Master's thesis

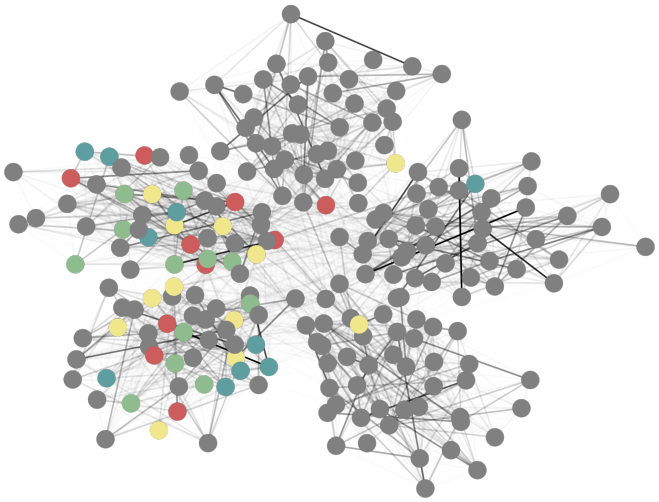
NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Sara Johanne Asche

Modelling the transmission of COVID-19 in primary schools

Using empirical interaction data to build a framework for modelling interactions and transmission of COVID-19 in primary schools

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



NTNU
Kunnskap for en bedre verden

Sara Johanne Asche

Modelling the transmission of COVID-19 in primary schools

Using empirical interaction data to build a framework for modelling interactions and transmission of COVID-19 in primary schools

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

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

NTNU
Kunnskap for en bedre verden

Preface

This thesis concludes my Master's degree in Biotechnology, with a specialisation in Systems Biology, at the Norwegian University of Science and Technology (NTNU). When I started studying biotechnology in 2017, it was due to my fascination with synthetic biology and molecular processes. Back then, I would never have imagined writing such a theoretical and computational thesis as this current work. However, during my first few years at NTNU, my gaze was quickly shifted toward systems biology and bioinformatics.

Eivind Almaas largely contributed to this shift of interest and allowed me to further explore computational biology through being the supervisor of this thesis. It is, therefore, with gratitude that I would like to sincerely thank Eivind for igniting the spark to work with both data and biology and for allowing me to be a part of the wonderful and supportive research group *Network Systems Biology*.

Furthermore, I would like to thank my co-supervisor, André Voigt, for his patience and words of encouragement throughout the work on my master's. André allowed me to explore and develop my computational skills at my own pace and supported me throughout the entire two-year process of working on this thesis.

I truly believe a picture is worth a thousand words, and I, therefore, ask the reader to be patient with the number of descriptive images included in this thesis.

Finally, I would like to thank my family, friends, and most importantly, my partner for enduring with and supporting me throughout the entire five-year process of getting my degree. Thank you!

Sara Johanne Asche,

16 th of May 2022

Abstract

When COVID-19 was declared a pandemic in early 2020, strict rules and regulations were quickly put into play to help mitigate and reduce transmission of the disease. Amongst those heavily affected by these measures were children and adolescents, having digital at-home schooling and being isolated from their peers during the first months of the pandemic. Furthermore, many close contacts between individuals belonging to different households occur in primary schools and may help amplify the overall transmission of the disease. Therefore, this Master's thesis aims to produce a novel framework for detailed agent-based modelling of social interactions in primary schools. In addition, the model should allow for simulation of COVID-19 transmission, specifically in primary school environments.

Through investigation of an empirical network with high-resolution interactions between peers in a primary school, collected and published by Barrat et al. [1], the main features and characteristics of the interactions were extracted and used to extrapolate a general model. The model produces a network representation of all occurring interactions and allows for tuning input variables, such as the size of a given school, the number of grades and classes, and the number of hours the students attend school.

The final interaction model closely resembles the interactions found in the empirical network, with the exception of the model having lower modularity and more interactions between individuals that were not in the same grades and classes compared to the empirical network. Disease transmission was run on the model to investigate the effect of the Norwegian government's traffic light levels and different testing strategies on disease mitigation in primary schools. Weekly testing of students with green traffic light levels was identified as the most efficient strategy, followed by yellow traffic light level in situations where there are not enough resources for weekly testing.

The novel framework allows for the generation of detailed interaction networks in primary schools and may be incorporated into the NTNU COVID-19 Taskforce's agent-based model. This framework thus creates a model that can be helpful for future exploration of different COVID-19 mitigation strategies in primary schools. However, the model was based on interaction data collected over two consecutive days at the same school. Therefore, to improve the quality of the model, more high-resolution data needs to be gathered from different primary school settings to help generalise the model.

Sammendrag

Da COVID-19 ble erklært en pandemi våren 2020, ble ulike regler og restriksjoner raskt implementert for å redusere spredning av sykdommen. Blant de som ble hardest rammet av disse tiltakene, finner man barn og ungdom, som ble isolert fra sine jevnaldrende gjennom digital hjemmeskole i løpet de første månedene av pandemien. Videre, har barn og unge på barneskolen ofte flere, mer langvarige og tette interaksjoner med sine skolekamerater, noe som gir mange potensielle smittebærende interaksjoner. Målet med denne masteravhandlingen er å produsere et nytt rammeverk for en detaljert agentbasert modellering av sosiale interaksjoner i barneskolen. Videre skal modellen kunne brukes for å simulere smittsomhet av COVID-19 på barneskolen.

Gjennom undersøkelser av et empirisk nettverk som inneholder detaljerte 20-sekunders interaksjoner mellom jevnaldrende på barneskolen, ble grunntrekkene i interaksjonenes varighet og karakteristikk identifisert og brukt til å ekstrapolere en generell modell for nettopp disse interaksjonene. Modellen produserer et nettverk av interaksjoner mellom individene på en skole og gir muligheter for å justere input-variabler som for eksempel størrelsen på en bestemt skole, antall trinn og klasser, og antall timer elevene er på skolen.

Det foreslåtte rammeverket produserte en interaksjonsmodell som genererer svært like interaksjoner sammenliknet med interaksjonene mellom elevene i det empiriske nettverket. Likevel produserer modellen flere interaksjoner på tvers av trinn og klasser sammenliknet med det empiriske nettverket. Videre ble sykdomsspredning simulert på modellen for å undersøke effekten av norske myndigheters strategi for å hindre sykdomsspredning på barneskoler. Strategier som ble sett på i denne oppgaven er trafikklysmodellen og ulike teststrategier. Ukentlig testing av elevene med grønt trafikklysnivå virker til å være den mest effektive strategien, etterfulgt av gult trafikklysnivå i situasjoner der det ikke er tilstrekkelig ressurser eller mulighet for å gjennomføre ukentlig testing.

Det nye rammeverket gjør det derfor mulig å generere detaljerte interaksjonsnettverk for barneskoler, og kan bli innarbeidet i NTNU COVID-19 Taskforce sin agentbaserte modell. Rammeverket gir dermed en unik mulighet til å undersøke ulike strategier for å hindre spredning av COVID-19 i barneskoler. Siden modellen kun er basert på interaksjonsdata fra to dager på en bestemt skole, vil forbedring av kvaliteten til modellen inkludere å samle inn mer interaksjonsdata fra forskjellige skoler med ulik struktur. Dette vil også hjelpe gjøre modellen mer generell.

Table of Contents

List of Tables

List of Figures

Abbreviations

ABM Agent-Based Modelling. 2, 3, 4, 8, 19, 20, 24, 25, 26, 30, 79

ACE2 Angiotensin-converting enzyme 2. 5

alphaCoV *Alphacoronavirus*. 4

BA Barabási–Albert model. 17, 18, 27, 50, 51, 71, 78

betaCoV *Betacoronavirus*. 4

CDF Cumulative density function. 15

COVID-19 Coronavirus Disease of 2019. 2, 3, 4, 5, 6, 7, 8, 75, 76, 77, 78, 79

ER Erdős-Rényi model. 17, 18, 27, 50, 71, 78

FHI Folkehelseinstituttet or The Norwegian Institute of Public Health. 2, 39

GIS Geographic Information System. 20

MERS *Middle East Respiratory Syndrome*. 4, 5

NPI Non-Pharmaceutical Interventions. 1, 2, 3, 7, 77

PDF Probability density function. 15, 16, 17

RBD Receptor-binding domain. 5

RF Radio Frequency. 21

RFID Radio Frequency Identification Device. 21, 72, 79

SARS *Severe Acute Respiratory Syndrome*. 4, 5

SARS-CoV-2 *Severe Acute Respiratory Syndrome Coronavirus 2*. v, 1, 2, 3, 4, 5, 6, 70

SEIR Susceptible, Exposed, Infected, Recovered. 38

SFO Skolefritidsordning, or After school program. 8

SIR Susceptible, Infectious, Recovered. 4, 9, 38

TMPRSS2 Transmembrane protease Serine 2. 5

UDIR Utdanningsdirektoratet or Norwegian Directorate for Education and Training. 7, 42

WHO World Health Organisation. 1, 2, 4, 6, 77

WS Watts-Strogatz model. 17, 27, 50, 71, 78

Chapter 1

Introduction

During the second wave of the Bubonic plague that ravaged Europe in the 15th and 16th century, the seaport of Alghero in Sardinia suffered a large epidemic outbreak. The disease, caused by the bacterium *Yersinia pestis*, was at the time believed to be caused by a pathogenic agent called miasma that polluted the surrounding air and could infect healthy individuals through their airways or the pores of the skin [2]. However, compared to other outbreaks during the second wave, the outbreak in Alghero was localised and did not spread to surrounding districts. This was primarily due to the effort of the elected *Protomedicus* of Sardinia, Quinto Tiberio Angelerio. The *Protomedicus*, which can be compared to a chief municipal physician, quickly implemented prophylactic measures to prevent further transmission of the disease in the seaport city.

In 1588 Angelerio published the booklet *Ectypa Pestilentis Status Algheriae Sardiniae* (Instructions on the Alghero, Sardinia, Plague Epidemic), where he wrote extensively on both the history of the 1582-1583 outbreak as well as measurements he had implemented to decrease the spread of the disease [2]. He detailed the comprehensive policies and measures put into play, such as isolating plague patients, quarantining individuals or goods that had been in contact with contamination, and disinfecting and burning items that were in direct contact with sick individuals. The city was also put into a lockdown, where arriving ships and crew were quarantined for 30 days before entering the seaport. Angelerio quickly became loathed among the population of Alghero due to his strict measures. However, the second wave of the Bubonic plague passed Alghero within eight months of arriving. Similar examples of systematically implementing regimes to hinder disease transmission can also be seen throughout history; see the historical overview *Infectious Disease epidemiology: Theory and Practice* by Nelson and Williams [3].

When declared a pandemic by the World Health Organisation (WHO) in March of 2020 [4], *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) caused similar Non-Pharmaceutical Interventions (NPI)'s to be implemented worldwide. To halt transmission of the highly contagious virus, strict infection control routines were implemented,

of which many impacted everyday life. City-wide lockdowns were implemented in locations with high incidents of Coronavirus Disease of 2019 (COVID-19) cases; quarantines were set in place for travellers, and mandatory use of masks in situations where one could not keep one meter apart was recommended. In Norway, Folkehelseinstituttet or The Norwegian Institute of Public Health (FHI) also implemented a resource-demanding disease tracking system to trace the source of outbreaks and quarantine individuals that may have been exposed to the virus.

Though this traditional handling of pathogens has been proven effective, the increased globalisation of the economy, culture, and populations during the 21th century has led to increased interactions across land borders [5]. In turn, halting inter-country travel and trading when implementing similar measures like the ones proposed by Angelerio has caused a significant financial and economic impact on businesses and individuals during the COVID-19 pandemic [6]. Furthermore, in 2020, WHO predicted that self-isolation and social distancing would increase loneliness, anxiety, and depression. This trend has also been shown by Kumar and Nayar, who surveyed the population during the lockdown in India, where they recorded a 20% increase in mental illness during the first months of the pandemic [7]. Similarly, De Figueiredo et al. highlighted that children and adolescents are more exposed to increased distress, feeling of hopelessness, domestic violence, and sensorial deprivation and neglect during the pandemic [8]. In the long term, this exposure can have grave consequences on the individuals leading to underdeveloped brain circuitry, obesity, psychiatric disorders, and more [8]. Although Angelerio's prophylactic measures were sufficient in the 16th century, implementation in a highly globalised world may give rise to more ethical dilemmas. It is therefore vital to not only take the direct consequences of COVID-19 transmission into account but also the impact of NPIs that followed the pandemic when considering the best strategy when facing transmissible pathogens.

As of April 2022, SARS-CoV-2 is still declared a pandemic and COVID-19 has claimed 6.2 million lives worldwide [9]. The strict restrictions and policies previously implemented have been discontinued by many countries due to the protection provided by vaccines. Since SARS-CoV-2 has the potential to become a seasonally re-emerging virus similar to influenza [10], it is still, however, important to continue investigating the mechanisms behind its transmission as well as other options to help mitigate disease transmission in the future. In order to simulate the spread of COVID-19 many models are proposed, ranging from simple compartmental models, see Ndaïrou et al. [11], to more complex Agent-Based Modelling (ABM), see the agent-based community spread model by the NTNU COVID-19 Taskforce [12] and the ABM by Cuevas [13] aimed at evaluating the risk of COVID-19 transmission in facilities. The value of such models is that different NPIs can be easily implemented and their effect quantified through running simulations.

Since the lockdown heavily impacted children and adolescents, it is also interesting to look into this group further. Barrat et al. highlights the role school children play as hubs in the community spread of respiratory diseases such as influenza [1]. This is due to their susceptibility to infection combined with sustained close contact in communities such as schools, playgrounds, and similar venues. Viboud et al. quantifies this risk and concludes that vaccination of children would prevent 32-38% of influenza cases caused by contact with a sick child's household [14]. Therefore, it is essential to investigate and overview

how children interact and spread disease in primary school environments to quantify the community spread through households.

Currently, most of the investigation of the role children in primary schools play in community transmission of COVID-19 is limited to retrospective studies using the frequency of COVID-19 cases recorded. For instance, Brandal et al. describes the approach of quantifying the transmission of SARS-CoV-2 by tracing and testing contacts of confirmed COVID-19 positive individuals [15]. Similarly, Xu et al. [16] conducted a systematic review and meta-analysis of articles and their associated studies discussing the transmission of SARS-CoV-2 in schools to try and quantify the role children play in disease transmission. Although there have been studies linking transmission in children with COVID-19 transmission in communities, see Goldstein et al. [17], few models attempt to create a simulated model for interactions in primary schools and investigate the impact of different regulations and NPIs may have on disease transmission.

Since primary schools may play a prominent role in the community transmission of SARS-CoV-2, **this thesis aims to create a novel framework for a model of interactions in primary schools and use this to investigate the effect of different NPIs have on the transmission of COVID-19**. By using high-resolution experimental data on interactions in primary schools, provided by Barrat et al.[1], this thesis will attempt to produce a framework for simulating interactions between primary school pupils on a day-to-day basis in a network. This network of interactions further allows for the simulation of disease transmission. Furthermore, implementing different policies such as routine testing and the Norwegian government's traffic light model on the simulated network can be done to assert their effect on mitigating disease transmission. Finally, the framework should produce general networks of interactions in primary schools so that it, at a later stage, may be implemented into the ABM by NTNU COVID-19 Taskforce [12].

Chapter 2

Theory

This chapter will provide a thorough overview of the underlying theory relevant to this thesis. Since the field of COVID-19 is rapidly expanding, it is worth mentioning that this section was written in March 2022 and that new findings may have been made afterwards. Firstly, section 2.1 will describe the structure and transmission of SARS-CoV-2 alongside measures and regulations that were put into place to hinder said transmission in Norway. Then, section 2.2 will detail fundamental disease spread models, ranging from differential equation Susceptible, Infectious, Recovered cohort models to complex Agent-Based Modelling. Finally, the paper and study by Barrat et al. [1], containing the empirical interaction data used in this thesis to characterise features of primary school interactions, will be described in detail in section 2.3.

2.1 SARS-CoV-2 and viruses of *coronaviridae*

Although SARS-CoV-2 is the first virus of the viral *coronaviridae* family to be declared a pandemic by the WHO, it is not the only known virus of the family to infect humans [18]. Table 2.1 describes the genus, name, lineage, and symptoms of the seven currently known viruses of the family to infect humans. The family is divided into two genera; The *Alphacoronavirus* (alphaCoV) and *Betacoronavirus* (betaCoV). Viruses of alphaCoV alongside betaCoV viruses of lineage A usually cause a common cold when they infect humans, while the remaining viruses of betaCoV cause some form of a respiratory syndrome [18].

Even though no diseases caused by this family have been declared a pandemic previously to SARS-CoV-2, emerging behaviour has been observed in both the *Severe Acute Respiratory Syndrome* (SARS) and *Middle East Respiratory Syndrome* (MERS) viruses. Both underwent zoonosis, a term that describes any pathogen crossing a species barrier hosts [19], and were just recently identified in human hosts. From November 2002 to August 2003, there were 8422 confirmed cases of SARS across 32 countries [20], and the virus is

Table 2.1: Description of viruses of *Coronaviridae* family that infect humans, alongside their lineage and associated symptoms. A line marks viruses that have multiple lineages [18]

Genus	Name	lineage	Symptoms
alphaCoV	HCoV-229E	—	Common cold
	HKU-NL63	—	Common cold
betaCoV	HCoV-OC43	A	Common cold
	HCoV-HKU1	A	Common cold
	SARS-CoV	B	Severe acute respiratory syndrome
	SARS-CoV-2	—	Severe acute respiratory syndrome
	MERS-CoV	C	Severe acute respiratory syndrome

believed to be of masked palm civet and bat origin [21]. Meanwhile, viral spillover from camels is what caused the MERS outbreak in April 2012, which led to 2496 documented cases from April 2012 to December 2019 [20]. Cases are still occurring in areas where human-camel interactions are high. It is important to keep these two viruses in mind, as they share characteristics and DNA with the virus that caused the COVID-19 pandemic. SARS-CoV-2 shares about 80% of its genome sequence with SARS and 50% with MERS [19]. This chapter will now go into more detail on the disease transmission characteristics of SARS-CoV-2.

2.1.1 The viral structure of SARS-CoV-2 determines tissue tropism and transmission rate

The viral structure of SARS-CoV-2 can be seen in Figure 2.1 and is a part of deciding the tissue tropism of the disease. The tissue tropism of any given disease denotes the types of tissues in which a virus can infect and replicate. The virus consists of a bilipid membrane, spike (S), membrane (M), and envelope (E) proteins and a nucleocapsid (NC) surrounding positive-stranded RNA (+ssRNA). The spike proteins determine which cells the virus will interact with and mediates viral entry into host cells. Several mechanisms help drive this viral entry. First, the spike protein must be primed by being proteolytically cleaved by the cellular Transmembrane protease Serine 2 (TMPRSS2). After this cleavage, the Receptor-binding domain (RBD), which is found towards the tip of the spike protein, is primed and able to interact with another cellular receptor, the Angiotensin-converting enzyme 2 (ACE2). The binding of the two facilitates viral envelope fusion of the virus, where cellular cathepsin L assists in entry into the cells. The genome of SARS-CoV-2 is then released and initiates the production of new viruses that can be further transmitted [19]. For more detail on SARS-CoV-2 transmission and pathogenesis, see Harrison et al. [19].

The main consequence of the viral entry of SARS-CoV-2 into cells is that it requires the presence of ACE2 and a protease, such as TMPRSS2. Harrison et al. determines that the tissue tropism of SARS-CoV-2 is the upper respiratory tract and airways (epithelial and endothelial cells). Furthermore, Xiao et al. identified a possible fecal-oral-transmission route through the gastrointestinal tracts [23]. This gives the following possible transmission pathways: respiratory droplets, aerosol, direct contact with contaminated surfaces or

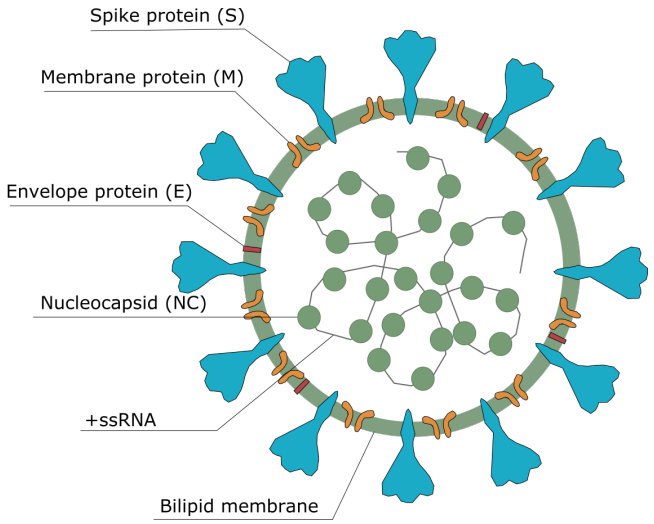


Figure 2.1: Viral structure of SARS-CoV-2 with its Spike protein (S), Membrane protein (M), Envelope protein (E), Nucleocapsid (NC), positive strand RNA (+ssRNA) and bilipid membrane. Modified from [22].

infected individuals, and the fecal-oral route [19].

During the first few months of the pandemic, many argued that SARS-CoV-2 might mutate more slowly compared to other RNA viruses such as influenza [24]. Epidemiologist William Hanage argued that in a susceptible population, the virus had low evolutionary pressure to adopt new and more competitive behaviour [24]. However, it was later discovered that SARS-CoV-2 could, in fact, generate and accumulate new mutations quickly [18]. Firstly, this is due to SARS-CoV-2 being an RNA virus and using an RNA-dependent polymerase without proofreading ability [18], causing the rapid emergence of new genotypes and mutations of the virus. Secondly, the virus can infect other species, further increasing the possibility of mutation and sharing of genes between viruses that affect different species. Finally, the accumulation of mutations can be sped up by the recombination of different variants. All these characteristics allow for rapid mutation of the virus, which can give diverse transmission and disease courses, making it likely to reemerge [18]. Multiple variants of the virus have already emerged, for instance, the newly discovered delta and omicron variants [25].

2.1.2 National measures against SARS-CoV-2 in primary schools in Norway

After SARS-CoV-2 was declared a pandemic in March 2020, the WHO called for all countries to implement regulations and measures to limit the transmission of COVID-19. Measures implemented include social distancing, limiting travel within and between countries, public information campaigns, closing or partially shutting down workplaces and schools, and finally imposing quarantines and isolation on individuals exposed to the

virus. On Thursday, the 12th of March 2020, the Norwegian government imposed multiple strict and comprehensive measures nationally [26]. To read more about the history and the Norwegian governments handling of the COVID-19 pandemic, see the 2021 report by Koronakommisjonen, the Norwegian Coronavirus Commission [26].

Of importance in this master, however, is the different governmental measures aimed to halt or reduce COVID-19 transmission, specifically in primary schools. Initially, the NPI introduced on 13th of March included no physical attendance at primary schools and a shift to digital teaching. On the 11th of May, the Norwegian government introduced a traffic light model to allow students to begin returning to their schools [27]. The traffic light model describes which infection control measures a given school should implement at different times. The model includes three levels of increasing disease spread in society, from green to yellow to red, like a traffic light. Figure 2.2 shows the measures implemented at primary schools at each step, as presented by Utdanningsdirektoratet or Norwegian Directorate for Education and Training (UDIR) [28].

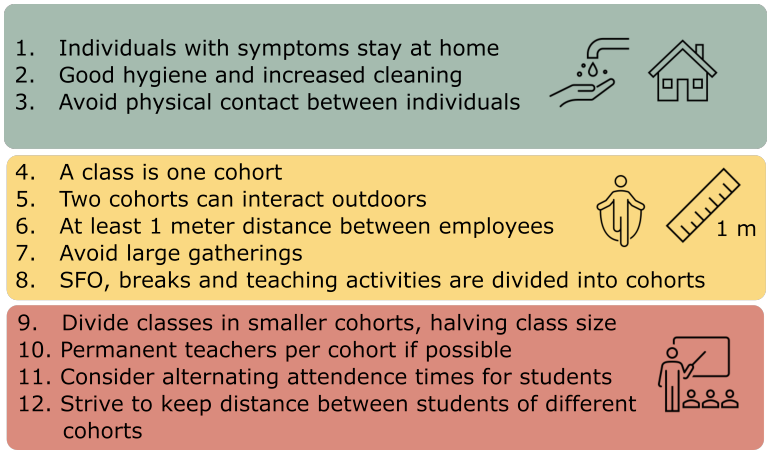


Figure 2.2: Traffic light model used in Norwegian primary schools to determine which infection control measures are implemented at different times. Ranges from green (low disease spread) to yellow (intermediate disease spread) to red (high disease spread) where each step increasing from green is a part of the next step. Modified from [28]

All levels of the traffic light model used for Norwegian schools are specific to the recent COVID-19 pandemic. The green level describes the laws and regulations schools should abide by in a situation with low levels of uncontrolled disease transmission. In such cases, the individuals with symptoms of COVID-19 are advised to stay at home, and good hygiene and increased cleaning are practised in the schools. Furthermore, there is a recommendation to avoid physical contact between individuals.

