the model, except for the initially infected ones, reside in this state when being initialized. Susceptible agents have two possible transitions to the following states:

- Exposed (E): A susceptible agent can acquire the virus when being in contact with an infectious agent. The probability of transmission is given by p_T.
- Recovered (R): A susceptible agent can obtain immunity towards the virus through vaccination. The probability of immunization is given by the product of vaccination probability p_V and vaccination efficacy p_E.

Transmissibility

The transmissibility p_T is a measure of how easily the virus transmits from one agent to another. As discussed in Section 2.1.1, the two key routes of SARS-CoV-2 transmission encompass close contact, where respiratory droplets are transferred in close proximity interactions, and indirect contact where the virus is transmitted through fomites. For the purpose of investigating how the pathogen spreads across the simulated temporal contact network, the latter route of transmission is neglected. The threshold value for contact detection in the empirical contact network of 1.5 meters correspond well with the mean value of the WHO and CDC definitions of close contact; 1 and 2 meters, respectively.

Some modeling frameworks investigating close-contact transmission of the influenza virus propose an expression where the transmissibility depends on the weight of the contacts between two individuals. This contact weight could for instance be the number or duration of contacts made in a given time period. Their proposed equation is presented in Equation (3.3.1), where c denotes the contact weight and the value 0.003 represents the transmissibility per contact. This value is calculated based on the attack rate of an influenza outbreak on a jet airliner in 1979 [170–172].

$$p_T(c) = 1 - (1 - 0.003)^c \tag{3.3.1}$$

The equation representing the influenza virus transmissibility serves as basis for the transmission probability p_T in this model. The contact weight c is defined as the total number of contacts made between two agents in the course of an hour. The relative contagiousness of SARS-CoV-2 compared to the influenza virus is accounted for by multiplying the baseline transmissibility value with a factor γ. If this value is set to two, for instance, the coronavirus transmits twice as easily as the influenza virus. Based on the presented R_0 values, the baseline value for γ is set to 2. Figure 3.3.2 illustrates how the probability of transmission evolves with number of hourly contacts. The three different colors correspond to three levels of contagiousness. The dotted lines represents a linear reference; $p_T = \gamma \cdot 0.003 \cdot c$.

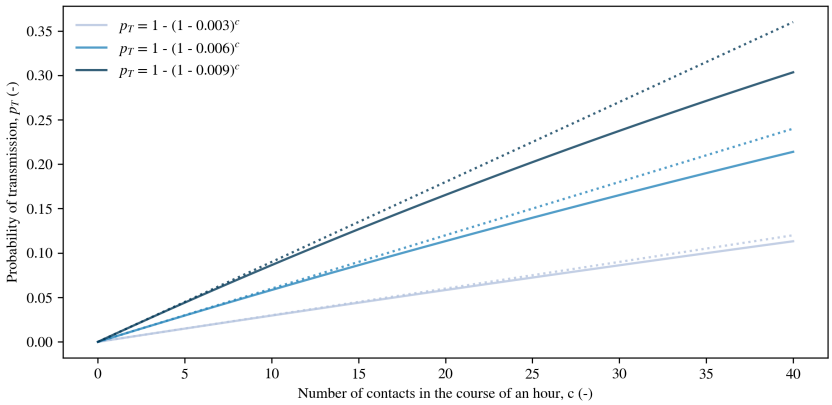

Figure 3.3.2: The solid lines present the probability of transmission of SARS-CoV-2 from a contagious agent to a susceptible agent as function of the hourly number of contacts. The three curves represent three cases where SARS-CoV-2 transmits one, two and three times as easily as the influenza virus, respectively. The dotted lines provides a linear reference; $p_T = \gamma \cdot 0.003 \cdot c$

Of the six epidemiological states in the model, only agents in the exposed (E), infected (I) and asymptomatic (A) states carry the virus. Agents in the infected state represent obvious candidates for transmission of the virus. As earlier introduced, however, see Section 2.1.2, some studies have shown that the novel coronavirus may transmit in the absence of symptoms. The level of contagiousness regarding presymptomatic and asymptomatic carriers relative to symptomatic individuals is not yet quantified. In order to allow for

exploring how the level of pre- or asymptomatic infectiousness affects the overall disease transmission, the factor α is introduced. The infectiousness of agents in state I, assumed to be constant and equal for all agents in the state, is used as a base; $\alpha_I = 1$. For non-carriers the value is set to zero. Since the virus transmits through droplets, it is natural to assume that agents without symptoms are less contagious than agents *with* symptoms. Hence, the value for α_A, representing presymptomatic agents in state E and asymptomatic agents in state A, is assumed to lie somewhere between 0 and 1. The parameters governing the relative contagiousness are summarized in Table 3.3.2. The baseline value for α_A is set to 0.5. Given contact between a susceptible and an infected agent, this value corresponds to an assumption of the infectiousness being halved if the infected agent does not have any symptoms.

Table 3.3.2: The relative infectiousness, α, varies with the state of disease and presence of symptoms. *Agents without symptoms include both agents with an asymptomatic course of disease and agents in incubation phase. †Agents in state E may be contagious only in parts of their incubation phase, see discussion in Section 3.3.2.

Relative contagiousness	Description	States
$\alpha_I = 1$	Symptomatic	I
$\alpha_A \in [0, 1]$	Asymptomatic*	A, E†
$\alpha_N = 0$	Non-contagious	N, R, D

So far, the expression describing transmission probability is supplemented with a factor γ describing the relative contagiousness of SARS-CoV-2 compared to the influenza virus, and a factor α capturing the relative contagiousness of agents lacking symptoms compared to agents with symptoms, see Equation (3.3.2). These values can be tuned for the model as a whole, but is assumed to be equal for all model agents. The probability of transmission is also affected by any infection control measures initiated by agent quarantine or isolation, which is discussed in Section 3.3.6.

$$p_T(\alpha, \gamma, c) = 1 - (1 - \alpha \cdot \gamma \cdot 0.003)^c \tag{3.3.2}$$

Vaccination

There is currently no approved vaccine providing immunity towards the coronavirus. However, for the purpose of evaluating the effect of a future vaccine on a possible resurgence, vaccination is included in the model. The vaccination coverage p_V denotes the proportion of agents who takes a vaccine. Regarding the seasonal influenza virus, the World Health Organization recommends a vaccination coverage of $75\,\%$ for all individuals in high risk groups, including elderly and individuals with chronic illnesses [173]. The vaccine efficacy p_E denotes the probability of a vaccine, once taken, providing full immunity towards the pathogen of interest. Again using the seasonal influenza as a reference, the value of p_E lies around $60\,\%$, varying slightly from year to year [174]. Based on these numbers and the current situation where no approved SARS-CoV-2 vaccine exists, the baseline values are set to $p_V = 0$ and $p_E = 0.60$.

3.3.2 Exposed state

The exposed state comprises all agents being in the incubation period. These agents carry the virus, but do not have an active infection. The initially infected agents, serving as patients zero in the model, start their incubation period when being initialized. When leaving the exposed state after the given time span t_E, agents have two possible transitions to the following states:

- Asymptomatic (A): An exposed agent can enter an asymptomatic course of disease with the probability p_A.
- Infected (I): If not entering an asymptomatic infection, an exposed agent enters a symptomatic course of disease. The probability is hence given by $1 - p_A$.

Duration of the Incubation Period

As introduced in Section 2.1.2, the incubation period for COVID-19 can vary greatly between individuals. In order to capture this variation, it seems reasonable to model the incubation period using a probability distribution. A lognormal distribution is chosen in order to capture the positive skew. This probability distribution is determined by a location parameter μ and a scale parameter σ, see Section 2.4.2. Based on the given incubation period estimates [56, 66–69], which provide information about both centrality and variance in the analyzed data, the following set of equations is proposed for deriving the lognormal parameter values: The first equation, see Equation (3.3.3), is based on the scale value, either the mean, $\bar{x}$, or the median value, $\tilde{x}$:

$$\tilde{x} = \exp(\mu) \qquad \bar{x} = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{3.3.3}$$

The second equation is based on the variation in the data, and utilizes that the sum of the area under the probability density curve (PDF) can be specified by choosing a given lower and upper bound: When a lower bound a and upper bound b is given, the lognormal parameters can be set so that the integral spanning the interval $[a, b]$ summarizes to a given sum s. For instance, Linton *et al.* estimated the mean incubation period to 5.6 days with a $95\,\%$ confidence interval spanning 2 to 14 days [68]. Hence, the lognormal parameters can be set so that there is a $95\,\%$ chance of drawing an incubation period t_E which lies within the interval $[2, 14]$ days. The equation is presented in Equation (3.3.4).

$$\int_a^b \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \mathrm{d}x - s = 0 \tag{3.3.4}$$

The set of equations can then be solved to determine the values of μ and σ. The code solving this set of equations based on the mean value is provided in Appendix B.2.6. Here is also a graphical illustration of how the given example above is solved numerically. The mean or median value, as well as the lower bound, upper bound and sum of integral determined from the given references are presented in Table 3.3.3.

Table 3.3.3: Overview of the estimated mean or median incubation time for a set of references. The integral of the lognormal probability density function between the lower and upper bounds a and b is set to summarize to the given sum. *WHO does not provide a confidence interval or percentage of cases falling within the given range. The sum is set to 0.99 in order to let the given range capture a high proportion of the cases.

Authors	Mean	Median	Lower	Upper	Sum	Citation
WHO	5.5	-	1	14	0.99*	[56]
Zhang *et al.*	5.2	-	1.8	12.4	0.95	[67]
Linton *et al.*	5.6	-	2	14	0.95	[68]
Lauer *et al.*	-	5.1	0	11.5	0.975	[66]
Backer *et al.*	6.4		2.1	11.1	0.95	[69]

A visualization of the resulting lognormal distributions based on the values given in Table 3.3.3 is presented in Figure 3.3.3. The solid lines represent the probability density function, whereas the dotted, vertical lines represent the median value.

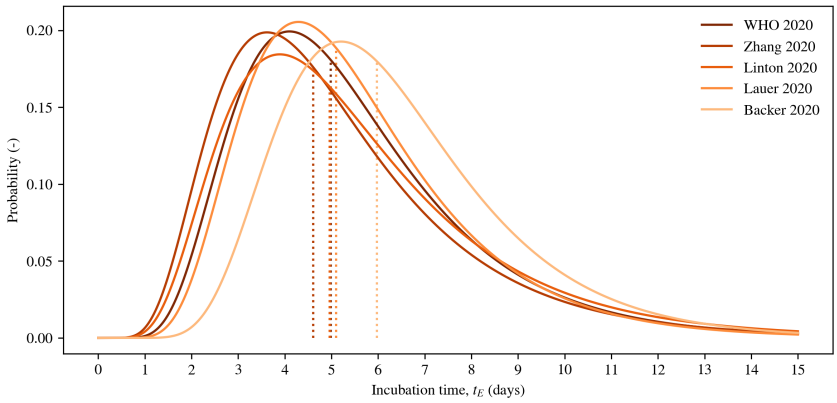

Figure 3.3.3: Five different lognormal distributions based on estimated incubation times. The solid curves represent the probability density function, whereas the dotted lines represent the calculated median value.

In the epidemiological model, the incubation time is drawn from the lognormal distribution based on the parameters derived from the WHO numbers, see Equation (3.3.5). As Figure 3.3.3 shows, the numbers used by the World Health Organization are supported by the estimates provided by several peer-reviewed articles. The distribution from which the incubation periods are drawn from can be shifted in both directions from the baseline by tuning the values for μ and σ.

$$t_E \sim \text{Lognormal}\,(\mu,\, \sigma^2), \qquad \text{where } \mu = 1.6067 \text{ and } \sigma = 0.4427 \tag{3.3.5}$$

Infectiousness in Exposed State

As discussed in Section 3.3.1, the level of infectiousness is assumed to differ between agents with and without symptoms. In order to investigate the impact of potential presymptomatic transmission, also agents in the incubation period can be infectious. The relative level of infectiousness is assumed to equal that of asymptomatic agents, α_A. Although the viral load has shown to vary in a curve-like fashion in the course of the disease, it is here, for simplicity, assumed to be a flat rate. Exposed agents become infectious when h hours of the incubation period remains, as illustrated in Figure 3.3.4. If the drawn duration of the incubation time is shorter than h, the agent is assumed to be infectious directly after transmission.

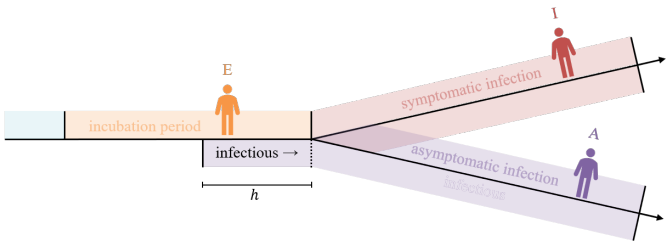

Figure 3.3.4: Timeline illustrating the infectiousness of an agent changing with its epidemiological state. An agent enters the incubation period upon transmission from an infectious agent. The agent itself becomes infectious when entering the last h hours of its incubation period, and is infectious until leaving either state I or A.

The value of h is assumed to be equal for all agents in the exposed state. The baseline value is based on the averaged estimates presented by Tindale *et al.* and He *et al.* [80, 82], see Equation (3.3.6).

$$h = \frac{2.55\,\text{day} + 2.89\,\text{day} + 2.3\,\text{day}}{3} = 2.58\,\text{days} \approx 62\,\text{hours} \tag{3.3.6}$$

Probability of Asymptomatic Infection

Despite the fact that COVID-19 materializes in a range of symptoms and degrees of severity, this model assumes a binary distinction between asymptomatic and symptomatic infection. As the rate of asymptomatic cases seems to be higher for younger individuals, it could be reasonable let the p_A parameter value differ for patients and staff in the ward. The estimated rate of asymptomatic infections proposed by Mizumoto *et al.*, $17.9\,\%$, was based on tests conducted among individuals where about three quarters were 60 years or older [74]. Assuming that most patients visiting the geriatric ward are in the same age range, the baseline value for patients is set to $p_A \approx 0.18$. The registered rate of asymptomatic cases among the infected residents of the Italian Municipality Vo', $43\,\%$ [75], serves as baseline value for agents belonging to any of the Staff roles, see Equation (3.3.7).

$$p_A = \begin{cases} 0.43, & \text{if STAFF} \\ 0.18, & \text{if PAT} \end{cases} \tag{3.3.7}$$

3.3.3 Asymptomatic State

The asymptomatic state comprises all agents being infected with SARS-CoV-2 without having any symptoms. Asymptomatic infections are therefore assumed to go unnoticed in terms of testing and initiation of control measures. The relative infectiousness compared to symptomatic cases is given by the parameter α_A, and is assumed to remain constant as long as the infection lasts. Since the fatality rate among asymptomatic agents is assumed to be zero, agents in this state only have one possible transition:

- Recovered (R): An agent with an asymptomatic course of disease enters a recovered state after a time span t_A.

Duration of Asymptomatic Infection

As asymptomatic COVID-19 cases easily pass undetected, it is difficult to establish the duration of asymptomatic courses of disease. Hence, the value of the t_A parameter in this model is a guesstimate based on the duration of mild cases, where symptoms usually pass in the course of one to two weeks, see Section 2.1.2. Similarly to the incubation period, the duration of an asymptomatic infection is drawn from a lognormal distribution. Figure 3.3.5 shows three examples with varying median values, where $99\,\%$ of the area under the curves falls within fourteen days.

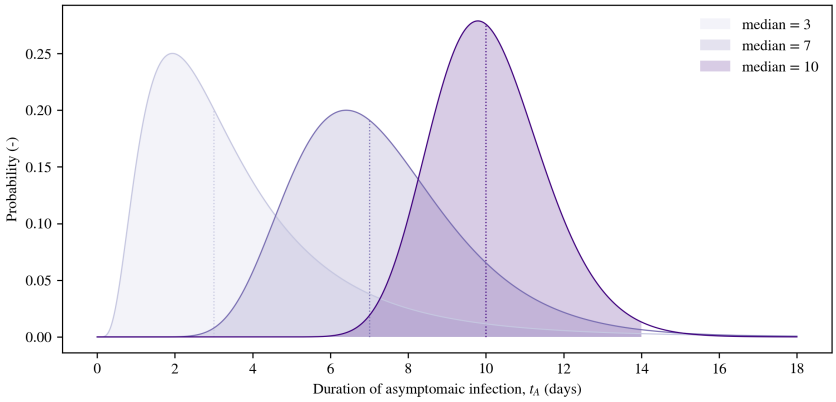

Figure 3.3.5: The solid curves represent three different lognormal probability curves with median values of 3, 7 and 10, respectively. $99\,\%$ of the area under the curves fall within zero and fourteen days. This area is shaded. The dotted, vertical lines represent the median value.

In the model, the baseline distribution for t_A corresponds to the lognormal probability function using a median value of seven days, see Equation (3.3.8). The distribution can easily be shifted by tuning the location and scale parameters.

$$t_A \sim \text{Lognormal}\,(\mu,\, \sigma^2), \qquad \text{where } \mu = 1.9459 \text{ and } \sigma = 0.2980 \tag{3.3.8}$$

3.3.4 Infected State

The infected state comprises all agents infected by SARS-Cov-2 who also have symptoms of the disease. All agents having symptoms are assumed to be equally infectious, with $\alpha_I = 1$. The presence of symptoms initiates several control measures, as will be described in Section 3.3.6. Agents belonging to this state have two possible transitions to the following states:

1. Recovered (R): An infected agent can recover from the disease when reaching the end of the time span t_I.

2. Diseased (D): An agent with severe symptoms may suffer a fatal outcome. The probability of a fatal outcome is given by the fatality probability p_D.

Duration of Symptomatic Infection

Compared to asymptomatic infections, the literature covering the duration of symptomatic courses of disease is somewhat more copious, see Section 2.1.2. In this model, the parameter t_I is based on estimates of the duration from symptom onset to death. When an agent comes to the end of this time span, the probability parameter p_D determines whether the outcome actually is fatal or not. Section 2.1.2 presented three estimates of the duration from symptom onset to death. Jung *et al.* provided parameters for a lognormal distribution [84], whereas Linton *et al.* and Wang *et al.* reported a median value and a given range wherein a given proportion of all cases fell within [25, 68]. The information provided by the two latter references were used to derive lognormal parameters. The three resulting probability density curves are plotted in the left panel of Figure 3.3.6.

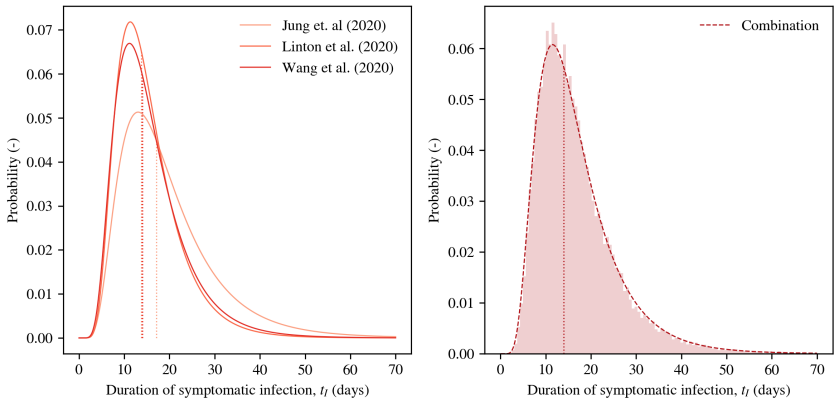

Figure 3.3.6: Left panel: Three lognormal distributions of the duration from symptom onset to death. The dotted vertical lines represent the median value of the distributions. Right panel: A histogram of 10 000 values drawn from each of the three distributions is plotted in red. The combined data are fitted by a log-normal distribution, represented by the dashed curve. Approximately 1.43 ‰ of the drawn values exceed 70 days and are omitted from the plot to improve readability.

The lognormal parameters determining the probability distribution of t_I is based on a combination of all three curves. 10 000 values were drawn from each of the three distributions presented in the left panel, and a new, combined lognormal curve was derived based on the mean and median values of the 30 000 drawn values combined. The right panel of Figure 3.3.6 shows the histogram of the drawn values together with a dashed curve representing the combined lognormal probability curve and a dotted, vertical line representing the median value. The baseline parameter values used to describe the duration of symptomatic courses of disease in the model is presented in Equation (3.3.9). This distribution has a median of fourteen days.

$$t_I \sim \text{Lognormal}\,(\mu,\, \sigma^2), \qquad \text{where } \mu = 2.6931 \text{ and } \sigma = 0.4016 \tag{3.3.9}$$

There are some issues regarding the assumption of using the duration from symptom onset to death as foundation for t_I, as this parameter is supposed to give a realistic picture of the duration of all symptomatic cases, not only the most severe. As introduced in Section 2.1.2, the duration of milder cases is shorter than the duration of more severe cases. In addition, the fatality rate, see below, is clearly linked to age. Hence, the model assumes that the staff in the ward, if infected, are likely to go through a milder and shorter course of disease with a lower probability of a fatal outcome than the hospitalized patients. The model allows for distinguishing between the duration of disease course between patients and staff by letting the user set the relative ratio r. This ratio goes from zero to one, and shifts the lognormal curve to the left by multiplying the median and mode by the provided ratio. The lognormal parameters are calculated as shown in Equation (3.3.10). Figure 3.3.7 visualizes the two distributions using a ratio $r = 0.5$. This value serves as baseline in the model.

$$\mu_* = \ln\left(r \cdot \exp\left(\mu\right)\right) \qquad \sigma_* = \sqrt{\mu_* - \ln\left(r \cdot \exp(\mu - \sigma^2)\right)} \tag{3.3.10}$$

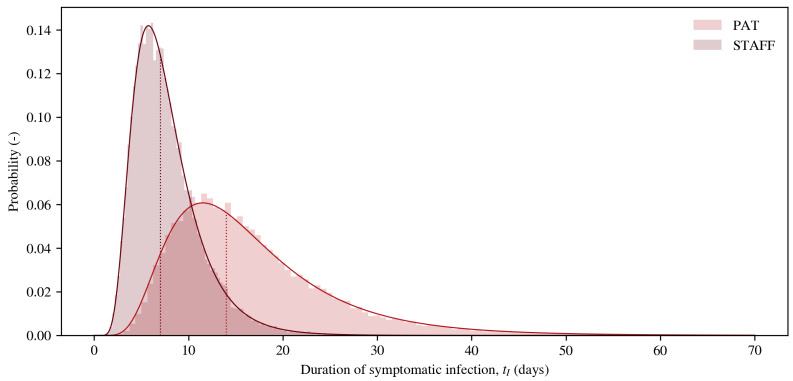

Figure 3.3.7: The two lognormal distributions describes the duration of a symptomatic course of disease and differs between patients and agents belonging to any of the staff roles. In this plot, the median and mode value of the distribution governing the staff t_I is half the values of median and mode governing the distribution for the patient t_I parameter.

Fatal Outcome

The observed case fatality rate (CFR) among individuals with a symptomatic COVID-19 infection has shown to vary greatly with age, see Figure 2.1.4. Hence, it becomes natural to distinguish between staff and patients when determining the value of the parameter p_D. Assuming that all agents belonging to any of the staff classes are between the age 20 and 60, and that patients are all 60 years or older, the following estimates for the fatality rate p_D are derived, based on the CFR among symptomatic individuals in Hubei and Italy:

Table 3.3.4: Age-aggregated estimates for the model parameter p_D based on estimated case fatality rates from Hubei and Italy. The values for the staff and patient classes are based on CFR values for individuals aged 20 to 60 years and 60 years and above, respectively.

	p_D^{STAFF}	p_D^{PAT}
Hubei	0.00975	0.219
Italy	0.00325	0.434
Mean	0.00650	0.327

The estimated means serve as baseline values in the model. If varying the probability of fatal outcome, it seems reasonable to let the p_D^{STAFF} and p_D^{PAT} parameters take values in the order of magnitude of 10^{-3} and 10^{-1}, respectively.