The yellow level describes the additional precautions a school is advised to implement when there is an intermediate spread of the disease. In addition to the measures implemented in the green model, the yellow model should also include dividing classes into cohorts and setting a maximum of two cohorts interacting outdoors simultaneously. The

employees at the school should also keep at least one meter distance from one another, and the school should avoid arranging events with large gatherings. Finally, all activities except interacting outdoors should be conducted in cohorts. The yellow level also includes after-school activities such as Skolefritidsordning, or After school program (SFO).

The red level describes how schools should be managed in situations with high disease spread. Smaller cohorts are advised, halving the class size. One teacher should stay permanently with one cohort if possible. Alternating attendance times for students should also be considered to decrease the number of students present at school at a given time. Furthermore, the school should strive to keep distance between students of different cohorts. Implementing these rules should help reduce the disease spread.

Table 2.2: Overview of the sensitivity of rapid antigen tests for COVID-19 when performed by different operators. The 95% confidence interval (CI) is also given [29]

Operator	Sensitivity	95% CI
Laboratory scientist	78.8%	72.4-84.3%
Trained health care worker	70.0%	63.5-75.0%
Self-trained individual	57.5%	52.3-62.6%

Later on, the strategy to keep COVID-19 transmission low also included having pupils perform rapid antigen tests at home. This could either be done weekly in areas with a lot of disease transmission or periodically, for instance, after a holiday [26]. Table 2.2 shows an overview of the sensitivity of using rapid antigen tests with different operators ranging from trained laboratory scientists to self-trained individuals [29]. Since the children either perform the tests themselves or have parents help them, the sensitivity of rapid antigen tests for COVID-19 in schools was calculated by Peto et al. to 57.5%.

2.2 Disease spread models

When creating epidemic models to predict or investigate disease transmission, there are several options. This section will detail multiple epidemic concepts and models widely used and is largely based on a comprehensive review by Duan et al. [30] on mathematical and computational approaches to epidemic modelling. Duan et al. describes the advantages and disadvantages of three different models that will be described in further detail in the following section. These include differential equation models (subsection 2.2.1), complex network models (subsection 2.2.2) and Agent-Based Modelling (subsection 2.2.3). The complexity and detail of each type of model increase as one moves from one subsection to the next.

2.2.1 Differential equation models - Reed-Frost and SIR

Differential equation models are the most straightforward and least computationally demanding epidemic models to formulate. One of the earliest examples of using epidemic models is from 1766 when Daniel Bernoulli developed a model to assess the effectiveness of the recent vaccination against smallpox [30]. The differential equation models

primarily focus on macroscopic regularities of epidemic spread, using assumptions and simplifications to calculate epidemic diffusion, threshold, and size. However, the assumption of a homogeneous, well-mixed population with simplified disease transmission limits the model's ability to accurately represent disease spreading in close detail [30].

During the 1920s, Lowell Reed and Wade Hampton Frost presented what is now known as the Reed-Frost model, which described how diseases spread across a population. It uses a binomial stochastic epidemic model to determine whether infection occurs after contact between two individuals. The most significant outtake from the model is that it laid the groundwork for assumptions and rules for numerous other epidemic models that are still in use nowadays. The assumptions described by the model are quoted verbatim from Abbey and are as follows [31]:

1. *"The infection is spread directly from infected individuals to others by a certain kind of contact (adequate contact) and in no other way"*
2. *"Any non-immune individual in the group, after such contact with an infectious person in a given period, will develop the infection and will be infectious to others only within the following time period, after which he is wholly immune"*
3. *"Each individual has a fixed probability of coming into adequate contact with any other specified individual in the group within one time interval, and this probability is the same for every member of the group"*
4. *"The individuals are wholly segregated from others outside the group"*
5. *"These conditions remain constant during the epidemic"*

Furthermore, a second cornerstone, of which many epidemic models are derivatives, is the Susceptible, Infectious, Recovered (SIR) model presented by Kermack and McKendrick in 1927. Based on the assumptions of the Reed-Frost model, it is a deterministic compartmental model where individuals progress from one compartment to another when the state of an individual changes. Creating compartments for each possible state an individual can be in helps form a dynamic model to assess the transmission of infectious diseases [30].

Figure 2.3: Dynamics of transition between compartments in the SIR model, where S details susceptible, I, infectious, and R, recovered. β describes the transition from S to I, whilst γ describes transition from I to R.

The population is assumed to be homogeneous and well-dispersed, and individuals are divided into compartments based on their health status (whether or not they are susceptible, infectious, or recovered). An individual will move from one compartment to another if

their health status changes, with differential equations describing the rate of this transition. Figure 2.3 explains the rates of transition between the compartments as one moves from susceptible to infected to recovered.

Equation 2.1, 2.2 and 2.3 for $\beta, \gamma > 0$ describes the transition of individuals to and from the three compartments. For instance, will Equation 2.2 describe the rate at which individuals transition into or from the infected compartment [30].

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \tag{2.1}$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \tag{2.2}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{2.3}$$

To further calculate R_0, the basic reproduction number, one must divide the number of individuals transitioning from susceptible to infected by the number of individuals moving from infected to recovered. The equation can be seen in Equation 2.4, and the range of R_0 is a part of determining the course of a disease. R_0, therefore, indicates how many individuals, on average, each infected individual infects before recovering. For instance, a R_0 below one details that more individuals recover from a disease without infecting other individuals, indicating a halt in the spread of the disease. On the contrary, $R_0 > 1$ suggests that the disease spreads faster than people are recovering, allowing it to spread across a susceptible population [30].

$$R_0 = \frac{\beta}{\gamma} \tag{2.4}$$

Although mathematical models allow for less complex computation and faster simulations, there are multiple drawbacks and limitations to using this method to predict the spread of disease. The assumption of homogeneous and well-mixed populations has gotten more accurate due to the recent globalisation; however, it is still limited with regard to representing different individuals in situations where they have microscopic behaviours and attributes that are a part of determining their interactions. In addition, a drawback to the simple models is their low number of variables which further limits the possibilities of simulating transmission for different scenarios with changed human behaviour. Furthermore, the models discussed in this section assume that all individuals are connected and can infect each other in a single step. Although this assumption is becoming more accurate in current the globalisation, geographical and social still exist. Finally, the parameterised variables primarily rely on averages that completely rule out the heterogeneous contagiousness and the time scale of epidemic progress [30].

2.2.2 Network models - Graph theory

Network models are introduced to study disease transmission while addressing the homogeneous population's assumption [30]. This section will detail graph theory and its

application when investigating epidemiology.

Definition of a graph

Graph, or network, theory is a term describing the study of graphs, which are mathematical structures consisting of two finite sets of nodes or vertices, **V**, and edges, **E**. The set **V** consists of multiple nodes that are entities that have a given relationship with one another. The relationship between two nodes is presented in the **E** set [32]. One example is a node-set consisting of the five following nodes, $\mathbf{V} = (0, 1, 2, 3, 4)$ where the relationship between the nodes is described by edges $\mathbf{E} = ((0, 1), (0, 2), (1, 3), (3, 2), (3, 4), (4, 2))$. In general, any tuple (a, b) describes that node a interacts with node b in some way [32].

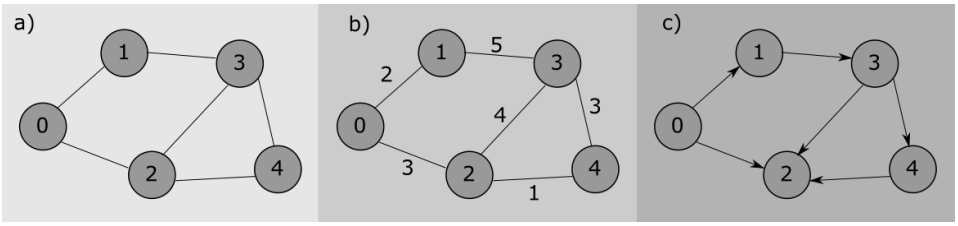


Figure 2.4: Illustration of a) unweighted and undirected network, b) weighted undirected network and b) directed unweighted network with five nodes and six edges.

Another way to represent the relationship between the nodes and edges is to visualise the network. Figure 2.4a) illustrates the same set of nodes, **V**, and edges, **E**. A node is represented as a circle where its respective number or ID is displayed. The different nodes are linked by an edge, passing from one node to the next for the nodes that have a relationship. This visual representation of a graph can provide insights by showing an overview of the interactions present and, for instance, by investigating the network's topology. The topology describes how the nodes and edges are spread and quantify the overall relationship [32].

As Figure 2.4 illustrates, there are multiple ways to both illustrate and denote the relationship between different nodes. For example, in Figure 2.4b) the edges between each node have a weight attribute associated with them. Adding weight to an edge could have a practical meaning, for instance, describing the capacity of a road between locations a and b or the number of times individuals a and b have interacted with one another in a day [30]. Furthermore, Figure 2.4c describes a directed network, where the edges between nodes are drawn with an arrow representing a directionality of the interaction [33]. This directionality could, for instance, represent how a protein a binds and causes conformational changes of a protein b. Since protein b does not cause changes in protein a, the interaction is directional.

A more abstract way of representing a graph is through an adjacency matrix. For a network with N nodes, the adjacency matrix is a square matrix with a shape of NxN [32]. An adjacency matrix representation, A_{ij} of Figure 2.4b can be seen in Equation 2.5.

$$A_{ij} = \begin{bmatrix} 0 & 2 & 3 & 0 & 0 \\ 2 & 0 & 0 & 5 & 0 \\ 3 & 0 & 0 & 4 & 1 \\ 0 & 5 & 4 & 0 & 3 \\ 0 & 0 & 1 & 3 & 0 \end{bmatrix} \tag{2.5}$$

Each element in the matrix describes whether or not there is an edge between two nodes i and j. On the diagonal $i = j$ and as there are no self-loops in the graph, the entry is 0 along it. As shown in Equation 2.5, the first row would describe node 0 in Figure 2.4b and its relationship between the other nodes in the graph. As it has no connections to itself, the first entry is 0. Meanwhile, it has an edge to node 1 with the weight of two and another edge to node 2 with the weight of three. Since the network depicted in the adjacency matrix is undirected, the matrix is symmetrical along the diagonal marked by zeros. Similarly, the first column shows how node 0 interacts with the other nodes in the graph [33].

Graphs with their respective visualisations are highly versatile. Still, they are used to denote the different components of networks, such as interactions between genes and their proteins, social interactions in high schools, the transmission of sexually transmitted diseases (STD), and other disease transmissions [30].

Graph analysis

Many analyses can be run on graphs to get more information on their characteristics. This subsection will detail a few measures used in this thesis: network degree, diameter, average shortest distance, clustering coefficient, average clustering coefficient, network density, and two different network centralisation measures.

The degree of a node represents how many edges that specific node has connected to other nodes in the network [33]. For instance, would the degree of node 1 in Figure 2.4a be two, while the degree of node 2 would be three. For a weighted network, the measurement weighted degree can also be helpful. The weighted degree also considers the weight of each edge a node has to other nodes in the network. For Figure 2.4b, this means that node 1 would have a degree of seven, while node 2 would have a degree of eight.

The diameter of a network is a measure of the maximum shortest path present in a network. In essence, for the shortest path between all pairs of nodes in a network measured, the network diameter is the largest value between any of the two nodes in the network. In Figure 2.4a, the network diameter would be 2, as there are two edges ($0 \rightarrow 2 \rightarrow 4$) separating node 0 from node 4. Similar to the network diameter, the average shortest path identifies all shortest paths between any two pairs of nodes and takes the average length of all the distances [33]. If a network has a low network diameter in relation to the size of a network, the network is said to display a small-world property. More detail on algorithms and ways to identify both network diameter and average shortest paths in any given network can be found in Barabási and Pósfai's Network science [33]; however, this is beyond the scope of this thesis.

The local clustering coefficient is a measure that is specific to each node in the graph. It measures to which extent a specific node's neighbours are linked to each other. For a node i with degree k_i the local clustering coefficient is calculated as follows [33].

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \tag{2.6}$$

Li represents the number of links found between the k_i nearest neighbours of node i. Therefore, if $C_i = 1$, all the neighbours of node i are all linked to each other, forming a cluster. If $C_i = 0$, none of the neighbours of node i are linked to each other. Consequently, C_i is a measure of the probability that the neighbours of a node i are linked to each other. For $C_i = 0.5$, there is a 50% chance of any two neighbours of node i being linked together. Meanwhile, for a network of N nodes, the global clustering coefficient, also called the average clustering coefficient $\langle C \rangle$, takes the average of C_i of all nodes $i = 1, ..., N$ in a network. Similarly to C_i, the global clustering coefficient denotes the probability that any two neighbours of any node in the network have an edge between them [33].

Network density describes how connected a network is compared to its full connection potential. It can be identified for any network by dividing the number of links found in a network by the maximum number of links possible. The maximum number of links a network can have is one where all nodes are linked pairwise. The network density then describes how dense a network is with its given set of nodes, compared to the potential density [33]. In real-life social networks, the density is generally on the lower spectrum. For example, Auber et al. describes a social network of actresses generated by using IMBD, where they collected a network of 419 actors and 5651 edges representing social contact resulting in a network density of 0.1794 [34].

Network heterogeneity measures the diversity of node degrees in a network. A heterogeneous network is a network where nodes are connected with other dissimilar nodes by having different degrees. Consequently, a homogeneous network is one where nodes with a similar degree are connected. Heterogeneity is often used to measure how robust a network can be to different perturbations, for instance, when introduced to transmissible diseases. The higher heterogeneity, the more robust a network is to random perturbations. Furthermore, a highly heterogeneous network will also be more vulnerable to targeted perturbations. For example, when introducing a disease into an interaction network, if a node with a high degree is chosen, it has a higher probability of transmitting the disease more quickly to the rest of the network than when choosing a low degree node. For a network of size N, with the average degree of the network $\langle k \rangle$ where node i has a degree of k_i, the heterogeneity VAR can be calculated as follows [35]:

$$\text{VAR} = \frac{1}{N} \sum_{i}^{N} (k_i - \langle k \rangle)^2 \tag{2.7}$$

Another way to measure if nodes are connected to similar nodes is to investigate assortativity. It is a measurement of the preference of a network's nodes to attach to other nodes

that are similar to themselves. The most frequent assortativity measurement used on networks is to see to which extent nodes connect to nodes that have a similar degree, meaning that, for instance, high degree nodes connect to other high degree nodes. This can be visualised by plotting the degree of node i as a function of node j for all interacting nodes (i, j) [36].

Multiple network centrality measures within graph theory can help quantify the nodes of heterogeneous networks. Since heterogeneous networks often have a significant difference in which nodes are connected to others, it is essential to have measures of the centrality of nodes so that the impact of specific nodes can be quantified [37]. Take, for instance, an interaction network of an office, where each node represents colleagues and edges between their interactions. Transmission of disease would transpire faster through the network if the human resource (HR) employee, who had frequent encounters with all the office workers, were patient zero instead of the company director, who was isolated in their own office. Therefore, it is a fundamental characteristic to investigate in networks where simulations are to be run. There are multiple different centrality measures, and Rodrigues provides an extensive overview of them [37]. However, for this thesis, the centrality measures described in detail are closeness centrality and betweenness centrality.

Closeness centrality is a local measure of each node's average distance to all other nodes in the network. The closer a given node has to all other nodes in the network, the higher their closeness centrality. For a network with N nodes, the closeness centrality (CC_i) of node i with neighbouring nodes j is [37]:

$$CC_i = \frac{N}{\sum_{j=1, j\neq i}^{N} d_{ij}} \tag{2.8}$$

Where d_{ij} is the shortest path between node i and node j. The closeness centrality is a simple measure that highlights the nodes with the shortest paths to the other nodes in the network and helps identify the most isolated nodes. The drawback of using closeness to measure centrality is that most complex networks have a small network diameter. Thus, it is more likely that several nodes may have the same closeness centrality making it difficult to separate them [37].

Therefore, it is also essential to explore other centrality measures, such as betweenness centrality. Instead of measuring the average distance to other nodes from a given node, betweenness centrality focus on locating "bridges" or nodes in the network where most traffic or load is passed through. To continue the office example, the HR employee may interact with many of the other employees in the office and have a high centrality measure. However, the company director may work with people in both the office and at a corporate level, thus functioning as a bridge between the office workers and the corporate workers. The company director would then have a higher betweenness centrality.

The load of a node, i, is defined as the number of shortest paths that pass through the node. Since multiple shortest paths often exist between any two given nodes, the load is often defined as the fraction of shortest paths connecting each pair of nodes (a, b) for which

a and b could be the nodes $1, \ldots, N$ that have a shortest path that pass through node i [37].

$$BC_i = \sum_{a,b} \frac{\eta(a, i, b)}{\eta(a, b)} \tag{2.9}$$

Equation 2.9 shows the calculation of the betweeness centrality. $\eta(a, i, b)$ denotes the number of shortest paths that are connecting the edge a and b through edge i, whilst $\eta(a, b)$ is the total number of shortest paths that exists between a and b. Network centrality is an important measurement, and both betweenness and closeness centrality can provide insights into the nodes of networks. Whereas closeness centrality would highlight the HR employee, which could be interesting to investigate how the disease may be transmitted in an office setting, betweenness centrality would highlight the company director who could describe how the disease could transmit from the office to the corporate workers. The value of looking into the betweenness centrality of a network is that it may highlight vulnerable nodes that, when compromised, may cause a connected network to be divided into components with no communication between the different components [37].

In nature, whether that be social interactions or regulatory proteins and genes interacting, networks tend to have a community of nodes that interact more with one another. This community is often called a module or cluster in network theory, where the individuals within one cluster tend to interact with other individuals of the same cluster more than nodes outside of it. Modularity is a way to measure the prominence of such clusters or modules in a network. For example, a network with N nodes and L links that is divided into n_c communities, with L_c number of links within the community c and k_c total degree of the nodes in the community c, modularity can be calculated as shown in Equation 2.10 [33].

$$M = \sum_{c=1}^{n_c} \left(\frac{L_c}{L} - (\frac{k_c}{2L})^2 \right) \tag{2.10}$$

The higher the modularity, the more prominent and connected the different nodes are to other nodes within their communities. In addition, this also means that there are sparse connections between nodes of different communities. For more information on calculation of modularity, see Newman [38].

Degree distribution

The degree distribution of a network, p_k denotes the probability that a randomly chosen node has a degree of k. p_k is normally plotted as a function of the degrees k present in the network [33]. The distribution can either be drawn as a Probability density function (PDF) or a Cumulative density function (CDF). In degree distributions, PDF denotes the probability that a random node in the network has a given degree k. Meanwhile, a CDF accumulates the probability that a random node in the network has a value of less than or equal to degree k. There are multiple degree distributions one can expect to encounter

when investigating networks. Figure 2.5 provides an overview of the PDF of the distributions that will be detailed in this subsection. This section will only address degree distributions used in this thesis.

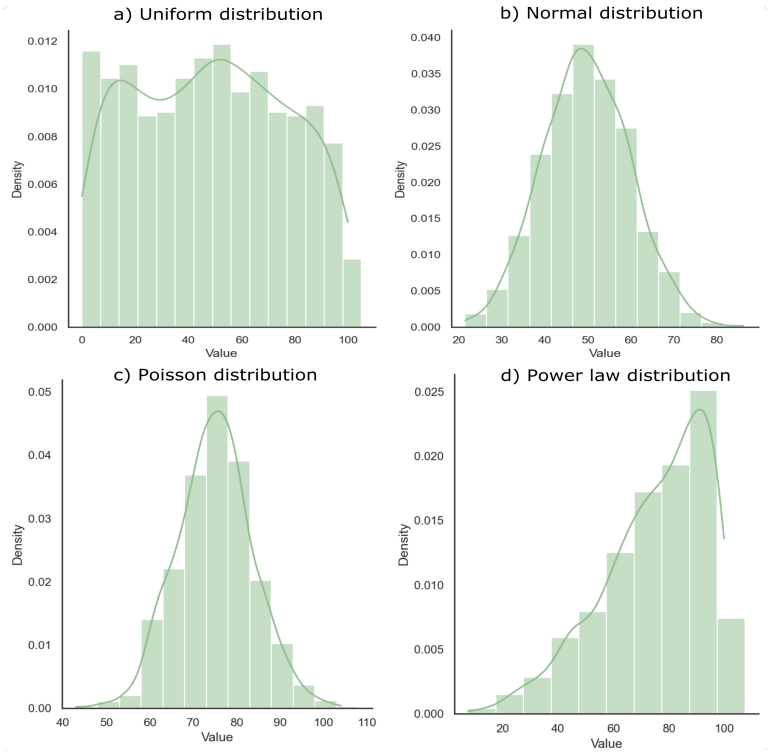


Figure 2.5: Density plot of the result of 1000 generated values between 0-100 for different distributions, where a) Uniform distribution, b) Normal distribution, c) Poisson distribution and c) Power law distribution.

The simplest form of degree distribution is the Uniform distribution. It can be seen in Figure 2.5a, and an example of its usage is in random variable generators. For a variable x in the interval [A,B] the PDF of a uniform function is given in Equation 2.11 [39].

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & \text{if } A \leq x \leq B \\ 0, & \text{elsewhere} \end{cases} \tag{2.11}$$

A normal distribution is a distribution has the symmetric bell-shape that is visible in Figure 2.5b. The parameters mean (μ) and standard deviation (σ) are used to determine the value of x in Equation 2.12 [39].

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{2.12}$$

Power law distribution is a continuous probability distribution often seen for social interactions, especially in scale free networks [40]. The density plot of the Power law can be seen in Figure 2.5d. Equation 2.13 describes the PDF of a Power law where α is a constant whilst k is the order of scaling [33].

$$f(x) = \alpha x^k \tag{2.13}$$

The density of the Poisson distribution is shown in Figure 2.5c, and it is a discrete probability distribution. It is often used to identify the probability of certain events occurring within a given interval of time. The equation for the Poisson distribution can be seen in Equation 2.14 [39].

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{2.14}$$

Random networks and topology

Since being introduced into the field of graph theory by Paul Erdős and Alfréd Rényi in the 1950s and 60s, random graphs have completely revolutionised graph theory; see Albert and Barabási for more detailed information on random networks [41]. Random networks aim to reproduce some characteristics found in real-life empirical networks, such as Power law degree distribution, high clustering, and small-world property by having a low network diameter [41]. In random networks, the number of nodes is often set, and various probabilities are used to determine the nodes and edges formed between them. This section will go into detail the generation, usefulness, and characteristics of the random networks Erdős-Rényi model (ER),Watts-Strogatz model (WS) and Barabási–Albert model (BA).

The ER model was the first random network model to be described. It takes in the parameters N, the number of nodes in the network, and V, the number of edges in the network. The edges are chosen randomly from the $N/(N-1)/2$ possible edges in a network with N nodes using p as the probability for edge formation between each pair of nodes. Thus the ER model starts with N nodes and then uses the probability p to determine whether or not an edge is drawn between any two pairs of nodes. For more details on how the graph is generated, see Albert and Barabási [41]. Properties of ER is the appearance of subgraphs (such as cycles or clusters), however, with an overall lower clustering coefficient than whats seen for empirical interaction networks. The degree distribution for a ER model is a binomial degree distribution that is approximated by Poisson for $N << \langle k \rangle$ [41]. In addition, ER networks tend to have a small network diameter, as long as $Np < 1$.

Furthermore, the WS model proposed in 1998 only uses the parameters N, the number of nodes, and p, the likelihood of rearranging the edges in the network. In order to keep the low network diameter and at the same time have a high clustering coefficient similar to that seen for empirical networks, Watts and Strogatz created a model that was an interpolation between a regular lattice and random graphs. First, all nodes were connected to their k closest neighbours in a ring lattice. Furthermore, each edge has a p likelihood of rewiring the edge to another node. This is similar to how individuals are more social with their

neighbours or the people they are closest to in school or at work. Likewise, people mostly socialise with individuals they are in close contact with but may also have some other acquaintances with other individuals with whom they have a more considerable distance. The rewiring of the edges in the ring lattice helps capture both the close contact interactions and the more rare acquaintance interactions. The degree distribution is similar to and approaching a Poisson distribution, like the ER model. In addition, rewiring also provides a small-world property, where the average path length between the nodes in the system is low and overall high clustering coefficient for the network [41].

The BA network is a scale-free network, meaning that the degree distribution follows a Power law. In contrast to ER, the model is not built by first displaying all nodes and then creating edges between them according to established rules. Instead, the idea for BA networks is that nodes are added through growth and preferential attachment. The World Wide Web network consisting of webpages as nodes and hyperlinks between them as edges is an example of how preferential attachment works. Each new webpage and its hyperlinks have a higher probability of linking to a more well-known webpage (with more hyperlinks) than an unknown one. BA generates networks by growing the network one node at a time, with the likelihood of a given node creating an edge to any other node being proportional to the degree of that given node. The concept of nodes having a bias for interacting with other nodes with a higher degree is called preferential attachment. As a result, the average path length of a BA model is smaller than for the other mentioned random graphs, and the model generates networks with Power law degree distribution [41]. BA models do not, however, have a high clustering coefficient as compared to real-life empirical networks.

Disease modelling using networks

Duan et al. divides complex network models into the following two categories: spreading dynamics in complex networks and numerical simulations of epidemics in complex networks. The first category allows for the simulation of spread on a created network. It uses differential equations, similarly to compartmental models, which determine which nodes move from one state to another (for instance, from susceptible to infected). The use of networks alongside the compartmental model provides more realistic than heterogeneous networks. Meanwhile, numerical simulation in complex networks integrates the Reed-Frost model to generate a network where a link between two nodes denotes the probability of infection being spread between them.

Complex network models also allow for interpretation of the network topology of a given situation and how this can impact epidemic spread. This has allowed several conclusions regarding epidemic spread on the system solely based on its network typology. A node with a higher degree in a social network will, for instance, have a higher probability of being infected compared to a node with a low degree. Furthermore, epidemics will spread faster in a scale-free network compared to small-world networks as the disease is in its early state [30]. Whether or not a network is weighted also plays a role in how the interactions between individuals are modelled and disease is spread. Finally, temporal networks, where interactions are accumulated over a specific time, may also give a new time dimension to disease modelling.

2.2.3 Agent based modelling

Agent-Based Modelling is a way to model more complex systems where each component, or agent, has a set of characteristics or behaviours taken into account. Each entity in a model is called an agent, and they act and behave according to a set of rules. Compared to the previously mentioned methods, it is the most complex, detailed, and computationally demanding way to model disease spread. However, it is valuable as it gives a deeper understanding of how each agent behaves in a particular situation and more accurately represents the heterogeneity in individual attributes and behaviours, which can further be used in disease modelling. Furthermore, it is a bottom-up approach, where each individual's small behavioural changes and attributes are used to predict and investigate the epidemic spread of a whole set of individuals. The recent advancements in increased computational power, data availability, and development of new algorithms have allowed the usage of ABM [30].

Macal and North describes steps and guidelines for creating an ABM in detail. Their steps will be used in the formulation of a model for this thesis and will be described in further detail in section 3.3 model. They describe the three following characteristics to implement into a typical ABM, quoted verbatim from Macal and North [42]:

1. *"A set of agents, their attributes and behaviours"*
2. *"A set of agent relationships and methods of interaction: An underlying topology of connectedness defines how and with whom agents interact."*
3. *"The agents' environment: Agents interact with their environment in addition to other agents"*

It is vital that the model is based on actual human contact patterns since these interactions will influence how diseases are transmitted. Different methods can obtain information on human contact behaviour through social questionnaires, diaries, and wearable sensors. From this, it is possible to observe patterns that can be used to determine the distribution of contacts, as seen in Figure 2.6. Edmunds et al. used an agent-based approach to model interactions between adults in the UK to identify contact patterns with epidemiological significance [43]. After analysing an experimental network collected by the participants, Edmunds et al. [43] found that the number of contacts per individual per day was approximately normally distributed with a mean of 16.8 and a standard deviation of 8.5. This was further used to generate a network [43].

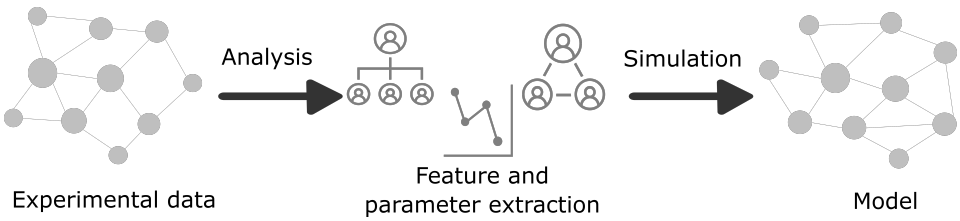


Figure 2.6: Thought process of moving from experimental data to simulation of an Agent-based model through a feature and extraction analysis step.

Another way to model the interactions, if they do not follow any known continuous distribution, is to determine interaction based on each agent's different attributes. Moon and Carley [44] wanted to generate a social network between terrorists and used multiple probabilities to generate interactions between individuals: relative similarity (RS), relative expertise (RE), social distance (SD), and spatial proximity (SP) [44]. Equation 2.15 describes the probability of two terrorists, i and j, interacting based on the four parameters. ω_1, ω_2, ω_3, and ω_4 describe weights of other factors.

$$P_{ij}^{Interaction} = \omega_1 RS_{ij} + \omega_2 RE_{ij} + \omega_3 SD_{ij} + \omega_4 SP_{ij} \tag{2.15}$$

Multiple simulation systems that allow for ABM simulation are also available online. Duan et al. mentions MASON, GeoGraph, EpiSims, BioWar, and FluTe. The mentioned large-scale systems rely on demographic data, and some also utilise Geographic Information System (GIS) techniques to visualise outbreaks in maps.

The fire model, described by Wilensky and Rand [45], details how interactions between agents can be modelled according to both their surroundings and behaviour in a spatial manner [45]. For this model, a spatial network of different agents is required. The model's objective is to simulate how a fire would spread through an area with varying densities of trees distributed heterogeneously. The chance that a particular tree will catch on fire is based on variables such as the type of wood, wind, and how close a tree's branches are together. The agent-based fire model approach can then help predict what would happen if a specific tree catches on fire and how it will spread to surrounding trees based on the characteristics and behaviour associated with the burning tree and its neighbours.

The advantage of using ABM is that they provide complex networks that can take into account mobility and contact patterns. This helps a model formulate a more detailed and realistic network where stochastic processes and probabilities can be used to simulate epidemic spread. The disadvantage of ABM is, however, that the more detailed an ABM is, the computational power that is required. Furthermore, even though the field is rapidly evolving, another setback is that it relays heavily on high-resolution collected data on human behaviour, psychology, and movements. Lastly, some of the models might end up being too complex to generalise and formalise [30].

2.3 SocioPatterns: Face-to-face contact patterns in a primary school

The data set used to describe interactions within primary schools in this thesis was collected by Barrat et al. [1] and made available through the interdisciplinary research collaboration Sociopatterns [46]. Sociopatterns utilise data and emerging technology to study human behaviour and activity. After the swine flu outbreak during the spring of 2009, the main motivation of the study by Barrat et al. [1] was to collect information on how children mix and interact in a primary school environment and use this information on an epidemiological scale. During the 1st and 2th of October 2009, Barrat et al. collected 77,602 contacts between 242 individuals in a French primary school in Lyon. The subjects

in the study consisted of 232 children between the ages of 6 to 12 and 10 teachers. All subjects wore Radio Frequency Identification Device (RFID) badges to detect close proximity face-to-face interaction within 1-1.5 meter distance, lasting longer than 20 seconds.

2.3.1 Collection of data

The data was collected using RFID badges that were placed on the chest of the pupils. RFID devices allow for specific identification of individuals with their own IDs, collection of data, and entering of the mentioned data into computer systems. In the study by Barrat et al. [1], the RFID badges detected other RFID ID's and reported to RFID readers placed in the environment of the primary school, which further connected to a central computer system that saved the interactions. The RF readers were placed to cover the entire school with receivers in all classrooms, the cafeteria, stairways, and playground. Even though Barrat et al. collected the location in which each interaction found place, this data was not made publicly available.

For Barrat et al.'s study, the exchange of radio packets between badges worn by individuals was only possible for individuals facing each other at the close proximity of about 1-1.5 meters [1]. This is mainly due to the human body working as a Radio Frequency (RF) shield for the carrier frequencies used in the study. Thus, the interactions that are detected by Barrat et al. are close proximity and face-to-face, mimicking how respiratory diseases might spread through, for instance, coughing, sneezing, or direct contact [1]. The parameters used in the RFID badges were tuned so that the proximity of two individuals wearing the badge can be assessed with a probability above 99% for 20 seconds. Conversely, the interval for detection of interactions is set to 20 seconds.

2.3.2 Features and metadata

All data on interactions and RFID badges were fully encrypted to ensure no personal information of the participants was shared. The interaction data set published and available through SocioPatterns, contained six columns; source ID, source ID's class target ID, target ID's class, and time interacted. All IDs were anonymous, and a complementary metadata file was also available, containing metadata on the gender associated with each ID. The columns of data available are represented in Table 2.3, and the high-resolution data captures the ID, class, gender, and time each interaction occurs.

The school in which the data was collected consisted of five grades (1^{st} to 5^{th}) with two classes (A and B) per grade. Of the 251 individuals attending the school, 232 children participated (96% coverage) and ten teachers (100% coverage). The size of each class varied from 22 to 26 students, with an average of 24 students per class.

On the 1^{st} of October 2009, the trackers collected data from 8.45am to 5.20pm, whereas on the 2^{nd} of October 2009, data was collected from 8.30am to 5.05pm. The schedule of the primary school is presented in Figure 2.7 including the time school starts, breaks, lunch break, and ends. In addition, the capacity of the playground and cafeteria did not hold the entire school, so the lunch break was broken into two consecutive turns where two or three grades had a break at the same time while the others ate in the cafeteria. This was usually

Table 2.3: Overview of the data available from Barrat et al.'s experiment in Primary schools. An interaction occurs between SourceID and a TargetID.

Data	Description	Example	Type of data
Timestamp	Seconds since midnight on the first test day	31220	Integer
SourceID	ID of interacting individual	1558	Integer
TargetID	ID of interacting individual	1567	Integer
SourceClass	Class of SourceID	3B	String
TargetClass	Class of TargetID	3B	String
SourceGender	Gender of SourceID	M	String
TargetGender	Gender of TargetID	F	String

divided so that grades one to three ate lunch together while grades four and five were in the playground. Afterwards, grades four and five entered the cafeteria, while grades one to three went into the playground. It is also important to keep in mind that some school children eat their lunch at home during the two-hour lunch break. Therefore, there is a drop in the number of individuals present during the lunch break.

Figure 2.7: Schedule of an eight-hour school day in a primary school in Lyon that starts at 08.30 and ends at 16.30. The schedule includes two half an hour breaks at 10.30 and 15.30 as well as a two-hour lunch break at 12:00.

2.3.3 Main findings reported by Barrat et al.

The average time spent between two persons in face-to-face proximity during day one was 3 minutes and 27 seconds, while it was 3 minutes and 56 seconds on day two. Barrat et al. highlights that there is a limited degree of heterogeneity across classes as well as across days. Meanwhile, there is a clear separation between individuals of different classes, and individuals of the same class are more likely to interact with one another. In addition, another separation that is apparent is the one between the lower grades (1-3) and upper grades (4-5), likely due to the grades having a shared lunch break. Figure 2.8 describes the degree of each individual at different time points. The spikes visible are during lunch and breaks that were described in Figure 2.7. In addition, the number of individuals present drops during the two-hour lunch break as individuals go home to eat lunch. Finally, Barrat et al. found that each child had, on average, 26 repeated contacts on the second day with children they interacted with on the first day and 20 new connections. Of 26 repeated interactions, 19 were in the same class as the individual, and seven were outside the class

[1].

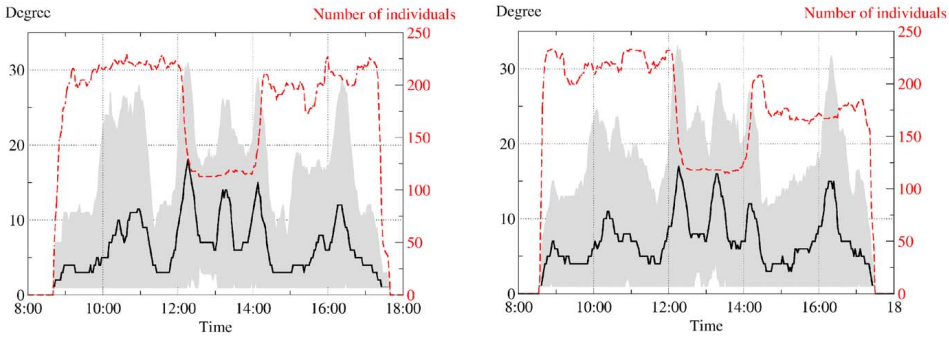


Figure 2.8: Degree of individuals in empiric network for 20 minute intervals for day one (left) and day two (right). The black line shows the mean for each time point whilst the 95% confidence interval is drawn in grey and the number of individuals at each time point is marked in red. From Barrat et al. [1]

Chapter 3

Methods and software

The objective of this thesis is to extract information and trends from the experimental interaction data collected by Barrat et al. [1] and use this to estimate parameters that can be used in an ABM of the same environment. Furthermore, disease transmission should be run on the model. Figure 3.1 describes the overall process of moving from experimental data to running disease transmission on a model.

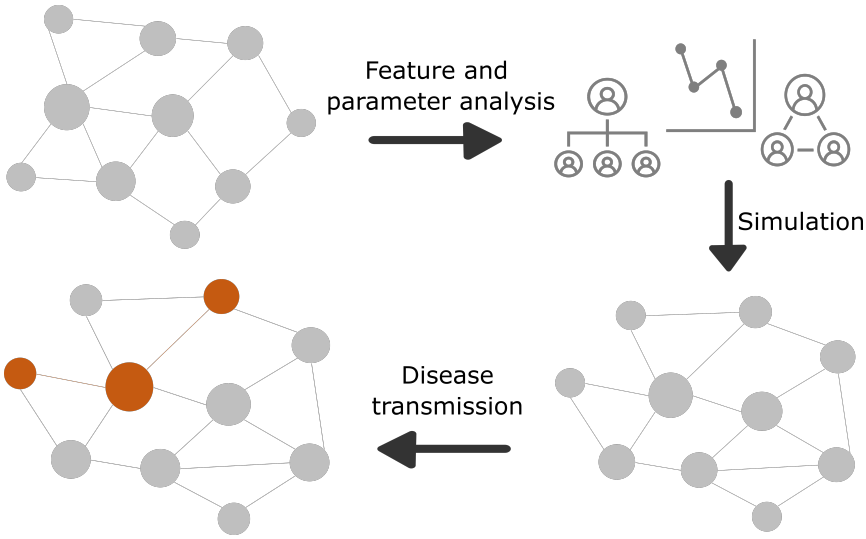


Figure 3.1: The approach of this thesis is to use experimental data to characterise features and metadata and use these to create new interactions that can be used in a disease transmission model. The size of the nodes represents their degree.

Figure 3.1 also describes the four parts this chapter is divided into. Firstly, the software

used to analyse and program the model is described in detail in section 3.1. Secondly, the approach for feature characterisation and analysis of the experimental data is described in section 3.2. Thirdly, said features and characteristics are used to create an ABM as described in section 3.3. Finally, disease transmission and preventative strategies were implemented into the model and are described in section 3.4.

3.1 Software

This section will go into further detail on the different software and programs used throughout this thesis. Some of the data generated for this thesis was, in addition, processed in the high-performance computing cluster Idun at the Norwegian University of Science and Technology, Trondheim, Norway.

The ABM, alongside most of the analysis for both the experimental and simulated networks, was created using the programming language Python, version 3.9.7 [47]. Python was chosen as it both allows for object-oriented programming and will make the remaining model compatible with the NTNU Taskforce model [12]. Python can be downloaded from `www.python.org`. In addition to the Python Standard Library, additional modules were imported and used to analyse and construct the networks. Table A.1, Appendix A, shows an overview of the installed modules.

Furthermore, Cytoscape version 3.9.0 [48] allowed for network analysis and visualisation of both the simulated and experimental networks. It is an open-source software environment originally produced to analyse and integrate biomolecular interaction networks with, for instance, expression data [48]. However, it is also a valuable tool for working interactively with networks and analysing them through installed modules. Cytoscape can be downloaded from `www.cytoscape.org`.

3.2 Analysis of Sociopattern's Primary school network

The initial analysis of the empirical network largely focused on recreating Barrat et al.'s [1] findings, which are summed up in section 2.3. First, the data was pre-processed as described in subsection 3.2.1. Secondly, subsection 3.2.2 describes the topological parameters investigated in a general network analysis. Then, random networks with similar features to the experimental networks were generated as described in subsection 3.2.4. Next, the dynamic of interactions between individuals in the data set was investigated, as shown in subsection 3.2.3. Finally, the degree and interaction distributions were investigated as described in subsection 3.2.6 and subsection 3.2.7. All analysis and pre-processing of data was done using python, and the code for the initial analysis can be accessed through section C.1, Appendix C.

3.2.1 Data preparation

As described in subsection 2.3.2, each interaction was recorded on a 20-second interval. In order to decrease computational time, it was decided that the framework of this thesis

should create a model for daily interactions. This would also make the model compatible with the ABM of the NTNU COVID-19 Taskforce. Since the model should describe the interaction dynamics on a day-to-day basis, the interactions had to be accumulated for each of the two days the study by Barrat et al. was conducted [1]. This was done by looping through all interactions registered for a single day and creating an edge attribute, the weight. The weight variable describes the number of 20-seconds-interactions registered between two individuals on a given day. To illustrate this, if, for instance, individuals 0 and 1 were to interact with each other one time on a given day for three consecutive 20-second intervals, their accumulated time spent together would be one minute, and the weight would be set to three.

Table 3.1: Overview of class as well as new and old IDs associated with the source and target individuals. The weight describes how many 20-seconds intervals the two individuals have been in close contact.

SourceID	Old sourceID	SourceClass	TargetID	Old targetID	TargetClass	Weight
0	1711	1A	1	1752	1A	16
0	1711	1A	41	1697	1B	2
0	1711	1A	179	1866	4B	2
0	1711	1A	221	1471	5B	2

The IDs generated and presented in the Sociopattern's data were randomly assigned to individuals regardless of their grades. They spanned a large scale (1426 to 1922) and were similar between individuals of the same classes, with the exception that some numbers were continually skipped. Therefore, it was decided that new IDs should be generated and that they would increase from 0 to the size of the school. 1st grade, class A, should have the lowest IDs beginning with 0. The IDs should then increase until the highest grade at a given school is reached. This sorting and creation of new IDs was done by sorting a pandas data frame containing all IDs for the school by their class and grade. For more information on the pandas module see section A.1, Appendix A. In addition, the students were sorted by their given IDs within each class. The final IDs were then mapped onto the interaction data set. Table 3.1 shows how the ID from the first individual, 0, in class "1A" had an old ID of 1771, while individual 221 of class "5B" had an old ID of 1471. This configuration allows for more intuitive IDs and the ability to sort by classes more easily.

The accumulated data set with new IDs for both source and target IDs was then used to generate a NetworkX Graph object. For more information on the NetworkX module, see section A.1 in Appendix A. The network was loaded as an undirected weighted network, where the accumulated interactions were used as weights. In addition, node attributes such as gender, grade, and class were added to the graph object, making an empirical network out of the interaction data.

3.2.2 General network analysis

First, an undirected, unweighted network analysis was done using Cytoscape's Analyzer tool and python's Networkx module. The description of each of the topological network parameters investigated in this general analysis is provided in section 2.2.2. Network

heterogeneity was calculated by Cytoscape's Analyzer tool, while the network's diameter, characteristic path length, clustering coefficient, network density, modularity, and network centralisation were calculated using Networkx's inbuilt analysis functions. See section A.1, Appendix A, for more information on the NetworkX module.

3.2.3 Investigating adjacency matrix and creating heatmaps

As described in section 2.2.2, an adjacency matrix is a way to represent all interactions between the nodes in a graph. By investigating the strength of each interaction between the individuals, it is possible to identify the underlying rules that govern how individuals interact. The NetworkX module was used to create an adjacency matrix based on a given graph when weight is a specified input argument.

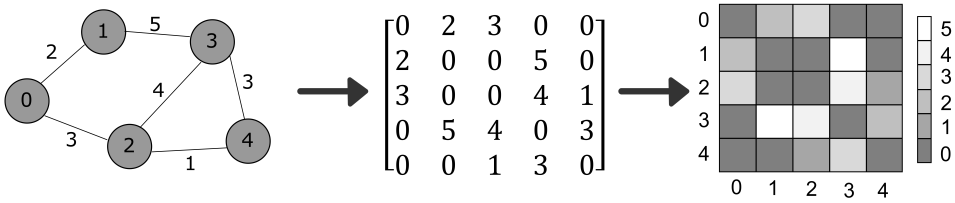


Figure 3.2: Example of the adjacency matrix and heatmap for an undirected weighted graph. The scale for the weighted interactions viewed in the heatmap can be seen on the left, with weights ranging from 0 to 5.

A heatmap was generated to visualize each interaction's strength as presented by the adjacency matrix. A heatmap takes in a value between two interacting nodes and visualises the interaction's strength using a colour scheme. For example, Figure 3.2 shows how an undirected weighted graph can be illustrated through an adjacency matrix in a heatmap. The strength of the colours represents how strong one interaction within the heatmap is. Seaborn, see section A.1, Appendix A, was used to generate all heatmaps presented in this thesis. Heatmaps of primary school interactions could then help define specific interaction patterns between individuals based on their similarities. For instance, they could highlight the difference in how individuals interact with their peers across different grades and classes.

3.2.4 Generating random networks with similar features

Furthermore, random networks, such as the Erdős-Rényi model, Watts-Strogatz model and the Barabási–Albert model were created. The mechanisms behind these models are described in section 2.2.2 and the networks were generated using the NetworkX functions and the parameters that are presented in Table 3.2. The values for the parameters used in each function were chosen by educated trial and error until a network with the same amount of nodes and edges as the experimental network was achieved.

Three random networks were generated using the values from Table 3.2. A general network analysis alongside degree distributions analysis was conducted to compare the random networks to the empirical ones. Comparing the analysis of well-documented random

Table 3.2: Overview of the functions, parameters, and values chosen for random generation of Erdős-Rényi, Watts-Strogatz, and Barabási-Albert models with similar traits to the empirical network on day one.

	Function	Parameters	Values
Erdős-Rényi	erdos_renyi_graph(n, p)	n - #nodes p - % of edge creation	$n = 237$ $p = 0.21$
Watts-Strogatz	watts_strogatz_graph(n, k, p)	n - #nodes k - each node is joined with its k nearest neighbours in a ring topology. p - % of rewiring each edge	$n = 237$ $k = 50$ $p = 0.4$
Barabási-Albert	barabasi_albert_graph(n, m)	n - #nodes m - #edges each node has	$n = 237$ $m = 28$

network models with similar nodes and edges to the empirical network can help identify the patterns and characteristics of the environment investigated.

3.2.5 Investigating assortativity

Then, the assortativity was assessed for the preference of students to connect to other students with a similar degree. In order to investigate the degree-related assortativity of the networks, for all the interactions in the network between node i and j, the degree of i was plotted as a function of j. Furthermore, the weight of the edge between the two nodes denotes the intensity of the point plotted. In addition, a Pearson correlation coefficient was generated to describe the relationship between the nodes linking to other nodes with a similar degree.

3.2.6 Degree distribution analysis

As described in section 2.2.2, the degree distribution is a representation of how connected nodes are in a graph. Essentially, it details the frequency of a node in the network having a given degree. Since the empirical network generated is weighted, the degrees used in the distributions will be weighted. The representation of degree distribution used throughout this thesis is a cumulative weighted degree distribution, $P(X \geq x)$, where P is the probability that a given node has a weighted degree, X, greater than or equal to a specific degree, x. The cumulative distribution is calculated by counting the number of times a given degree for a node has appeared in a graph and then dividing each of the different node's values by the total number of nodes. This gives a cumulative normalised distribution. The cumulative weighted degree distribution will be referenced as the degree distribution for the rest of the thesis. The degree distribution was plotted for each of the two days recorded by Barabási and Pósfai.

Then, to investigate whether the same individuals had a similar degree on day one as on day two, the degree of each node on day one was plotted as a function of day two. The aim was to investigate whether the same individuals interacted a lot on both days or if it was random.

3.2.7 Interaction distribution analysis

The distribution of the weight of each interaction between students can also provide additional information on network topology and be used in characterising the network's features. In this thesis, the term interaction distribution denotes the distribution of the cumulative frequency of each weighted interaction. The interaction distribution is then calculated by creating a list of all the weights present in the network and using this to calculate the frequency of each weight, similarly to how degree distribution is calculated. Then, by plotting a cumulative distribution, similar to the one described in subsection 3.2.6, the distribution of interactions or weights can be generated.

3.2.8 Interaction distribution of single nodes

To further investigate the relationship of the interaction between students, the distribution of all edges with weights for each node was investigated. This was done by plotting the cumulative distribution of each node's weighted edges on a log-log scale and collecting the R^2 for the linear regression of the distribution. The distribution was investigated for all the layers identified in subsection 3.2.9. A high R^2 would indicate a Power law distribution for the edges of the single nodes, while a low R^2 would indicate the distribution was not Power law distributed.