The model also includes the probability of an agent dying from a secondary infection, independent of the outcome of the primary infection. Ideally, the spread of an opportunistic pathogen would be implemented as a distinct spreading agent with its own epidemiological states and transitions. Due to time constraints, however, the current version of the model implements the risk of fatal secondary infections as a series of Bernoulli trials. The line of thinking is presented in Figure 3.3.8. The baseline values for the model, estimated from the theory presented in Section 2.1.2, are summarized below.

1. Probability of secondary infections: Zhou *et al.* determined the proportion of survivors and non-survivors who had secondary infections: Out of 137 survivors, only 1 person had a secondary infection compared to 27 out of 54 non-survivors. However, these observations yield estimated probabilities of secondary infection given fatal and non-fatal outcomes, not the probability of acquiring a secondary infection given a primary infection. The baseline value of $p_{\text{secondary infection}}$ is therefore guesstimated to 0.5.

2. Antibiotic treatment: Zhou *et al.* reported that $95\,\%$ of the hospitalized patients were treated with antibiotics. $p_{\text{antibiotics}} = 0.95$.

3. Resistant bacteria: The prevalence of antibiotic resistant bacteria varies greatly across Europe, see Table 2.1.4. As for Norway, the baseline values is set to $p_{\text{resistant}} = 0.009$, but could for instance be raised to 0.34 to mirror the situation in Italy.

4. Successful treatment: In case of an antibiotic-treated secondary infection with a susceptible bacterial strain, the chance of a fatal outcome is based on the average of the estimated case fatality rates for MSSA, see Table 2.1.5: $p_{\text{fatal, susceptible}} = 0.19$.

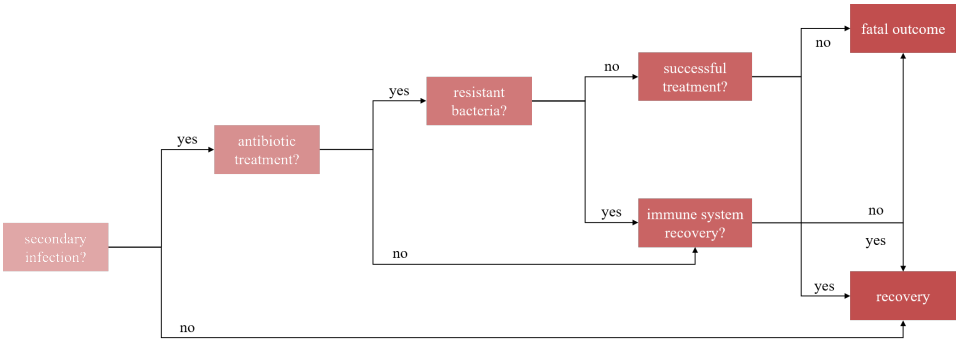

Figure 3.3.8: Flowchart representing the Bernoulli trials governing the probability of dying from a secondary infection for COVID-19 patients in the hospital ward.

5. Successful immune system recovery: In the case where a secondary infection is not treated with antibiotics or antibiotics are used in the attempt of treating a resistant bacterial strain, the agent depends on a successful immune system recovery in order to survive. In both cases, the probability of a fatal outcome is based on the average value of estimated case fatality rates for MRSA: $p_{\text{fatal, resistant}} = 0.40$.

Since the given data is based on hospitalized patients, only agents belonging to the PAT class in the model can die from secondary infections.

3.3.5 Recovered state

The recovered state comprises all agents who have received immunity toward the virus, either through an infection or through vaccination. These agents cannot be re-infected with SARS-CoV-2. Agents in this state have one possible transition to the following state:

- Susceptible (S): A recovered individual may return to a susceptible state if it does not gain immunity after going through an infection. The probability is given by p_S.

Probability of Loss of Immunity

Whether an individual who has gone through a COVID-19 infection can be re-infected is still debated. In this model, the potential loss of immunity following the recovery from the viral infection is captured by including a returning transition to state S. The probability p_S is assumed to be equal for agents recovering from both symptomatic and asymptomatic infections. The baseline value is set to $p_S = 0.01$. Agents who gain immunity through vaccination are assumed to remain immune and cannot return to state S.

3.3.6 Implementation of Control Measures

As introduced in Section 2.1.3, several control measures have been implemented (in the real world) in order to limit the spread of SARS-CoV-2. This section describes how some of the most central interventions are implemented in the epidemiological model.

Social Distancing

The model implements social distancing as a reduction of the total number of contacts in each simulation step. By increasing the inter-individual distance, fewer CPIs are detected. Hence, if the parameter value is set to 0.1, only $90\,\%$ of the originally drawn contacts for the given hour occur. One should keep in mind, however, that increasing the social distancing to 1.0 would eliminate all potential spreading in the ward due to the total absence of inter-individual contacts, although this is not practically feasible in a hospital ward.

Testing

Due to limited testing capacity, not all individuals with symptoms of COVID-19 have been tested. Based on the prioritized list presented in Section 2.1.3, however, both patients and staff in the model ward are included as priority item 2 and 3, respectively. The model therefore assumes that all infected agents in the hospital are tested, given that they have one of the symptoms fever, cough or breathing difficulties. The probability of each of these symptoms are given in Table 2.1.2. Thus, the following algorithm applies to testing of the model agents:

1. When entering a symptomatic infection, the agent is assigned with symptoms. Each of the three above-mentioned symptoms are drawn from Bernoulli distributions, as demonstrated for fever:

$$f(\text{fever}; 0.879) = \begin{cases} 0.879 & \text{if } x = \text{ fever} \\ 1 - 0.879 & \text{if } x = \text{ no fever} \end{cases} \tag{3.3.11}$$

   Hence, an infected agent may be assigned with zero to three symptoms.

2. If an agent has any of the three key symptoms, a test is conducted immediately. An agent awaiting test results is set in quarantine, see discussion below. The waiting time from a test is conducted until the test results are available is set by the user. The baseline value is set to 36 hours.

3. Healthcare employees with milder symptoms are considered for testing if an acute respiratory infection lasts for more than two days, see Section 2.1.3. In the model, "milder symptoms" are interpreted as a symptomatic infection, but where all three key symptoms are absent. Hence, if an agent belonging to any of the staff roles are in state I (symptomatic infection) but lacks all of the three symptoms, a timer is set to count down from 48 hours. If the agent still has an infection, i.e. remains in state I, a test is conducted.

Once a positive test is reported, two control measures are implemented; contact tracing and isolation, see discussion below.

Contact Tracing

When an agent has tested positive for COVID-19, contact tracing is initiated in order to capture potential cases of presymptomatic transmission. As the empirical network forms, the temporal contact information for each agent is stored as an agent attribute. The growing list contains information about who the agent has been in contact with, referred to as neighbors, and number of contacts made with the given neighbor for each time-step.

The magnitude of the contact tracing in the model can be tuned by manipulating three different parameters, illustrated in Figure 3.3.9: Firstly, the duration of the critical time window describes how many hours before symptom onset contacts with potential infectees are traced. The baseline value in the model is set to 24 hours, see Section 2.1.3. Secondly, the proportion of remembered agents denote how many unique neighbors the agent remembers having contact with. The baseline value is set to 0.5. Lastly, the minimum required number of contacts mirrors the threshold value for defining the interaction as a close contact. The baseline value is set to 20 contacts. It is important to emphasize that this parameter value captures the *total* number of contacts between the confirmed COVID-19 case and the given neighbor in the course of the time window.

Figure 3.3.9: The contact tracing following a confirmed COVID-19 case depends on the duration of the critical time window preceding symptom onset, the proportion of neighbors an agent remembers and the minimum number of contacts required to be defined as a close contact.

These three parameters can be tuned for the model as a whole, but applies equally for all agents in the simulation. Discharged patients are not included in the contact tracing. The algorithm below describes the contact tracing following a positive COVID-19 test for an arbitrary agent, using the baseline values described above for illustration.

1. The critical time window is calculated as an hour interval spanning 24 hours from before symptom onset to the hour when the symptoms appear.
2. For each of the hours in this time window, information about neighbors and the number of contacts are examined: For each neighbor, the total number of contacts and the last hour of exposure are updated. The resulting data contains an overview of all agents the agent has had contact with, including the total number of contacts and the last hour of exposure.
3. Of all the neighbors listed, the agent only remembers having contact with half of them. The updated list of remembered neighbors is drawn randomly from the original list, and does for instance not take the number of contacts made with the respective neighbors into account.

4. In the remaining list of neighbors, only those who have had 20 or more contacts with the infected agent during the critical time window are notified. Neighbors who are already in quarantine or isolation, are not notified. The duration of the resulting quarantine is calculated based on the last hour of exposure.

Quarantine and Isolation

Both quarantine and isolation control measures have been mentioned in the context of testing and contact tracing. These measures affect the staffing of shifts and the believably the pathogen transmissibility between agents due to stricter hygiene measures.

Agents may enter a quarantine in one of three possible scenarios:

- The agent has symptoms in accordance with the testing criteria and enters a quarantine pending the test results.
- The agent has an acute respiratory infection, but lacks the key symptoms. The infection is therefore categorized as "mild" and the agent enters a quarantine lasting until 24 hours after the symptoms have passed
- The agent has been in close contact with a confirmed COVID-19 case and enters a quarantine lasting for fourteen days after the last exposure.

Quarantined patients are not discharged, but stay in the ward until the restriction is lifted. Members of the staff cannot work when being in quarantine. If they enter a quarantine during a shift, the user can choose if they are sent directly home or finishing their shift by switching a boolean parameter. Agents only enter isolation if they test positive for COVID-19. The isolation lasts until seven days after symptoms are gone. Similar as for quarantine, patients are not discharged when being isolated and staff members have to stay home. Also here, the user can choose if a member of the staff finishes its shift hours or is sent directly home upon positive test results.

Members of the staff who stay home during a quarantine or isolation, do not pose any risk of infection in the ward as they are not present. Patients, however, who necessarily must uphold some degree of contact with the staff members in order to sustain the medical treatment, may contribute to the spread of the virus. However, quarantine and isolation of patients would believably reduce the probability of transmission due to more stringent infection control procedures. Bringing back the transmissibility p_T, see Equation (3.3.2), the reduced transmission probability is reflected in the variable β: This parameter denotes the relative risk of transmission between a situation with no restrictions, $\beta_N = 1$, and a situation with quarantine or isolation, where β_Q and β_I take values between zero and one, respectively. The baseline values are guessed to $\beta_Q = 0.5$ and $\beta_I = 0.1$. The final expression for transmissibility is presented in Equation (3.3.12),

$$p_T(\alpha, \beta, \gamma, c) = 1 - (1 - \alpha \cdot \beta \cdot \gamma \cdot 0.003)^c, \tag{3.3.12}$$

where α captures the presence or absence of symptoms, β the potential restrictions in terms of quarantine and isolation, γ the relative infectiousness of SARS-CoV-2 compared to the seasonal influenza and finally c which denotes the number of contacts between an infector and a potential infectee in a given hour.

3.3.7 Implementation of the Epidemiological Part of the Model

Before summarizing the model parameters, the following section provides a brief introduction to how the epidemiological part of the model is implemented. The discussion revolves around the two remaining steps of the list presented in Figure 3.2.29, see Figure 3.3.10, namely the pathogen spread on the contact network, including the incubation time, infection period and test results waiting time following a COVID-19 infection, and lastly the summary step collecting and storing data on the epidemiological state of the system.

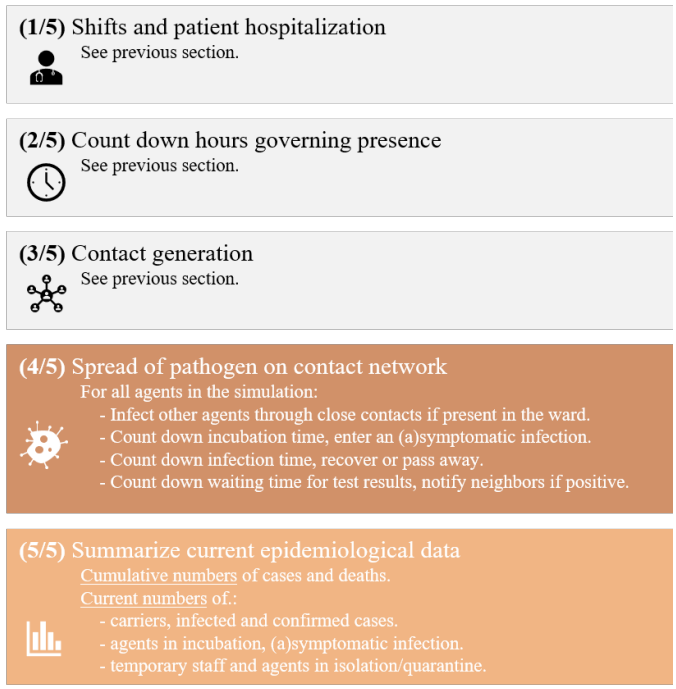

Figure 3.3.10: Each simulation step is divided into five main parts: The three governs the presence and contacts made in the hospital ward(s). The fourth step handles the spread of SARS-CoV-2 on the generated contact network in addition to counting down the remaining hours of incubation periods, infections or test result waiting times. The fifth step summarizes current and cumulative data on the epidemiological state in the system.

4. **Spread of pathogen on contact network**: For each simulation hour h, this step loops through all simulation agents who are current carriers of the virus.

   (a) **Infect neighbors**: If the agent is present in the ward and infectious, the model runs through all susceptible neighbors the agent has had contact with in the given hour. The transmissibility p_T is calculated as shown in Equation (3.3.12), choosing the restriction factor β based on the strictest restriction measure among itself and the neighbor. Whether the neighbor is infected or not is drawn from a Bernoulli distribution where $p = p_T$. Infected neighbors enter

state E and are no longer susceptible. They are assigned with a incubation period t_E.

(b) **Count down incubation time**: If the agent is in the incubation period (E), the remaining hours of t_E are reduced by one for each simulation step. For each step, the model checks if the remaining hours correspond to the time point when presymptomatic transmission begins. At this point, the α_N is changed to α_I. When zero hours of the incubation period remains, the agent enters an infection: The course of disease, symptomatic or asymptomatic, is drawn using a Bernoulli distribution with $p = p_A$ and the duration of the infection is drawn from the corresponding distributions. Based on the potential presence of symptoms, testing and quarantine measures are implemented as described in Section 3.3.6.

(c) **Count down remaining infection hours**: If the agent is in a state of infection (A or I), the remaining hours of infection are counted down. When reaching zero, the outcome is determined by the following factors: In case of a symptomatic infection, a potential fatal outcome is drawn from a Bernoulli distribution where $p = p_D$. If the agent belongs to the PAT role, the potential fatal outcome from a secondary infection is determined as described in Section 3.3.4. If not dying from either infections, the agent may return to state S depending on the outcome of a Bernoulli distribution where $p = p_S$. If not losing its immunity, the agent recovers. In any case, the agent is no longer a carrier of the virus and the infectiousness is reset to α_N.

(d) **Wait for test results**: If the agent is awaiting test results, the remaining waiting time is reduced by one for each hour time step. When the test results are available, the potential control measures regarding isolation and contact tracing is initiated as explained in Section 3.3.6. For an agent with a mild infection, the remaining waiting hours before a test is conducted are counted down. When reaching zero, the agent is tested if the infection persists.

5. **Summarize current epidemiological data**: The model keeps track of the epidemiological states by lists and map structures counting the number of agents satisfying certain criteria for each simulation step. For instance, the number of temporary staff working the given hour is found by sorting the agents based on if they are currently present in the ward and that they are not part of the regular staff. A list the summarized data is presented in the next section.

After the initialization step, described in the previous section, a given number of the initialized agents can be infection in order to function as patient zero in the ward. The number of agents, which roles they belong to and if they should belong to the same ward or not, are parameters set by the user. In the baseline parameter set, the initially infected agents consist of two NUR agents belonging to the same ward.

3.4 Parameter Values

This chapter is concluded by listing the parameter values required to run the model. The following table presents model parameters which can be defined by the user. The table provides information about the parameter name, the data type, a short description and the baseline value used in the model. An overview of the constant parameter values, many of them already presented in this section, is presented in Appendix B.2.7.

Table 3.4.1: Overview of the model parameters based on user input. The parameter are divided into five categories in order to make it easier to navigate. The columns present the parameter name and type, in addition to a brief description and the baseline value. *The presented Python built-in data types are described briefly in Table B.1.2.

Parameter	Type*	Description	Baseline
(1 / 5) General			
plot_results	bool	Summarizes main results in graphs	True
log_as_network	bool	Stores the evolving contacts in a network file	True
print_to_console	bool	Prints messages and progression bar to console	True
measure_elapsed_time	bool	Measures the elapsed simulation time	True
number_of_days	int	Number of simulation days	90
number_of_wards	int	Number of distinct hospital wards	2
beds_per_ward	int	Number of beds per hospital ward	19
(2 / 5) Contacts and shifts			
contact_randomization	float	Share of all contacts within a ward drawn randomly	0.0
ward_isolation	float	Share of all contacts that are made within the ward	1.0
shift_assignment	str	Sets shift assigned ('random' or 'percentage')	'percentage'
resting_time	bool	Sets if resting hours for NUR shifts are taken into account	True
(3 / 5) Control measures			
social_distancing	float	Proportion of avoided contacts	0.0
t_quarantine	int	Quarantine days after last exposure	14
t_isolation	int	Isolation days after symptoms are gone	7
t_milder	int	Quarantine days after mild symptoms are gone	1
test_results_waiting_time	int	Hours of waiting before results are available	36
close_contact_definition	int	Number of contacts required for close contact	20
t_notify_contacts	int	Critical window for contact tracing	24
p_remember	float	Proportion of unique individuals remembered	0.5
t_quarantine_exception	int	Potential quarantine exception made for HCW	14
send_staff_directly_home	bool	Forces staff to leave the ward if quarantine/isolation	True
(4 / 5) COVID-19			
incubation_time	dict	Lognormal parameter values for t_E	{'mu': 1.6067, 'sigma': 0.4427}
asymptomatic_time	dict	Lognormal parameter values for t_A	{'mu': 1.9459, 'sigma': 0.2980}
infection_time	dict	Lognormal parameter values for t_I	{'mu': 2.6931, 'sigma': 0.4016}
infection_ratio	float	Relative symptomatic infection rate, STAFF : PAT	0.5
p_asymptomatic	dict	Probabilities for asymptomatic infection	{'STAFF': 0.43, 'PAT': 0.18}
p_death	dict	Probabilities for fatal outcome among symptomatics	{'STAFF': 0.00650, 'PAT': 0.327}
p_vaccination	float	Probability of being vaccinated	0.0
p_efficacy	float	Probability of obtaining immunity from vaccination	0.60
p_susceptible	float	Probability of not obtaining immunity after infection	0.01
t_presymptomatic	float	Number of infectious days before symptom onset	2.58
alpha	dict	Relative infectiousness to symptomatic infection	{'N': 0.0, 'A': 0.5, 'I': 1.0}
beta	dict	Relative infectiousness based on restriction level	{'N': 1.0, 'Q': 0.5, 'I': 0.1}
gamma	float	Relative infectiousness to the seasonal influenza	2.0
patient_zero	dict	Information about initially infected agents	{'n': 2, 'role': 'NUR', 'ward': True}
(5 / 5) Secondary infection			
p_secondary	float	Probability of acquiring a secondary infection	0.50
p_antibiotics	float	Probability of being treated with antibiotics	0.95
p_resistant	float	Proportion of bacterial infections being resistant	0.32
p_D_sensitive	float	Probability of dying given a sensitive strain	0.19
p_D_resistant	float	Probability of dying given a resistant strain	0.68

Chapter 4

Results and Analysis

The previous chapter gave a detailed introduction to the logic and the assumptions underlying the network generation and epidemiological model. The resulting modeling framework functions like a black box where the input parameters culminate in a set of output data based on a range of algorithms. It is therefore important to test if the machinery works as intended. This chapter is introduced with an overview of the developed modeling framework, including its main structure and output types. The second part presents the results from a selection of self-consistency tests. The third part assesses the model stability and compares the variability in outbreak sizes. Lastly, the fourth part compares temporal snippets of the simulated network with the original, empirical contact network through static and temporal analyses.

4.1 The Modeling Framework

The modeling framework is built in Python and can be downloaded from GitHub (`https://github.com/signesaevareid/covid-19`). The code is separated in three different files: ward.py serves as the main modeling unit and imports all parameters from parameters.py and the classes governing the behavior of all agents in the ward from classes.py. The output files are stored in a created output folder.

The model can be divided into three main part, illustrated in the Figure 4.1.1: The input consists of parameters which the user can tune freely and constant parameters derived from the literature or the analysis of the empirical contact network. After these parameter values are set, the simulation step includes the initialization of agents followed by the main simulation. In each simulation step[1] the contact network evolves allowing for the transmission of SARS-Cov-2 through the generated inter-individual contacts. The model output is stored in four different files: The main model output is listed in a text file. This file contains the calculated case fatality rate and infection fatality rate, followed by several dictionaries listing epidemiological data, for instance the number of carriers or confirmed

[1] In order to reduce running time, the simulation is stopped when there are no carriers present in the ward.

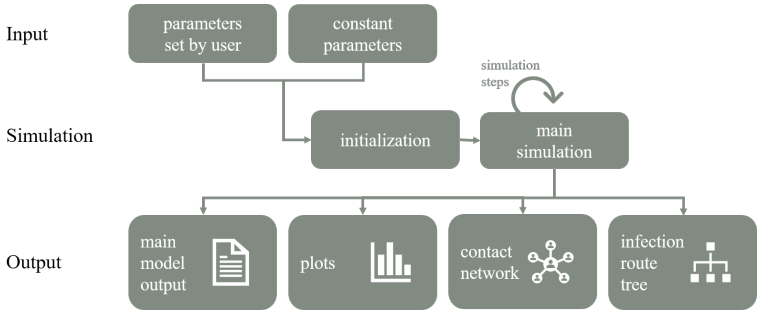

Figure 4.1.1: Workflow of the model simulating the spread on SARS-CoV-2 on a temporal contact network: The input parameters consist of those set by the user and other constant parameters. After agents are initialized, the main simulation proceeds with a given number of simulation steps in which the contact network evolves and the pathogen is spread. The model output is summarized in a text file, several plots and two networks presenting the resulting contact network and a tree network representing the routes of infection.

cases, for each time step. The file also lists the cumulative number of confirmed cases and deaths, see Table C.1.1 for a complete overview. The second model output is a plot consisting of four panels, see Figure 4.1.2. They portray graphical illustrations of the

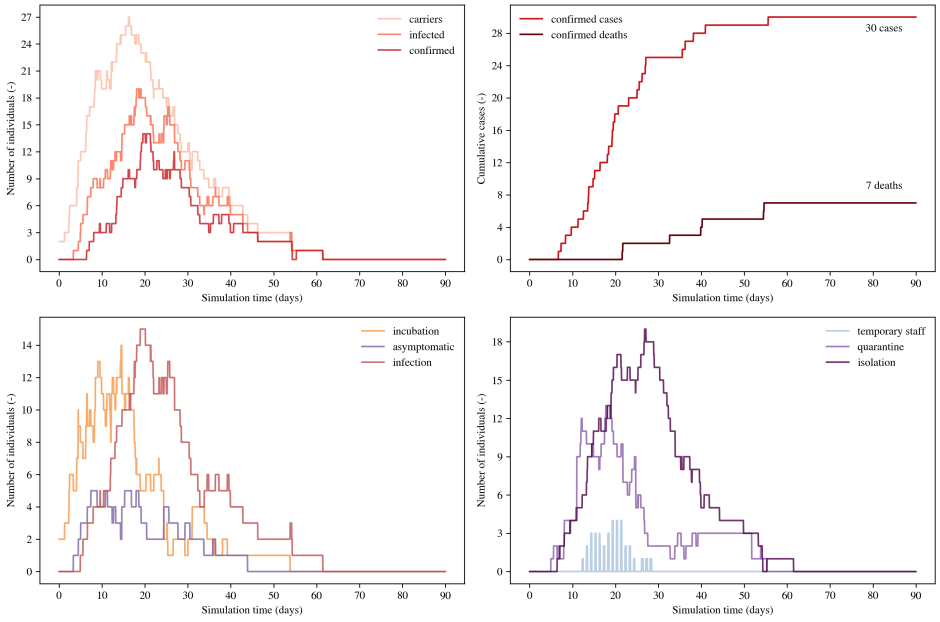

Figure 4.1.2: An example of a model output plot. The four different panels present graphical illustrations of how the key epidemiological data vary in the course of the simulation days.

epidemiological modeling output given in the text file. The upper left panel presents the number of agents carrying the virus (comprising states E, A and I), the number of infected agents (states A and I) and lastly the confirmed cases (state I with positive test). The upper right panel shows how the cumulative numbers of confirmed cases and deaths evolve with time. The lower left panel presents a dissection of the 'carriers' curve in the panel above, where each line represents agents in incubation period, asymptomatic and symptomatic infection, respectively. Lastly, the lower right panel shows how the number of quarantined and isolated agents, as well as the number of temporary staff, varies with time.