3.2.9 Identifying and extracting three layers of interaction

The heatmap analysis, as described in subsection 3.2.3, revealed four layers of interaction that accumulate to form the interactions seen in a primary school. The reasoning behind the choice of the layers will be reserved for subsection 4.1.3; however, a simple illustration of the three layers can be seen in Figure 3.3. As illustrated, the underlying interactions are off-diagonal. Furthermore, grade interactions consists of both off-diagonal and grade interactions before grade interaction are defined by interactions in all the layers.

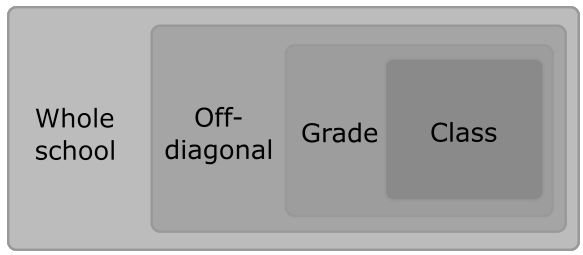


Figure 3.3: The identified layers of interaction in the data collected by Barret et al; whole school, off-diagonal, grade and class interactions.

The exact way that each layer was generated and divided from one another can be seen in section C.1, Appendix C. In short, all interactions were filtered according to the similarity between the two individuals interacting. For instance, a class interaction would occur between two individuals that interacted and were placed in the same class. Similarly, an off-diagonal interaction is an interaction between two individuals who do not belong to the same grade or class. Grade interactions contain all interactions between individuals in the same grade, excluding those in the same class. Thus, analysis of for instance degree and interaction distributions was run on subgraphs of the empirical network for each of the described layers.

An additional division within the layers, is that the off-diagonal layer is influenced by when the different groups have lunch. Therefore, the off-diagonal layer is divided into interactions happening regardless of similarities between individuals as well as interactions happening between individuals in the same lunch groups.

3.2.10 Hourly investigation of the network

Since the time stamps were provided in the empirical data, it allowed for a more detailed and high resolution look into interactions on an hour-by-hour time-scale. Figure 2.7 shows the hour-by-hour schedule of the school. A detailed hour-by-hour plot was generated where the number of interactions each hour was plotted separately for each grades. This was done to see how the grades' activity changed throughout the day.

Even though an hour-by-hour analysis was conducted, it was decided that too much detail may contribute to overfitting the model making it less applicable for integration into the COVID-19 Taskforce model. Therefore, homogeneous hour-by-hour interactions are used when generating daily interactions in the model presented in this thesis.

Furthermore, Barrat et al. [1] described that the two-hour lunch break was divided into two locations; the cafeteria and the playground. This was investigated to quantify off-diagonal interactions further. As a result, the lunch division was added as an additional possible interaction layer. The three following steps were done to examine the effect of the lunch break division:

1. Extract all interactions that occur during the two hour lunch break
2. Investigate heatmap of the interaction
3. Investigate degree distribution of the lunch break

3.3 The primary school model

This section will go into detail on the framework for a model of interaction between pupils in primary schools. The framework was created using the traits discovered by the analysis presented in section 3.2. First, by answering Macal and North's [42] method to produce an ABM, as described in subsection 2.2.3, and then by describing the overall logic of the model. Following Macal and North rules for creating an ABM, the set of agents, relationships, and environment of the model are:

1. **A set of agents**: students with attributes ID, gender, age, grade, class, and lunch group
2. **A set of agent relationships and methods of interaction**: Community driven hierarchic structure where similar individuals have a higher probability of interacting with each other.
3. **The agents environment**: students interact with their peers

To satisfy the requirements by Macal and North the model needs first to contain a set of agents [42]. The model's agents described in this thesis are the pupils that attend a given school. Each agent has different attributes associated with themselves, such as their ID, age, gender, grade, class, and which lunch group they belong to. Furthermore, strict rules should govern the relationships and interactions between the agents based on their attributes. Depending on the similarity between the attributes of the agents or students, an interaction between two agents may occur. As previously mentioned, this depends mainly on the layer in which the interaction between the two individuals is placed in. Furthermore, to address the environment of the individuals, the higher the similarity, the more alike the environment of the two students is. Thus the more likely they are to interact. Therefore, the agents' environment denotes all the peers with which a given agent may interact.

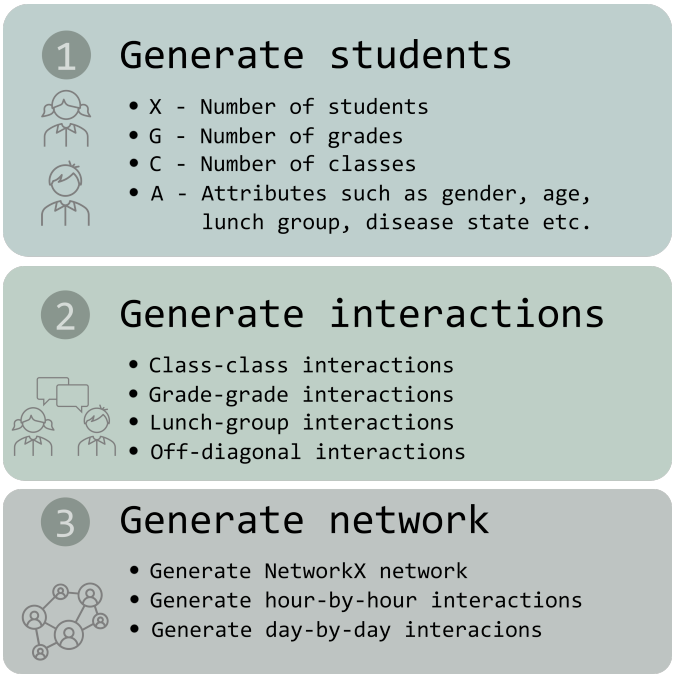


Figure 3.4: Overview of the three steps needed to generate the agent-based model; Generate students, interactions, and a network.

Figure 3.4 shows an overview of the three steps needed to create the model. Firstly, students are generated according to the number of students at a given school, X, before being

divided into grades and classes. The number of grades and classes at a given school are denoted by G and C, respectively. For the primary school in Lyon that was investigated by Barrat et al. [1], $G = 5$ and $C = 2$, as there are two classes, A and B, per five grades. Furthermore, for each student generated, an additional set of attributes, A, should also be associated with the student. The attributes include static and demographic data such as age and gender in addition to data to keep track of students' behaviour, as it may affect how they interact with their peers. Such behavioural attributes are dynamic and include information on when the students have lunch and their disease state. See section subsection 3.4.1 for more information on the different disease states. With this in mind, each agent is distinct and has specific characteristics involved in deciding the agent's behaviour and interactions.

The next step in Figure 3.4 is generating interactions. As described in subsection 3.2.9, different rules govern how the pupils at a school interact with each other, depending on the similarity between each pair of students. Each layer represents a different interaction mechanisms and must be modelled differently to accurately replicate the structures identified in the experimental data. Because of this, the types of interactions are divided into four different layers, as shown in Figure 3.4. The parameters used to define when interactions occurs, will be further described in subsection 3.3.2.

Lastly, the different students' interactions with each other need to be accumulated into a network. Each student should be added to the network as a distinct node with their respective attributes. By looping through each possible pair of students, the network should add hourly weighted links between the students according to the similarity between the students and following a given set of interaction rules. Furthermore, the hour-by-hour interactions should be compiled into a day-by-day network with weighted links denoting the number of times two individuals have interacted. The precise mechanism of creating the different weights will be described in further detail in subsection 3.3.2.

3.3.1 The build of the model

The model comprises three classes; Person, Interaction, and Network. In addition, an Analysis class is also incorporated to run analysis and compare the simulated to the empirical network. Finally, an Age_group file was also added to keep track of the age group of the Person objects. This file was not used specifically in this framework, but was implemented to ensure compatibility with the NTNU COVID-19 Taskforce model. The following section will detail the different classes and how they interact with each other. An overview of the classes that form the interaction model can be seen in Table 3.3.

The Person class stores all information of a single student in a Person object and contains the attributes associated with that person, such as ID, age, grade, and class. A detailed description of all attributes associated with each object can be viewed in Table A.2, in Appendix A. These attributes are used to model how that specific individual will interact with their peers in a primary school environment. In addition, each Person objects are initialised and divided into grades, classes, and age depending on how many grades and classes are at a particular school. Since the data collected by Barrat et al. [1] only contained ten teachers that were assigned one class each, it was decided that the framework presented

in this thesis should only model interactions between students. The Person class also provides methods for getting and setting said attributes and other formatting functions. The full documentation of the class can be seen in the GitHub of the model, described in section C.2, Appendix C.

Table 3.3: Description of all object-oriented classes involved in generating the primary school model; Person, Age-group, Interaction, Network, and analysis.

Class	Description
Person	Represents the agents (individuals) of the model and could in theory be either teachers or students. This model only contains students. Contains the attributes such as ID, age, grade, class, etc.
Age_group	Enum that keeps track of the age group of the different ages. Not specifically used in this model, however it is implemented to ensure compatibility with the NTNU COVID-19 Taskforce model
Interaction	Keeps track of interaction between two Person objects. Stores the two interacting Person objects as well as the attribute count, which keeps track of how many times the two individuals have interacted
Network	Generates a nx.Graph object from the Interaction objects. Also stores attributes such as students, which is a list of all Person object attending a certain school
Analysis	Contains multiple functions that can be used to run analysis on Network objects. Functions includes generating heatmaps, degree distribution, interaction distribution and other measurements that can be helpful when comparing the model to the experimental data

Of most importance are the biases generated and associated with each Person object. The bias quantifies how likely a person is to interact with other individuals; in other words, their bias for interacting with any given person. Once each individual is generated, their bias will be set, and each individual will have a unique bias for interacting with any other individuals. Table 3.4 shows an overview of the three biases included in this model: base (b_i^b), grade (b_i^g), and class (b_i^c) bias. Each Person object will have their own b_i^b, b_i^g, and b_i^c generated upon initialisation. Similar to the layers described in subsection 3.2.9, the base bias describes the bias an individual has for interacting with individuals that are off-diagonal contacts, meaning they are not in the same grade nor class. Likewise, grade bias is an individual's bias for interacting with individuals in their grade, while class bias describes their bias for interacting with peers in the same class. The bias, alongside the P_{ij}, which will be described in the next section, is what helps decide which interaction occurs. The exact calculation of the different parameters of the biases will be provided in subsection 3.3.2.

Table 3.4: Overview over base, grade and class bias with their respective descriptions.

Bias	Description
Base (b_i^b)	Base bias for all individuals to interact with each other
Grade (b_i^b)	Additional bias for interaction between individuals of the same class
Class (b_i^b)	Additional bias for interaction between individuals of the same class

The Person class also includes the functions for generating the P_{ij}-value, which utilises the bias (b_i and b_j) and similarity (p_{ij}) between two individuals, i and j, to decide how likely they are to interact with each other. The P_{ij}-value is higher for individuals with more in common, for instance, those in the same grade or class. Figure 3.5 shows how the P_{ij}-value is calculated for three different interactions between students i and j by the addition of different p_{ij}. Similarly to the biases, the p_{ij} are specific for each layer, divided into off-diagonal (p^b_{ij}), off-diagonal lunch (p^l_{ij}), grade (p^g_{ij}) and class (p^c_{ij}). The same bias is assumed for creating both lunch and off-diagonal interactions. The more two individuals have in common, the more P_{ij} increases due to more p_{ij} multiplied by their respective biases being added.

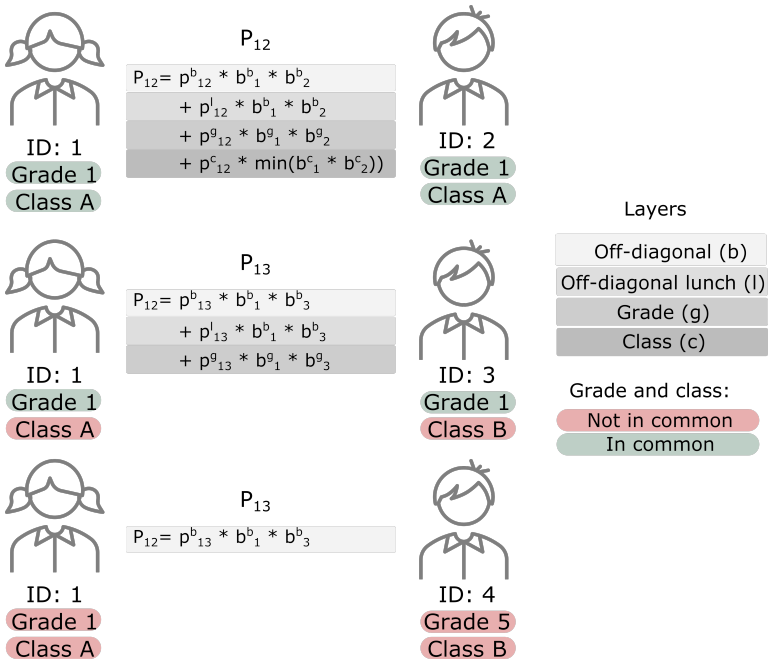


Figure 3.5: P_ij-value calculated for three different interactions; individuals of the same class, individuals of the same grade and individuals that only share off-diagonal interactions.

The first interaction, between Person 1 and Person 2 in Figure 3.5, is between two individuals of the same class. Therefore, the P_{12}-value includes the interaction types off-diagonal, off-diagonal lunch, grade, and class. The P_{ij}-value increases the more interaction types two individuals share. Similarly, P_{14}-value for the third interaction between Person 1 and Person 4, where the students are in different grades, classes, and lunch groups, will most likely be lower. For each characteristic two individuals have in common, the P_{ij}-value increases. The reasoning behind, and functions for the parameters seen in Figure 3.5, will be described in further detail in subsection 3.3.2.

Furthermore, the Interaction class stores interactions between two Person objects. It also keeps track of how many times (i.e., the weight) two individuals have interacted and has

accessible functions to extract the two interacting people. After each interaction is generated, weight is added hourly, and daily edges between individuals are added to the Network class.

The Network class allows for the collection of all Interaction objects to generate and add nodes and edges to an nx.Graph object. Each node is a Person object, and a weighted edge between two nodes indicates how much those two Person objects have interacted with each other. Upon initialisation, the Network class utilises calls to functions within the Person and Interaction classes. See Figure 3.6 for a complete run of a Network object where two consecutive days are generated. When a Network object is initialised, it calls for generating a list of Person objects in generate_students, where the P_{ij}-value is generated between all pairs of individuals at the school. Based on the P_{ij}-values, a Poisson distribution is used to generate the hourly weight of each interaction. If the weight is higher than a set threshold, the weight is added to an edge between the two Person objects. Then, accumulating all hourly interactions of a day, a final daily network is produced. More detail on the exact functions for the bias and P_{ij}-values as well as calculation of weight is given in subsection 3.3.2 in addition to documentation of the different threshold being presented in the GitHub described in section C.2, Appendix C.

In Figure 3.6, a Network object is first created for a school containing 236 students divided over five grades with two classes per grade. The first part of the initialisation of a Network object is to create students with their respective biases. Furthermore, P_{ij} are generated between all students in the generate_a_p_vector(). The interactions are created based on the P_{ij} values and are stored in Interaction objects, before the interactions are added to the network. Similarly, when running multiple days of interaction on the same network, the same biases and daily generated P_{ij} values are used to simulate interactions.

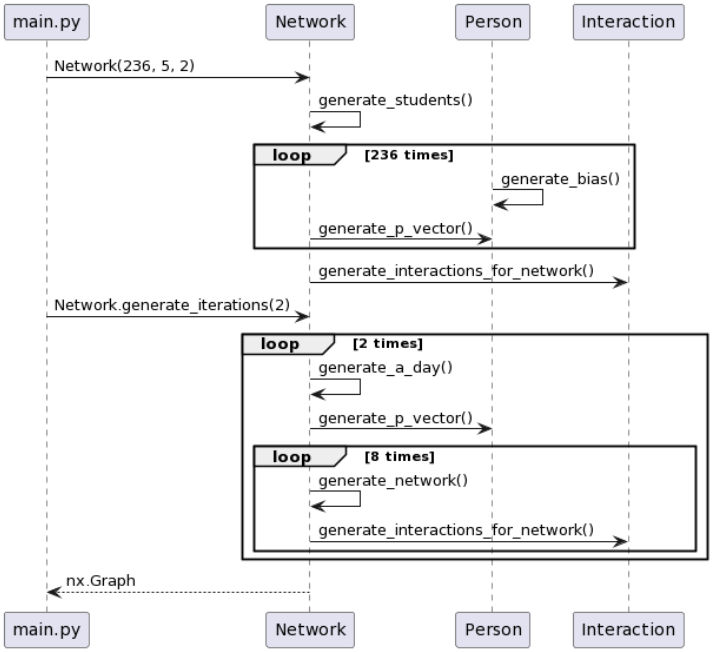


Figure 3.6: Simplified sequence diagram of an initialisation of a Network object in addition to the call of the Network's function generate_iterations(2). Dashed lines represents returned values, whilst boxes indicates loops. The function generate_p_vector() returns P_{ij} between all students, while generate_bias() creates the bias associated with each Person object

3.3.2 Model parameterisation

The aim of the model was for it to be general enough to be incorporated into the NTNU COVID-19 Taskforce model [12], for multiple schools of different sizes. Therefore, the approach for extracting model parameters was heavily influenced by a need for reproducible functions, for instance, to model interactions for different sized schools. The approach used to estimate all biases and P_{ij} values can be generalised by the following three iterative steps:

1. Extract bias from degree distribution
2. Investigate degree distribution
3. Investigate interaction distribution
4. Investigate network properties
5. Adjust parameters until satisfactory
6. Satisfactory parameters

Each iteration was conducted to fit degree and interaction distributions for all the identified interaction layers of the model. Since the outermost layer, the off-diagonal, is also included in grade and class interactions, a change in the parameters for the off-diagonal would also yield changes in grade and class interactions. Therefore, it was decided that the estimation of parameters should be done layer by layer, starting with the outermost layer. By following the steps and investigating one layer at a time, one degree of freedom was addressed at a time. If the degree distribution did not fit with the given parameters and functions, new values for the parameters were chosen. Meanwhile, if the degree and interaction distribution fit nicely for a set of parameters, it indicated that the selected bias and P_{ij} value captured the variation within the students' behaviour and interactions.

Table 3.5: Function and parameters estimated for the interactions between nodes i and j. The biases for node i are base bias (b_i^b), grade bias (b_i^g) and class bias (b_i^c). Likewise, biases for j is calculated. The bias is then used to generate the P_{ij} value between the two individuals i and j. Finally weight is calculated using the value P_{ij}.

	Function	**Parameter**
Base bias (b_i^b)	$\alpha \log_{10}(\frac{1}{x})$	$\alpha = 20$
Grade bias (b_i^g)	$\alpha \log_{10}(\frac{1}{x})$	$\alpha = 17$
Class bias (b_i^c)	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	$\mu = 100, \sigma = 5$
Base (p_{ij}^b)	$\alpha x^\beta * b_i^b * b_j^b$	$\alpha = 2, \beta = 0.04$
Lunch (p_{ij}^l)	$\alpha x^\beta * b_i^b * b_j^b$	$\alpha = 0.01, \beta = 0.01$
Grade (p_i^g)	$\alpha x^\beta * b_i^g * b_j^g$	$\alpha = 2, \beta = 0.04$
Class (p_i^c)	$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}\ min(b_i^c, b_j^c)$	$k = 2, \theta = 0.04$
P_{ij}	$p_{ij}^b + p_{ij}^l + p_{ij}^g + p_{ij}^c$	—
Weight	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	$\lambda = 0.01 * P_{ij}$

The parameters and functions presented in Table 3.5 were then decided for the primary school interaction model. The x for each function was generated by random.random(), which draws values from 0 to 1 using a uniform distribution. As seen from the table, the interaction patterns seen in Primary schools can be modelled using a combination of Power law, Gamma, Normal, Log-normal, and Poisson distributions. The exact choice for these distributions will be described in further detail in section 4.1 and section 4.2.

Each individual i has a bias for interacting with all other individuals (Base bias (b_i^b)), individuals of the same grade (Grade bias (b_i^g)) and individuals of the same class (Class bias (b_i^c)). The likelihood, P_{ij}, of two individuals i and j interacting is calculated by multiplying their appropriate biases with a p_{ij} value. The biases and p_{ij} values chosen are a result of how similar two individuals are, and multiplications of them are added to the final P_{ij} value. Section 4.1, will describe the analysis of the empirical network, and explain the premises for the parameters.

3.4 Simulating disease spread

As described in Figure 3.1 in chapter 3, the model should then be used to simulate how a disease will spread across a primary school network. Following the interaction network generated by the Primary school model, the aim is first to see how disease spreads unhinged in the environment and then test how different measures and policies can help mitigate disease transmission. First, subsection 3.4.1 will describe in detail how the different compartments, or disease states, of the model are defined. Secondly, subsection 3.4.2 will describe the parameters that govern how individuals move between the different parameters. Combined, the states and disease parameters will help create a view of how disease can be transmitted in a primary school environment. Then, disease transmission is run on the experimental data in subsection 3.4.4, which can be used to compare and validate the model according to the experimental data. Finally, policies and measures are implemented into the model, as described in subsection 3.4.5 to see how they affect the overall disease transmission. The code for the disease transmission is described and can be accessed on GitHub as detailed in Appendix C, section C.2.

3.4.1 Disease states

The disease transmission model of this thesis uses an extension of Susceptible, Exposed, Infected, Recovered (SEIR), which is a version of the SIR models described in subsection 2.2.1. The extended SEIR of this thesis also includes the disease states infected asymptomatic (I_a), infected pre-symptomatic (I_p), and infected symptomatic (I_s). Figure 3.7 shows the different states and how individuals move between them. As the model only focuses on Primary schools and their role in disease transmission, a closed system is modelled, meaning only one person is set as patient zero, and no outside contamination is introduced into the model.

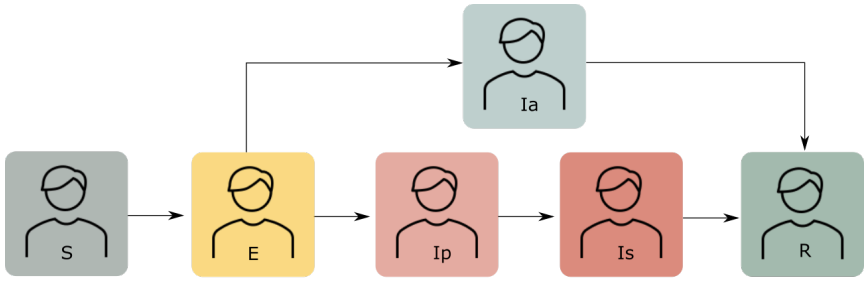


Figure 3.7: Overview of the disease states and movement between them. Where susceptible (S), exposed (E), infected asymptomatic (Ia), infected presymptomatic (Ip), infected symptomatic (Is) and recovered (R) are the disease states.

In the default disease transmission model, all Person objects are initiated in the susceptible state, where they have not yet contracted the disease and are susceptible to getting it. If a susceptible individual comes into contact with an infected (either asymptomatic, presymptomatic, or symptomatic) individual, they have a set probability of becoming exposed. Individuals in the exposed state have an incubation time before moving on to either

an asymptomatic or a presymptomatic disease course. However, this does not necessarily mean that every person an infected individual interacts with gets exposed. Section 3.4.2 will go into more detail about the parameters that determine whether or not an infected individual exposes a susceptible individual. The infectious states asymptomatic, presymptomatic and symptomatic are all capable of changing the states of the susceptible people they have interacted with to exposed. Table 3.6 provides a detailed overview of the duration of the different disease states in addition to their relative likelihood of infecting other individuals relative to symptomatic transmission. These values are gathered from Ferretti et al. [49], and are the same values the FHI uses in their models.

Table 3.6: Overview of the disease states exposed, presymptomatic, symptomatic, and asymptomatic with duration and infection relative to symptomatic [49].

State	Duration (days)	Infection (relative to symptomatic)
Exposed	3	—
Presymptomatic	3	1.3
Symptomatic	4	1.0
Asymptomatic	6	0.1

3.4.2 Disease transmission parameterisation

The modelling of disease transmission used in this thesis is similar to Wilensky and Rand's [45] fire model that was described in subsection 2.2.3. Similarly to how neighbours of a burning tree can catch on fire based on the density of their leaves or how close they are to each other, disease may be transmitted from infected individuals to their neighbours based on how many times they have interacted and the infectiousness of the sick individual. This subsection will detail how a threshold was set to regulate transmission and the mechanisms of transitioning from one disease state to another.