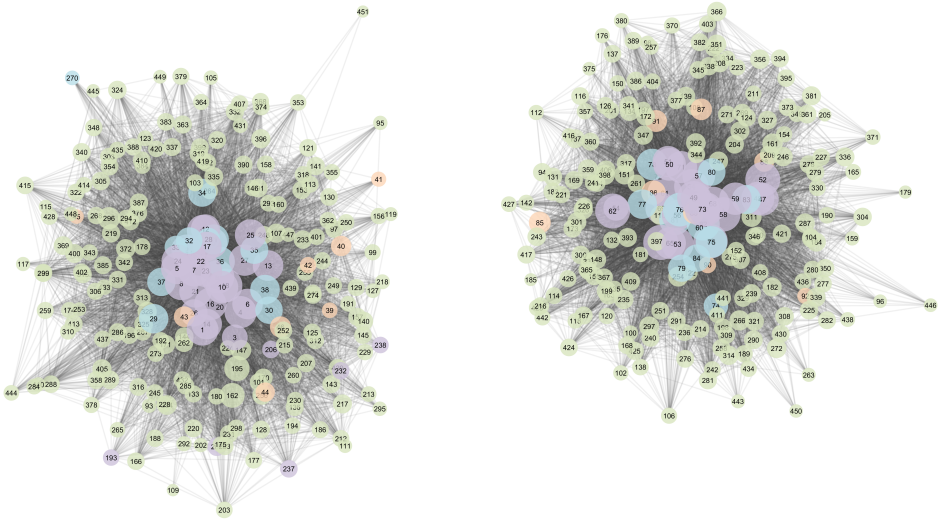

Figure 4.1.3: An example of the third model output; the generated contact network. This static aggregation shows two hospital wards with no cooperation across the them. The network therefore consists of two distinct components. The links are undirected and the node size increases with increasing node degree. The colors correspond to the different roles, where nodes belonging to the NUR role are colored in purple, MED in blue, ADM in orange and PAT in green. The simulation is run with with the baseline set of parameter values.

The third and fourth model outputs consist of network files: The former network contains the generated contact network where each node contains information about the agent identification number, as well as the given role and ward number. The links contain information about the total number of contacts and the listed hours of when these contacts were made. Figure 4.1.3 shows an example generated from the same simulation run as the plots presented above. Figure 4.1.4 presents an example of the fourth model output, namely a directed network showing the infection routes in the ward. As this specific example shows, the two NUR agents with identification numbers 18 and 25 were initially infected. Only one of them passed the infection on to other agents in the ward, resulting in a total number of 46 cases. In addition to the identification number, role and ward,

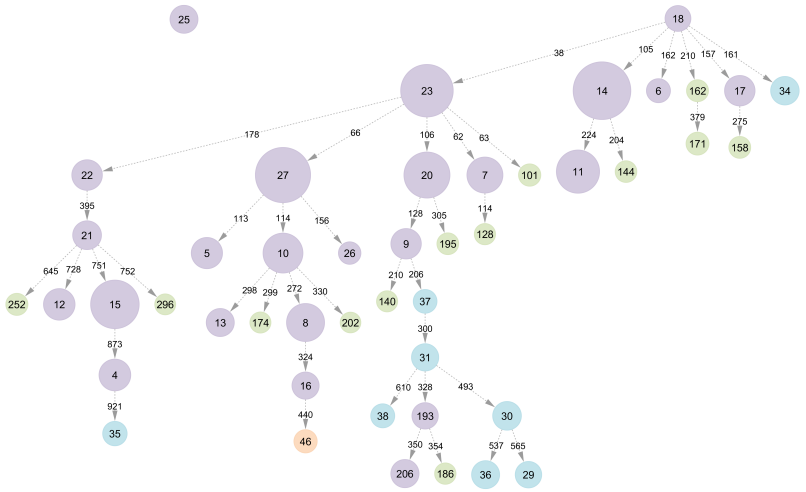

Figure 4.1.4: An example of the fourth model output; an infection tree illustrating the routes of infection in the ward. The network is directed and the node size increases with an increasing degree of extroversion. The colors correspond to the different roles, where nodes belonging to the NUR role are colored in purple, MED in blue, ADM in orange and PAT in green. The hour of which the agent was infected is denoted on the link. The simulation is run with with the baseline set of parameter values.

the node contains information about the degree of extroversion and the hour of which the agent caught the virus. In the network visualization, the node size is set to increase with the degree of extroversion.

4.2 Model Validation and Self Consistency

In order to get an indication of whether the black box machinery works as intended, a few self consistency tests were conducted. In general, they were conducted by setting a parameter to an extreme value and observe if the model output was as expected.

The Effect of Not Obtaining Immunity

In the baseline parameter set, the probability of not obtaining immunity after going through an infection is set to 0.01, indicating that one out of 100 agents, on average, return to a susceptible state after recovering from the infection. Given this parameter set, most simulation runs result in an outbreak that eventually dies out, see the example presented in Figure 4.1.2. If this parameter value of p_S is set to 1.0, indicating that no agents obtain immunity, reoccurring waves of infection are expected, in addition to that the cumulative number of confirmed cases and deaths should keep increasing instead of flattening out. In order to test this hypothesis, a simulation was run over 300 simulation days where the p_S value was set to 1.0. As a reference, another simulation was run with $p_S = 0.0$. The resulting output plots are given in Figure 4.2.1.

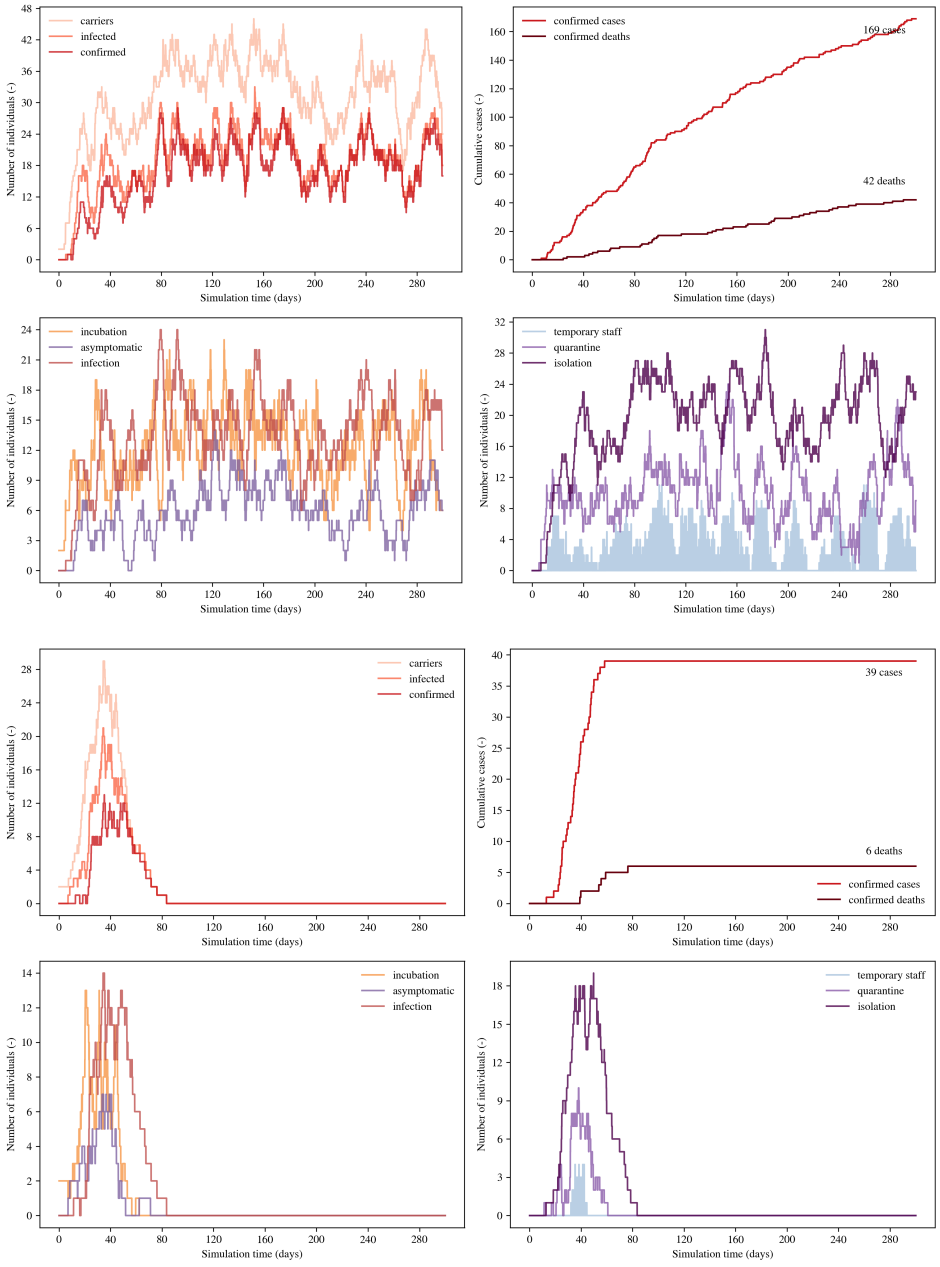

Figure 4.2.1: The upper four panels present the model output plot in the extreme case where all agents return to a susceptible state after recovering from COVID-19, that is $p_S = 1.0$. The lower panels present the model output plot from a reference run where $p_S = 0.0$. Except from the p_S parameter and the number of simulation days which is set to 300, all other parameters are set to their baseline values.

In the very upper left panel of Figure 4.2.1, representing the situation where no agents obtain immunity, the number of carriers, infected and confirmed cases reaches a top after approximately three weeks, then decreases for a few days before the numbers start increasing again. This pattern is observed throughout the 300 days long simulation run, and is also reflected in the two lower panels. The number of confirmed cases, see the upper right panel, increases the most in the beginning. The increase somewhat levels off, but the number of cases keeps increasing throughout the simulation. The number of deaths seem to follow a fairly steady slope. In comparison, the upper left and two lower panels of the reference run shows one outbreak peak, before the infection dies out after around 80 days. The cumulative numbers of cases and deaths stops increasing when there is no longer carriers left in the ward. These numbers end up one order of magnitude below the observed number of cases and deaths in the extreme scenario.

Social Distancing

The social distancing parameter, which reduces the total number of hourly contacts, is set to 0.0 in the baseline parameter set. If it is set to 1.0, it is expected that the only carriers in the hospital wards are those who were initially infected, hence giving an infection route tree with two isolated, purple-colored nodes. It is also expected that the resulting contact network consists of a number of isolated nodes, since 100 % of contacts are avoided. This hypothesis was tested by performing 50 simulation runs with the social distancing set to

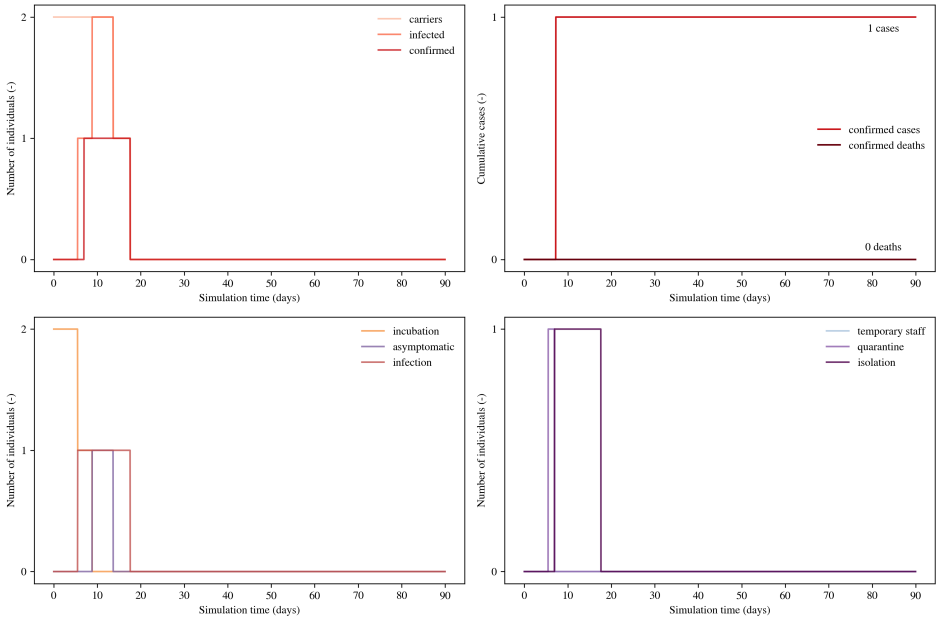

Figure 4.2.2: The model output plot from a simulation where the social distancing is set to its maximum value, 1.0. All other parameters are set to their baseline values.

its maximum value and all other parameter set to their baseline values. The resulting output plot from one of these runs is presented in Figure 4.2.2. The corresponding contact network and infection route tree are presented panels A and B of Figure 4.2.3, respectively.

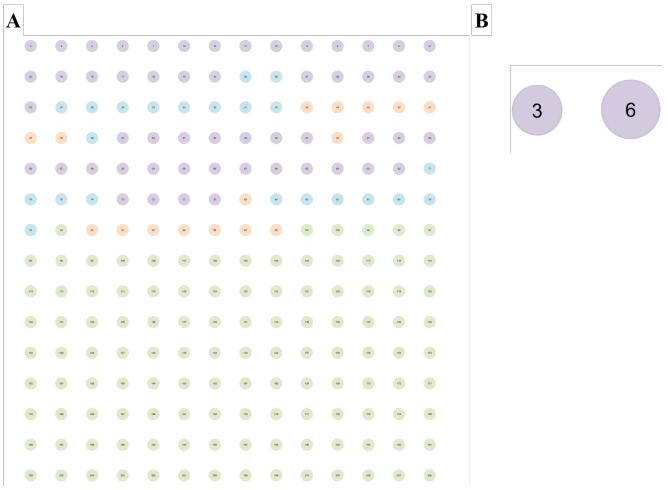

Figure 4.2.3: Panel A shows a contact network generated from a simulation where the social distancing was set to its maximum, 1.0. This network consists of 210 isolated nodes, each corresponding one of the simulation agents. Panel B shows the tree of infection routes belonging to the the same simulation run. All parameters, except from the social distancing, were set to their baseline values.

All of the 50 simulation runs resulted in two SARS-CoV-2 carriers, equivalent to the number of initially infected agents. As Figure 4.2.2 demonstrates, after these two agents had finished their courses of disease, there was no longer any carriers left in the ward. In this specific run, only one of the two agents developed symptoms. Figure 4.2.3 presents the corresponding network outputs: A contact network consisting of 210 isolated nodes and a tree of infection routes presenting the two initial carriers without any infectees.

Setting the Relative Infectiousness to 100

In the baseline parameter set, the relative infectiousness compared to the seasonal influenza is set to $\gamma = 2$. Here, it is tested at $\gamma = 100$ in order to investigate how the model behaves in response to an extremely infectious spreading agent. Since the probability of transmission depends on several factors, for instance reduced in the absence of symptoms or in case of quarantine or isolation, it is expected that the majority, but not necessarily all agents, belonging to the ward where the virus is spread become infected. This hypothesis was tested by running a simulation with $\gamma = 100$ and all other parameters set to their baseline value. The resulting output plot is presented in Figure 4.2.4 and the corresponding infectious route tree in Figure 4.2.5.

As the model output plot shows, the high relative infectiousness results in an extensive outbreak where the number of confirmed cases reaches a top between two and three weeks after the infection was introduced to the ward, before decreasing again. Following

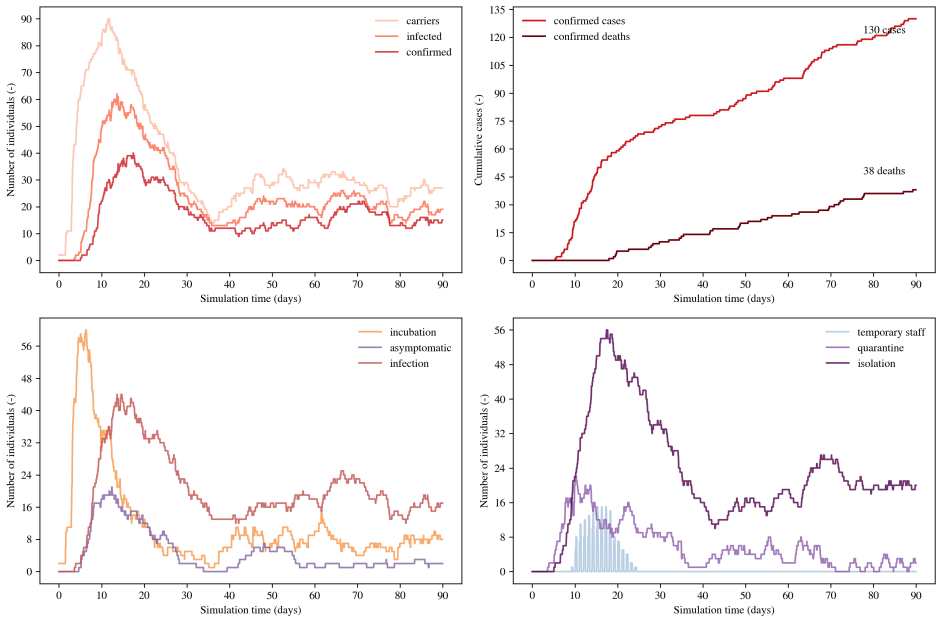

Figure 4.2.4: The model output plot from a simulation where the relative infectiousness is set to $\gamma = 100$. All other parameters are set to their baseline values.

the cumulative number of confirmed cases in the upper right panel, the curve seems to start flattening out in the subsequent period, but does not level off as observed when the infection died out. Instead, the number of cases persists in the ward with smaller increases and decreases, but without resulting in a new, large outbreak. The component of the contact network corresponding to nodes belonging to the ward where this outbreak took place consists of 229 nodes, of which 203 nodes (88.65 %) are present in the resulting infection route tree, see Figure 4.2.5. Thus, almost nine out of ten individuals who were present in

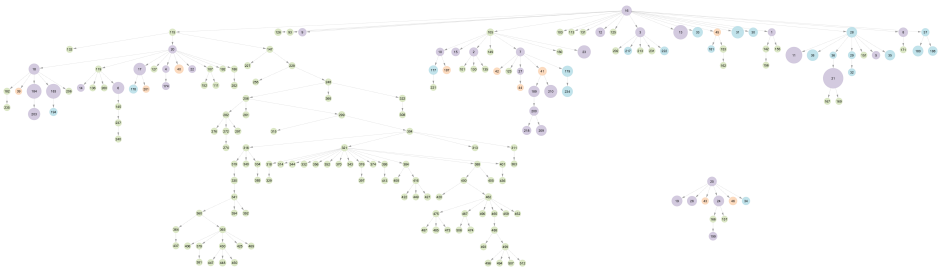

Figure 4.2.5: An infection route tree from a simulation where the relative infectiousness is set to $\gamma = 100$. The network consists of 203 nodes. The hours denoting the time of infection are omitted for enhancing readability.

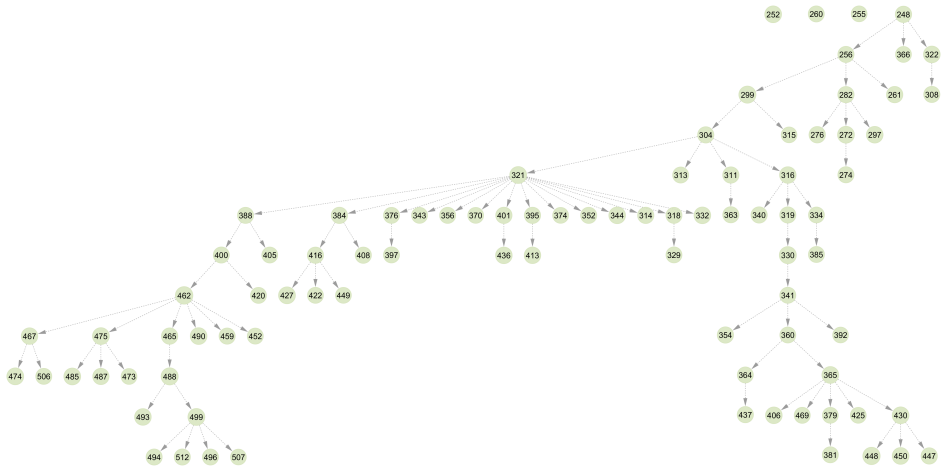

Figure 4.2.6: A filtered infection route tree where all nodes correspond to agents that were infected in the last two-thirds of the simulation, from day 25 to 90. This network consists of 88 of which all belong to the PAT role. The hours denoting the time of infection are omitted for enhancing readability.

the ward in this simulation were infected with SARS-CoV-2. It is also worth noticing the curve showing the number of temporary staff present in the ward in the lower right panel of Figure 4.2.4: The number of temporary staff present in the ward is considerable between approximately 10 and 25 days before laying flat at zero in the remaining simulation time. This indicates that all the shifts after the initial outbreak was covered by the regular staff in the ward, even though there were still a considerable number of infection cases in the ward. This result points towards that all regular staff members obtained immunity through the initial outbreak, and that the continuing cases resulted from the admission of susceptible patients. To examine this hypothesis, the infection route tree was filtered according to the hour of infection: Figure 4.2.6 shows the nodes in the infection route tree corresponding to agents that were infected *after* 25 simulation days. This network consists of 88 nodes, all corresponding to agents belonging to the PAT role.

4.2.1 Stochasticity and Model Stability

The developed modeling framework is stochastic. Thus, feeding two simulations with identical input parameters could result in two outputs differing significantly from each other. In order to investigate the model stability, 100 simulations with identical parameters were run. The parameter set was equal to the baseline parameters, except for the number of wards which was set to one. The number of SARS-CoV-2 carriers was used as a model indicator. Figure 4.2.7 shows the average number of carriers present in the ward at each time step in the simulation period plotted as a solid line. The colored bands represent the standard deviation error bars, and span the interval $\bar{x} \pm s$. As the figure shows, the width of the band is narrow in the very beginning of the simulation period before it widens with the

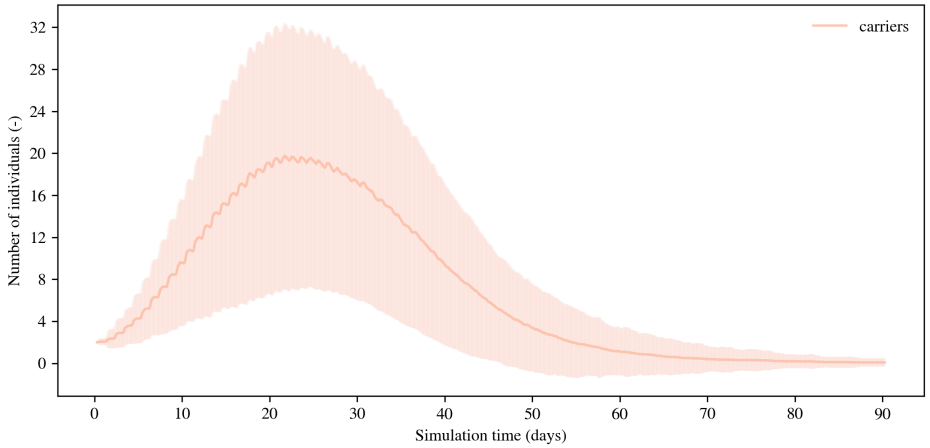

Figure 4.2.7: The figure represents data obtained from 100 simulation runs using the baseline parameter set. The solid line represents the average number of SARS-CoV-2 carriers in the ward calculated from these runs. The bands correspond to error bars representing the standard deviation and span the interval $\bar{x} \pm s$.

increasing mean value. Towards the end of the simulation, the bands narrow as the average number of carriers approaches zero. Corresponding curves for the number of infected and confirmed cases are presented in Figure C.2.1.

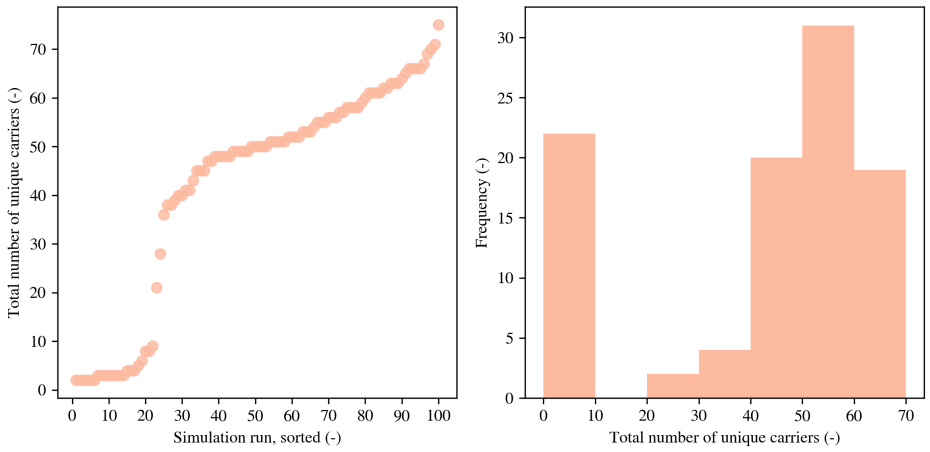

Figure 4.2.8: Illustration of the total number of unique SARS-CoV-2 carriers in the ward obtained from 100 simulations using the baseline parameter set. The left panel shows a scatter plot of the total number of unique carriers registered in the ward, sorted in ascending order. The right panel presents the corresponding histogram.

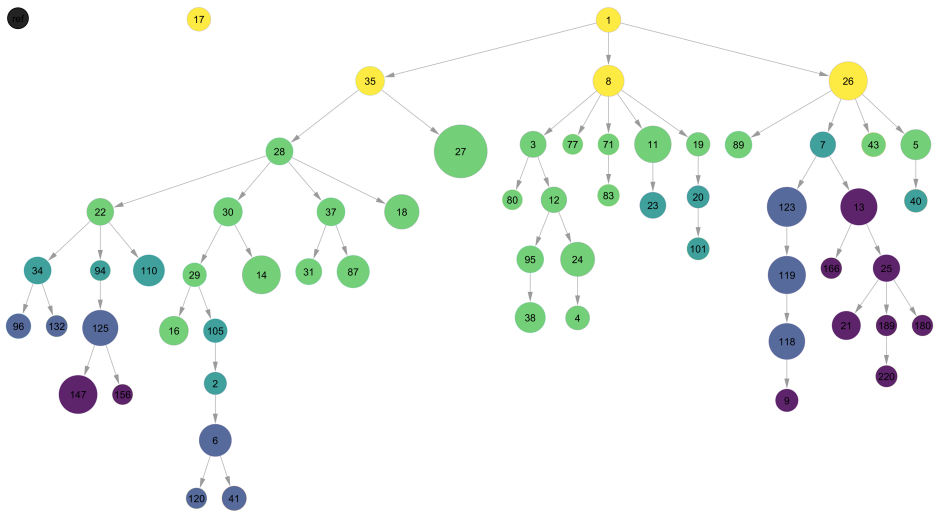

Figure 4.2.9: Infection route tree for a simulation run resulting in 60 unique carriers of SARS-CoV-2 in the ward. The colors of the node correspond to the week of infection, where yellow denotes week number one, green week number two, turquoise week number three, blue week number five and purple week number six or later. The black-colored node is included for size reference across figures. baseline parameter set, except from the number of wards which is set to 1.

The output results from the 100 simulation runs were examined in order to identify key differences between the simulations resulting in differing output results. Firstly, the total number of unique carriers in each run was examined. The left panel of Figure 4.2.8 shows a scatter plot where the number of unique carriers for all runs were sorted in ascending order. The right panel in the same figure shows the corresponding histogram. As these results show, roughly one-fifth of the simulations resulted in less than ten SARS-CoV-2 carriers in the course of the 90 simulation days. When the number of total carriers exceeds ten, an abrupt jump is observed, resulting in larger outbreak sizes.

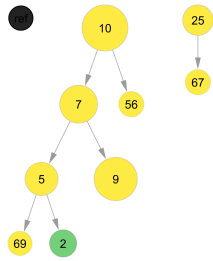

Figure 4.2.10: Infection route tree for a simulation run resulting in 9 unique carriers of SARS-CoV-2 in the ward. The colors of the node correspond to the week of infection, where yellow denotes week number one and green week number two. The black-colored node is included for size reference across figures. baseline parameter set, except from the number of wards which is set to 1.

Figures 4.2.9 and 4.2.10 show two examples of infection route trees obtained from the stability assessment. The node colors reflect the week of which the agents were infected. The former figure illustrates a run which resulted in 60 unique carriers. Of these, 5 agents were infected in the course of the first week, 26 in the second week, 10 in the third week, 9 in the fourth week, 4 in the fifth week and 6 in the sixth week or later. In comparison, the latter figure shows the infection route tree of a run resulting in 9 unique carriers. Of these 8 and 1 agents were infected in the first and second week, respectively. Of the nodes infected in the course of the first week, the average number of infectees per infector was 2.8 and 0.875, respectively.

4.2.2 Spread Across Several Wards

The model allows for up-scaling of the hospital capacity, both in terms of wards and beds per ward. In addition, the user can determine the degree of cooperation across the wards by tuning the ward isolation parameter. To demonstrate how the interaction between agents belonging to different wards may affect the spreading, a simulation was run by setting the number of wards to 4 and the ward isolation to 0.9. Thus, on average, an agent will have a ten times higher chance of interacting with another agent from its own ward than agents belonging to other wards. Agents belonging to the PAT role cannot interact with other PAT

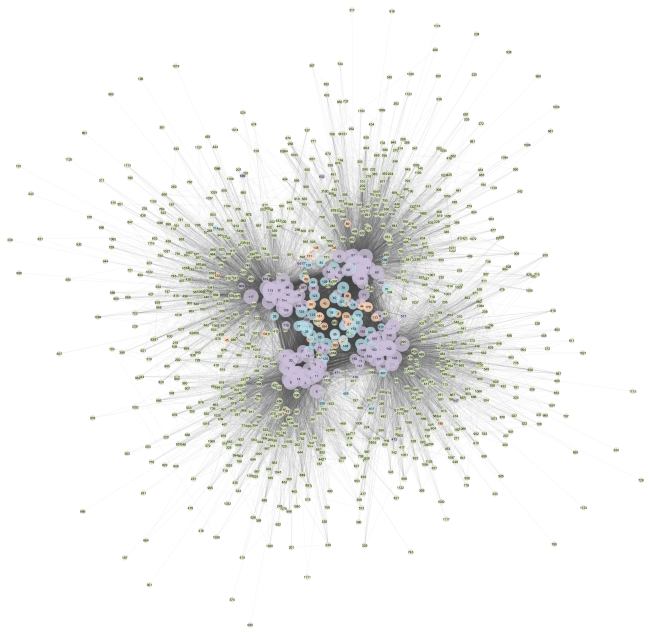

Figure 4.2.11: The contact network a simulation across four different hospital wards where the ward isolation is set to 0.9. Links representing fewer than five interactions have been removed and isolated nodes are deleted from the network. The node size increases with increasing node degree and the link thickness increases with number of interactions.

Figure 4.2.12: The thinned contact network generated from running a simulation over four wards yields four distinct components when filtering the nodes based on the PAT role.

agents across the wards.

The static aggregation of the resulting contact network is dense, with 55 993 links interconnecting 1133 nodes. In order to make the network a bit more sparse, all links representing interactions with fewer than five contacts were deleted. This resulted in approximately half the number of links. Figure 4.2.11 presents the connected component of the thinned network, where all isolated nodes are removed. As the figure shows, nodes representing staff agents are centered in the middle, with nodes corresponding to PAT agents spring out towards the four different corners. To check whether agents of the PAT role were interacting across the wards, the contact network presented in Figure 4.2.11 was

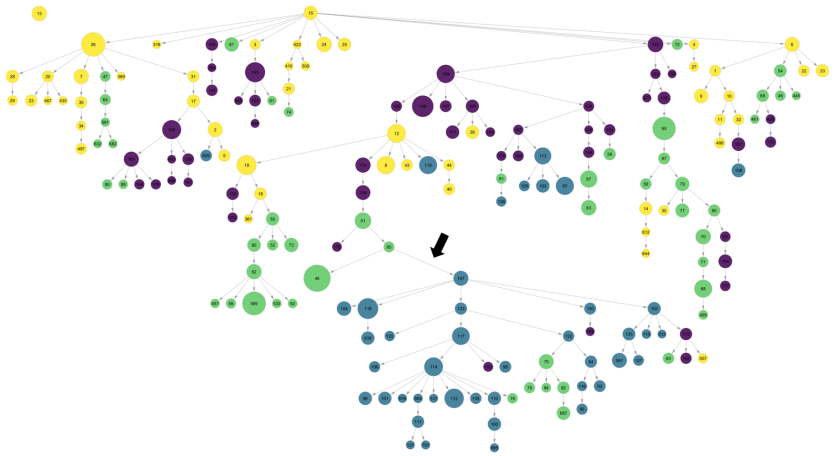

Figure 4.2.13: A tree of infection routes illustrating the spread of SARS-CoV-2 from one hospital ward (yellow-colored nodes) to three other wards through inter-ward contacts. The arrow points at an example of a transmission which results in a large number of cases in the the ward of which the infectee belongs to.

filtered selecting for the PAT role. Isolated nodes were deleted. The resulting network consists of four distinct components corresponding to each of the four wards, as seen in Figure 4.2.12.

The given simulation resulted in 193 SARS-CoV-2 carriers, of which 122 were confirmed by laboratory tests. Figure 4.2.13 shows the tree of infection routes. Here, however, the colors represent each of the four wards and not the roles: Yellow-colored nodes correspond to agents belonging to the ward in which the infection started. The green-colored nodes represent agents belonging to ward number two, blue-colored nodes to ward number three and lastly purple-colored nodes to ward number four. The figure shows that the virus has spread from ward number one to all other wards in the hospital. The arrow in the figure points at a link representing a transmission resulting in a large number of cases in ward number three: This transmission results in 35 new cases in ward number three.

4.3 Comparison of Empirical and Simulated Network

The simulated contact networks may span several weeks, or as long as the user desires. The original, empirical contact network, on the other hand, spans 97 hours. In order to compare these two networks, one could therefore extract a temporal snippet from the simulated network, containing contacts starting from 13:00 on the first simulation day and lasting for 97 hours. Thus, for all links in the generated network, the time-points indicating hours of contact are filtered, retaining contact hours $\in [13, 109]$. All other time-points of detected contacts are removed from the list and links where no contacts falling within this time window are deleted from the filtered network.

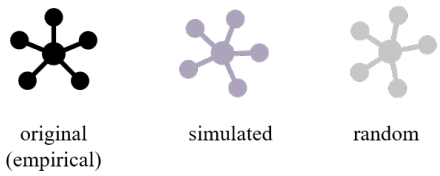

Figure 4.3.1: The original, empirical contact network is compared to temporal snippets of simulated contact networks, using randomized contact networks as reference.

In the following analysis, the empirical network is compared to the snippets of simulated contact networks using randomized contact networks as a reference, see Figure 4.3.1. When running simulations for investigating the structure of the resulting contact networks in this very short time span in the very beginning of the simulation, the simulations are run without any spreading infection. Thus, the number of initially infected agents is reduced from 2 to 0 compared to the baseline parameter set. In addition, since there is only one ward in the empirical network, the number of virtual wards is also set to 1. In the randomized reference network, the following additional parameter values deviate from the baseline set: The contact randomization is set to 1.0, the shift assignment to 'random' and the parameter governing the NUR role resting time is set to false. The hourly number of

contacts are still intact in the randomized network. In the following, a few static and temporal analyses of the empirical and simulated network are performed using a randomized network as control.

4.3.1 Analysis of Static Aggregation

For the purpose of comparing the simulated with the empirical contact network, ten simulations were run using the parameters described above. After filtering out the contacts corresponding to the time window of interest, the degree of each node was calculated. In Figure 4.3.2, the black line shows the degree of each node in the static aggregation of the empirical network, sorted in descending order. The colored lines represent the node degrees in the simulated networks, whereas the gray lines represent the randomized networks. From this figure, the simulated network seems to overestimate the node degrees for agents belonging to the NUR, MED and PAT roles, whereas a an underestimation or a partially overlap is observed for the ADM role. In the upper left panel, the degrees of the simulated network is somewhat entangled in the degrees of the random network, although they on average seem to be positioned closer to the node degrees of the empirical network. In the three other panels, the simulated network does a noticeably better job in

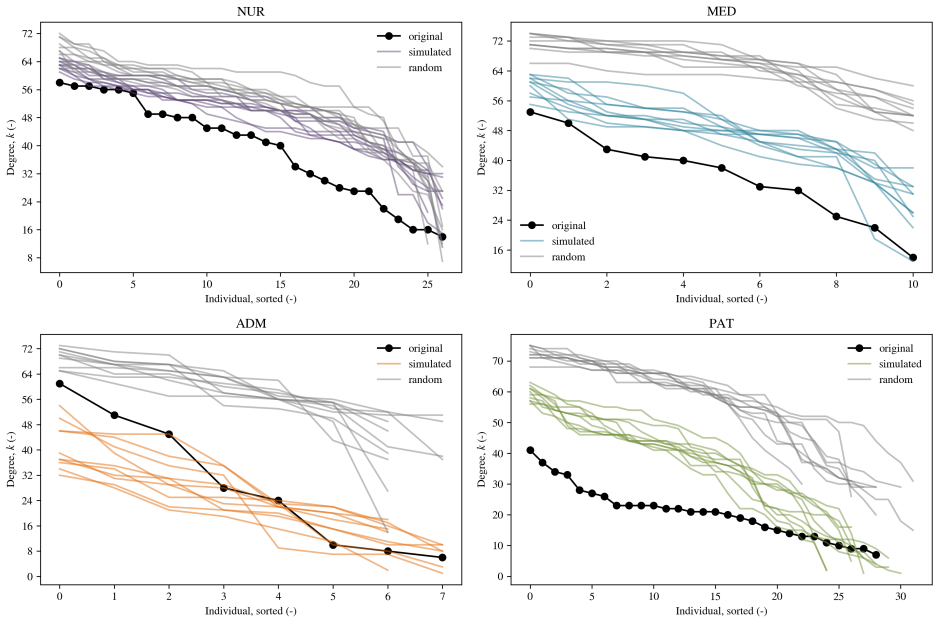

Figure 4.3.2: Comparison of the degree distribution of nodes comparing the empirical contact network, 10 simulated networks and 10 randomized reference networks. Each panel represents the degrees of the nodes in the static aggregation of the empirical contact network (black line with circular markers), and temporal snippets of ten realizations of simulated (colored lines) and randomized (gray lines) contact network. The degrees are sorted in descending order.

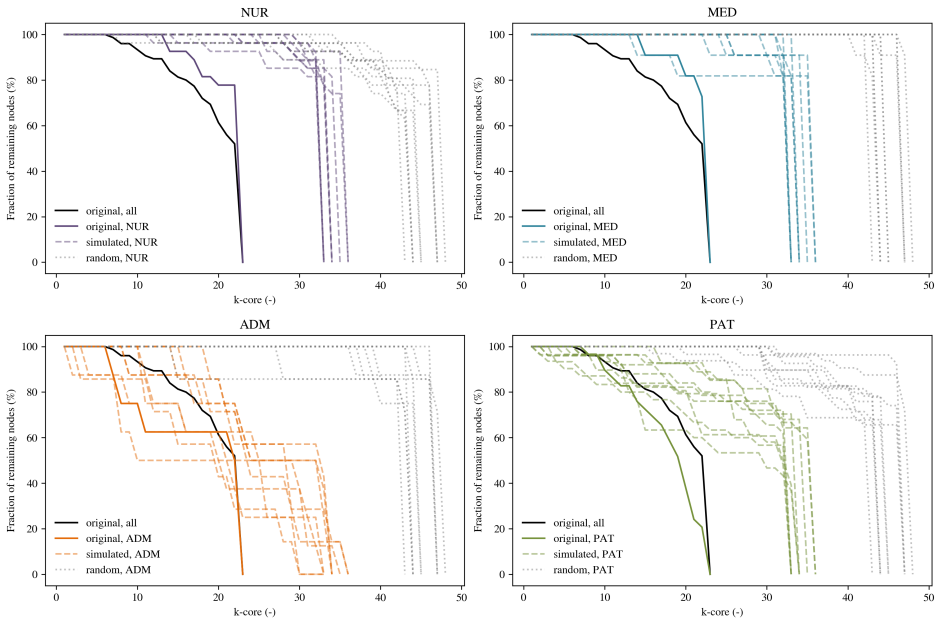

Figure 4.3.3: Comparison of the fraction of remaining nodes as function of k-core comparing the empirical contact network, 10 simulated networks and 10 randomized reference networks. Each panel represents how the fraction of remaining nodes in the network evolves with the k-core for each of the four ward roles.. The black line in all four panels represents the fraction of remaining nodes in the static aggregation containing nodes from all four roles. The solid, colored lines represent the fraction of remaining nodes belonging to the given role. Similarly, the dashed, colored lines represent the role-specific nodes remaining in the simulated networks and the dotted, gray lines correspond to the randomized reference.

approximating the node degrees than the randomized reference.

A note on the registered number of individuals in each plot: Whereas the number of patients present in the ward during this time window depend on the drawn lengths of hospitalization, the number of initialized agents belonging to the staff roles are kept constant. As the figure shows, however, the numbers of NUR and ADM individuals do in some simulation runs deviate from the initialized number of agents. One explanation may be that lacking agents were not chosen for any of the shifts in the time window, or that its corresponding node in the contact network did not have any links to other nodes with contacts registered within the same hour interval.

Another analysis which was performed based on the node degrees of the static aggregations was a k-core analysis. In this analysis, the number of nodes with degrees less than k was recursively removed, resulting in a reduction in number of remaining nodes in the network with an increasing k. Figure 4.3.3 presents the percentage of remaining nodes in the respective networks as function of the k-core: The black, solid line in all panels represent all nodes in the empirical network, whereas the solid, color-coded lines represent the number of remaining nodes belonging to the given role in the empirical network. The colored,

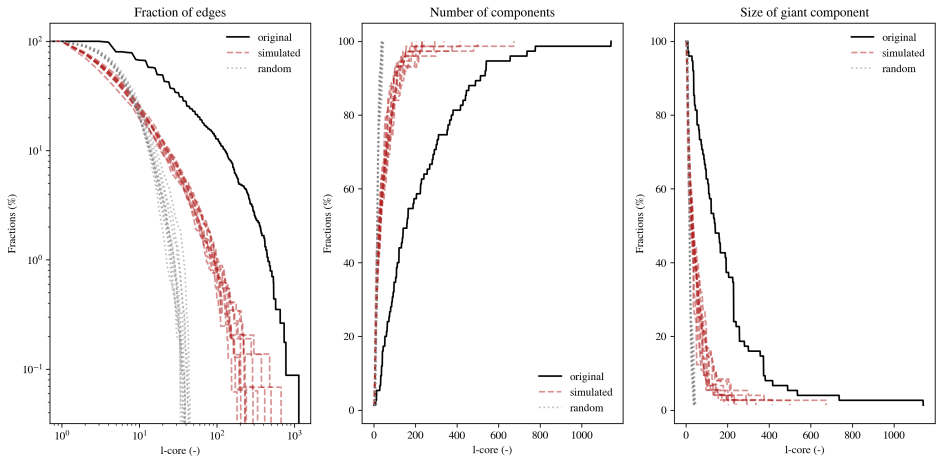

Figure 4.3.4: The panels present the fraction of remaining links, the number of distinct components and the relative size of the giant component as function of the minimum number of contacts in all links in the network. The black lines correspond to the empirical network, the dashed, red lines to the simulated network and the dotted, gray lines to the randomized reference.

dashed lines represent the simulated networks, whereas the dotted gray lines correspond to the randomized reference networks. In general, the empirical network breaks down faster than the simulated network does, which again breaks down faster than the randomized reference. Thus, the same pattern as above is observed: The curves of the simulated networks lie closer to the empirical network curves than the randomized curves do. In particular, the percentages of remaining nodes for the ADM role are partially overlapping the observed evolution in the original data. However, none of the plotted percentages of remaining nodes in the k-core fall as quickly as observed for the empirical data. Thus, the static aggregations of the simulated (and randomized) networks are more interconnected than the empirical network.

Finally, a similar analysis was performed by breaking the network down based on the total number of registered contacts in each link, here referred to as the l-core. Starting by deleting all links with 1 registered contact, the fraction of remaining links, the number of distinct components and the relative size of the giant component was calculated. The procedure was repeated by increasing the number of required contacts until no links were left in the network. The resulting plots are presented in Figure 4.3.4: The black lines correspond to the empirical network, the dashed, red lines to the simulated networks and the dotted, gray lines to the randomized reference network.

The results presented in Figure 4.3.4 show that some of the nodes in the empirical network are interconnected by links with considerably higher numbers of contact than both the simulated and randomizes networks, although the behavior of the simulated network better resembles the behavior of the empirical network than the randomized network does. The results show that the links in the random network contain few contacts.

4.3.2 Temporal Analysis

In order to get an indication of the temporal network properties, analyses of the network reachability ratio was performed on the empirical network, one randomly chosen simulated network and one randomly chosen randomized reference network. For each of these networks, the reachability ratio was calculated as function of both the maximum allowed waiting time and the link weight corresponding to the number of registered contacts in the given hour. The larger maximum allowed waiting time, the longer each node can infect other nodes. Therefore, the reachability ratio is expected to increase when increasing this parameter value. The link weight corresponds to the minimum number of contacts required for the infection to spread from one node to another. Thus, the reachability ratio is expected to decrease with an increasing number of required contacts. For each network, the maximum allowed waiting time was run from 0 to 99 hours with a step of 1 hour. The link weight corresponding to the number of required contacts was run from 1 to 29, also using a step of 1. The resulting threedimentional plots, each consisting of 2900 calculated reachability ratios, is presented in Figure 4.3.5.

All three figures show that the reachability ratio increases with increasing maximum allowed waiting time and decreasing link weight. Comparing the simulated and randomized reference network, the former results in temporal characteristics that better resembles the original network. Most notably, the reachability ratio in the randomized network remains low for most link weights before an abrupt jump is observed when the requirement reaches low weight values. Thus, the simulated network shows to be able to better replicate time-respecting paths in the original network than the randomized network.