Algorithm 1: Disease transmission on a network

$P_0 \leftarrow$ generate_patient_zero()
for *neighbour* N *of* P_0 **do**
  $p_{inf} \leftarrow 1 - (1 - p_0)^{w_{0,i} * I_{P_0}}$
  $r \leftarrow random.random()$
  if $r < p_{inf}$ **then**
    $N.state \leftarrow$Disease_state.Exposed
  end
end
Update day and state for all students

The simplified algorithm presented in Algorithm 1 helps illustrate how transmission of disease spreads from an infected patient zero to susceptible individuals. The scaling variable, p_0, shown in Algorithm 1, helps set the overall transmission of the disease. It was set to 0.001 as running simulations with this value gave an overall R_0 of 2-3, making it pos-

sible to test the effect of different disease mitigation strategies. Meanwhile, $w_{0,i}$ denotes the weight between patient zero, P_0, and their neighbour, i, while I_{P_0} is the infection rate of P_0. Both the weight of interactions and the infectiousness of the infected individuals determine how high the possibility of infection, p_{inf} is. Then, by comparing the p_{inf} to a random variable r between 0 and 1, the neighbour is either moved into an exposed state or kept in the susceptible state.

Thus, the multiplication of weight with the infectiousness associated with the disease state of the infected individual helps sway the p_{inf}. Algorithm 1 only describes how transmission is modelled from patient zero to their neighbours on the first day; however, the same mechanisms are used to simulate transmission from any infected individual to their neighbours. The code for disease transmission can be found in section C.2, Appendix C, where the function infection_spread() is responsible for checking the neighbours of infected individuals and determining if the disease spreads. Meanwhile, run_transmission() models transmission for the number of days specified as an argument. A new network is generated and used to predict how the disease is spread each day.

3.4.3 Calculating r_0 and generating epidemic plots for analysis

In order to assess disease transmission on both the model and the empirical networks, both epidemic plots and estimated R_0 were calculated. The basic reproduction number R_0, is, as described in subsection 2.2.1, a measurement of disease transmission across a susceptible population. Therefore, an estimated value for R_0 is introduced in this thesis to describe transmission on a smaller sample of a population. Similarly to the basic reproductive number, r_0 describes how many individuals an infected individual may infect. In this thesis, two different measures of r_0 are used, with their own notation so as not to be confused with R_0.

The first measurement is the r_{00}, which describes how many individuals patient zero infects before recovering. It is a useful measurement, as it describes how disease transmits in the early stages of the transmission without being affected by the population becoming immune. By running multiple iterations of disease transmission with the same set of parameters, the number of individuals infected by patient zero can be saved for each iteration and be used to calculate an expected value, R_{00}. Figure 3.8 shows how the r_{00} is calculated from the number of people infected by patient zero. subsection A.3.3 and subsection A.3.1 in Appendix A shows the approach and calculation of R_{00} for different disease mitigation strategies.

The second measurement is r_{0N}, and it describes how many individuals each infected person has infected before recovering. In contrast to r_{00}, it takes in all the recovered individuals at the end of the simulation and returns a list of how many each one infected before they recovered. Therefore, it yields more values and combined with multiple iterations, an expected R_{0N} can be calculated. Similarly, Figure 3.8 describes how the r_{0N} captures each individual each infected person has transmitted the disease to and stores the values in a list. Appendix A.3.2 in Appendix A shows the approach for the calculating R_{0N}.

Epidemic plots are also used to display how disease transmission develops over time. In this thesis, they are plotted by the number of recovered as a function of days since the

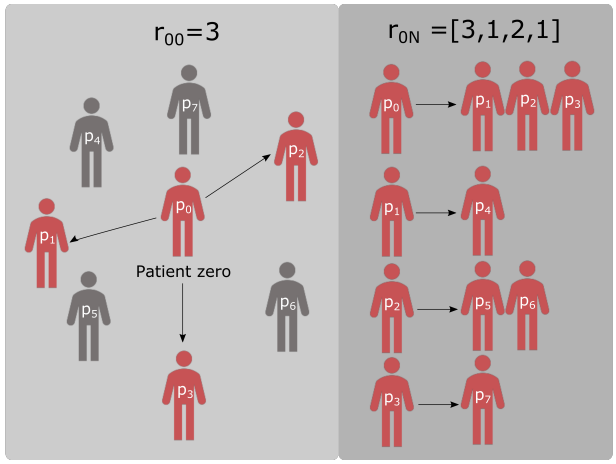


Figure 3.8: Description of r_{00} and r_{0N} in a network with seven individuals. For r_{00}, the number of people patient zero, (p_0), infects is r_{00}. Whilst for r_{0N}, a list of how many individuals each infected person has infected before recovery is given.

disease was introduced. When individuals recover, they become immune to the disease; thus, the number of recovered individuals will accumulate over time. When either all individuals in the model are infected or the disease has stopped being transmitted ($R_0 = 0$), the curve will flatten out.

3.4.4 Disease transmission on the empirical network

Another measurement that can help validate the primary school model described by this thesis in section 3.3 is to compare the transmission of disease on the model with the transmission on the empirical network. The same rules and parameters govern the transmission on the empirical network as for the model.

Since Barrat et al. [1] provided interaction data for two days, transmission of the disease was first run separately on the two days. Then, the transmission was run on a network that alternated between the two networks each day, called the switch. The reasoning behind this was to investigate whether the model accurately could predict disease transmission and see how the changing of the two empirical networks every other day affected transmission. By alternating transmission on the two empirical networks, one can also get a more realistic view of how interactions may change from one day to the next and how it impacts disease transmission. In addition, transmission of the disease was run on both a static and a dynamic version of the simulated model. For the static model, the same interaction network was used for each day of disease transmission. This was done to see if the static model showed similar disease transmission as the empiric networks. Furthermore, the dynamic model changed the interaction network daily, making it more realistic to run disease transmission on it for multiple consecutive days. The code for transmission on empirical and model networks can be viewed is provided in the file *empiric_to_person_object.py* in the GitHub described in section C.2, Appendix C.

3.4.5 Incorporation of policies and measures

This section will detail how different policies and measures were incorporated into the disease transmission model. The policies addressed by this thesis are the effect of the different levels of the traffic light model and the incorporation of weekly testing.

Traffic light

As described in subsection 2.1.2, the Norwegian government incorporated a traffic light model in primary schools to try and mitigate disease transmission. The three levels of the traffic light model, green, yellow, and red, were described in Figure 2.2, and a goal of this thesis was to investigate the effect that implementing them would have on the simulated model. Each step of the traffic light model, increasing from green to red, includes implementing its previous actions. Therefore, all measures initiated for the yellow and green levels are also implemented at the red level. Since the detail level of the model is restricted to interactions between people with attributes such as grade, class, age, and sex, only some of the measures could be implemented. This section will describe how each level was implemented, their effect on interactions within the model, and detail which measures were not applicable. The exact implementation of the different traffic light levels can be seen in the GitHub described in section C.2, Appendix C.

For the green level, three control measures were described by UDIR and are shown in Figure 2.2. The first one was that "Individuals with symptoms should stay at home", meaning that individuals with respiratory symptoms should be isolated from the rest of the pupils. This could easily be implemented by isolating symptomatic individuals from the model, discarding the current interactions they have with other individuals, and preventing new interactions with them from being formed. Furthermore, the second step of maintaining "Good hygiene and increase cleaning" includes both the students, with increased handwashing, and the environment by disinfecting. This could be implemented by decreasing the scaling variable p_0, causing the overall transmission to lower due to the cleaner environment. Finally, the third level in the green level was to "avoid physical contact between individuals". Since the model only measures face-to-face interactions that are of 1-1.5 meter distance, it is not possible to directly implement this control measure. Therefore, the following two implementations of the green level are plausible to implement for the scope of this thesis:

1. Isolate symptomatic individuals
2. Decrease transmission

Furthermore, the yellow level includes the implementation of an additional five policies. The fourth policy is that each "class is divided into one cohort", where interactions outside the cohort are limited. The model implemented this by adding a new attribute to the Person object that stores which cohort the object belongs in. Interactions between Person objects that are not in the same cohort are limited in the model by decreasing the duration and likelihood of those interactions occurring. More specifically, this is implemented into the model by having a higher threshold for interactions occurring outside cohorts within the generate_interactions_for_network() function in the Network class. Furthermore, as "only

two cohorts can interact outdoors", the cohorts were paired according to grade to allow for an average level of interactions between the two cohorts. The pairing of classes with similar grades was assumed, as those are the cohorts most likely to be interacting.

The sixth policy within the yellow level was that there should be "at least one-meter distance between employees". Since the model does not include employees or teachers, this policy was not implemented. Furthermore, the seventh policy to "avoid large gatherings" has already been addressed as interactions between individuals that are not in the same cohorts nor paired cohorts are decreased. In addition, the policy of having "SFO, breaks, and teaching activities divided into cohorts" has already been addressed by the implementation of paired cohorts. With this in mind, the yellow level incorporates the following policies:

4. Each class is divided into a cohort and interactions outside cohorts are decreased
5. Cohorts were paired according to class and grade, and interactions within paired cohorts were lower than for cohorts but higher than for the rest of the pupils in the school

The red level has increased strictness, and in addition to implementing the measures introduced by the yellow level, it also implements an additional three policies. The first one is to further "divide classes into smaller cohorts, halving class size", which is easily implemented into the model by assigning individuals in a class to two different cohorts and storing their respective cohorts as an attribute on the Person objects. Furthermore, implementing a "permanent teacher if possible" is discarded as this model does not include teachers. Similarly, to "consider alternating attendance times for students" is not possible to implement into the model as the disease transmission is predicted on a daily basis. Finally, the policy to "strive to keep distance between students of different cohorts" is implemented by further decreasing the duration and likelihood of interactions occurring. This allows for the implementation of the two following policies for the red level:

6. One class is divided into two cohorts
7. Further decreased interaction between individuals of different cohorts

Weekly testing

Weekly testing was implemented as it was one of the policies the Norwegian government used to mitigate disease transmission during the later stages of the pandemic (i.e., when rapid-antigen tests became readily available). Therefore, it was also decided to implement the possibility of weekly and biweekly testing in the model to allow for the isolation of individuals who were either presymptomatic or asymptomatic. Since the sensitivity of rapid-antigen tests differs depending on the operator, as described in Table 2.2, the likelihood of each infected individual in state presymptomatic and asymptomatic testing positive was set to 60%. This is due to self-trained individuals having a 57.5% sensitivity as described by Peto et al. [29]. Implementing testing into the disease transmission functionality was done by having a function weekly_testing() run every N number of days. The function was run every seven days for weekly testing, while it was run every 14 days for biweekly testing. Each individual in the presymptomatic or asymptomatic state would

then have a 60% chance of testing positive and be isolated from the interaction network. The individuals that tested positive then updated their is_tested attribute, associated with each Person object, to True. Individuals with is_tested set to True are not included in the interaction network until they have recovered. However, it is important to note that weekly testing in this model does not register the individuals in the exposed state, as it is assumed that the viral load is not high enough to be detected.

3.4.6 Investigation the effect of different initial disease states on transmission

Since the initial disease state could transition from exposed to a presymptomatic or asymptomatic disease course, it was decided to further investigate the effect the two initial disease states has on R_{00}. This was done by running 100 iterations where patient zero was set as either asymptomatic or presymptomatic and recording the frequency of r_{00}. In practice, the code easily allows for setting the different disease states for patient zero, by changing an input argument, sympt, in the run_transmission() model that denoted whether or not patient zero should have an asymptomatic or a presymptomatic course of the disease. When sympt = True, patient zero is symptomatic; otherwise, they are asymptomatic. This simulation with choosing different initial states of patient zero was only done for this specific analysis and is not relevant for the overall disease transmission run on the model or for the different disease mitigation policies.

Chapter 4

Results and Analysis

Similarly to chapter 3, this chapter will be divided into three sections regarding analysis of the experimental data in section 4.1, the creation and analysis of the interaction model in section 4.2 and finally the results of running disease transmission on the model in section 4.3. An overview is provided in Figure 4.1.

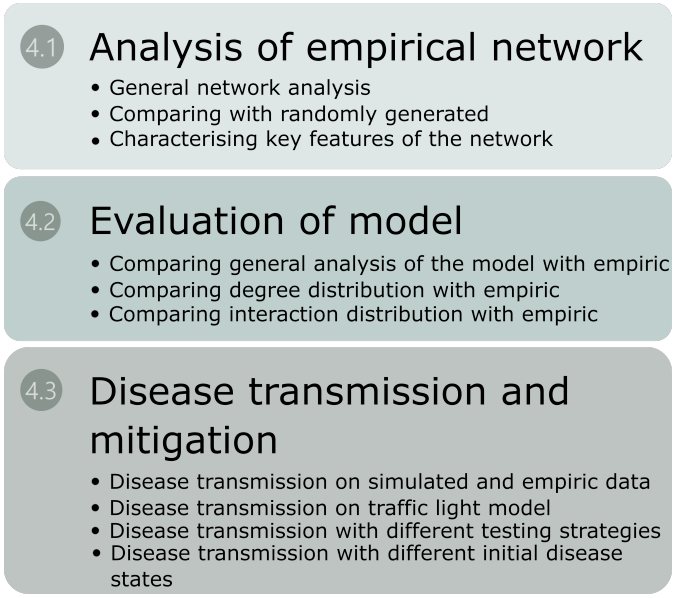


Figure 4.1: Overview of the results and analysis chapter divided into Analysis of empirical network, Evaluation of model, and Disease transmission and mitigation.

4.1 Analysis of the primary school interaction data set

This section will detail an extensive analysis of Barrat et al.'s interaction network [1]. First, subsection 4.1.1 will describe a general analysis of the network on both days, while subsection 4.1.2 will compare how similar the networks are to three chosen random network models. Then, a closer look at the heatmap for the network on both days subsection 4.1.3 will reveal the layers mentioned in subsection 3.2.9. Network assortativity is then explored in subsection 4.1.4 before characterisation of degree and interaction distribution will be described in subsection 4.1.5 and subsection 4.1.6, respectively.

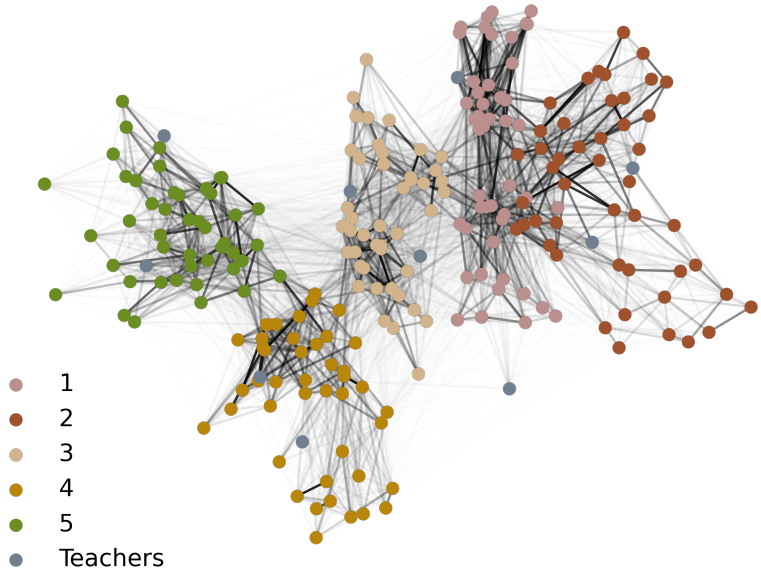


Figure 4.2: Spring-layout of the network on day one with node coloured by their grouping; either teacher or the grades (1-5) they attend. The weight of interactions are proportional to the strength of the edges drawn.

Figure 4.2 shows the overall network on day one. It is apparent that individuals interact most with other individuals in the same grade, and a clear distinction can be seen between the five grades. Meanwhile, the teachers, coloured in grey, seem to be associated with one class, having two teachers assigned per grade. In addition, it seems that grades one to three interact more with each other than with grades five and four. This pattern will be examined more closely in subsection 4.1.3.

4.1.1 General network analysis

After the interaction data was accumulated into a network with weight denoting how many 20-second intervals two individuals have interacted, as described in subsection 3.2.1, an analysis of a handful of general network features was performed. As described in subsection 3.2.2, different network features were measured, and the results for the two days

recorded are presented in Table 4.1.

Table 4.1: Network analysis of topological parameters of a network conducted for day one and day two using Networkx and Cytoscape's Analyzer.

Parameters	**Day one**	**Day two**
#nodes	236	238
#edges	5901	5541
Average degree	513.8	547.5
Network diameter	3	3
Average shortest path length	2.2	2.4
Clustering coefficient	0.502	0.560
Weighted clustering coefficient	0.0079	0.011
Network density	0.213	0.196
Network heterogeneity	0.378	0.426
Average betweenness centrality	0.00367	0.00398
Modularity	0.62	0.62

Across the two days, the number of nodes and edges stays relatively similar. The number of nodes can be attributed to the different pupils who participated on the two days, whereas 228 pupils attended the second day and only 226 the first. All ten teachers attended both days. Meanwhile, the number of edges is higher on the first day compared to the second. This could be due to higher activity on the first day as the badges were introduced and may have encouraged the students to interact more with new individuals. The average weighted degree of the networks was 513.8 on day one and 547.5 on day two. This means that each pupil spends, on average, about 2 hours and 51 minutes on day one interacting and about 3 hours on day two. Furthermore, indicating that individuals interacted on average with more new people for shorter periods and created more edges on day one. In contrast, individuals on day two interacted on average longer with fewer individuals.

The low network diameter and average shortest path indicate that the network displays a small-world property on both days. A diameter of three means that each individual is only three connections away from any other peer or teacher in the school on any given day. Furthermore, the average shortest path between any two individuals on both days is 2.2 on day one and 2.4 on day two. Similarly to the empirical networks described in section 2.2.2, this indicates that the networks have a small-world property with an overall low distance between all nodes in the network.

Moreover, the high average clustering coefficient, $\langle C \rangle$, of 0.502 on day one and 0.560 on day two indicates that the individuals tend to form clusters within the network. As described in section 2.2.2, in a complete cluster, all nodes are connected to one another through edges. The global clustering coefficient describes the probability that any neighbours of a given node are connected. A $\langle C \rangle$ of 0.5 thus indicates that there is a 50% chance of any two neighbours of a given node in the network being liked together. As seen in Figure 4.2, there is a tendency for individuals in the same class to interact more with one another, thus increasing the likelihood that neighbours in the same grade interact with

one another. Furthermore, the high clustering could also be attributed to social cliques or friend groups interacting. However, since the network was only recorded for two days, it is challenging to further identify and quantify said cliques.

The network density, or average closeness centrality, further indicates a highly connected network on the two days. The network density measures the proportion of potential connections possible in a network in relation to the actual connections. It is relatively high on both days, with 0.213 on day one and 0.196 on day two. This indicates that about 20% of all possible connections are made in the network, making it similar to the density of actors' social interaction network described in section 2.2.2. Day one seems to be more connected than day two, further emphasising that there may have been some bias to interact with more individuals on day one than on day two.

Furthermore, the heterogeneity reveals students' interaction bias. Heterogeneity measures to which extent nodes are connected with other nodes that have different degrees. The higher the heterogeneity, the more prominent the appearance of hubs is in the network. A network heterogeneity of around 0.4 indicates some diversity or unevenness in the distribution of node degrees in the network, with pupils not exclusively interacting with other nodes that have a similar degree or, in terms of primary schools, popularity. Thus, it may seem like some individuals act as hubs and are interacting with other nodes with a lower degree. Therefore, the heterogeneity indicates a bias for some individuals to interact more with their peers.

Average betweenness centrality further describes the mechanisms that drive the connected network. As mentioned in section 2.2.2 betweenness centrality focus on locating the nodes of the network where many interactions are passed through. An average of the betweenness centrality for all nodes can indicate how dependent a network is on interactions through bridge nodes, for instance connecting individuals from different grades. The betweenness centrality is low at around 0.004 for both days. This indicates that the presence of central hubs connects and decreases the network diameter and that the flow of interactions is not evenly spread throughout the network. As seen in Figure 4.2 there are clear divisions between the grades. Therefore, the highly connected network can be attributed to the presence of hubs or nodes connecting the grades.

The modularity further highlights the divisions between each grade. The higher the modularity, the more prominent and connected the different nodes are to other nodes within their communities. By choosing the five different grades as modules, modularity was calculated to 0.62 on both days. This relatively high value indicates that there is a preference for individuals to interact with individuals in the same grade or module, further confirming the groupings seen in Figure 4.2.

To get an overview of how individuals interacted across the two days, the degree of each node on the two days studied was plotted against one another as shown in Figure 4.3. The relationship between the degree a node has for two days seems to be increasing, indicating that the same nodes have the same amount of activity on day one and day two. A Pearson correlation coefficient was computed to assess the linear relationship between the degree of each individual across the two days and was found to be positively correlated, $r(230) = 0.54$, $P < 10^{-3}$. Therefore, individuals that interact a lot on day one are more likely to

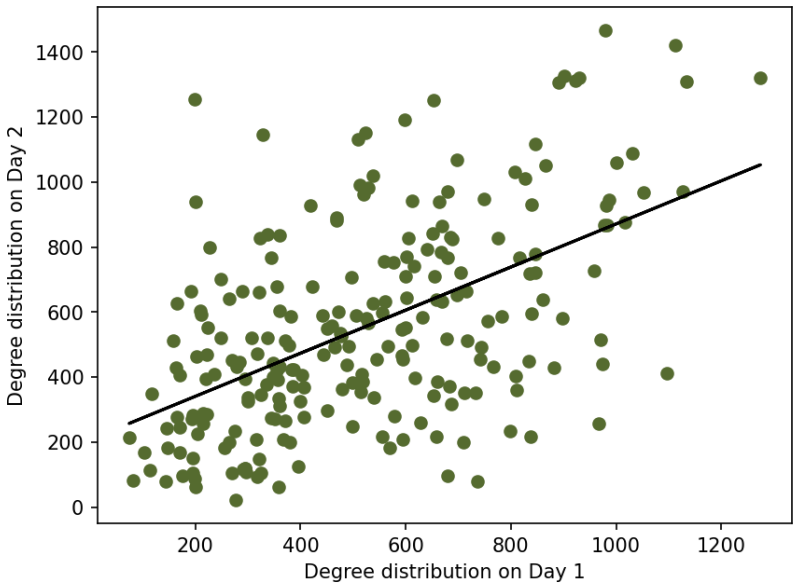


Figure 4.3: The degree of specific nodes on day one and against day two. The line represents a linear regression of the values.

have increased interactions on day two. Due to the pattern of some individuals being more social across multiple days than others, the bias that was detailed in subsection 3.3.1 was implemented into the interaction model.

4.1.2 Comparing the empirical networks with random models

As described in subsection 3.2.4, three random networks were generated and compared with the average of the two days investigated. The reasoning behind this comparison is to check whether there were any prominent shared features between the random networks and the empirical ones and to see whether or not a random network would be satisfactory when modelling primary school interactions.

Table 4.2 shows an overview of the topological network parameters investigated for the random networks and a mean of the parameters from the empirical data set on day one and day two. The random networks were generated using the parameters described in Table 3.2, and thus all networks have the same number of nodes, $N = 237$. The numbers of edges present in the networks are also relatively similar across all compared networks with a slight variation. Additionally, the network diameter and network density are the same for all networks, as they all share a small-world property. Finally, the average betweenness centrality is also similar across all investigated networks.

In contrast, the average degree for the randomly generated networks is slightly higher than for the mean of the two days. This indicates that, on average, the nodes in the random

Table 4.2: Comparison of analysis of topological parameters of the mean of day 1 and day 2 ($\overline{1\&2}$) and the random networks Erdős-Rényi (ER), Watts-Strogatz (WS) and Barbasi-Albert (BA).

Parameters	$\overline{\textbf{1\&2}}$	**ER**	**WS**	**BA**
#nodes	237	237	237	237
#edges	5721	5895	5925	5852
Average degree	48.3	49.7	50	49.4
Network diameter	3	3	3	3
Average shortest path length	2.3	1.79	1.79	1.79
Average clustering coefficient	0.531	0.211	0.285	0.311
Network density	0.205	0.211	0.212	0.210
Network heterogeneity	0.402	0.122	0.078	0.478
Average betweenness centrality	0.00383	0.0034	0.0034	0.0034

networks either have a higher maximum degree of nodes or fewer low degree nodes. The distribution of degrees will be described in more detail later in this section. Furthermore, the average shortest path is shorter for the random networks, which may indicate the presence of more high degree hubs that decrease the distance across the network since the network diameter is the same for all investigated networks.

The average clustering coefficient for the random network is lower than the mean of the two days. Since WS and BA both are known for having a clustering coefficient similar to that of real-world social interactions, this indicates that the empirical networks have higher clustering than what is typically seen for social interaction networks. The higher clustering coefficient could be explained due to the social interactions in a school environment being more controlled by the physical separation of individuals into classrooms. Therefore, one would expect higher clustering in classes compared to an undisturbed social interaction network. Thus, none of the random networks tested in this thesis captures the high clustering characteristic of the empirical model.