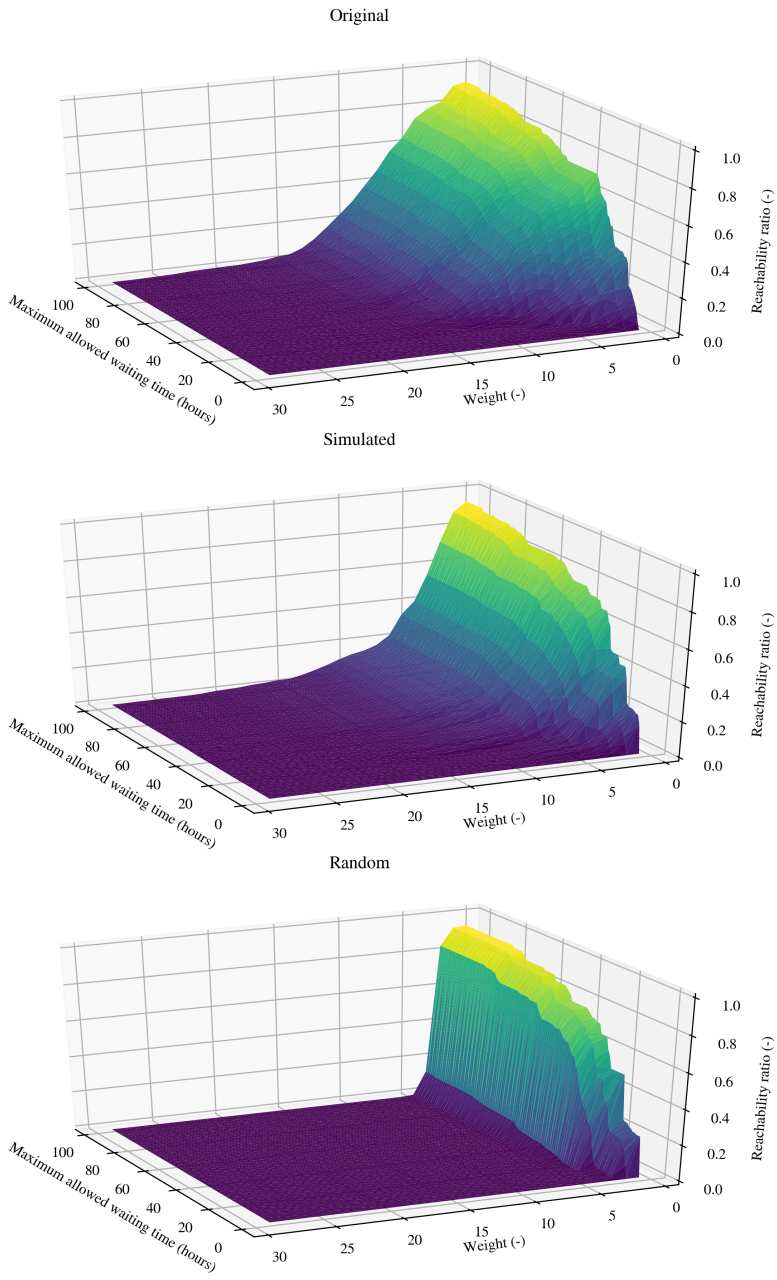

Figure 4.3.5: The plots present calculated reachability ratios as function of the maximum allowed waiting time and the link weight corresponding to the minimum number of registered contacts required for transmission from one node to another. The panels represent the empirical contact network, a simulated contact network and a randomized reference network.

Chapter 5

Discussion

As any other model, this modeling framework represents a simplified representation of a real life situation [145]. Numerous assumptions and simplifications were described when presenting how the model was built in Chapter 3. This section focuses on the main assumptions and the challenges they imply.

As previously introduced, agent-based models are suitable for exploring how patterns on the macro-level emerge from micro-level rules [20]. This also implies that the model output is vulnerable to uncertainties in these micro-level rules. High-quality data are required in order to obtain realistic agent behavior in the model. In the developed modeling framework, several of the mechanisms, both governing the contact network and the epidemiological spread, are based on conclusions drawn from rather scarce data.

The size of the data set describing the empirical network on which the simulated contact network is based on, poses such a challenge: The data were collected in the course of roughly four days (97 hours) and the 75 participants were divided into four categories yielding even fewer individuals per role. Thus, the derived rules and distribution parameters were based on very few data points. There are several issues related to this process: Firstly, when the hourly contact numbers are based on observations from four days only, any irregularities in one of these days may shift the overall pattern. For instance, Kulisiewicz *et al.*, who examined the entropy in several temporal contact networks, pointed out that that the very first day of contacts in the empirical hospital ward network stood out from the other days, and questioned whether Saint Nicolas Day celebrated on December 6 could have influenced the dynamics [175]. Secondly, the choice of a suitable statistical distribution becomes rather arbitrary when the data points are few. Choosing a lognormal distribution over a normal distribution for describing the relative contact rates, for instance, would facilitate for super spreaders in the network.

The challenge of deriving model rules from scarce empirical data applies also for the epidemiological model. Several of the parameters, for instance the rate of asymptomatic cases for patients and staff ward, were based on one estimated value only. Other parameter values were based on several estimates, although the estimates could be biased: As an example, Lauer *et al.* pointed out that the publicly reported cases of COVID-19 could

contain an overweight of severe cases relative to milder cases, and thus that the estimated incubation periods not necessarily covered all degrees of severity [66].

Modeling an ongoing crisis has been somewhat challenging in terms of obtaining a wide and still in-depth literature foundation. Even though the knowledge and accessible data resources have increased throughout the last six months, parts of the modeling framework were developed before empirical evidence supporting the choice of model mechanisms was published. An example is the assumed flat rate of infectiousness both for presymptomatic, asymptomatic and symptomatic individuals. The article where He *et al.* presented a curve-shaped temporal infectiousness profile was first published on April 15. Due to time restrictions, the model algorithms were not updated to include this dynamics, although implementing the temporal infectiousness profile could have increased the realism of the model. It is also worth mentioning that several of the underlying estimates for epidemiological model are based on unpublished preprints obtained from archives as arXiv, medRxiv and BioRxiv. These are not peer-reviewed. However, due to lack of published literature related to this ongoing pandemic, these preprints constituted important contributions to the developed framework.

To summarize, both the temporal contact network and the epidemiological model are based on assumptions and numerical values estimated from a scarce data basis, either in terms of few data points or due to the limitation of available literature. In addition to the stochastic variability in the model, these assumptions and estimated input parameters contribute to uncertainty in the model output. An improved future version of the model would desirably incorporate updated parameter values taking the increasing knowledge of COVID-19 into account.

5.1 Key Assumptions

In addition to the assumptions and simplifications made in regard to the temporal contact network and the epidemiological characteristics of the spreading agent, some other key assumptions have been made. They are discussed in the following:

- **No imported COVID-19 cases**: The model only contains one source of infection, namely the initially infected agents. No new cases are imported into the hospital wards in the course of the simulation run. Among other things, this implies that the infection may die out if the initially infected agents do not infect any other agents before they recover. In a real life scenario, new cases could probably have been introduced by admitted patients being infected, members of the staff that could have been infected outside the ward or by visitors carrying the virus. The probability of importing new infection cases into the hospital wards in the course of the simulation could for instance have been implemented through a simple parameter value mirroring the national or regional level of infection.

- **Transmission solely through inter-individual contacts**: In the model, transmission through indirect contact is neglected. That implies that the transmission of SARS-CoV-2 is restricted to close proximity interactions between the individuals,

and that potential routes of transmission via contaminated door knobs, light switches or the buttons of any microwave ovens are disregarded. Based on the enhanced real-life hygiene measures in terms of frequent cleaning and disinfection, it is believed that transmission through such indirect contacts would have contributed to the overall spread also in the virtual hospital. One way to implement the risk of being infected through indirect contact with an infectious individual could be to introduce an infection probability that is independent of contacts and which applies equally to all susceptible agents in the ward. This parameter value could for instance increase proportionally with the general level of infection in the ward. Alternatively, one could add a spatial dimension to the model, as have been previously done when examining the spread of antibiotic resistant pathogens [176]. Figure 5.1.1 shows a dummy illustration of how an infected person touching several objects could contribute to the overall spread of the virus.

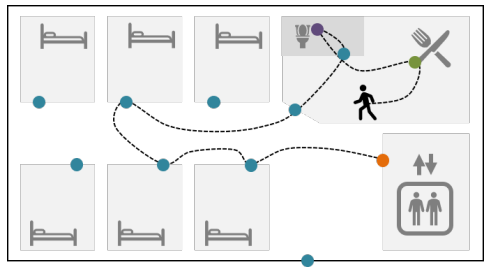

Figure 5.1.1: Example of a dummy virtual hospital ward where an infected person have touched the elevator buttons (orange), several patient room door knobs (blue), the water tap in a toilet (purple) and the common kitchen crockery, leaving contaminated fomites facilitating spread through indirect contact for individuals touching the same objects subsequently.

- **Equal transmission rates among all role pairs**: The model takes the presence or absence of symptoms, as well as the the enhanced hygiene measures in case of quarantine or isolation into account when calculating the probability of transmission in a close contact setting. It assumes, however, that the probability of transmission is equal among all role pairs: Given two identical situations, the model assumes equal transmissibility in an interaction between two agents belonging to the PAT and NUR role as two agents belonging to the PAT and ADM role. Keeping in mind the nature of the work of a nurse relative to an individual belonging to the administrative staff, it would be natural to assume that the former situation involves closer contact than the latter. One could also imagine that the transmission probability not necessarily is equal in both directions. Duval *et al.* investigated the transmission of MRSA on a temporal contact network in a long-term care hospital in France and did for instance find that the transmission rate from patient to staff was in an order of magnitude 10^2 times higher than the transmission rate from staff to patient. The transmission rate from one staff member to another was found to be 10^4 times as large as between two patients [177]. One could either incorporate these numbers directly into the expression describing p_T, or at least introduce a set of parameters allowing for exploration of how the variations in transmissibility among the ward roles affect the

overall transmission dynamics.

- **Transmission through CPIs**: The whole model builds upon the assumption that close-proximity interactions are able to capture the inter-individual transmission of SARS-CoV-2. Given the threshold for contact detection of 1.5 meters and the WHO and CDC definitions of close contact of 1 and 2 meters, respectively, it seems reasonable to assume that transmissions are captured when model agents are within this range. Earlier, CPIs have been demonstrated to capture the spread of MRSA, but not Enterobacteriaceae [177, 178]. The latter bacterial family includes several pathogenic species such as *Salmonella*, *Escherichia coli* (E. coli) and [179].
- **All individuals comply to the given control measures**: The model implements several control measures, such as social distancing, contact tracing, quarantine and isolation. The model also assumes that all individuals comply to these measures. In a real life scenario, however, it would be possible that a few individuals shorten their quarantine or ignores the social distancing. Since agent-based modeling allows for heterogeneity among agents, it could be interesting to let a proportion of the agents ignore the control measures in order to investigate how many dissidents the system can tolerate before, presumably, the spreading increases.

5.1.1 Further Work

There exists an endless number of possibilities in terms of investigating the model functionality, analyze the static and temporal characteristics of the generated contact networks, explore the epidemiological parameter space and evaluate the effect of several key control measures. Some suggestions are listed below:

- Compare the generated temporal contact network with reference models where the time-points of contact are randomly permuted in order to study the effect of order of events.
- Explore the N dimensional parameter space by running the model multiple times with randomly drawn values in a given range of each of the N parameters.
- Determine the optimal control measure strategy, for instance the number of quarantine days or number of hours before symptom onset included in the contact tracing, by using an evolutionary algorithm, such as particle swarm optimization, to find the parameter combination that gives the lowest fatality *and* the lowest usage of temporary staff, for instance.

Chapter 6

Conclusion

The ongoing coronavirus pandemic is not the first pandemic to ravage the world, and will most likely not be the last. Computational modeling provides an important tool for obtaining a better understanding of the underlying spreading mechanisms of human pathogens. In this thesis, a novel agent-based modeling framework was developed to explore the transmission of SARS-CoV-2 on a temporal network mirroring the inter-individual contacts in a virtual hospital.

The contact network facilitating the spread of SARS-CoV-2 was simulated based on agent and system attributes derived from an empirical hospital ward network, including the total number of registered contacts per hour, the relative contact rate for each individual and the pairwise role interactions in the ward. The resulting temporal contact network evolved simultaneously with the simulation and laid the foundation for the viral transmission. The baseline parameter set governing the epidemiological spread was derived from the available SARS-CoV-2 literature, although a wide range of the parameters are manipulable. For instance, the user is free to tune epidemiological parameters such as the incubation time and pathogen infectiousness, or control measures such as social distancing and quarantine duration. The developed framework responded as expected to several extreme parameter values, confirming the model functionality. The simulated contact network was able to better replicate several empirical network properties than a randomized reference network.

A future objective would be to explore the model parameter space in order to obtain a firm understanding of key parameters crucial for the spreading mechanism of SARS-CoV-2. The developed modeling framework constitutes a flexible modeling tool which easily can be modified and utilized for capturing the spread of future human pathogens.

Bibliography

1. Hays, J. N. *Epidemics and pandemics: their impact on human history* ISBN: 1-85109-663-9 (Ringgold Inc, 2005).
2. Dobson, A. P. & Carper, E. R. Infectious diseases and human population history. *BioScience* **46,** 115. ISSN: 0006-3568 (1996).
3. Wolfe, N. D., Dunavan, C. P. & Diamond, J. Origins of major human infectious diseases. *Nature* **447,** 279. ISSN: 0028-0836 (2007).
4. Larsen, C. S. The agricultural revolution as environmental catastrophe: Implications for health and lifestyle in the Holocene. *Quaternary International* **150,** 12–20. ISSN: 1040-6182 (2006).
5. Fukuda-Parr, S. New Threats to Human Security in the Era of Globalization. *Journal of Human Development* **4,** 167–179. ISSN: 1464-9888 (2003).
6. Klovdahl, A. S. Social networks and the spread of infectious diseases: The AIDS example. *Social Science & Medicine* **21,** 1203–1216. ISSN: 0277-9536 (1985).
7. Thomas, J. C. *Epidemiologic Methods for the Study of Infectious Diseases* ISBN: 9780199747757 (Oxford University Press, 2001).
8. Bridges, C. B., Kuehnert, M. J. & Hall, C. B. Transmission of influenza: implications for control in health care settings. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **37,** 1094–1101. ISSN: 10584838 (2003).
9. Gopinath, S., Carden, S. & Monack, D. Shedding light on Salmonella carriers. *Trends in Microbiology* **20,** 320–327. ISSN: 0966-842X (2012).
10. Benelli, G. & Mehlhorn, H. Declining malaria, rising of dengue and Zika virus: insights for mosquito vector control. *Parasitology Research* **115,** 1747–1754. ISSN: 0932-0113 (2016).
11. Perry, R. D. & Fetherston, J. D. Yersinia pestis – etiologic agent of plague. *Clinical Microbiology Reviews* **10,** 35. ISSN: 0893-8512 (1997).
12. Haensch, S. *et al.* Distinct Clones of Yersinia pestis Caused the Black Death. *PLoS Pathogens* **6,** e1001134. ISSN: 1553-7366 (2010).

13. Colizza, V., Barthélemy, M., Barrat, A. & Vespignani, A. Epidemic modeling in complex realities. *Comptes rendus - Biologies* **330,** 364–374. ISSN: 1631-0691 (2007).

14. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 2015. ISSN: 0027-8424 (2006).

15. Van Den Dool, C., Haenen, A., Leenstra, T. & Wallinga, J. The Role of Nursing Homes in the Spread of Antimicrobial Resistance Over the Healthcare Network. *Infection control and hospital epidemiology* **37,** 761–767. ISSN: 0899823X (2016).

16. Vanhems, P. *et al.* Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLoS ONE* **8,** e73970. ISSN: 1932-6203 (2013).

17. Marathe, M. & Vullikanti, A. Computational epidemiology. *Communications of the ACM* **56,** 88–96. ISSN: 00010782 (2013).

18. Caudill, L. & Lawson, B. A unified inter-host and in-host model of antibiotic resistance and infection spread in a hospital ward. *Journal of Theoretical Biology* **421,** 112–126. ISSN: 0022-5193 (2017).

19. Balcan, D. *et al.* Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of Computational Science* **1,** 132–145. ISSN: 1877-7503 (2010).

20. Abar, S., Theodoropoulos, G. K., Lemarinier, P. & O'hare, G. M. P. Agent Based Modelling and Simulation tools: A review of the state-of-art software. *Computer Science Review* **24,** 13–33. ISSN: 1574-0137 (2017).

21. Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 7280–7287. ISSN: 00278424 (2002).

22. Gorochowski, T. E. Agent-based modelling in synthetic biology. *Essays in Biochemistry* **60,** 325–336. ISSN: 0071-1365 (2016).

23. World Health Organization. *Antibiotic resistance* Feb. 5, 2018. `https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance`.

24. Sævareid, S. O. *Agent-based modeling of evolution and spread of antibiotic resistance on networks: a literature review* tech. rep. (Norwegian University of Science and Technology, Trondheim, 2019).

25. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *The Lancet* **395,** 470–473. ISSN: 0140-6736 (2020).

26. Chan, J. F. *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* **395,** 514–523. ISSN: 0140-6736 (2020).

27. Ren, L. *et al.* Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chinese medical journal* **133,** 1015. ISSN: 0366-6999 (2020).

28. Li, J. *et al.* The epidemic of 2019-novel-coronavirus (2019-nCoV) pneumonia and insights for emerging infectious diseases in the future. *Microbes and Infection* **22,** 80–85. ISSN: 1286-4579 (2020).

29. Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian Journal of Pediatrics* **87,** 281–286. ISSN: 0019-5456 (2020).

30. World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020* Press release. Mar. 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

31. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet. Infectious diseases.* ISSN: 1474-4457. https://arcg.is/0fHmTX (2020).

32. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* **395,** 1054–1062. ISSN: 0140-6736 (2020).

33. Fehr, A. R. & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods in molecular biology* **1282,** 1–23. ISSN: 1940-6029 (2015).

34. Virology: Coronaviruses. *Nature* **220,** 650. ISSN: 0028-0836 (1968).

35. Tyrrell, D. A. J. & Fielder, M. *Cold Wars: The Fight Against the Common Cold* ISBN: 0-19-263285-X (Oxford University Press, 2002).

36. Ahmad, A., Krumkamp, R. & Reintjes, R. Controlling SARS: a review on China's response compared with other SARS-affected countries. *Tropical Medicine & International Health* **14,** 36–45. ISSN: 1360-2276 (2009).

37. Park, C. *et al.* An Updated Assessment of the Economic Impact of COVID-19. *ADB BRIEFS.* ISSN: 2218-2675 (2020).

38. Mahase, E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* **368.** ISSN: 1756-1833 (2020).

39. Kim, K. H., Tandi, T. E., Choi, J. W., Moon, J. M. & Kim, M. S. Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology, characteristics and public health implications. *Journal of Hospital Infection* **95,** 207–213. ISSN: 0195-6701 (2017).

40. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310,** 676. ISSN: 0036-8075 (2005).

41. Chu, D. K. W. *et al.* MERS coronaviruses in dromedary camels, Egypt. *Emerging Infectious Diseases* **20,** 1049. ISSN: 1080-6040 (2014).

42. Lai, C., Shih, T., Ko, W., Tang, H. & Hsueh, P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents* **55.** ISSN: 0924-8579 (2020).

43. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395,** 565–574. ISSN: 0140-6736 (2020).

44. Guo, Y. *et al.* The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Military Medical Research* **7,** 11–11. ISSN: 20957467 (2020).

45. Li, X., Zai, J., Wang, X. & Li, Y. Potential of large "first generation" human-to-human transmission of 2019-nCoV. *Journal of Medical Virology* **92,** 448–454. ISSN: 0146-6615 (2020).

46. Weber, T. P. & Stilianakis, N. I. Inactivation of influenza A viruses in the environment and modes of transmission: A critical review. *Journal of Infection* **57,** 361–373. ISSN: 0163-4453 (2008).

47. Lewis, D. Is the coronavirus airborne? Experts can't agree. *Nature* **580,** 175–175. ISSN: 00280836 (2020).

48. *Q&A on coronaviruses (COVID-19)* World Health Organization, Apr. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses.

49. *How COVID-19 Spreads* Centers for Disease Control and Prevention, Apr. 2020. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html.

50. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *The New England Journal of Medicine* **382,** 1199–1207. ISSN: 0028-4793 (2020).

51. *COVID-19 – Virtual press conference - 27 March, 2020* World Health Organizaiton, Mar. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/press-briefings/.

52. Liu, Y. *et al.* Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature.* ISSN: 00280836 (2020).

53. Wu, D., Wu, T., Liu, Q. & Yang, Z. The SARS-CoV-2 outbreak: What we know. *International Journal of Infectious Diseases* **94,** 44–48. ISSN: 1201-9712 (2020).

54. Van Den Driessche, P. & Watmough, J. in *Mathematical Epidemiology* 159–178 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008). ISBN: 9783540789109.

55. Dietz, K. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research* **2,** 23–41. ISSN: 0962-2802 (1993).

56. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)* World Health Organization, Feb. 2020. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.

57. Guerra, F. M. *et al.* The basic reproduction number (R0) of measles: a systematic review. *The Lancet Infectious Diseases* **17,** e420–e428. ISSN: 1473-3099 (2017).

58. Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M. & Finelli, L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infectious Diseases* **14.** ISSN: 1471-2334 (2014).

59. Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine* **26,** 1–5. ISSN: 1078-8956 (2020).

60. Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368.** ISSN: 00368075 (2020).

61. Riou, J. & Althaus, C. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* **25,** 1. ISSN: 1025496X (2020).

62. Zhao, S. *et al.* Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases* **92,** 214–217. ISSN: 1201-9712 (2020).

63. Zhang, S. *et al.* Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *International Journal of Infectious Diseases* **93,** 201–204. ISSN: 1201-9712 (2020).

64. Sanche, S. *et al.* High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging infectious diseases* **26.** ISSN: 1080-6059 (2020).

65. *A dictionary of epidemiology* (ed Porta, M.) (Oxford University Press, 2014).

66. Lauer, S. A. *et al.* The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine.* ISSN: 0003-4819 (2020).

67. Zhang, J. *et al.* Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet. Infectious diseases.* ISSN: 1473-3099 (2020).

68. Linton, N. M. *et al.* Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of clinical medicine* **9.** ISSN: 2077-0383 (2020).

69. Backer, J., Klinkenberg, D. & Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance* **25,** 1. ISSN: 1025496X (2020).

70. Guan, W. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England Journal of Medicine* **382,** 1708–1720. ISSN: 0028-4793 (2020).

71. Hu, Z. *et al.* Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China. Life Sciences* **63,** 1–6. ISSN: 1674-7305 (2020).

72. Ludvigsson, J. F. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta paediatrica* **109,** 1088. ISSN: 0803-5253 (2020).

73. Norwegian Institute of Public Health. *Facts about the virus and COVID-19 disease* Feb. 2020. https://www.fhi.no/en/op/novel-coronavirus-facts-advice/facts-and-knowledge-about-covid-19/facts-about-novel-coronavirus/?term=&h=1.

74. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro surveillance : European communicable disease bulletin* **25.** ISSN: 1025496X (2020).

75. Lavezzo, E. *et al.* Suppression of COVID-19 outbreak in the municipality of Vo, Italy. *medRxiv* (2020).

76. John, T. *Iceland lab's testing suggests 50% of coronavirus cases have no symptoms* CNN, Apr. 2020. https://edition.cnn.com/2020/04/01/europe/iceland-testing-coronavirus-intl/index.html.

77. Bai, Y. *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* **323.** ISSN: 0098-7484 (2020).

78. Zou, L. *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *The New England journal of medicine* **382,** 1177. ISSN: 0028-4793 (2020).

79. Shi, Y. *et al.* COVID-19 infection: the perspectives on immune responses. *Cell Death and Differentiation* **27,** 1–4. ISSN: 1350-9047 (2020).

80. Tindale, L. *et al.* Transmission interval estimates suggest pre-symptomatic spread of COVID-19. *medRxiv* (2020).

81. Du, Z. *et al.* Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerging infectious diseases* **26.** ISSN: 10806040 (2020).

82. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine.* ISSN: 1546-170X (2020).

83. Ki, M. Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea. *Epidemiology and health* **42,** e2020007–e2020007. ISSN: 2092-7193 (2020).

84. Jung, S. *et al.* Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. *Journal of clinical medicine* **9.** ISSN: 2077-0383 (2020).

85. Wölfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature.* ISSN: 00280836 (2020).

86. To, K. K. *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *The Lancet Infectious Diseases* **20,** 565–574. ISSN: 1473-3099 (2020).

87. Onder, G., Rezza, G. & Brusaferro, S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA*. ISSN: 00987484 (2020).

88. Wu, Z. & Mcgoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* **323.** ISSN: 00987484 (2020).

89. Streeck, H. *et al.* Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *medRxiv* (2020).

90. Jordan, R. E., Adab, P. & Cheng, K. K. Covid-19: risk factors for severe disease and death. *BMJ* **368.** ISSN: 1756-1833 (2020).

91. Hauser, A. *et al.* Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: a modelling study in Hubei, China and northern Italy. *medRxiv* (2020).

92. Ruan, Q., Yang, K., Wang, W., Jiang, L. & Song, J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive care medicine* **46,** 846. ISSN: 0342-4642 (2020).

93. Madigan, M., Martinko, J., Bender, K., Buckley, D. & Stahl, D. *Brock biology of microorganisms* 14th ed., Global ed. ISBN: 9781292018317 (Pearson, Harlow, 2015).

94. Sintchenko, V., Iredell, J. R. & Gilbert, G. L. Antibiotic therapy of ventilator-associated pneumonia—a reappraisal of rationale in the era of bacterial resistance. *International Journal of Antimicrobial Agents* **18,** 223–229. ISSN: 0924-8579 (2001).

95. Joseph, C., Togawa, Y. & Shindo, N. Bacterial and viral infections associated with influenza. *Influenza and other respiratory viruses* **7,** 105–113 (2013).

96. Grabowska, K., Högberg, L., Penttinen, P., Svensson, Å. & Ekdahl, K. Occurrence of invasive pneumococcal disease and number of excess cases due to influenza. *BMC Infectious Diseases* **6,** 58–58. ISSN: 1471-2334 (2006).

97. Gullberg, E. *Selection of Resistance at very low Antibiotic Concentrations* PhD thesis (Uppsala University, 2014).

98. Cosgrove, S. The Relationship between Antimicrobial Resistance and Patient Outcomes: Mortality, Length of Hospital Stay, and Health Care Costs. *Clinical Infectious Diseases* **42,** S82–9. ISSN: 10584838 (2006).

99. Naber, C. Staphylococcus aureus Bacteremia: Epidemiology, Pathophysiology, and Management Strategies. *Clinical Infectious Diseases* **48,** S231. ISSN: 10584838 (2009).

100. Lowy, F. D. Staphylococcus aureus infections. *The New England Journal of Medicine* **339,** 520. ISSN: 0028-4793 (1998).

101. *Surveillance of antimicrobial resistance in Europe 2018* European Centre for Disease Prevention and Control, Nov. 2019.

102. Hanberger, H. *et al.* Increased mortality associated with meticillin-resistant Staphylococcus aureus (MRSA) infection in the Intensive Care Unit: results from the EPIC II study. *International Journal of Antimicrobial Agents* **38,** 331–335. ISSN: 0924-8579 (2011).

103. Gastmeier, P. *et al.* Mortality Risk Factors with Nosocomial Staphylococcus aureus Infections in Intensive Care Units: Results from the German Nosocomial Infection Surveillance System (KISS). *Infection* **33,** 50–55. ISSN: 0300-8126 (2005).

104. Blot, S., Vandewoude, K., Hoste, E. & Colardyn, F. Attributable mortality in critically ill patients with bacteremia involving methicillin susceptible (MSSA) and methicillin resistant Staphylococcus aureus (MRSA). *Critical Care* **5,** P090. ISSN: 1364-8535 (2001).

105. Amanat, F. & Krammer, F. SARS-CoV-2 Vaccines: Status Report. *Immunity* **52,** 583–589. ISSN: 1074-7613 (2020).

106. Lan, L. *et al.* Positive RT-PCR Test Results in Patients Recovered From COVID-19. *JAMA* **323.** ISSN: 0098-7484 (2020).

107. Bao, L. *et al.* Reinfection could not occur in SARS-CoV-2 infected rhesus macaques. *BioRxiv* (2020).

108. Fafi-Kremer, S. *et al.* Serologic responses to SARS-CoV-2 infection among hospital staff with mild disease in eastern France. *medRxiv* (2020).

109. Lewnard, J. A. & Lo, N. C. Scientific and ethical basis for social-distancing interventions against COVID-19. *The Lancet. Infectious diseases.* ISSN: 14733099 (2020).

110. Petherick, A. Developing antibody tests for SARS-CoV-2. *The Lancet* **395,** 1101–1102. ISSN: 0140-6736 (2020).

111. *Test criteria for coronavirus* Norwegian Institute of Public Health, Apr. 2020. https://www.fhi.no/en/op/novel-coronavirus-facts-advice/advice-to-health-personnel/test-criteria-for-coronavirus/.

112. *Flowchart for COVID-19 testing for acute respiratory tract infections* Accessed on 01.04.20. Norwegian Institute of Public Health, Apr. 2020. https://www.fhi.no/en/publ/posters/flowchart-for-COVID-19-testing/.

113. *Social distance, quarantine and isolation* (accessed on 05.04.20). Norwegian Institute of Public Health, Apr. 2020. https://www.fhi.no/en/op/novel-coronavirus-facts-advice/facts-and-general-advice/social-distance-quarantine-and-isolation/.

114. *Definitions of COVID-19 cases and close contacts* Accessed on 01.04.20. Norwegian Institute of Public Health, Feb. 2020. https://www.fhi.no/en/op/novel-coronavirus-facts-advice/advice-to-health-personnel/definitions-of-probable-and-confirmed-cases-of-coronavirus-covid-19-and-con/.

115. Barabási, A. *Network Science* (ed Cambridge university press) http://networksciencebook.com/ (2016).

116. Bowen, J. T. & Laroe, C. Airline networks and the international diffusion of severe acute respiratory syndrome (SARS). *Geographical Journal* **172,** 130–144. ISSN: 0016-7398 (2006).

117. Anderson, R. M., Gupta, S. & Ng, W. The significance of sexual partner contact networks for the transmission dynamics of HIV. *Journal of acquired immune deficiency syndromes* **3,** 417–429. ISSN: 0894-9255 (1990).

118. Balthrop, J., Forrest, S., Newman, M. & Williamson, M. Technological Networks and the Spread of Computer Viruses. *Science* **304,** 527–529. ISSN: 0036-8075 (2004).

119. Serrano, E., Iglesias, C. & Garijo, M. *A Novel Agent-Based Rumor Spreading Model in Twitter* in *Proceedings of the 24th International Conference on world wide web* (2015), 811–814. ISBN: 9781450334730.

120. Gao, S., Ma, J., Chen, Z., Wang, G. & Xing, C. Ranking the spreading ability of nodes in complex networks based on local structure. *Physica A: Statistical Mechanics and its Applications* **403,** 130–147. ISSN: 0378-4371 (2014).

121. Miller, J. C. & Hyman, J. M. Effective vaccination strategies for realistic social networks. *Physica A: Statistical Mechanics and its Applications* **386,** 780–785. ISSN: 0378-4371 (2007).

122. Karkada, U., Adamic, L., Kahn, J. & Iwashyna, T. Limiting the spread of highly resistant hospital-acquired microorganisms via critical care transfers: a simulation study. *Intensive Care Medicine* **37,** 1633–1640. ISSN: 0342-4642 (2011).

123. Newman, M. E. J. *Networks : an introduction* ISBN: 9780199206650 (Oxford University Press, Oxford, 2010).

124. Opsahl, T., Agneessens, F. & Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks* **32,** 245–251. ISSN: 0378-8733 (2010).

125. Eidsaa, M. & Almaas, E. Investigating the relationship between k-core and s-core network decompositions. *Physica A: Statistical Mechanics and its Applications* **449,** 111–125. ISSN: 0378-4371 (2016).

126. Seidman, S. B. Network structure and minimum degree. *Social Networks* **5,** 269–287. ISSN: 0378-8733 (1983).

127. Holme, P. & Saramaki, J. Temporal networks. *Physics Reports* **519,** 97. ISSN: 0370-1573 (2012).

128. Masuda, N. & Holme, P. Predicting and controlling infectious disease epidemics using temporal networks. *F1000prime reports* **5,** 6–6. ISSN: 2051-7599 (2013).

129. Pan, R. & Saramäki, J. Path lengths, correlations, and centrality in temporal networks. *arXiv.org* **84.** ISSN: 15502376 (2011).

130. Mastrandrea, R., Fournet, J. & Barrat, A. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *arXiv.org* **10.** ISSN: 19326203 (2015).

131. Smieszek, T. *et al.* How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC infectious diseases* **14,** 136–136. ISSN: 1471-2334 (2014).

132. Read, J. M., Edmunds, W. J., Riley, S., Lessler, J. & Cummings, D. A. T. Close encounters of the infectious kind: methods to measure social mixing behaviour. *Epidemiology and Infection* **140,** 2117–2130. ISSN: 0950-2688 (2012).

133. Barrat, A., Cattuto, C., Tozzi, A. E., Vanhems, P. & Voirin, N. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection* **20,** 10–16. ISSN: 1198-743X (2014).

134. Stehlé, J. *et al.* High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLoS ONE* **6,** e23176. ISSN: 1932-6203 (2011).

135. Fournet, J. & Barrat, A. Contact patterns among high school students. *arXiv.org* **9.** ISSN: 19326203 (2014).

136. Génois, M. *et al.* Data on face-to-face contacts in an office building suggests a low-cost vaccination strategy based on community linkers. *arXiv.org* **3.** ISSN: 20501242 (2015).

137. Kiti, M. *et al.* Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Science* **5,** 1–21. ISSN: 2193-1127 (2016).

138. Régis, C. *et al.* Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine* **9,** 87. ISSN: 1741-7015 (2011).

139. Isella, L. *et al.* What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology* **271,** 166–180. ISSN: 0022-5193 (2011).

140. Saramäki, J. & Holme, P. Exploring temporal networks with greedy walks. *The European Physical Journal B* **88,** 1–8. ISSN: 1434-6028 (2015).

141. Rodríguez, J. P., Ghanbarnejad, F. & Eguiluz, V. M. Risk of Coinfection Outbreaks in Temporal Networks: A Case Study of a Hospital Contact Network. *Frontiers in Physics* **5.** ISSN: 2296-424X (2017).

142. Colman, E., Holme, P., Sayama, H. & Gershenson, C. Efficient sentinel surveillance strategies for preventing epidemics on networks. *PLOS Computational Biology* **15.** ISSN: 1553734X (2019).

143. Bernoulli, D. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad., Roy. Sci.(Paris) avec Mem,* 1–45 (1760).

144. Kermack, W. O. & McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character (1905-1934)* **115,** 700–721. ISSN: 0950-1207 (1927).

145. Bandini, S., Manzoni, S. & Vizzari, G. Agent Based Modeling and Simulation: An Informatics Perspective. *Journal of Artificial Societies and Social Simulation* **12.** ISSN: 1460-7425 (2009).

146. Opatowski, L., Guillemot, D., Boëlle, P. & Temime, L. Contribution of mathematical modeling to the fight against bacterial antibiotic resistance. *Current Opinion in Infectious Diseases* **24,** 279–287. ISSN: 0951-7375 (2011).

147. Rahmandad H. & Sterman, J. Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. *Management Science* **54,** 998–1014. ISSN: 0025-1909 (2008).

148. Hethcote, H. W. The Mathematics of Infectious Diseases. *SIAM Review* **42,** 599–653. ISSN: 0036-1445 (2000).

149. Novozhilov, A. S. Epidemiological Models With Parametric Heterogeneity : Deterministic Theory for Closed Populations. *Mathematical Modelling of Natural Phenomena* **7,** 147–167. ISSN: 09735348 (2012).

150. McCallum, H., Barlow, N. & Hone, J. How should pathogen transmission be modelled? *Trends in Ecology & Evolution* **16,** 295–300. ISSN: 0169-5347 (2001).

151. Klein, E. Y., Chelen, J., Makowsky, M. D. & Smaldino, P. E. in *Mathematical and Computational Modeling* 121–134 (John Wiley & Sons, Inc, Hoboken, NJ, 2015). ISBN: 9781118853986.

152. Caudill, L. & Lawson, B. *A hybrid agent-based and differential equations model for simulating antibiotic resistance in a hospital ward* in *2013 Winter Simulations Conference* (IEEE, 2013), 1419–1430.

153. Kouyos, R., Klein, E. & Grenfell, B. Hospital-community interactions foster coexistence between methicillin-resistant strains of Staphylococcus aureus. *PLoS Pathogens* **9.** ISSN: 1553-7366 (2013).

154. Grimm, V. *Individual-based modeling and ecology* ISBN: 0-691-09665-1 (Princeton University Press, Princeton, 2005).

155. Hanappi, G. Agent-based modelling. History, essence, future. **70.** ISSN: 20373635 (2017).

156. Schelling, T. C. Dynamic models of segregation. *The Journal of Mathematical Sociology* **1,** 143–186. ISSN: 0022-250X (1971).

157. Dada, J. O. & Mendes, P. Multi-scale modelling and simulation in systems biology. *Integrative Biology* **3,** 86–96. ISSN: 1757-9694 (2011).

158. Bruch, E. & Atwell, J. Agent-Based Models in Empirical Social Research. *Sociological Methods & Research* **44,** 186–221. ISSN: 0049-1241 (2015).

159. Duan, W., Fan, Z., Zhang, P., Guo, G. & Qiu, X. Mathematical and computational approaches to epidemic modeling: a comprehensive review. *Frontiers of Computer Science* **9,** 806–826. ISSN: 2095-2228 (2015).

160. Upton, G. & Cook, I. *A dictionary of statistics* ISBN: 9780191726866 (Oxford University Press, 2008).

161. Walpole, R. E., Myers, R. H., Myers, S. L. & Ye, K. *Probability & statistics for engineers & scientists* ISBN: 978-0-321-62911-1 (Pearson, 2012).

162. Montgomery, D. C. & Runger, G. C. *Applied statistics and probability for engineers* ISBN: 978-0-470-05304-1 (Wiley, 2011).

163. Zhang, H. H. *Bernoulli Distribution* 2018.

164. Burkardt, J. *The truncated normal distribution* 1–35 (Department of Scientific Computing Website, Florida State University, 2014).

165. Tsagris, M., Beneki, C. & Hassani, H. On the Folded Normal Distribution. *Mathematics* **2,** 12–28. ISSN: 2227-7390 (2014).

166. Adler, J. & Parmryd, I. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A* **77,** 733–742. ISSN: 1552-4922 (2010).

167. Wang, Z. & Bovik, A. C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine* **26,** 98–117. ISSN: 1053-5888 (2009).

168. Python Software Foundation. *Python, version 3.7.3* `https://www.python.org/`.

169. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13,** 2498–2504. ISSN: 1088-9051 (2003).

170. Smieszek, T., Lazzari, G. & Salathé, M. Assessing the Dynamics and Control of Droplet- and Aerosol-Transmitted Influenza Using an Indoor Positioning System. *Scientific Reports* **9.** ISSN: 2045-2322 (2019).

171. Salathé, M. *et al.* A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences* **107,** 22020. ISSN: 0027-8424 (2010).

172. Moser, M. *et al.* An Outbreak of Influenza Aboard a Commercial Airliner. *American Journal of Epidemiology* **110.** ISSN: 0002-9262 (1979).

173. Jorgensen, P. *et al.* How close are countries of the WHO European Region to achieving the goal of vaccinating 75% of key risk groups against influenza? Results from national surveys on seasonal influenza vaccination programmes, 2008/2009 to 2014/2015. *Vaccine* **36,** 442–452. ISSN: 0264-410X (2018).

174. *Influensavaksine - veileder for helsepersonell* Norwegian. Norwegian Institute of Public Health, Oct. 2019. `Influensavaksine%20-%20veileder%20for%20helsepersonell`.

175. Kulisiewicz, M., Kazienko, P., Szymański, B. & Michalski, R. Entropy Measures of Human Communication Dynamics. *arXiv.org* **8.** ISSN: 20452322 (2018).

176. Finstad, N. R. *Computational SIS Modeling of the Spread of Antibiotic Resistance within Bacterial Metapopulation Networks* MA thesis (Norwegian University of Science and Technology, 2018).

177. Duval, A. *et al.* Close proximity interactions support transmission of ESBL-K. pneumoniae but not ESBL-E. coli in healthcare settings. *PLoS Computational Biology* **15,** e1006496. ISSN: 1553-734X (2019).

178. Obadia, T. *et al.* Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals. *PLoS computational biology* **11,** e1004170. ISSN: 1553-734X (2015).

179. Edwards, P. & Ewing, W. *Identification of Enterobacteriaceae* (1972).

Appendix

Appendix A

Theory Supplementary

A.1 Theory Presented in Project Report

Parts of Chapter 2 are obtained from or based on material presented in my previous project report "Agent-Based Modeling of Evolution and Spread of Antibiotic Resistance on Networks: A Literature Review" [24]. The following list provides an overview of the relevant sections:

- Section 2.1.2: The theory related to mechanisms of antibiotic resistance is based on material presented in [24].
- Section 2.2: The introductory text regarding network and network spreading phenomena is obtained from [24], with minor textual alterations.
- Section 2.2.1: The material covering the adjacency matrix and centrality measures of undirected, unweighted networks (node degree, closeness centrality, betweenness centrality) is obtained from [24].
- Section 2.2.2: The first paragraph is obtained from [24]
- Section 2.2.3: The introduction on empirical contact networks and how they can be generated based on self-reporting systems, external observers or electronic proximity sensors is based on theory presented in [24].
- Section 2.3: This section is obtained from [24], with some adaptations to SARS-CoV-2 and minor textual alterations.
- Section 2.4.2: The theory related to Bernoulli distributions is based on [24].

In addition, the quote by Daniel Bernoulli presented in the very beginning is reused from the project report.

Appendix B

Method Supplementary

B.1 Python Modules and Data Types

Python Modules

Python modules utilized in the modeling work are presented in Table B.1.1. The table presents both modules from the Python Standard Library and installed modules.

Table B.1.1: Imported modules from the Python Standard Library and installed modules.

Module	Description	Version
math	mathematical functions	
os	operating system interfaces	
shelve	object persistence	
time	time-related functions	
typing	type hints	
random	generates pseudo-random numbers	
matplotlib	visualisation	3.1.2
matplotlib.venn	venn diagrams	0.11.5
more_itertools	iterables	8.2.0
networkx	complex networks	2.4
numpy	scientific computing	1.18.1
pandas	data analysis	0.25.3
scipy	statistics	1.4.1
tabulate	printing tabular data	0.8.6

Built-In Data Types

Some central Python built-in types is presented in Table B.1.2, see the Python documentation for a more detailed introduction.

Table B.1.2: Some of the most central built-in data types in Python.

Type	Description	Example
int	Numeric type, integers	12
float	Numeric type, floating point numbers	3.14
bool	Boolean value, takes only one of two values	True
str	Text sequence type	“This is a string”
list	Sequence type, stores collections of items	[A, B, C, D, E, F]
dict	Mapping type, dictionary with keys and values	{’name’: ’John’, ’age’: 100}

B.2 Temporal Contact Pattern

B.2.1 Contact Duration Distribution

The relative frequency of contact duration is presented in Table B.2.1. The absolute frequency is presented as a histogram in Figure B.2.1.

Table B.2.1: The relative frequency of duration (in seconds) per contact.

Duration (s)	Relative frequency (%)
20	55.9188
40	20.1496
60	8.9601
80	4.8362
100	2.9701
120	2.1510
140	1.2037
160	0.8405
≥ 180	2.9700

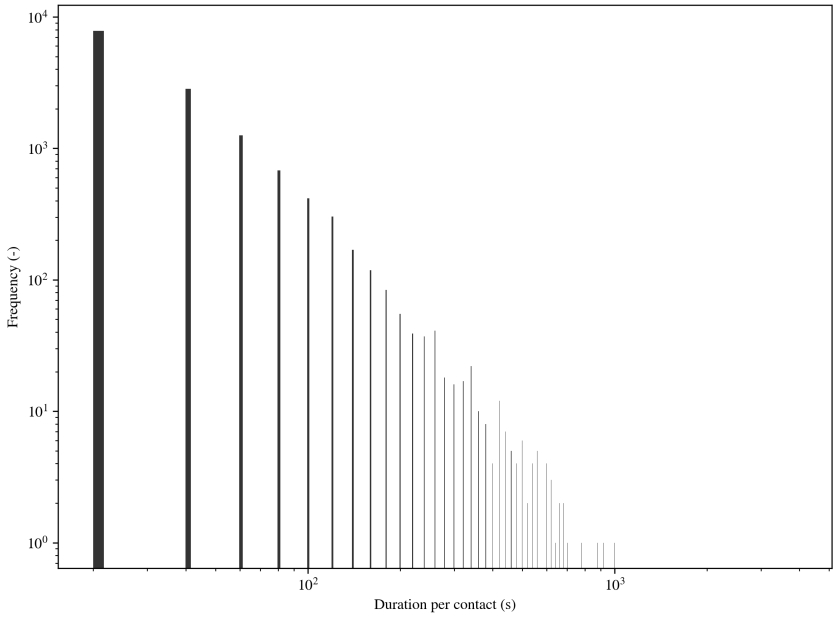

Figure B.2.1: Histogram showing the frequency of duration per contact. Both axes are logarithmic.

B.2.2 Comparing Normal, Truncated and Folded Distribution

In order to determine which of the normal, truncated normal and folded normal distributions was best suited for describing the number of contacts per bed for each hour in the model, the mean of $1\,000\,000$ drawn values for each distribution was calculated for each hour. All negative values drawn from the former distribution was set to zero. The resulting list of mean values for each distribution was compared to the list of empirical means by calculating the mean squared error. The numerical results are presented in Table B.2.2.

Table B.2.2: The mean squared error (MSE) was calculated based on the empirical mean of contacts per bed for each hour and the mean of $1\,000\,000$ values drawn from each of the three distributions. *All negative values were set to zero.

Distribution	MSE
Normal*	5.0881×10^{-3}
Truncated	9.4768×10^{-2}
Folded	1.9430×10^{-2}

B.2.3 Contacts per Hour

The total number of contacts registered for each hour in the study period, as well as corresponding averages and standard deviations, are presented in Table B.2.3. Contacts spanning more than one hour are registered in the hour where the contact began.

Table B.2.3: Total number of registered contacts per hour in the study period. The average number of contacts per hour and the sample standard deviation are listed in the two rightmost columns.