Meanwhile, the heterogeneity of the BA network is similar to that of the mean of the experimental ones; both ER and WS have a lower heterogeneity. This difference is due to mechanisms that drive the generation of the network for the different models. As both ER and WS have a given percentage of either drawing or rewiring an edge between two nodes, the connections are drawn regardless of the degree of each node. However, BA uses preferential attachment when adding new edges and nodes, causing new additional nodes to form edges with nodes that have a higher degree. Thus, BA networks will have a higher heterogeneity. This could indicate that the empirical networks also have a form of preferential attachment with regard to node degree.

Furthermore, none of the investigated random networks described have a similar enough degree distribution to the empirical network. As seen in Figure 4.4 both ER and WS have a Poisson degree distribution that deviates strongly from the degree distribution of the empirical network, and most nodes in these networks have the same degree with few highly connected nodes. However, for the degree distribution of both empirical networks and BA

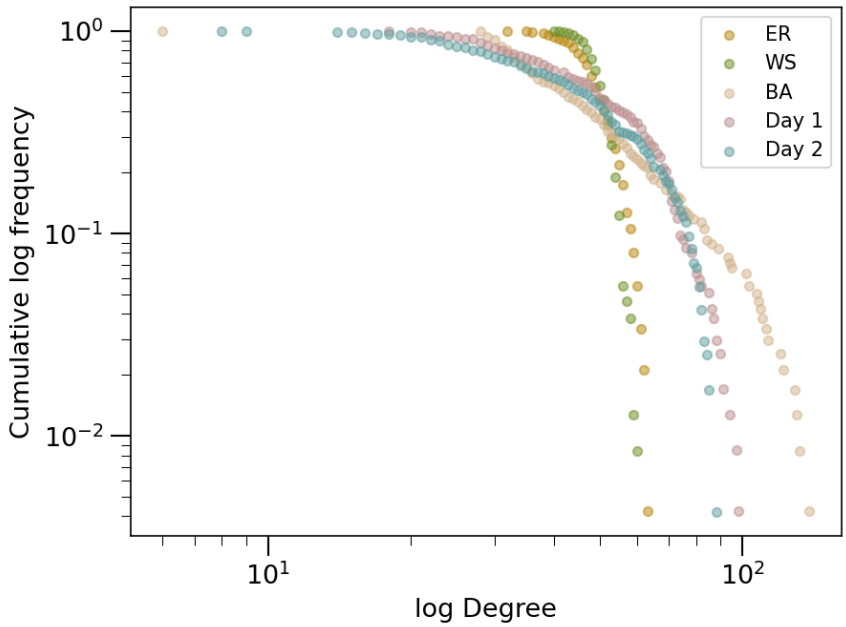


Figure 4.4: Cumulative degree distribution P($X < x$) on day one and day two and the random networks Erdős-Rényi (ER), Watts-Strogatz (WS) and Barbasi-Albert (BA).

many nodes have a low degree while a few hubs have a high degree. Even though BA network, with a Power law distribution, has the closest degree distribution to the empirical networks, it still deviates. This is because the empirical network has a more curved shape and has hubs of smaller size than for BA.

Overall, the two empirical networks were the most similar to the BA network. However, the major drawback to modelling primary school interactions using BA is the low clustering as well as the deviations seen for the degree distributions. Hence, none of the investigated models were satisfactory for the simulation of primary school interactions. However, they highlighted that the empirical network has a higher clustering coefficient than what is expected for social interactions and has a degree distribution that is the most similar to Power law.

4.1.3 Identifying interaction types through heatmaps

To further look into how individuals interact with each other in a primary school environment, a heatmap of the adjacency matrix of the empirical network was created as described in subsection 3.2.3. Figure 4.5 shows the heatmap for all interactions on the first and second day. Each pixel in the map increase in colour intensity according to how many times two individuals have interacted, and the heatmap is sorted according to grade. As shown in the figure, individuals tend to interact with other individuals in the same class. In addition, many interactions occur between individuals in the same grade. The remaining interac-

tions are between individuals less similar to one another, meaning between individuals who are not in the same class or grade.

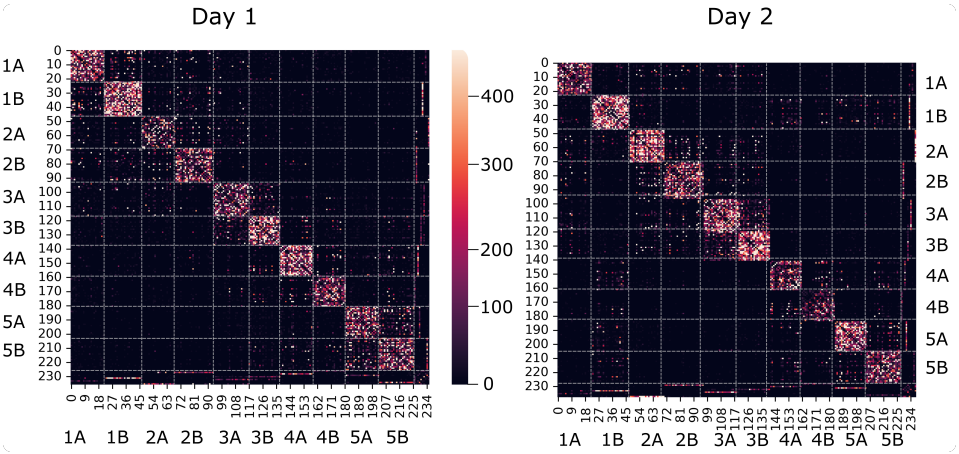


Figure 4.5: Heatmap representation of the adjacency matrix on day one and day two of the survey. Helplines describe the start and end of IDs for individuals of the same class.

Therefore, from the figure, it is possible to identify three different types of interactions; grade, class, and off-diagonal interactions. The off-diagonal interactions describe interactions that are not along the diagonal, where class and grade interactions are localised, meaning that they are between individuals less similar to one another. They occur more rarely and with lower duration than grade and class interactions and are observed between less similar individuals. Furthermore, the off-diagonal interactions capture interactions between individuals that have no scheduled classes or activities together and therefore represents how individuals interact with others during breaks. Since there is a significant difference between the interaction length and intensity during regular breaks and the lunch break as described in subsection 2.3.3, the effect of the lunch break division is further investigated.

Firstly, a hour-by-hour plot of the number of interactions per grades, as seen in Figure B.2 in Appendix B, revealed that most interactions occurred during breaks. Figure 4.5 shows that there are more interactions on the off-diagonal between the grades that have lunch at the same time, as seen by the division between interactions in grades one to three and grades four and five. Furthermore, extracting and creating a heatmap of all interactions during the two-hour lunch break further highlighted the division. See the heatmap in Figure B.3 Appendix B. Therefore, an additional sublayer was added to describe lunch break interactions, primarily to help increase the off-diagonal interactions between grades one to three and four and five.

Figure 4.6 shows heatmaps of the four layers into which the school is divided. Dividing the layers into different heatmaps allows the investigation of the less extreme interactions without them being skewed by the high-intensity class interactions. This next section will focus on quantifying the degree distribution for each layer. The grade interaction layer

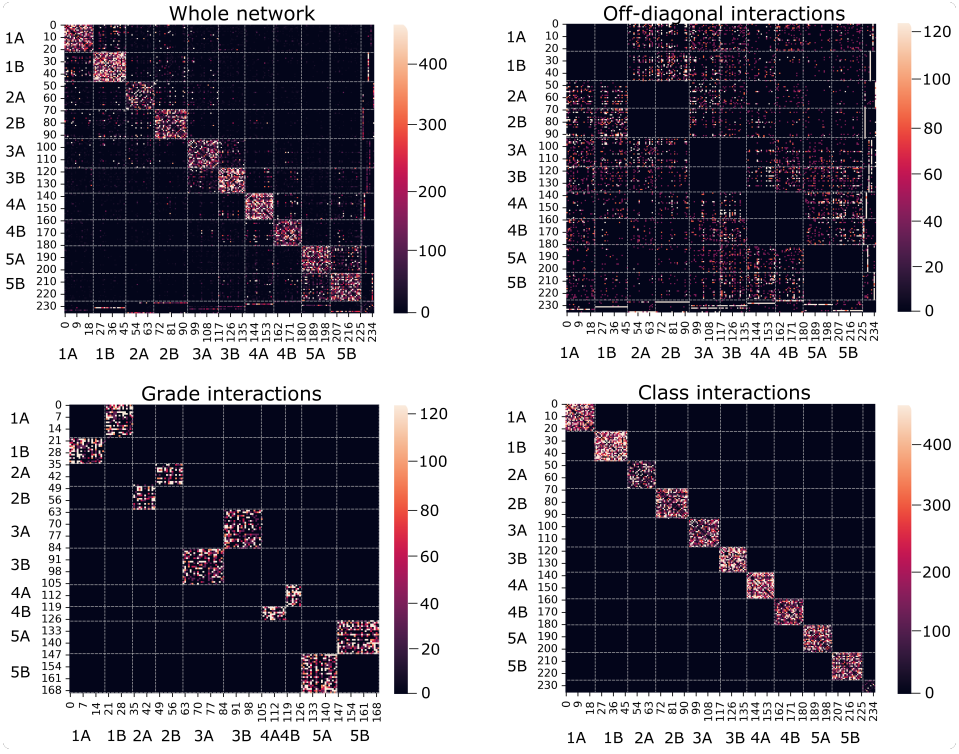


Figure 4.6: Dynamics of interaction between individuals for the whole network, exclusively grade interactions, off diagonal (all interactions except those of the same grade) and finally interactions between individuals in the same class. Helplines describe the start and end of IDs for individuals of the same class.

shows that the number of individuals interacting in each grade changes per grade; therefore, the number of pixels described for grade interactions varies. This gives the network a hierarchic structure, where both the chance and strength interactions differ depending on the similarity between the two interacting nodes.

4.1.4 Investigating network assortativity

Assortativity in a network denotes the preference of a network's nodes to interact with other similar nodes. Figure 4.7 describes the amount of times edges have occurred between nodes of different degrees. The figure is a mirror image along the diagonal. A small but visible trend is that the edges connect to similar degree nodes, with a Pearson correlation coefficient of 0.19 and 0.20 on days one and two, respectively. This is especially apparent for the lower degree nodes. This means that there is a slight tendency for nodes to connect to other nodes with similar degrees, or in the sense of school children, the same popularity.

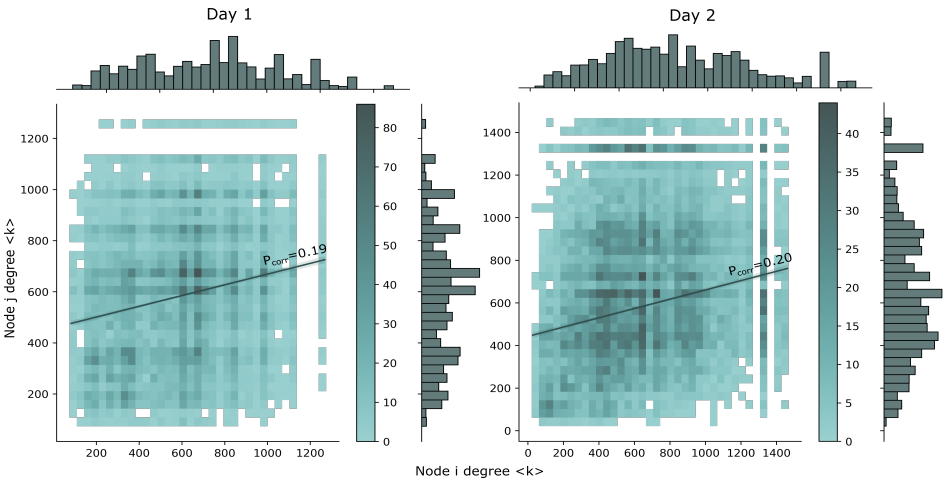


Figure 4.7: The node degree of both nodes i and j present in every edges (i,j) in the graph are plotted against each other, and the colour scheme represents how many times nodes of a certain degree interacts with another node and their respective degree.

4.1.5 Determining degree distribution

The distribution of degrees among the nodes in the network can reveal how different students interact with other students, and is an important characteristic of a model to help determine the distribution of the empirical network. As described in subsection 3.2.6 a cumulative distribution was plotted and can be seen in Figure 4.8.

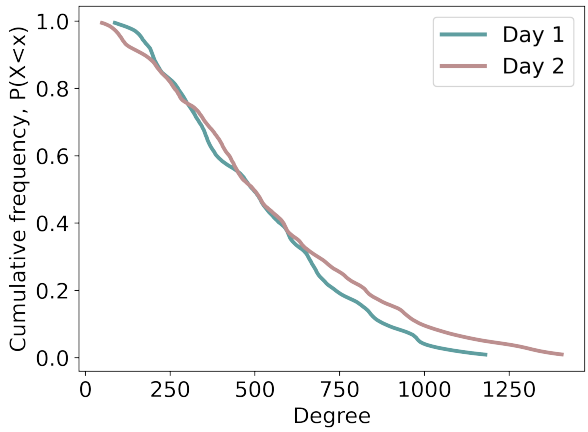


Figure 4.8: Cumulative degree distribution on day one and day two of the empirical networks. Different colours denote the two days and degree is plotted as a function of the cumulative frequency $P(X < x)$

No nodes in the network had a degree of zero; all nodes had a degree higher than about 100. The relationship between the frequency of the different degrees decreases linearly until about degree 750, where the frequency is around 0.2. This happens towards the tail of the graph, and the behaviour is visible for both day one and day two. The network thus has many nodes with an intermediate degree and few nodes with a high degree. In addition, the relationship between the nodes' degree and frequency is close to linear for degrees 100-750 and is then flattened out afterwards.

As concluded when comparing the network to other random models in subsection 4.1.2, the distribution does not seem to follow a specific distribution function. Instead, it looks like it is a combination of different distributions, much like the overall network consists of different interaction types (i.e., the different interaction layers). Therefore, the following sections will investigate the distribution of each of the interaction layers.

Section 4.1.3 describes the different layers the network is divided into. Investigating the degree distribution amongst the different layers may help explain the combined behaviour that was observed in Figure 4.8. Therefore, Figure 4.9 shows an overview of the degree distributions seen for the off-diagonal, grade, class, and whole network interactions. At first glance, it is possible to deduce that most interactions within the networks occur between individuals of the same class. They have a higher frequency of interacting with each other for all the degrees and have a maximum degree of about 900. Grade and off-diagonal interactions have a similar relationship between frequency and degree for the first 100 degrees. Then the off-diagonal interactions tail off with one node with a degree of about 500. Finally, the distribution of grade interactions stays steep, and the node with the highest degree has a degree of about 300.

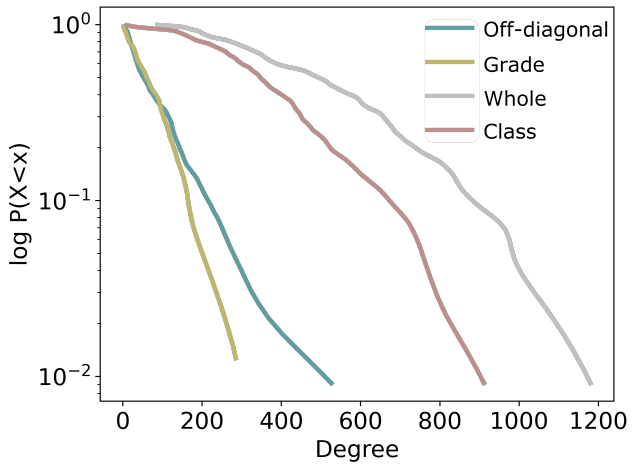


Figure 4.9: Cumulative degree distribution on day one for the three layers of the empirical network, with the degree as a function of the log cumulative frequency ($P(X < x)$). Different colours denote degree distribution for the layers off-diagonal, grade, whole and class.

Degree distribution of off-diagonal and grade interactions

The degree distribution of the off-diagonal interactions contributes to the overall interactions that are happening outside of those determined by grade and class. It is the outermost layer that impacts almost all interactions in the network, and all other interactions contain a level of off-diagonal interactions. Meanwhile, the grade degree distribution details interactions between individuals in the same grade but not the same class. It seems that the distribution for both off-diagonal and grade degrees is approximately linear on a semi-log scale. In order to quantify the distribution, the distribution of off-diagonal and grade interactions is denoted as shown in Equation 4.1, with different variables for c.

$$y = \alpha * \log_{10}(\frac{1}{x}) \tag{4.1}$$

Since $x = 300$ corresponds to 10^{-1} for the off-diagonal layer, the degree distribution is visible in Equation 4.2.

$$y = 300 * \log_{10}(\frac{1}{x}) \tag{4.2}$$

Furthermore, $x = 200$ corresponds to 10^{-1} for the grade layer, and the degree distribution is visible in Equation 4.2.

$$y = 200 * \log_{10}(\frac{1}{x}) \tag{4.3}$$

Degree distribution of class interactions

Similarly to the whole class interactions, it is difficult to find a function for the degree as a function of the cumulative frequency. Unlike the grade and off-diagonal layers, the class layer does not form a single line and is therefore not easy to estimate. A trial-and-error approach with different distributions and parameters was then used to match this degree distribution. The final parameters used to generate interactions between individuals of the same grade can be seen in Table 3.5.

4.1.6 Investigating interaction distribution

The interaction distribution can be investigated by plotting the distribution of the weight of every interaction between individuals as described in subsection 3.2.7. In essence, this means plotting the cumulative frequency of the intensity of each pixel shown in the heatmap in Figure 4.5. The result can be seen in Figure 4.10, where the log of the frequency is plotted as a function of the log of the weight for each interaction. The figure thus contains the weighted interaction distribution for each layer.

Of notice is that the weight distribution for the whole network compared to that of the class-class interactions is relatively similar. For the interactions with the highest weights,

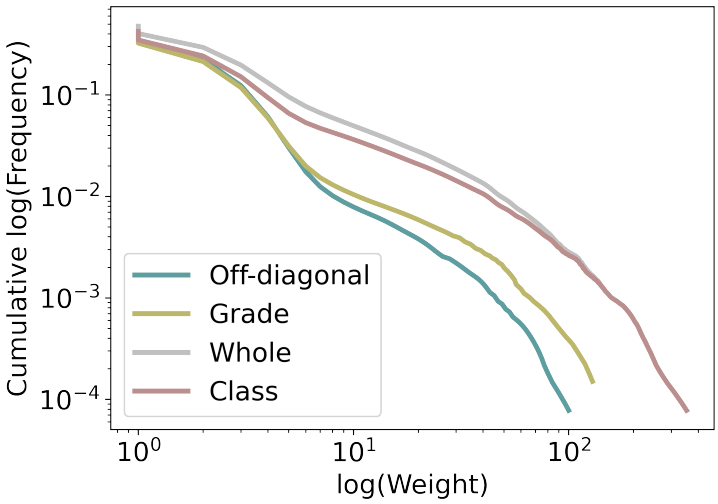


Figure 4.10: Cumulative log-log interaction distribution of all interactions for the three layers off-diagonal, grade and class in addition to the whole network. The cumulative log frequency is plotted as a function of the log weight.

the whole network interactions are all class interactions, which is seen by the class distribution overshadowing the whole network distribution for the high weights. This makes sense, as students spend the most time with individuals in their class, and thus the highest weights in the network should be of class-class interaction origin.

Meanwhile, just like the degree distribution, the grade and off-diagonal distribution of weights are similar. The distribution appears to be bimodal, with two distinct peaks. Furthermore, the lower weights registered in the network overlap entirely until about the weight of ten is reached. This is where the two starts to deviate. The grade weights maintain a higher frequency for higher weights than the distribution of the off-diagonal weights.

Since the network in itself does not display any definitive interaction distribution similar to the most common ones described in Figure 2.5, it was decided to also look into the distribution of interactions per single node. Appendix B.1.1 in Appendix B describes the approach of how the distributions of single nodes were plotted on a log-log scale and taken a linear regression of. Since this was performed for all nodes, Figure B.1 in Appendix B shows a excerpt of how each layers approaches linear on a log-log scale. Table 4.3 further highlights that the R^2 values are high for all interaction types. This is especially true for off-diagonal and grade interactions. Class interaction has a lower R^2 average. This indicates that each person's interactions with other individuals are distributed and closely resemble a Power law, at least for off-diagonal and grade interactions.

Table 4.3: The average and standard deviation of R^2 for a linear regression of each node's interactions with other individuals in the whole network, off-diagonal, grade or class for day one and two.

	$\mathbf{R^2}$ **day 1**	$\mathbf{R^2}$ **day 2**
Whole network	0.934 ± 0.0480	0.912 ± 0.0559
Off-diagonal	0.944 ± 0.0480	0.943 ± 0.0559
Grade	0.927 ± 0.066	0.923 ± 0.0767
Class	0.906 ± 0.0574	0.885 ± 0.0690

4.1.7 Summary of characteristics of the empirical data

The characteristic to be implemented into the model is then a high clustering coefficient compared to that of social interaction networks, small world property, slightly assortative network, and a hierarchic structure. In order to create a framework for such a network, there was an attempt to match degree and interaction distributions.

As mentioned in subsection 3.3.1 the likelihood of interactions happening is based on individual biases as well as the similarity between the two individuals interacting. Therefore, both bias for different layers as well as p_{ij} value, i.e., how likely individuals are to interact, was calculated. First, Power law was assumed for base, lunch, and grade p_{ij} value due to the distribution displayed by single node interactions. Then, the base and grade bias was set to the function described by Equation 4.1 before changing the α until a similar distribution was reached. Next, for generating class interactions, a combination of a normally distributed bias with a gamma-distributed p_{ij} was used. Finally, a Poisson distribution of the p_{ij} was used to simulate the weight of each interaction.

4.2 Evaluation of the simulated model

This section will detail the evaluation of the simulated interaction model. This will primarily be done by comparing the simulated model with the empirical one, either through general network analysis and heatmap as shown in subsection 4.2.1, comparing the degree distributions, as seen in subsection 4.2.2 or by comparing the interaction distribution as seen in subsection 4.2.3

4.2.1 General comparison between empirical and simulated network

Table 4.4 describes the overall differences between the simulated model and the mean of the values for the two empirical networks, according to general network analysis. Since the model was generated based on matching the degree and interaction distributions for the different layers, a general network analysis can help shed light on how successfully it replicated the other characteristics of the empirical model. Firstly, the number of edges, average degree, average shortest path length, and network density are lower for the simulated model than for the empirical one. This indicates that the simulated model overall has fewer connections and that its links are more evenly spread out throughout the network,

decreasing the average shortest path.

Table 4.4: Network analysis of topological parameters of a network conducted for day one and day two using Networkx and Cytoscapes Analyzer.

Parameters	1&2	Model
#nodes	237	237
#edges	5721	5172
Average degree	48.3	43.8
Network diameter	3	3
Average shortest path length	2.3	1.8
Clustering coefficient	0.531	0.400
Network density	0.205	0.187
Average betweenness centrality	0.00383	0.00357
Modularity	0.62	0.31

The lower modularity for the simulated model further highlights that interactions of the model are not as secluded and involved in forming modules within classes as what is seen for the empiric network. This indicates that more interactions are, for instance, occurring on the off-diagonal or outside of the class interactions in the model as compared to the empiric network. Meanwhile, the model has a slightly lower clustering coefficient than the empirical model. This shows that even though modularity has decreased, the model still maintains higher clustering than what is seen for other real-life networks.

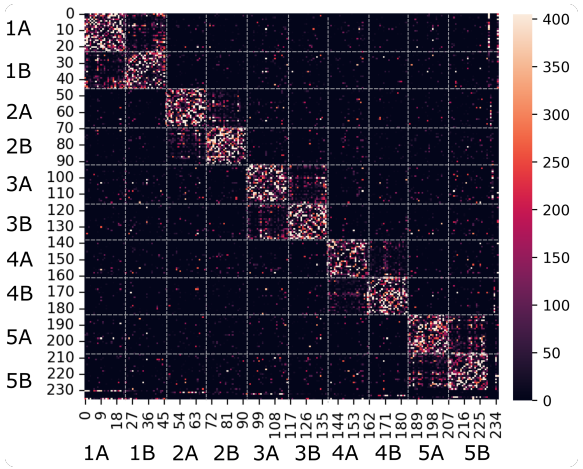


Figure 4.11: Heatmap representation of the adjacency matrix of a simulated network. The IDs alongside which class each individual is in is shown, with helplines dividing each class from one another.

Figure 4.11 describes the heatmap of the simulated model. Similar to the empirical network and of most importance, is the fact that the class-class interactions are the most

prominent. The strongest and highest frequency of interactions, with a maximum value of 400, occur within classes, which was a part of the hierarchical structure the framework aimed to replicate. Similarly, there are more interactions between individuals of the same grades than how they interact with individuals outside of their grade. The most significant difference between the empirical networks shown in Figure 4.5 and the simulated one, is the off-diagonal interactions. They are more evenly dispersed throughout the heatmap, and are not creating as strong of a preference between grades one to three and four and five as they are in the empirical network on day one.