Hour	Mon	Tue	Wed	Thu	Fri	Average	SD
00:00		4	0	0	1	1.25	1.8930
01:00		1	0	1	1	0.75	0.5000
02:00		2	0	13	2	4.25	5.9090
03:00		3	0	13	3	4.75	5.6789
04:00		0	0	0	0	0.00	0.0000
05:00		38	6	44	8	24.00	19.7990
06:00		27	37	54	47	41.25	11.7863
07:00		161	199	183	198	185.25	17.7459
08:00		250	235	226	234	236.25	10.0125
09:00		289	365	330	356	335.00	34.0686
10:00		467	448	445	421	445.25	18.8746
11:00		470	591	392	458	477.75	82.9232
12:00		395	485	372	456	427.00	52.4532
13:00	20	397	295	317	136	233.00	152.1299
14:00	113	206	247	191		189.25	56.0736
15:00	84	336	197	120		184.25	111.6076
16:00	109	260	203	114		171.50	73.1141
17:00	130	196	215	68		152.25	66.9446
18:00	171	196	94	56		129.25	65.3369
19:00	101	139	44	17		75.25	55.0659
20:00	84	122	69	25		75.00	40.1082
21:00	6	3	39	10		14.50	16.5831
22:00	33	0	59	16		27.00	25.2323
23:00	48	0	13	10		17.75	20.9185

Figure 3.2.4 shows how the number of registered contacts in the ward evolves with time, and demonstrates that the pattern behaves similarly from day to day. This figure is based on the *total* number of contacts in the ward. The following figures shows how the number of contacts evolve with time when being filtrated based on role and pair of roles. Figures B.2.2 to B.2.5 present contact numbers as function of time when at least one of the involved individuals belong to the NUR, MED, ADM and PAT category, respectively. Figures B.2.6 to B.2.15 presents contact numbers as function of time given specified roles for both individuals involved in the contact, such as NUR-NUR or MED-PAT.

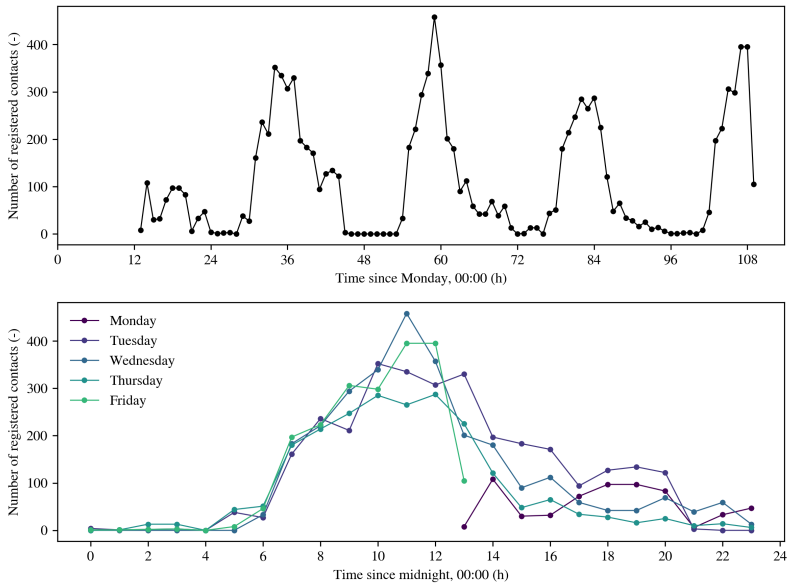

Figure B.2.2: Number of registered contacts per hour, where at least one of the involved individuals belong to the **NUR** category.

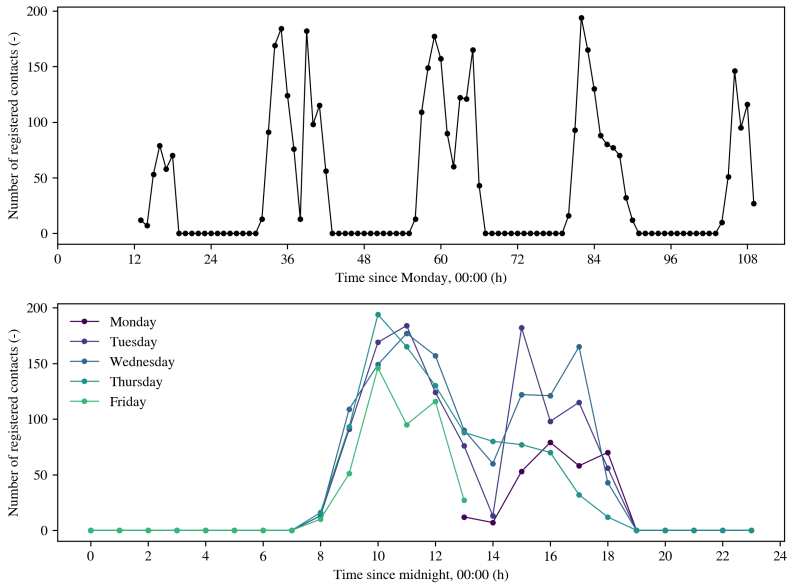

Figure B.2.3: Number of registered contacts per hour, where at least one of the involved individuals belong to the **MED** category.

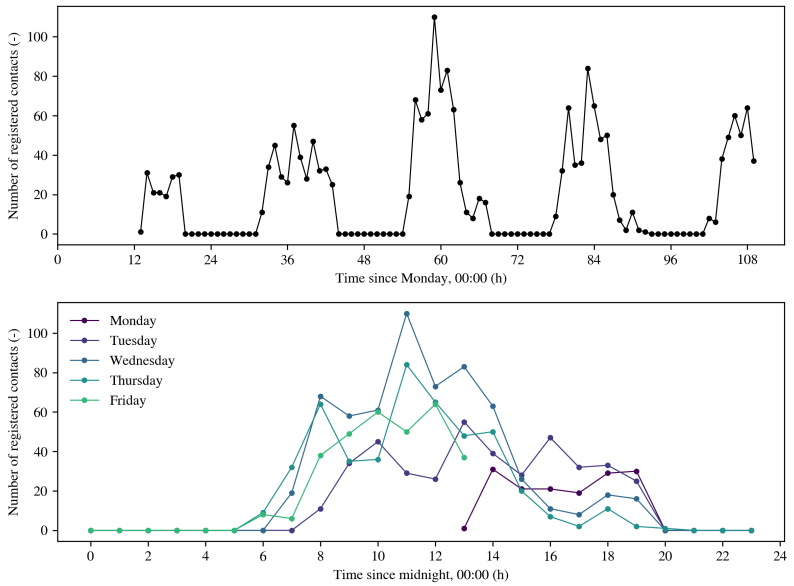

Figure B.2.4: Number of registered contacts per hour, where at least one of the involved individuals belong to the **ADM** category.

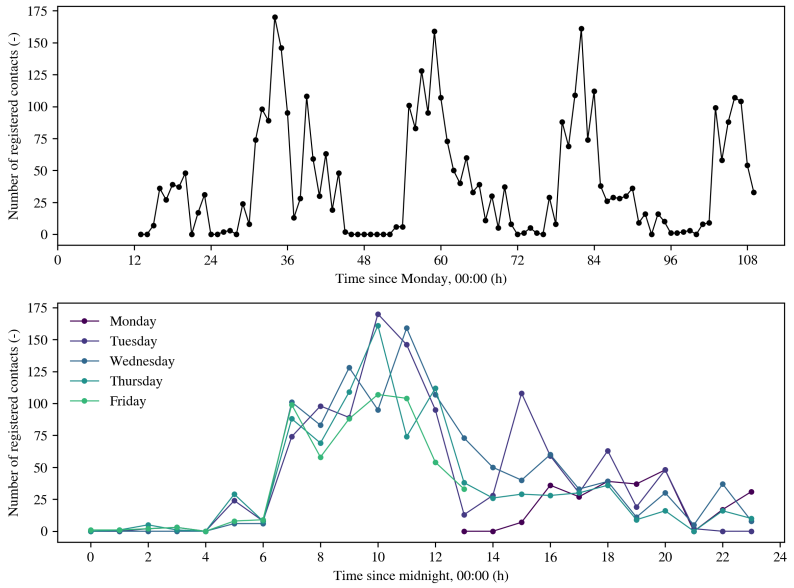

Figure B.2.5: Number of registered contacts per hour, where at least one of the involved individuals belong to the **PAT** category.

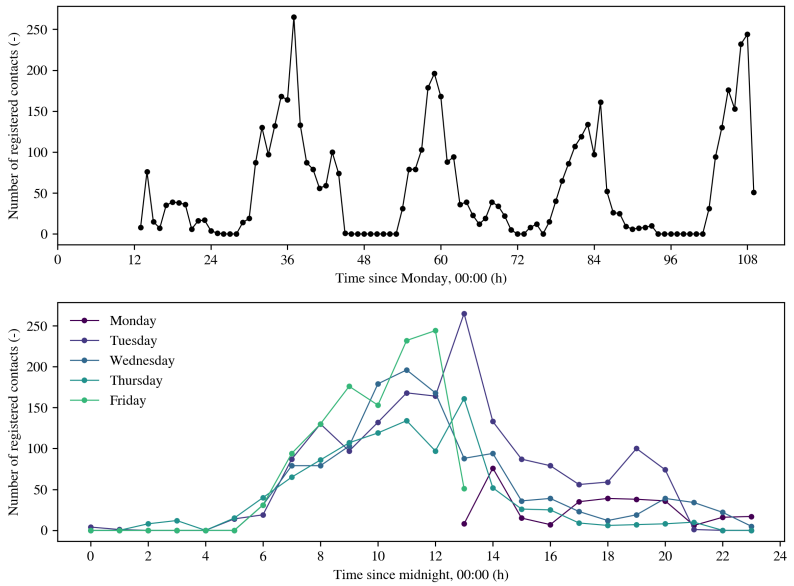

Figure B.2.6: Number of registered contacts per hour, where both of the involved individuals belong to the NUR category: **NUR-NUR**.

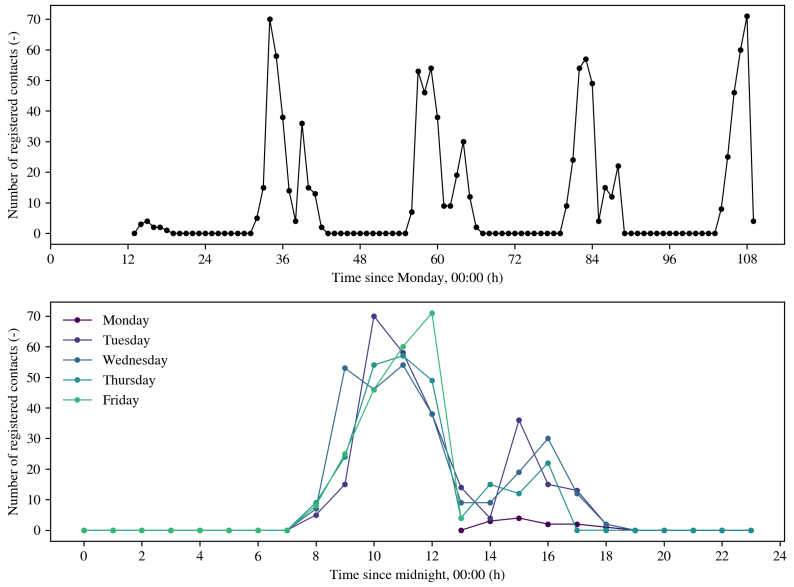

Figure B.2.7: Number of registered contacts per hour, where the two involved individuals belong to the NUR and MED categories: **NUR-MED**.

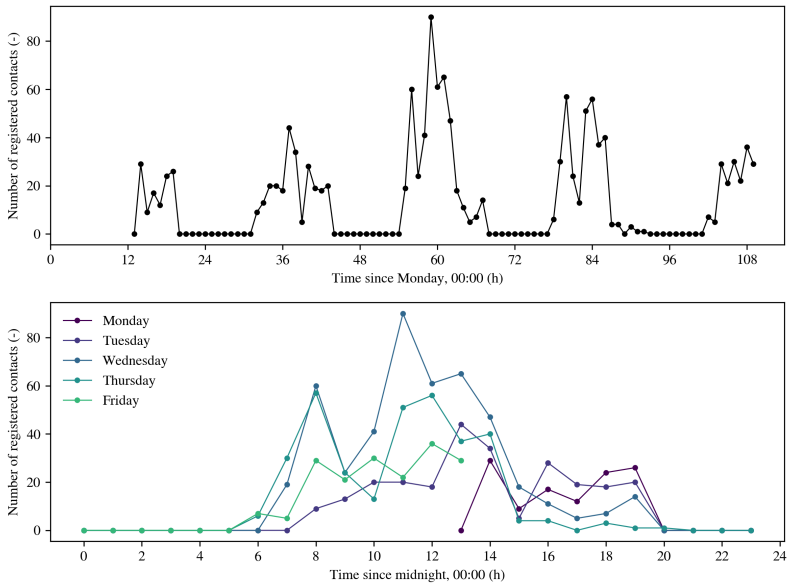

Figure B.2.8: Number of registered contacts per hour, where the two involved individuals belong to the NUR and ADM categories: **NUR-ADM**.

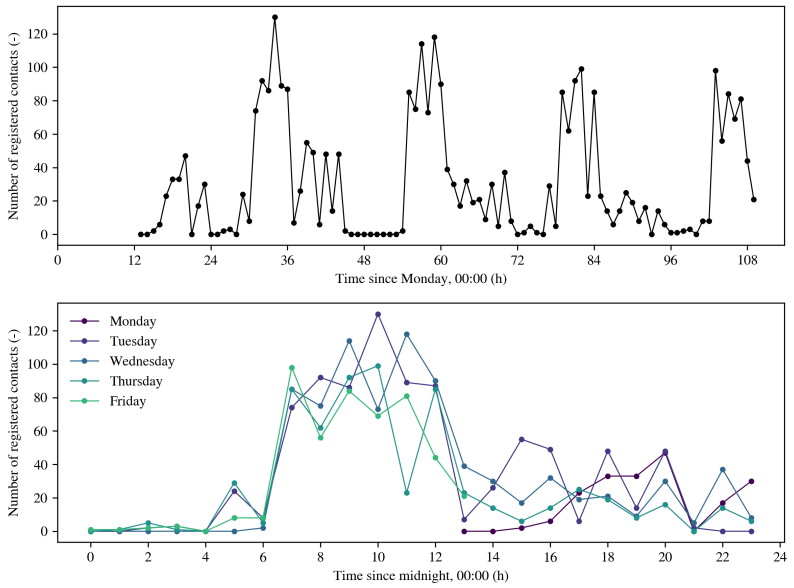

Figure B.2.9: Number of registered contacts per hour, where the two involved individuals belong to the NUR and PAT categories: **NUR-PAT**.

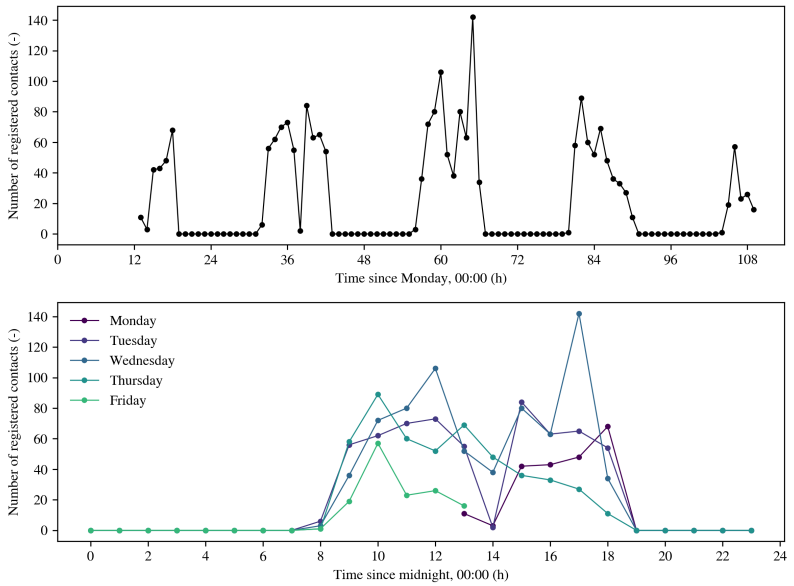

Figure B.2.10: Number of registered contacts per hour, where both of the involved individuals belong to the MED category: **MED-MED**.

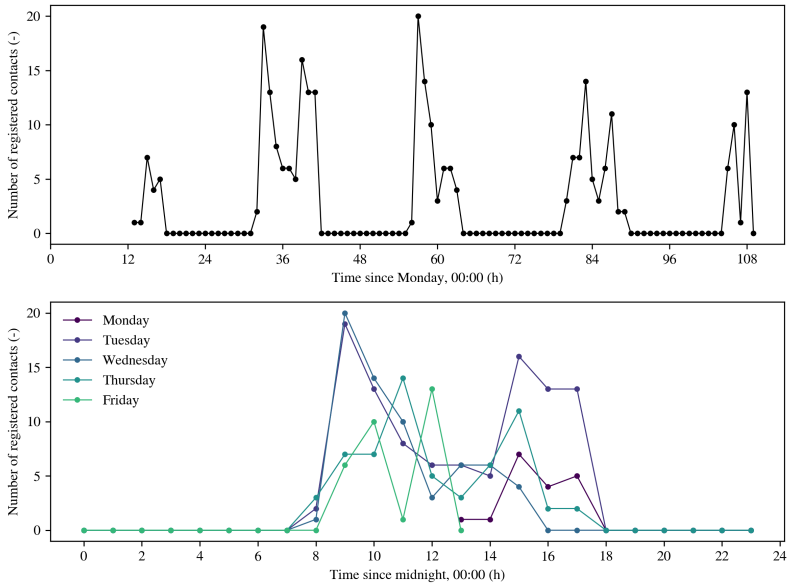

Figure B.2.11: Number of registered contacts per hour, where the two involved individuals belong to the MED and ADM categories: **MED-ADM**.

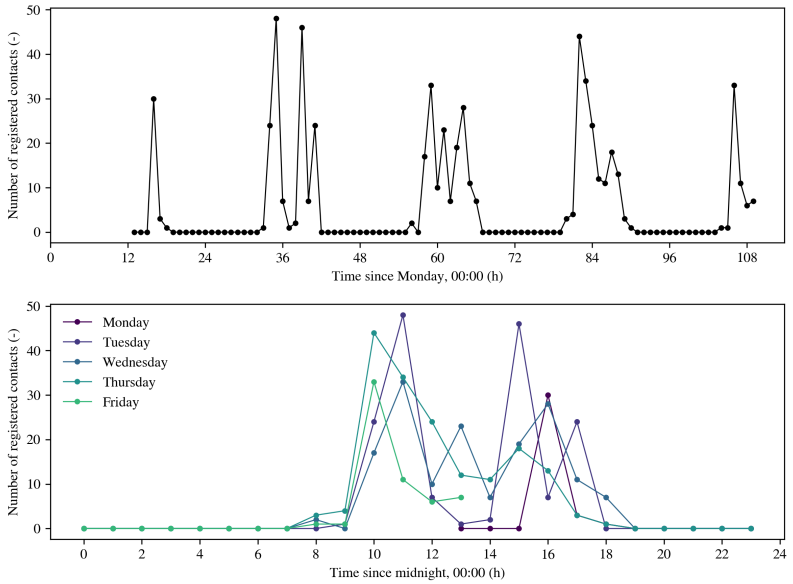

Figure B.2.12: Number of registered contacts per hour, where the two involved individuals belong to the MED and PAT categories: **MED-PAT**.

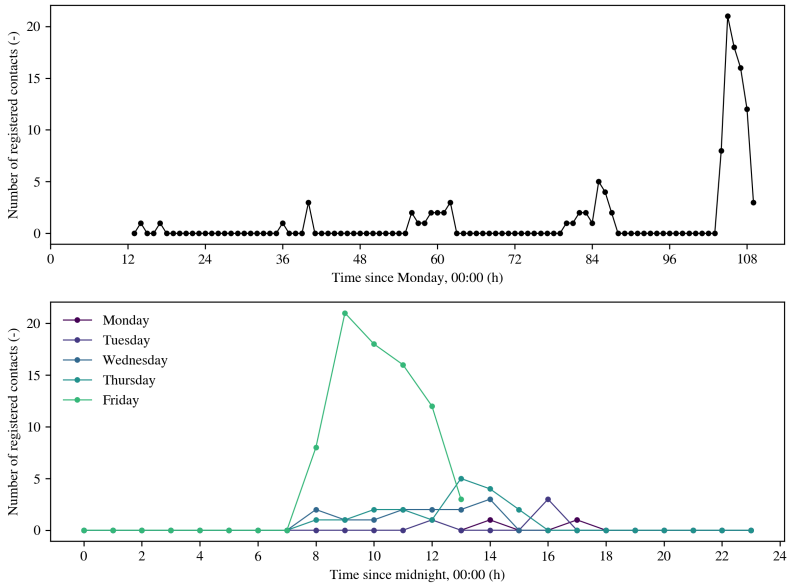

Figure B.2.13: Number of registered contacts per hour, where both of the involved individuals belong to the ADM category: **ADM-ADM**.

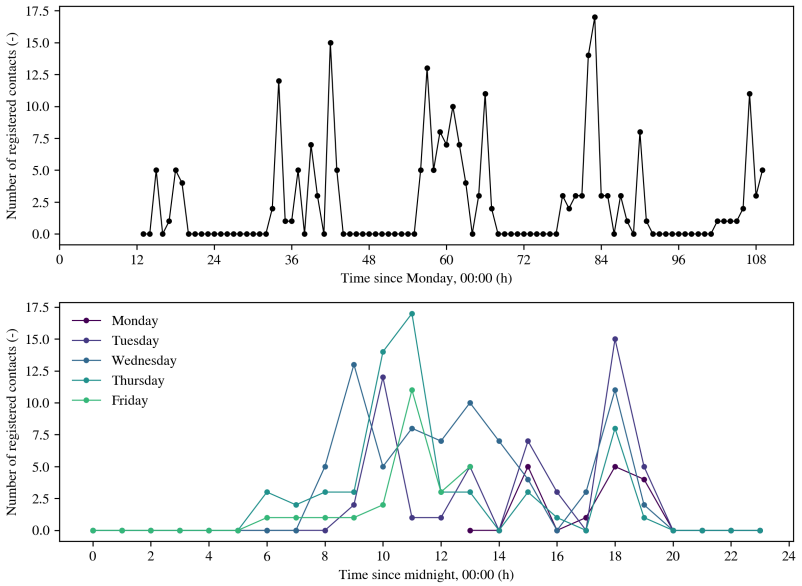

Figure B.2.14: Number of registered contacts per hour, where the two involved individuals belong to the ADM and PAT categories: **ADM-PAT**.

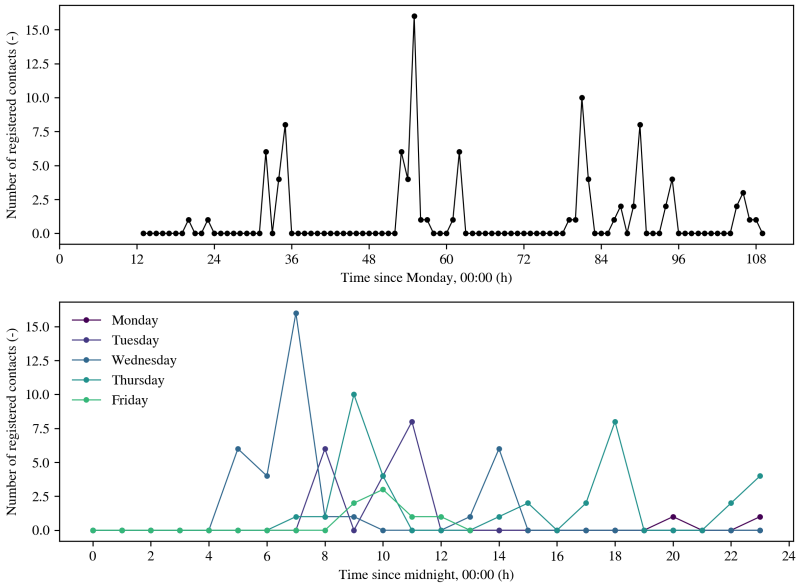

Figure B.2.15: Number of registered contacts per hour, where both of the involved individuals belong to the PAT category: **PAT-PAT**

B.2.4 Patient Hospitalization

In order to determine the duration of patient hospitalization, the patients were assumed to be present in the ward from the beginning if they had registered contacts within the first 12 hours. Similarly, the patients were assumed to be present in the ward until the end of the study period if they had registered contacts in the course of the 12 last hours. The upper panel of Figure B.2.16 shows how the bed coverage varies with time given these assumptions. The bar plot in the lower panel shows the distribution of bed coverage in the course of the 97 hours.

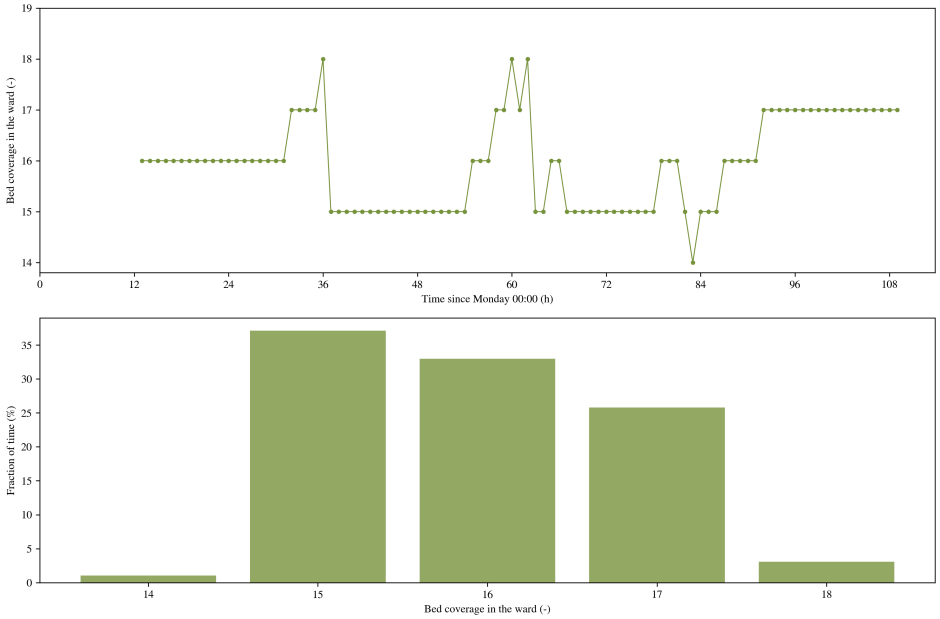

Figure B.2.16: The upper panel shows the bed coverage as function of time given the assumptions on patient admission and discharge. The lower panel shows the same data summarized in a bar plot.

B.2.5 Contact Pattern

Accumulated Number of Contacts

The total number of contacts per individual serves as basis for the contact parameter extraction discussed in Section 3.2.3. A stacked area plot visualizing how the number of accumulated contacts for each role category evolves with time is presented in Figure B.2.17. The colors reflects the total number of contacts registered for the given individual over the whole study period.

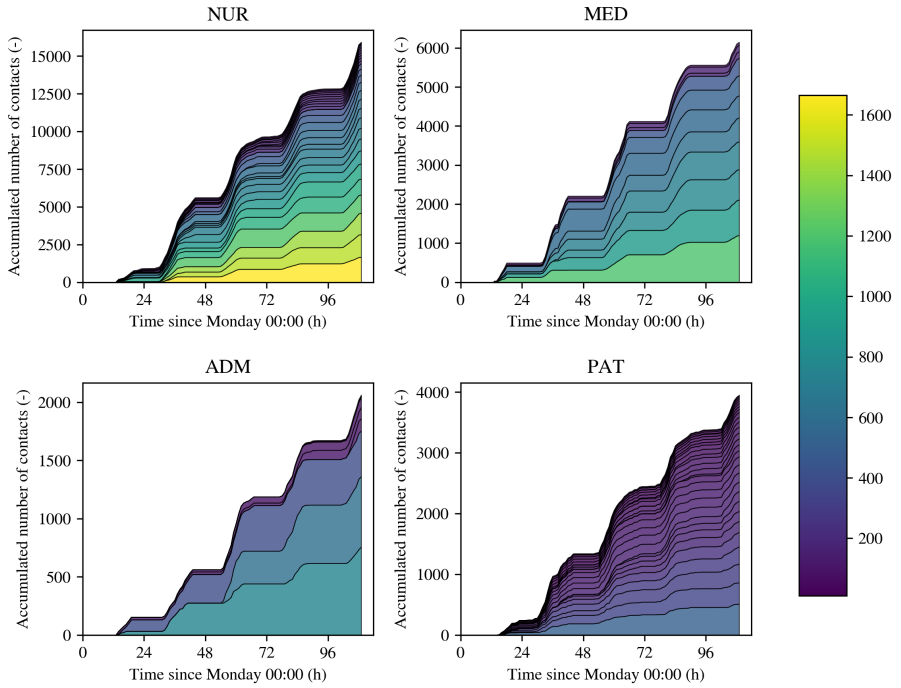

Figure B.2.17: Stacked area plot showing how the accumulated number of contacts evolve with time for the four different roles.

Relative Contact Rate

The contact rate per hour is calculated based on the total number of registered contacts and the estimated number of hours present in the ward. Table B.2.4 presents the the total number of contacts, estimated hours of presence and the calculated contact rate per hour for each individual.

Table B.2.4: Overview of the total number of registered contacts, the estimated duration of presence in the ward, and the calculated contact rate per hour for each of the 75 individuals constituting the nodes in the empirical contact network. The table is divided in the four different roles, and [innbyrdes] sorted in descending order based on the total number of contacts.

Role	ID	Contacts in total	Estimated hours of presence	Contact rate per hour
NUR	1115	1665	30	55.5000
	1210	1493	29	51.4828
	1295	1424	30	47.4667
	1207	1216	14	86.8571
	1164	1053	25	42.1200
	1193	1000	22	45.4545
	1109	881	23	38.3043
	1149	788	14	56.2857
	1196	704	19	37.0526
	1181	693	30	23.1000
	1190	628	22	28.5455
	1205	607	28	21.6786
	1114	593	14	42.3571
	1625	535	14	38.2143
	1245	468	20	23.4000
	1202	403	14	28.7857
	1629	361	7	51.5714
	1105	202	9	22.4444
	1108	194	26	7.4615
	1142	174	17	10.2353
	1485	170	22	7.7273
	1100	156	21	7.4286
	1613	138	11	12.5455
	1246	121	11	11.0000
	1261	109	14	7.7857
	1238	78	11	7.0909
	1116	67	7	9.5714
MED	1157	1195	41	29.1463
	1144	913	41	22.2683

Continued on next page

Table B.2.4 – *Continued from previous page*

	1221	778	30	25.9333
	1159	710	36	19.7222
	1260	615	35	17.5714
	1191	563	16	35.1875
	1148	516	41	12.5854
	1130	442	31	14.2581
	1660	175	25	7.0000
	1168	159	41	3.8780
	1152	79	11	7.1818
ADM	1098	757	41	18.4634
	1658	606	26	23.3077
	1179	392	25	15.6800
	1671	105	6	17.5000
	1232	95	21	4.5238
	1209	85	36	2.3611
	1535	13	16	0.8125
	1525	10	16	0.6250
PAT	1365	509	97	5.2474
	1378	352	97	3.6289
	1352	307	97	3.1649
	1383	287	97	2.9588
	1391	238	97	2.4536
	1362	180	72	2.5000
	1307	166	74	2.2432
	1702	165	55	3.0000
	1401	160	48	3.3333
	1393	156	48	3.2500
	1374	151	50	3.0200
	1363	135	69	1.9565
	1395	131	50	2.6200
	1327	100	50	2.0000
	1701	97	50	1.9400
	1460	97	31	3.1290
	1547	97	23	4.2174
	1385	95	44	2.1591
	1469	87	26	3.3462
	1320	79	54	1.4630
	1769	77	52	1.4808
	1416	52	45	1.1556
	1377	48	25	1.9200
	1399	47	24	1.9583

Continued on next page

Table B.2.4 – *Continued from previous page*

1305	38	24	1.5833
1784	35	18	1.9444
1323	30	26	1.1538
1373	24	78	0.3077
1332	11	24	0.4583

Correlation Between Total Contact Number and Presence in the Ward

Figure B.2.18 presents scatter plots where the total number of contacts regiseterd in the ward is plotted against the estimated number of present hours. The Pearson correlation coefficient, ρ, is presented in the upper left corner of all four panels.

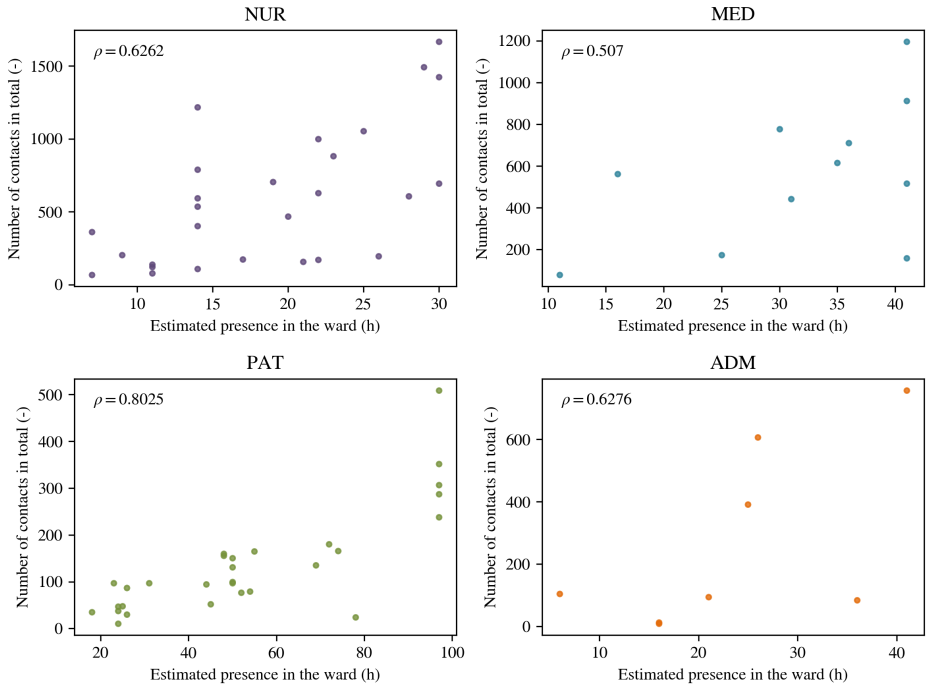

Figure B.2.18: The total number of contacts registered for an individual plotted against the estimated hours of presence in the ward. The Pearson correlation coefficient, a measure of degree of linear correlation between the two variables is presented in the upper left corners of the plots.

Relative Contact Data Fitted by Lognormal Distribution

Figure B.2.19 presents eight panels corresponding to those presented in Figure 3.2.25. There are two panels for each role: The panel to the left shows a histogram of the original relative contact data fitted with a lognormal distribution depicted with a dotted curve. The panel to the right shows the same probability function together with 100 simulations where n numbers are drawn from the given distribution. The amount n corresponds to the number of individuals belonging to each class, for instance 29 for the patient category. These numbers are drawn without any upper limit. As a result, some of the drawn numbers are multiple times higher than the highest contact rates in the original data. The magnitude of the numbers on the x-axis mirrors the large differences.

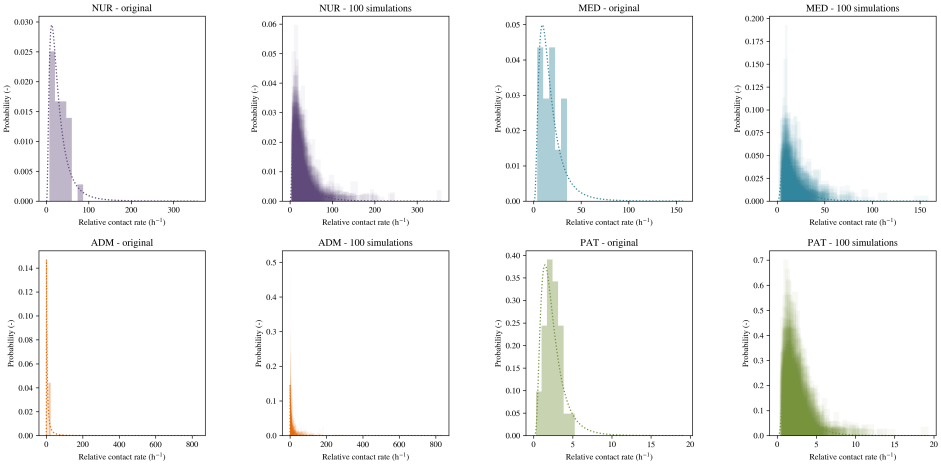

Figure B.2.19: Out of the eight panels, two and two represents one role. Of these, the panel to the left represents a histogram corresponding to the calculated relative contact rate and a lognormal distribution based on the mean and standard deviation of these data. The panel to the left presents the same distribution and 100 simulations with drawn values from this distribution. The number of drawn values in each simulation equals the number of individuals belonging to each role.

B.2.6 Solving Lognormal Parameter Equations

The epidemiological model depends on several parameters drawn from lognormal distributions. The parameters determining the shape of these distributions are either given directly from the literature or found by solving a set of equations using provided data describing the estimated centrality and variance. This set of equation is solved numerically, as illustrated in the code snippet below. The code runs through a range of σ values between 0 and 1, and calculates the corresponding value for μ. Then, the difference between the desired sum of the integral and the value of the cumulative distribution function (CDF) is calculated. The CDF equals the integral of the probability density function (PDF). The sigma value is found by choosing the lowest value of the absolute differences.

</> get_lognormal_parameters </>

```
import numpy as np
import math as m


def get_lognormal_parameters(mean_value, min_value, max_value, integral):
    min_value = 1E-10 if min_value == 0 else min_value
    t = {}

    for s in np.arange(0.00001, 1, 0.00001):
        mu = m.log(mean_value) - 0.5 * (s ** 2)

        cdf = (0.5 + 0.5*m.erf(np.divide((m.log(max_value) - mu), (2**0.5*s))))\
            - (0.5 + 0.5*m.erf(np.divide((m.log(min_value) - mu), (2**0.5*s))))\
            - integral
        t[s] = abs(cdf)

    sigma = min(t, key=t.get)
    mu = m.log(mean_value) - 0.5 * (s ** 2)

    return {'avg': mu, 'std': sigma}
```

Figure B.2.20 shows a graphic illustration of an example where the following input is given: The mean value is set to 5.6, minimum value to 2, maximum value to 14 and the integral sum is set to 0.95. These numbers correspond to an incubation period of COVID-19 where the mean value is estimated to 5.6 days with a range spanning form 2 to 14 days with $95\,\%$ confidence. The red-colored curve in the figure shows how the difference between the CDF and the sum of integral varies with different values of σ, where the black, dashed line marks the minimum. The grey, dashed line shows the corresponding μ value.

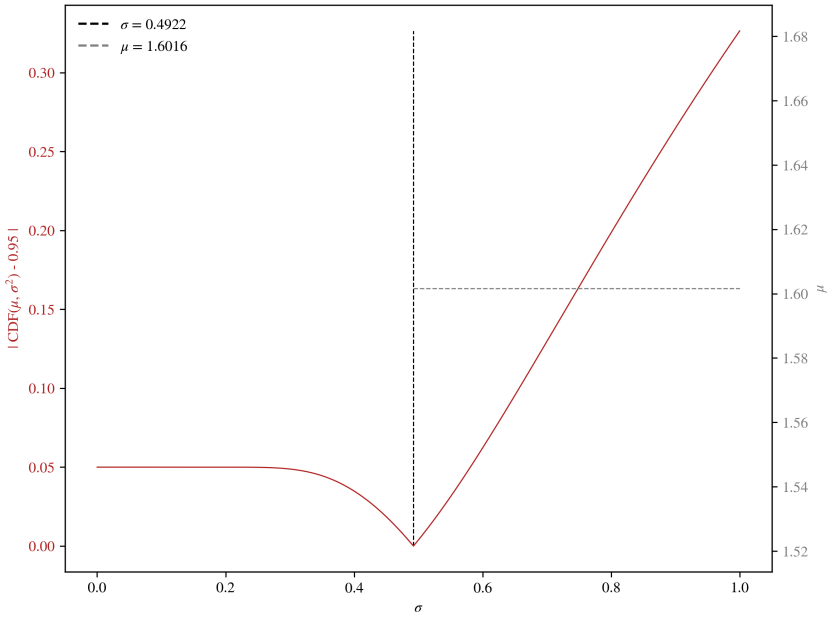

Figure B.2.20: Graphic illustration of how the set of equations is solved in order to determine the values of the lognormal parameters σ and μ. The red-colored line presents the difference between the CDF and the sum of integral. The black, dashed line marks the minimum value where the difference is closest to zero.

B.2.7 Model Parameters Values

This section gives an overview of all parameter *not* set by the user.

Table B.2.5: Location and scale parameter values for the normal distributions describing the number of contacts per bed for each simulation hour.

Hour	Location, μ	Scale, σ
0	0.0658	0.0996
1	0.0395	0.0263
2	0.2237	0.3110
3	0.2500	0.2989
4	0.0000	0.0263
5	1.2632	1.0421
6	2.1711	0.6203
7	9.7500	0.9340
8	12.4342	0.5270
9	17.6316	1.7931
10	23.4342	0.9934
11	25.1447	4.3644
12	22.4737	2.7607
13	11.4000	8.8798
14	9.9605	2.9512
15	9.6974	5.8741
16	9.0263	3.8481
17	8.0132	3.5234
18	6.8026	3.4388
19	3.9605	2.8982
20	3.9474	2.1110
21	0.7632	0.8728
22	1.4211	1.3280
23	0.9342	1.1010

Table B.2.6: Location and scale parameter values for the lognormal distributions, describing the relative contact rate for agents belonging to the different roles. The upper limit gives the maximum relative contact rate an agent can be assigned.

Role	Location, μ	Scale, σ	Upper limit (h^{-1})
NUR	3.1568	0.7804	104.2285
MED	2.6911	0.6861	42.225
ADM	1.6713	1.4697	27.9692
PAT	0.7410	0.6019	6.2969

Table B.2.7: Relative contact rates among each of the four roles in the ward.

	NUR	MED	ADM	PAT
NUR	0.5000	0.2547	0.6843	0.7662
MED	0.0962	0.5332	0.1401	0.1497
ADM	0.1256	0.0681	0.0590	0.0589
PAT	0.2782	0.1440	0.1165	0.0252

Table B.2.8: Numbers determining the percentages of employment in the ward.

Percentage	NUR	MED	ADM
0.2	5	0	1
0.4	6	2	2
0.6	7	2	3
0.8	6	3	1
1.0	3	4	1

Table B.2.9: Other parameter values set for the simulation.

Parameter	Type	Value
`symptoms`	dict	{'fever': 0.879, 'cough': 0.677, 'dyspnea': 0.186}
`hospitalization_duration`	dict	{'mu': 147.44, 'sigma': 52.78}
`new_shifts`	list	[7, 8, 9, 13, 20]
`occupancy`	float	0.8421
`rest`	int	18
`min_hospitalization`	int	8
`t0`	int	7

Appendix C

Results Supplementary

C.1 Text File Summarizing Output

The model output is summarized in four different files. One of them, a .txt file, summarizes the main model output as a text file, see Table C.1.1.

Table C.1.1: Overview of the model output which is stored in the output .txt file.

Output	Example value
cfr	23.3333
ifr	15.2174
carriers	{7: 2, 8: 2, 9: 2, 10: 2, 11: 2, 12: 2, ...}
infected_all	{7: 0, 8: 0, 9: 0, 10: 0, 11: 0, 12: 0, . . . }
infected_confirmed	{7: 0, 8: 0, 9: 0, 10: 0, 11: 0, 12: 0, . . . }
cumulative_cases	[0, 0, 0, 0, 0, 0, ...]
cumulative_deaths	[0, 0, 0, 0, 0, 0, ...]
incubation	{7: 2, 8: 2, 9: 2, 10: 2, 11: 2, 12: 2, ...}
asymptomatic	{7: 0, 8: 0, 9: 0, 10: 0, 11: 0, 12: 0, ...}
symptomatic	{7: 0, 8: 0, 9: 0, 10: 0, 11: 0, 12: 0, ...}
temporary_staff	{7: 0, 8: 0, 9: 0, 10: 0, 11: 0, 12: 0, ...}
quarantine	{7: 0, 8: 0, 9: 0, 10: 0, 11: 0, 12: 0, ...}
isolation	{7: 0, 8: 0, 9: 0, 10: 0, 11: 0, 12: 0, ...}
contacts	{7: 406, 8: 472, 9: 798, 10: 907, 11: 773, 12: 883, . . . }

C.2 Model Stability Assessment

In order to assess the stability of the stochastic model, 100 simulations with identical input parameters were run; the base line parameter set with one instead of two hospital wards. The average values with bands corresponding to the standard deviations, $\bar{x} \pm s$, are presented in Figure C.2.1. The panels present the number of carriers, infected and confirmed cases, respectively.

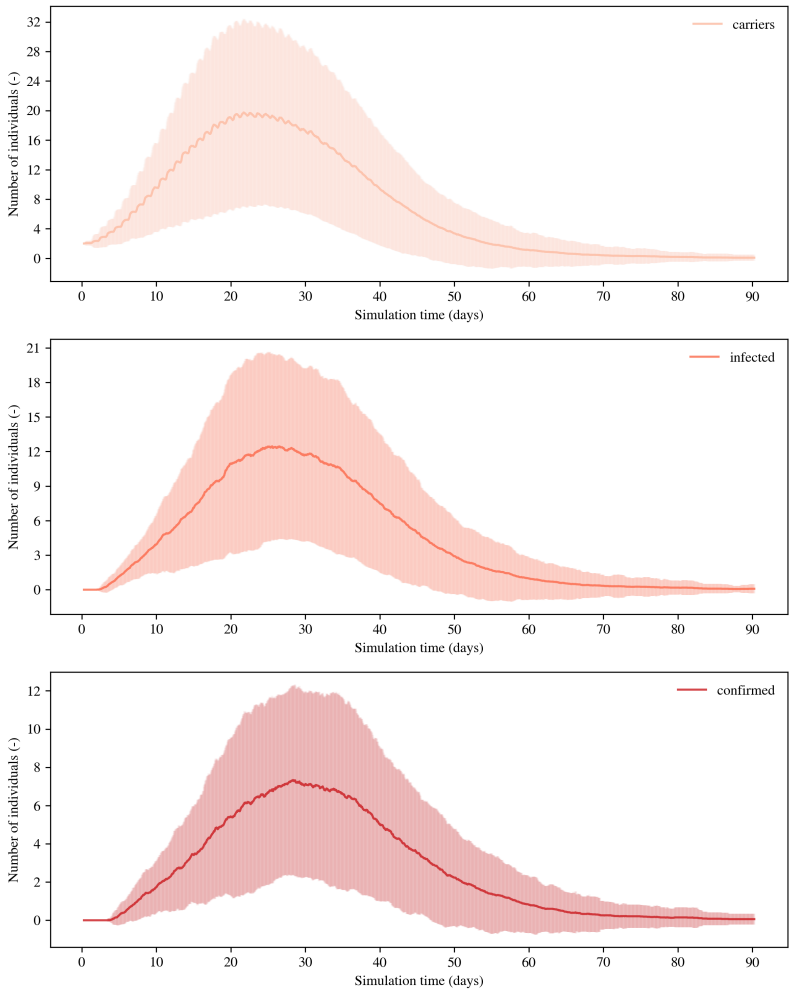

Figure C.2.1: The three panels illustrate the number of carriers, infected and confirmed cases, respectively, in 100 simulation runs where the number of wards is set to one, all other parameters to their base line value. The solid lines represent the mean value, whereas the bands are made up of error bars representing the calculated standard deviation in both directions, $\bar{x} \pm s$.

Signe Onstad Sævareid

NTNU
Norwegian University of
Science and Technology