4.2.2 Comparing degree distribution for empirical and model network

Figure 4.12 describes the degree distribution for all the layers as well as the whole graph. As described in subsection 3.2.6, the distribution of degrees is plotted on a semi-log cumulative scale.

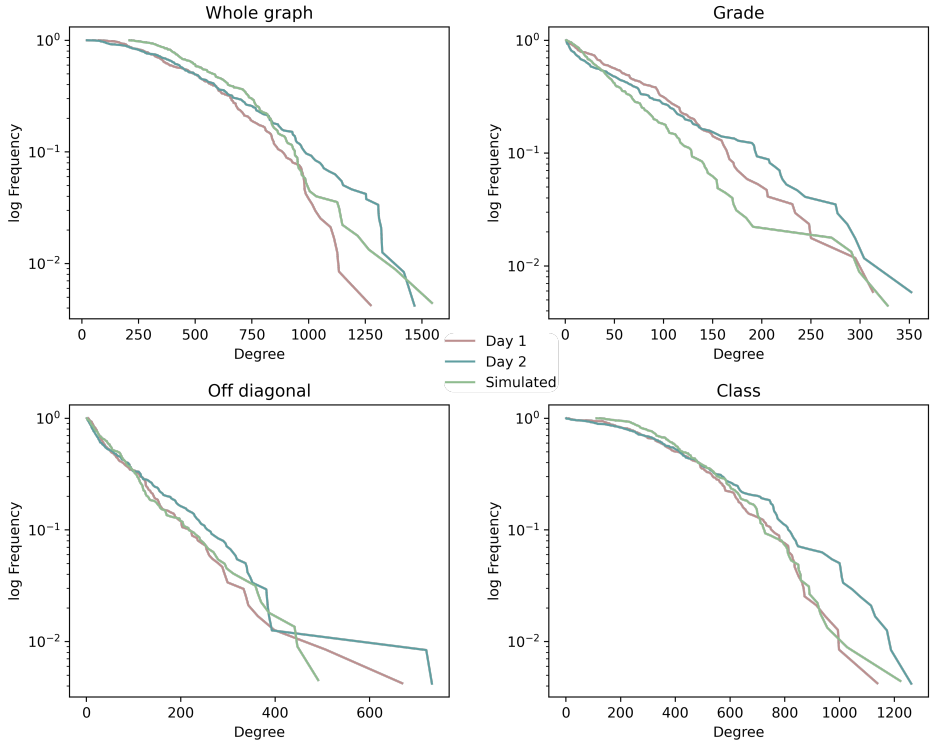


Figure 4.12: Plot of degree distribution for grade, off diagonal and class interactions as well as whole graph. The plot is semi-log where the log frequency is plotted as a function of degree.

The first distribution to be set was the off-diagonal, as it is the outermost layer of the model. As seen, the off-diagonal differs between day one and day two quite significantly,

especially towards the tail of the function. This is due to fewer data points, as the frequency of individuals with such a high degree is rare both on days one and two. This indicates that some deviance from day one and day two for degree distribution should be accepted for the simulated model. The simulated model follows day one and day two for the first few 100 degrees, with high frequency. Then, the simulated model has a consequently higher frequency for the nodes with a degree of 100-350 before tailing off at around 500 degrees. The simulated model thus follows the empirical one but deviates more for the larger degrees. This is similar to what is seen for the empirical networks and is likely because it is highly stochastic.

The grade interactions were the next layer to be set. As seen in the grade function of Figure 4.12, the simulated model has a lower frequency for all degrees below 300 than compared to the two empirical networks. Day one and day two of the empirical networks also deviate from each other for degrees higher than 150, where the degrees on day two have a higher frequency than the same degrees for day one. Overall, the grade distribution for the simulated model seems to follow the empirical ones; however, it is slightly more curved as opposed to the grade distribution that has an overall convex shape on both days.

The distribution of degrees within the class layer is seen for Figure 4.12, class. It is the innermost layer and is therefore affected by both off-diagonal and grade distributions. With the given off-diagonal and grade degree distributions, the degree distribution of classes is quite similar to its distribution on day one. The two functions follow each other through the frequency for the first 1000 degrees, and then both tail off towards the end. For the class interactions, day two deviates from day one and the empirical network.

Overall, the degree distributions for the entire network are seen in Figure 4.12, whole graph. For the frequencies of degrees below 750, the three distributions follow each other except that the simulated model has a higher frequency for the lower degrees than the empirical networks.

4.2.3 Comparing interaction distribution for empirical and model network

Figure 4.13, shows the cumulative distribution of weighted interactions as described in subsection 3.2.7. For the off-diagonal interactions of the figure, the simulated and empirical networks have an overall similar interaction distribution. All three distributions have a curved shape and end on the same maximum weighted interactions for the entire network. However, some difference is visible. Day one and day two's log frequency for the lower interactions is higher than for the simulated model. While the distribution of day one and day two follow each other for most interactions, the simulated network has a higher log frequency for the intermediate interactions than the simulated ones.

The grade interactions seen in Figure 4.13 differ to some extent between the empirical networks and the simulated ones. While the empirical networks only deviate from one another for the highest numbers of interactions in the model, the simulated distribution is more curved. Similarly to the off-diagonal distributions, the simulated starts at a lower nor-

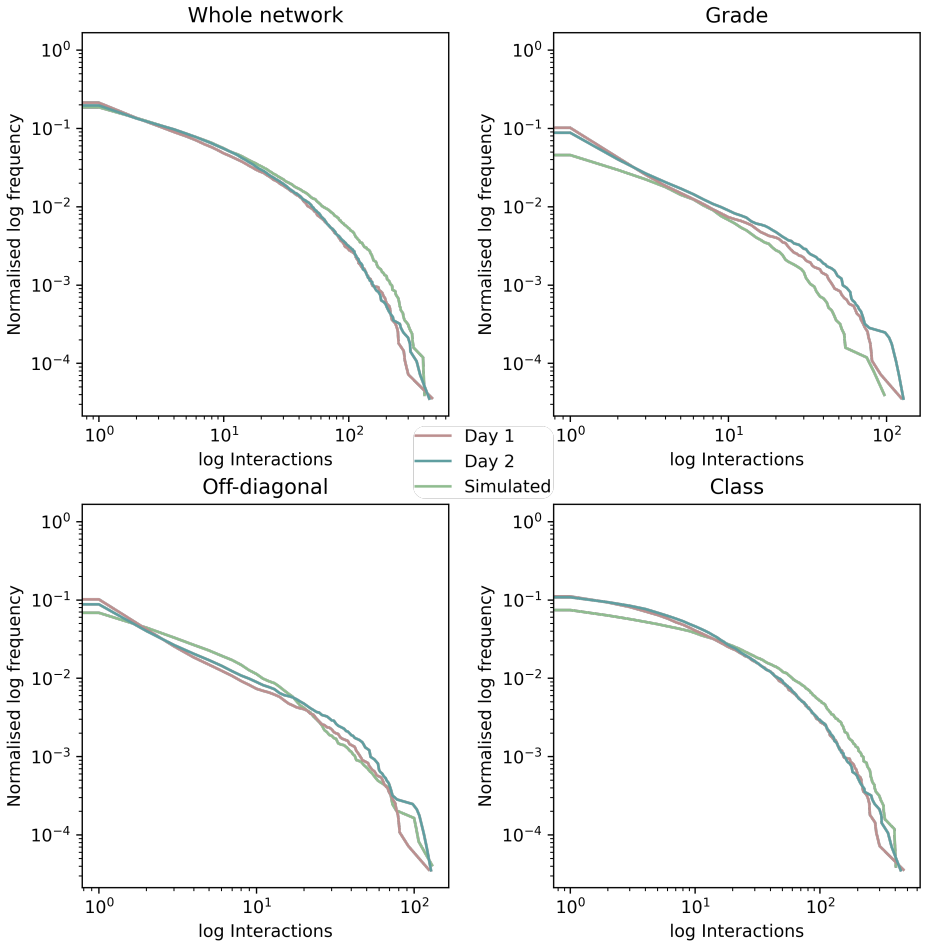


Figure 4.13: Log-log plot of cumulative distribution of weighted interactions for the simulated and empirical networks for the whole network, off-diagonal, grade, and class interactions.

malised frequency. In addition, it has a smaller cut-off for maximum weighted interaction in the network.

The class interactions in Figure 4.13 follow the same curved shape; however, it seems that the simulated is shifted towards the right. While the distributions for days one and two almost entirely overlap until the highest interactions are reached, the simulated curve first has a lower frequency for the low weighted interactions and a higher frequency for the higher weighted interactions.

Finally, the whole graph interaction distribution is similar for both the empirical and simulated networks. For the lowest weighted interactions, the distribution is spot on. Mean-

while, the simulated model has a higher normalised log frequency for the high-weighted interactions compared to the empirical networks. All in all, however, the interaction distributions are overlapping and end on similar maximum weighted interactions for the network.

4.2.4 Model sensitivity

Since the aim of creating this framework is that the model should be general, the model's sensitivity was tested using different input parameters. First, a scenario similar to that of many Norwegian primary schools was entered into the model, with seven grades and three classes. Then, five grades with four classes each were entered into the model. The result can be seen in Figure 4.14.

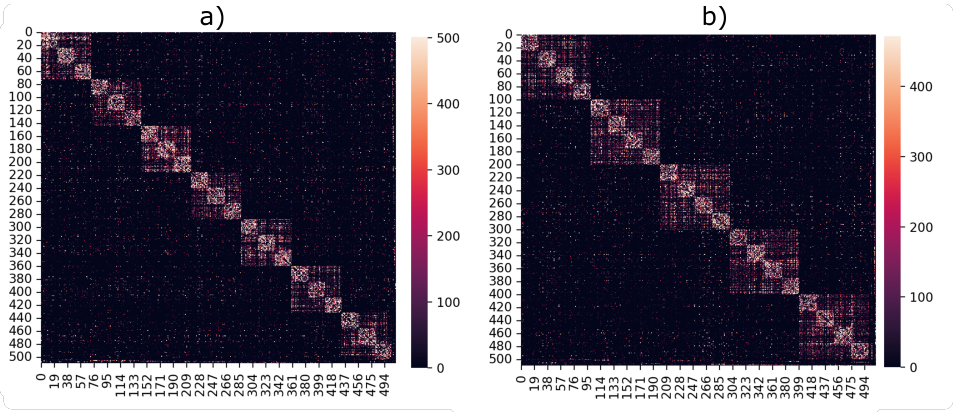


Figure 4.14: Heatmap of the interactions generated by the model when the input is a) seven grades and three classes and b) five grades and four classes

As seen from Figure 4.14, both inputs retained the overall characteristics one would expect to find in primary schools. The seven grades are visible as seven large boxes along the diagonal, with three smaller boxes within representing the three classes. Similarly, the model with five grades and four classes input shows the same hierarchical structure. Interestingly, the grades one to three and four and five separation is not visible.

4.3 Disease transmission and mitigation

The current section will go into detail on the disease transmission model that is implemented as described in section 3.4. subsection B.3.1 in Appendix B shows how disease is transmitted on a simulated model. First, subsection 4.3.1 will describe disease transmission runs on the empirical network compared to the simulated one. Then, subsection 4.3.2 will investigate the effect of implementing the different traffic light levels on the model. Furthermore, subsection 4.3.3 will describe how weekly testing affects disease transmission on the model. Finally, the effect of asymptomatic versus symptomatic disease course on r_{00} will be investigated in subsection 4.3.4.

4.3.1 COVID-19 transmission in empirical and simulated model

To further test the validity of the simulated model, the transmission of disease on the empirical networks was compared to transmission on the modelled network. The result is seen in Figure 4.15, where an average of the number of recovered individuals are plotted as a function of days since the disease was introduced into the system. The fewer infected individuals and thus the fewer recovered, the less disease transmission occurs.

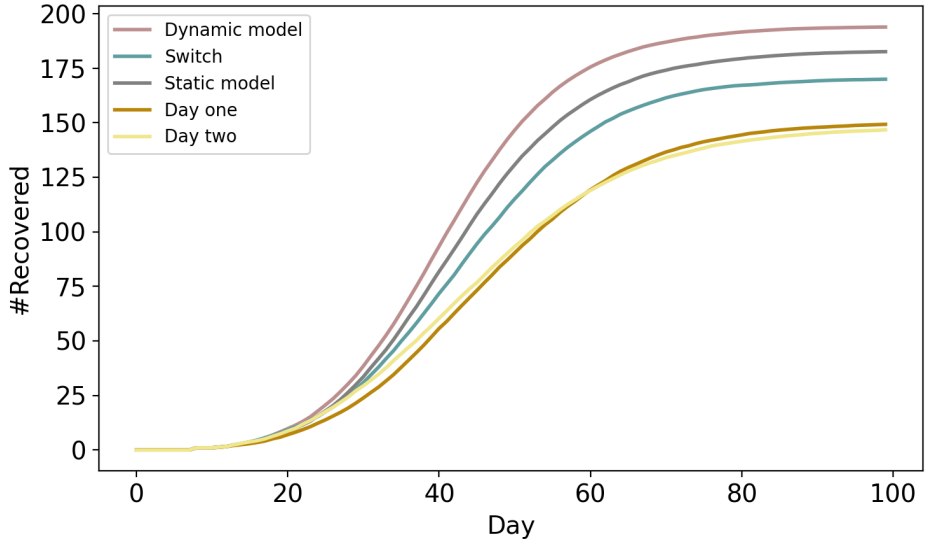


Figure 4.15: Disease transmission on the empirical network for two days in addition to the simulated network. Switch indicates that the interaction pattern on the two empirical days were alternating. Days since infection is plotted as a function of the number of recovered (#Recovered).

Transmission on alternating days between day one and day two networks has a higher disease transmission than that of only day one and day two in the empirical network. For the first 30 days, transmission on day one, day two, and the switch between the two days increased simultaneously before the number of recovered increased more rapidly for the switch. The average number of infected at the end of the epidemic is 160 for the switch and about 150 for day one and day two, indicating that the transmission on the switch is higher than running transmission solely on day one and day two empirical networks. This could suggest that the difference in interaction patterns seen between the two days plays a role in increasing transmission.

Meanwhile, transmission on the model caused more disease transmission in the network. Disease transmission on the static model, where transmission is run on the same network for all 100 days, is lower than in the dynamic model, where the network changes every day. Ideally, one would expect the static model to have a similar level of disease transmission as what was seen on day one and day two. However, the increased interactions on the off-diagonal and the lower average shortest path length may have made

the model more susceptible to disease transmission. Since off-diagonal interactions often contribute to moving infection from one cluster or module to another, it makes sense that increased off-diagonal interactions increase transmission. Similarly to what was seen for the switch where transmission was run by alternating the two empirical networks, having a dynamic model did increase transmission for the model. Therefore, disease transmission on the static model ends at 180 for static while the dynamic model ends with 190 recovered.

4.3.2 Effect of traffic light level on r_{00}

Using the measures and regulations described in section 3.4.5, in primary schools, the model predicted the transmission of disease in the different traffic light states as seen in Figure 4.16. Once again, the average number of recovered is plotted as a function of days since the disease was introduced into the system. Similarly to what one would expect, green is the traffic light level where most transmission occurs, followed by yellow and then the red level. Overall, the implemented traffic light measures effectively mitigate disease transmission when a disease is introduced into the system. To see the margin of error for each level, represented by plotting 10 randomly selected iterations, see Figure B.7, Figure B.8 and Figure B.9 in Appendix B.

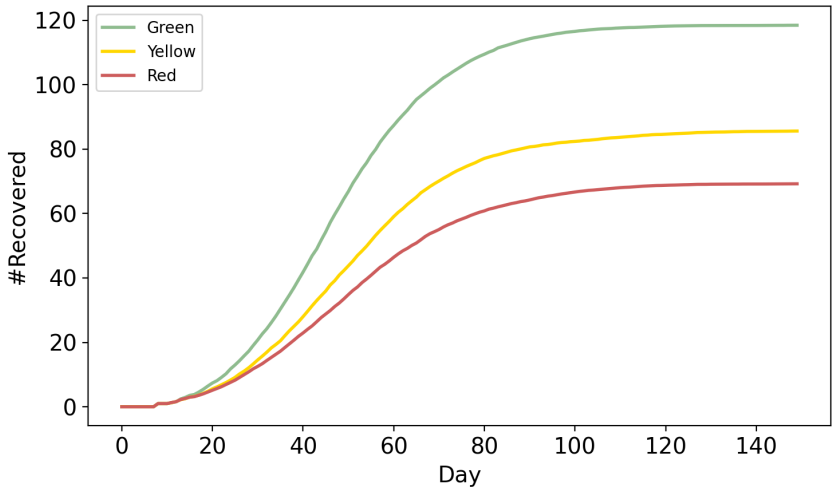


Figure 4.16: Days since infection as a function of recovered for the disease states green, yellow and red. Each curve is an average of 100 iterations

Figure 4.15 showed that disease transmission on the dynamic model ended with 190 infected and recovered individuals. In comparison, transmission on the network with the green level implemented ended with 120 recovered individuals. Therefore, the decrease from 190 to 120 recovered individuals when implementing green level measures indicates that it decreases transmission. Furthermore, the transmission rate is lower, and the amount of days it takes to reach a stable number of recovered with no new infections is around 100

days for the green level as opposed to 90 for the dynamic model.

Furthermore, the yellow and red levels of the traffic light model also have a lower number of infected/recovered at the end of the epidemic, with red at about 70 cases in the system and yellow at 85. The effect of the measures in both yellow and red levels of the traffic light successfully reduced the number of infected individuals. Of notice, however, is that going from yellow to red traffic light level does not contribute equally as much to decreasing transmission compared to moving from the green to yellow level. One explanation for this behaviour, may be that the yellow level heavily affects off-diagonal interactions by dividing classes into cohorts and limiting interactions outside the cohorts. Therefore, implementing the yellow level effectively limits off-diagonal interactions and, in turn, decreases transmission significantly. Meanwhile, the red level tries to limit connections within classes, thus reducing disease transmission, but not to the same extent as decreasing off-diagonal interactions.

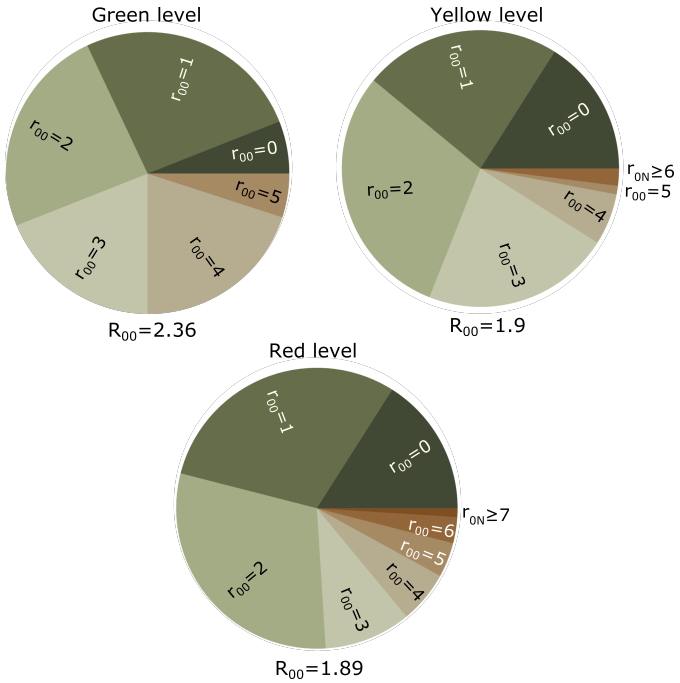


Figure 4.17: R_0 calculated for green, yellow and red traffic light level. The overall R_0 as detailed at the bottom of the figure describes its expected value. Frequency based on averages of 100 iterations.

Figure 4.17 describes the proportions of R_{00} for the 100 iterations run over 100 days for different traffic light levels. The expected value, R_{00} based on the frequencies of R_{00} calculated in subsection A.3.1, Appendix A, is also provided in the figure. The R_{00} of the green traffic light level is set to 2.36, while yellow has a R_{00} of 1.9 and red of 1.89. When compared to an R_{00} of 3.16 for a network where disease spreads unhinged, which will be described in subsection 4.3.4, the R_{00} is lowered for all traffic light levels. While the green level has a much higher R_{00}, the values for the red and green levels

are almost identical. This further highlights the difference in transmission between the yellow and red traffic light levels and is possibly attributed to the yellow level's significant effect on decreasing off-diagonal interactions. Interestingly none of the states had an R_{00} value below 1, indicating that the implemented measures did not mitigate transmission but merely slowed it down.

4.3.3 Weekly testing and effect on transmission

The next policy investigated was every week (weekly) and every other week (biweekly) testing to see how it would affect disease transmission, and the method for simulating testing is described in section 3.4.5. Figure 4.18 shows how the disease was spread across primary schools with different levels of testing on a green traffic light level. During the first 20 days, transmission on not tested, weekly tested, and biweekly tested networks moves slowly and at the same rate. Afterwards, the number of recovered cases for not tested and biweekly tested increases rapidly before stabilising at 70 and 120, respectively. Meanwhile, weekly testing results in a halted increase before stabilising around 30 recovered individuals. To see the margin of error, see Figure B.10, Figure B.11 and Figure B.12 in Appendix B.

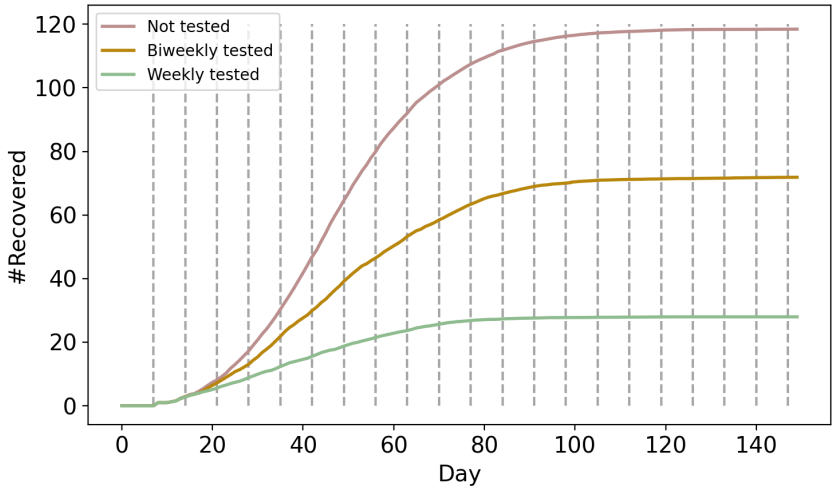


Figure 4.18: Recovered as a function of the day for not tested, weekly testing and biweekly testing on a green traffic light level. Helplines show every seven days, and the data is the average of 100 iterations of each testing type

Both weekly and biweekly testing seems to reduce transmission. The amount of recovered individuals seems to end at about 50 less recovered at the end of the epidemic when moving from non-tested to biweekly testing and by 40 when moving to weekly testing. This indicates there might be a proportional decrease in transmission when implementing weekly and biweekly testing.

Figure B.13, in Appendix B, shows the frequency of R_{0N} for each testing strategy. Since

this measurement R_{0N} denotes how many individuals each infected pupil infected before recovery, it is not comparable to the R_{00} of, for instance, the traffic light analysis. Therefore, the number can only be used when comparing the different testing strategies. See subsection 3.4.3 for the detailed explanation of the two. The R_{0N} of not testing and biweekly testing was 0.98, while weekly testing had a R_{0N} of 0.97. Since disease transmission is halted and individuals recover without infecting others, the overall R_{0N} is lower compared to R_{00}.

As expected from Figure 4.18 the lowest R_{0N} is for weekly testing, while biweekly and not testing had slightly higher R_{0N}. This shows that testing does affect disease transmission by lowering R_{0N} and that weekly testing has a higher effect on reducing transmission than biweekly. In combination with the green traffic light, weekly testing successfully decreases the time before the amount of recovered individuals is saturated. On average, the transmission was saturated and mitigated at day 70, with about 30 infected individuals per iteration.

4.3.4 The effect of initial disease state on r_0

Furthermore, the effect of choosing different initial disease states was investigated by comparing which effect patient zero's disease state had on overall disease transmission. The R_{00} was calculated for multiple iterations of disease transmission where patient zero was introduced as either symptomatic or asymptomatic. These calculations are shown in subsection A.3.3, in Appendix A. Figure 4.19 shows the proportion of r_{00} for the two initial disease states. For individuals with the initial disease state of symptomatic, the expected R_{00} was calculated to 3.16. In contrast, the expected R_{00} for asymptomatic patient zeros was 1.01.

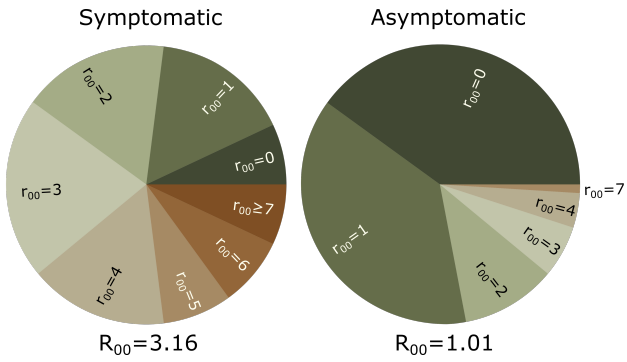


Figure 4.19: The frequency of r_{00} calculated for symptomatic and asymptomatic individuals. The expected value, R_{00}, is shown at the bottom of the pie charts. The frequencies are gathered from 100 iterations per initial disease state.

At first, the goal of investigating the effect each initial disease state had on R_{00} was to compare it to an expected ratio of R_0 for asymptomatic and symptomatic based on the disease parameters used for infectiousness and duration. However, this investigation led to a deeper understanding of the mechanisms of how disease transmission in primary school

networks. Algorithm 1 describes the mechanisms behind how transmission is modelled for this framework. A first-order approximation of the probability of transmission for low values of $w_{ij} * I_p$ can be seen in Equation 4.4.

$$p_{inf} = 1 - (1 - 0.001)^{w_{ij} * I_p} \approx 0.001 * w_{ij} * I_p \tag{4.4}$$

Following this approximation, subsection A.3.3 describes that the calculation of R_0^A and R_0^P should be as described in Equation 4.5, where an increase in $R0$ is proportional to an increases in I_p and $days_p$.

$$\frac{R_0^A}{R_0^{PS}} = \frac{I_a * days_A}{I_p * days_P} \tag{4.5}$$

The predicted ratio of R_0^A/R_0^{PS} was then calculated to 0.154, as described in subsection A.3.3. This indicates that each symptomatic individual is expected to infect 0.154 times more individuals than an asymptomatic individual. However, the ratio between the R_{00} of asymptomatic and symptomatic seen for the simulated model was calculated to 0.320.

This indicates that the assumption that the values of $w_{ij} * I_p$ are low is not valid and that the strength of the interactions between individuals is high. Therefore, Equation 4.4 is not fulfilled, and the ratio of R_0^A/R_0^{PS} can not be compared to the ratio generated by simulation.

Chapter 5

Discussion

This thesis aimed to create a model for primary school interactions and use it to investigate the transmission of SARS-CoV-2 in such an environment. By identifying key features and characteristics of Barrat et al.'s [1] interaction data set, a framework to produce networks of the same environment has been proposed. Section 5.1 will go into detail on the evaluation of the model in relation to the empirical data before section 5.2 addresses the assumptions of both the model and disease transmission simulated. Finally, section 5.3 aims to discuss and provide some advice for future pandemics based on the results of the disease transmission.

5.1 Evaluation of the model

The analysis of the empirical network revealed a heterogeneous small-world network with a high clustering coefficient and hierarchic structure, see section 4.1. Through matching interaction and degree distributions, the proposed novel framework for interactions in primary schools aims to reproduce these characteristics.

Since the approach for creating the model depended on matching degree distributions for the interaction layers of the primary school, the degree distribution is similar for the model and the empirical networks. Of notice, however, is that nodes with intermediate degrees have a higher frequency in the simulated model than in the empirical network. This trend was further highlighted by the increased intensity of off-diagonal interactions revealed by the heatmap of the simulated model. Furthermore, as discussed in subsection 4.2.2 the degree distribution is highly variable both between the two empirical days and for each day generated by the model, especially for the frequency of the high degree nodes. Since the frequency of these degrees is low, stochastic changes are more apparent. Overall, the degree distribution of the model closely resembles that of the degree distribution for the empirical networks.

Similarly, the interaction distribution of the model matches the interaction distribution of the empirical network. Few interactions are long-lasting, while most interactions are brief. This division makes sense in a school environment where individuals may have many brief encounters in the schoolyard while having more prolonged interactions with other individuals in the classroom.

Although some of the topological parameters calculated for the random network models ER, WS and BA matched the empirical interactions better than the parameters for the simulated model, the simulated model better represents a network for disease transmission in a primary school environment. The average degree and network density were more accurately predicted for the random networks. This is most likely due to the mechanisms of how they are generated. Since the parameter of the random networks were tweaked so that the network had the same number of degrees and nodes as the empirical network, the average degree and network density were easily replicated.

Meanwhile, both ER and WS failed to capture the high heterogeneity of the empirical network, while all the generated random networks were unable to replicate the high clustering coefficient and correct degree distribution. Since these exact parameters are the topological network parameters that affect transmission the most, the simulated model gives a more accurate network for disease transmission. High heterogeneity allows the disease to disperse throughout the network and amplifies the transmission as soon as hubs are infected. Dezső and Barabási also describes that scale-free networks such as the BA model do not have an epidemic thresholds due to the influence of large hubs that can help drive transmission [50]. Therefore, having a similar heterogeneity and degree distribution as the empirical network is essential for accurate disease transmission. Furthermore, clustering can help amplify transmission in a group of individuals. Therefore, the proposed framework captures the essential network topological parameter relevant for disease transmission.

Furthermore, the simulated network's lower average degree and average shortest path compared to the empirical network may be due to the presence of more off-diagonal interactions. Since the framework focused on creating a hierarchical structure with high clustering, primarily through matching degree and interaction distributions, one side effect was that other measurements did not completely match the model. Furthermore, the higher off-diagonal interactions in the model compared to the empirical network may have contributed to a decreased average path length. In turn, this could cause higher disease transmission on the model compared to on the empirical network.

The difference in the model's modularity compared to the empirical network's modularity is of more significant concern. Since one of the key characteristics of the empirical data is the hierarchical division of classes and grades, the model aimed to reproduce the high modularity of the empirical data. However, the modularity in the simulated model was halved compared to the empiric network. Romano et al. highlights that most interactions and thus transmission in social networks occurs for intermediate modularity [51]. In networks with high modularity, repeated interactions occur within each module and increase the probability of infections within the module. Furthermore, since contact between two infected individuals does not cause further transmission, the number of infectious contacts

is reduced in highly modular networks. Therefore, the lower modularity of the simulated model might contribute to an increase in disease transmission compared to transmission on the empirical networks. This is also a result of increased off-diagonal interactions, and helps disease spread faster on the model created by the framework in this thesis compared to the empirical network.

Furthermore, the framework aims to be general so that it can easily implement models of new schools of different sizes by having the possibility to input the number of grades, classes, and individuals to be simulated for a model. The investigation of model sensitivity to different input arguments, as described in subsection 4.2.4, is an attempt to see whether or not this generalisation is accomplished. With the same functions for creating biases and p_{ij}-values as extracted from the five-grade, two-class school system investigated by Barrat et al. [1], a hierarchic structure was visible for the different inputs shown in Figure 4.14. Since no high-resolution RFID data of primary school interactions in different sized schools with other grade and class divisions was available when writing this thesis, the model sensitivity results cannot be verified against real-life interaction data. However, the hierarchic structure follows the trend one would expect given the strict schedules of primary schools, making individuals more likely to interact with other individuals in the same class and grades during teaching activities.

The disease transmission simulated on the model and empirical networks as described in subsection 4.3.1 confirmed that disease transmission occurs faster and reaches a higher amount of infected individuals in the model compared to transmission run on the empiric network. This further confirms the effect of increased off-diagonal interactions and lower modularity in the simulated model. Thus, the lower the modularity of the model significantly affected and increased disease transmission.

The comparison also revealed higher transmission for models where the new interactions are generated each day. Since disease transmission on the empirical network alternating between day one and day two increased compared to running transmission on one of the days, the new contact generation each day helps amplify transmission. Like real-life interactions, the interactions of the model are updated every day, and thus the dynamic model with a network changing every day has the highest transmission. This further illustrates the importance of predicting disease transmission on dynamic networks and justifies the daily interaction generation of the model.

Furthermore, disease transmission on the model using the traffic light levels and weekly testing behave as expected for the model created in this thesis. The higher disease transmission of, for instance, the green level, compared to the red level, shows that the models' interactions are vulnerable to different parameter changes in a predictable way. However, one should keep in mind that this model only introduces one patient zero; thus, the system may behave entirely different for continued outside interaction with community spread. Therefore, it is important to keep the model's assumptions in mind when trying to extrapolate meaningful results from the different disease mitigation strategies.

Finally, the investigation of the expected ratio of R_0 between asymptomatic and symptomatic initial disease states revealed that the percent increase in R_0 is not proportional to the increase in I_p and I_{inf} for strong interactions in the model. High weighted interac-

tions can then give a different percent increase; for instance, I_p having an 80% increase may only increase R_0 by 30%. This highlights the importance of being specific when discussing increases in infectivity and separating between different disease parameters. Similarly to the strong interactions seen for some individuals in primary schools, this tendency may also be true for other highly weighted interactions, such as those between household members.

5.2 Addressing the assumptions of the model and its limitations

As with any model that is only an abstraction and aims to represent some features of the real-life phenomenon, all results should be interpreted within the context in which a model is built. A model seldom captures all the characteristics of a real-life phenomenon, and assumptions help limit the model to specific situations. This section will now go into further detail on how realistic the different key assumptions of the primary school model are and address the model's limitations.

Firstly, as the study by Barrat et al. [1] was conducted in a primary school in France, the model's applicability elsewhere may be restricted. The schedule at the primary school seen in Figure 2.7 describes an eight-hour-long day, with a two-hour lunch break at 12 pm. In contrast, Norwegian primary schools usually have five to six-hour days with around 45 minutes to one-hour lunch break [52]. Similarly, primary schools in Norway usually consist of grades one to seven, in contrast to the five-grade system seen in French primary schools. Since the model is based on interactions recorded for a typical primary school in France with two classes per grade and longer days, any input outside of eight-hour days and five grades with two classes each should be further investigated. The lunch breakdown and additional sublayer are also specific to the lunch break at the Primary school in Lyon and should be excluded when trying to model the interactions of other schools.

The model allows input of different length days and alternating the number of grades and classes; however, this functionality has not been appropriately investigated due to a lack of high-resolution interaction data for primary schools. Therefore, even though subsection 4.2.4 on sensitivity showed promising results regarding the use of different input parameters, this needs to be further investigated and compared to additional experimental data before the model is used for different school systems.

The lack of data also contributed to the model being based on data collected for two consecutive days at a small school with 242 individuals. This further contributes to limiting the model to the specific school. Because of this, it is possible that overfitting of the degree and interaction distributions occurred. Since there is no available data to compare the results to, the possibility of overfitting can neither be definitively confirmed nor denied in this thesis. However, the results interpreted from this model should keep this limitation and small sample size in mind.

Furthermore, even though the interaction data showed high hourly variation due to lunch breaks and class, the model was created by accumulating homogeneous hour-by-hour in-

teractions for each day. Therefore, an assumption of the model is that even though hourly differences exists, generating homogeneous hours into a daily network will sufficiently capture the same characteristics. This was shown to be true for degree and interaction distributions, however, implementations of more high detailed hour-by-hour contacts may have helped replicate the modularity.

Despite the fact that implementing detailed hour-by-hour contacts might have improved some of the general network topological parameters of the model for the specific empirical network, it would sacrifice the usability of the model. The use of hour-by-hour generation of a day was initially implemented so that the model could be used to simulate shorter school days and tailor the model to the schedule of a given school. Furthermore, using the same set of rules for each hour generated at a school should allow the model to be general enough to be run at different schools regardless of their daily schedule. Therefore, even though the data allowed for detailed hour-by-hour investigation, implementing it into a framework like the one presented in this thesis would further limit its usage to the specific primary school of Lyon.

Following the same logic, the model assumes that the interactions in primary schools are divided into three different layers; off-diagonal (including lunch), grade, and class. Furthermore, it was assumed that most interactions within each layer were of a distinct distribution. This is, of course, an oversimplification, and it does not capture all the dynamics of a real-life interaction network. If more data over many days was available, it could be interesting to further investigate other possible layers. For instance, investigate interactions in social cliques, or friend groups, and see how they impact disease transmission.

The model also assumes that each interaction between two individuals is motivated by both the bias the two individuals have for interacting with any other person, as well as the similarity (P_{ij}) between the individuals. In practice, this means that each individual has a set ”social” bias upon generation. The implementation of bias was inspired by the significant positive relationship seen between the degrees of nodes on day one and day two. The assumption of bias thus seems to hold, which makes sense considering individuals generally have a different degree of sociability. Furthermore, the premise that pupils with more in common interacted more with one another came from the hierarchical division of the interactions into layers by investigating heatmaps. Similarly, this assumption is also reasonable for real-life interactions since children in the same class are ”forced” to interact and become familiar through teaching activities. Furthermore, having the probability of two individuals interacting be determined by how social and similar they are, is also reasonable to assume in normally sized schools. Limitations to these assumptions could, for instance, be in schools in small villages where classes or even grades may have smaller sizes, consisting of, for example, ten individuals. Then, it is more likely that individuals interact more with other grades and classes. This should be kept in mind if the framework of this model is considered for running simulations in schools with fewer pupils.

Following the assumption of interactions being created by a combination of bias and similarity, it is also assumed that each individual has a different bias for interacting with individuals across the different layers. The reasoning behind this assumption is that there is a significant difference in how individuals interact within the layers observed, and using the

same bias for all interactions caused degree and interaction distributions to not accurately match the empirical network. However, the idea that some individuals of the same class are less ”shy” and approachable to one another than how they interact with individuals outside of their class is not entirely unreasonable. Therefore, the model presented by this framework is only valid for schools where there is a division between individuals of different grades interacting. As mentioned earlier, this could, for instance, occur in smaller schools where individuals interact more with one another across grades.

The model also assumes that accumulating the 20-second interactions into weighted daily interactions is a detailed enough approach to capture the high-resolution data set collected by Barrat et al. [1]. By accumulating all interactions of one day, the model disregards the possibility of increased transmission for continuous interactions compared to other interactions that may be on and off during the day. One could, for instance, argue that one long subsequent interaction has higher transmission than the accumulation of many short interactions of the same length. Folkehelseinstituttet defined that a close contact occurred between individuals that have either had direct physical contact with an infected individual or their excretions or been within 2-meter proximity of the infected individual for more than 15 minutes [53]. The accumulation of all interactions to daily weighted contacts thus removes this temporal aspect. However, the data only measures face-to-face contact within 1-1.5 meter distance. Therefore, more spatial data on when direct contact occurs between two individuals need to be collected before separating between direct and long-lasting close proximity interactions can be distinguished.

Since no spatial dimension is present in the model, the model also does not directly address direct transmission that may occur due to, for instance, direct contact or the fecal-oral route. Instead, the model assumes that the environment of students who interact a lot with each other is similar, capturing some of the contact transmission that may occur. This assumption may overshadow indirect transmission routes between individuals that do not interact, for instance, when a susceptible individual touches a doorknob that an infected individual has previously contaminated. Incorporating the spatial position of each individual at all times thus may help capture this indirect transmission. Meiksin incorporates this indirect transmission in a differential equation extended SIR model, and future expansion of this framework may aim to include indirect contact similar to their approach [54].

The transmission of the disease is based on assumptions, with a set stochastic transmission rate. Since the COVID-19 pandemic is still recent, exact disease transmission parameters are yet to be determined. Therefore, the model called for an approximation of the infectiousness of the different disease states. This led to the assumption that transmission increases as a function of exposure. The infectiousness per contact was decided by changing p_0 while running disease transmission until a R_0 of around 3 was reached. The reasoning for choosing to model an R_0 of this value was that disease transmission below one would not yield enough transmission to observe the effect of implementing different disease mitigation strategies. Conversely, a R_0 of higher than three would cause the transmission to reach the saturation of recovered individuals too quickly, making it challenging to observe the effect of the different strategies. In practice, the R_0 in schools may be lower than one, while community transmission is higher. Take, for instance, a child that infects 0.6 individuals at school, one at home, and 0.4 in other environments. The child would have

a total R_0 of 2. However, evaluating the model in an isolated situation like the framework presented in this thesis does not allow for testing of scenarios like this, and thus a set R_0 was introduced. These effects should be further investigated to increase the accuracy of this model. They can, for instance, be incorporated into the agent-based community spread model of the NTNU COVID-19 Taskforce.

Furthermore, the time spent in each disease state was deterministic. This approach was chosen due to the stochastic nature of the probability of transmission and interactions, making it easier to follow how the disease would spread in the network. However, this approach is not realistic compared to real life, where there is individual variance in how long individuals spend in different disease states. Using Weibull [49] to determine disease generation time instead of having them predetermined might have increased the accuracy of transmission on the model. Another option is to use Poisson distribution to determine the length of each disease state, as Voigt et al. did for the NTNU COVID-19 Taskforce model [12].

In addition, the environment of the individuals is assumed to be the same across all days where disease transmission is simulated. However, the environment may also impact how the disease is transmitted. Liu et al. emphasises that COVID-19 transmission changes with seasons, with the colder winter season having increased transmission [10]. Therefore further work on this model should aim to reproduce these seasonal changes in the environment, especially if the model is simulated over multiple months.

All disease transmission was run on an isolated system with only one patient zero generated each time. This was mainly done to test transmission on the model alone before implementing it into a community spread model such as the NTNU COVID-19 Taskforce model. Therefore, the model assumes that no new cases are introduced after patient zero. While this may be true for some isolated boarding schools, most primary schools do not house their students, and thus the disease may be transmitted from the community. As of now, the model only addresses transmission from one patient zero; however, functions within the model do allow for multiple patients zero.

The final assumption is that the role of teachers in disease transmission is minimal. Since only ten teachers were present in the network, and their interactions were limited to pupils in their class, no additional teacher interaction layer was added to the model. First off, this completely rules out any transmission between teachers during their breaks. Secondly, in Norwegian primary schools, the role of the teacher seems to differ from one school to another. One teacher may be teaching multiple classes in some schools, while other schools have one specific teacher per class. A mobile teacher moving between classes may have an essential role in disease transmission, as they may contribute to spreading the disease across different modules. Therefore, the effect teachers have on transmission should be addressed in future works, especially if the school allows for teachers to teach various classes.

5.3 The impacts of the result

Similarly to Angelerio, the *Protomedicus* of Sardinia, being tasked with implementing NPIs to control disease transmission of the Bubonic plague in 1682, the WHO was recently in charge of advising governments on which strategies and measures to implement during the COVID-19 pandemic. This section will now try and provide some advice based on the result of the investigation of the effect of some of the measures implemented in across primary schools in Norway in order to mitigate COVID-19 transmission. Keeping in mind the evaluation and assumptions of the model, the following advice may contribute to the strategy of disease mitigation in primary schools when the next respiratory virus emerges.

Winje et al. describes that implementing mitigation measures such as testing and the implementation of the traffic light model in Norwegian schools limited overall transmission related to primary schools. Similarly, all levels of the traffic light model successfully reduced the overall disease transmission compared to unhindered disease transmission for the simulations run in this thesis. Therefore, implementing these seems to be beneficial for reducing the transmission of respiratory viruses. Furthermore, the simulated transmission on the model seemed to indicate that moving from green to yellow traffic light levels causes significantly less transmission than moving from yellow to red. Since the cost of moving from yellow to red level includes halving class sizes [28], more resources such as teachers and classrooms have to be made available. Therefore, high transmission in the population should be present to justify the transition by weighing the cost versus the effect of moving from yellow to red traffic lights. In turn, if the resources for testing are not available, alternating between the green and yellow levels of the traffic light model may help reduce transmission in primary schools.

Furthermore, weekly testing decreased transmission on the green traffic light level. Therefore, if rapid antigen tests are readily available, testing and isolating positive cases while in green traffic light levels were effective at disease mitigation in the model presented in this thesis. Keep in mind that the predicted transmission decrease is dependent on continued testing even after transmission has ended. In areas with less community spread of the disease, one can also consider implementing biweekly testing as it was also found to reduce transmission. Further investigation of combinations of weekly and biweekly testing with different traffic light levels may also provide further advice on combinations of strategies.

However, since the data collected by Barrat et al. [1] and investigated in this was tailored to a French primary school, more research on high-frequency interaction data in Norwegian primary schools should be collected and compared to the model before the given advice can be considered. Of most importance is the fact that all measures successfully mitigated disease transmission. Thus, implementing the strategies presented in this thesis indicates that there is a premise for keeping schools open during pandemics.

Chapter 6

Conclusion/Outlook

This thesis provides a novel framework for generating interactions in primary schools as well as running disease transmission on said interactions. First, an extensive analysis of the empirical data collected by Barrat et al. [1] was conducted, laying the base for the interaction dynamics of the framework. Then, using both disease state parameters from Ferretti et al. [49] and estimates, the transmission of COVID-19 was simulated on the model produced by the framework.

The extensive analysis included an investigation of selected topological parameters and the relationship between degree distribution on two days, in addition to a comparison of the accumulated network of the empirical data with the random network models ER, WS and BA. These analyses revealed that the network has a high clustering coefficient and network heterogeneity and displays a small-world property by having a low average shortest path length. The interactions of the network were then divided into the layers off-diagonal, grade, and class by investigating a heatmap representation of an adjacency matrix of the empirical network.

The degree and interaction distribution of the network was then investigated, revealing a highly complex distribution that was further divided into the layers previously identified. Using a combination of the degree and interaction distribution allowed specific biases and likelihoods of interaction to be generated for each interaction layer.

The evaluation of the simulated model in comparison to the empirical network revealed that the generated model had lower modularity and average degree compared to the empirical network. In addition, the heatmap of the simulated model showed more high-intensity off-diagonal interactions than what was seen for the empirical network. Therefore, it is believed that the increased off-diagonal interactions is a consequence of the lower modularity and lower average degrees. Furthermore, the degree and interaction distribution were similar for the simulated and empirical networks, and an investigation of model sensitivity revealed that the model displays a similar hierarchical structure for different inputs

of the school structure.

Then, disease transmission of COVID-19 was run on both days of the empirical network, including a version that alternated between the two days and compared to transmission on dynamic and static models. The transmission was higher for both the static model, only containing the same network across all 100 days, and the dynamic model compared to the empirical network. This increased transmission was primarily attributed to the simulated model's higher number of off-diagonal interactions.

Then, the effect of imposing different traffic light levels as well as testing strategies were investigated. The green level decreased transmission in comparison with the transmission on a network with no measures implemented. Furthermore, implementing a yellow traffic light level significantly decreased transmission, lowering the average number of infected by 50 over 100 iterations. Finally, the red traffic light level further decreased transmission compared to the yellow level. Weekly testing was the most efficient at reducing transmission; meanwhile, biweekly testing also reduced transmission compared to transmission on an untested population.

6.1 Future work

First, as a stand-alone model, the primary school model runs quickly for transmission on one school at a time. However, for it to be implemented into a broader model, such as the ABM by NTNU COVID-19 Taskforce, further optimisation of the code is needed to increase run-time. Potential routes for this optimisation include decreasing the time it takes to generate interactions, as this was the code that took the longest to run.

Since the model was based on two days of interaction, to make the model more accurate, the first order of business should be to conduct more RFID interaction measurements studies for more extended periods of time and in differently structured primary schools. For instance, one could follow the interactions at a given school for 100 days to extract more detailed modules in class for more resolution and the addition of another layer. In addition to recording the interactions, one could also aim to collect spatial data on where individuals have interacted or moved throughout the day. This can help detect or predict the possibility of indirect transmission.

In addition, since the framework of the model was created so that it could be implemented into the NTNU COVID-19 Taskforce model, it could be interesting to see what effect the detailed model can contribute with. Furthermore, comparing the impact the size of a school has on disease transmission on a model generated by the detailed framework of this thesis with the transmission in the abstract way primary schools are presented in the agent-based NTNU COVID-19 Taskforce model may yield interesting results. For instance, one could observe whether the model's modularity created by the framework presented in this thesis hindered or increased transmission in contrast to the modularity of the primary schools simulated by the Taskforce model. This could further help adjust the model of this framework to predict more accurate interactions.

This thesis only had the scope and time to investigate the effect of the traffic light model

and biweekly and weekly testing. However, there are numerous other strategies and possibilities that can be explored by using the framework presented in this thesis. For instance, it could be possible to decrease the time individuals are in the school and see how this may affect transmission. Furthermore, investigating the effect of shifting the schedule for when different grades attend school on disease transmission could be interesting. Implementing different traffic light levels and testing strategies also allows for more combinations to be investigated. For instance, could the result of biweekly or even monthly testing in addition to a yellow or red traffic light level help provide answers as to how testing affects transmission. Finally, since vaccinations are now available for children aged 5-11, investigating the effect of different percentages of individuals at the school being vaccinated may also provide interesting results on herd immunity in primary schools.

Bibliography

[1] Alain Barrat, Ciro Cattuto, Alberto. E. Tozzi, Philippe Vanhems, and Nicholar Voirin. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection*, 20(1):10–16, 2014. doi:10.1111/1469-0691.12472.

[2] Raffaella Bianucci, Ole Jørgen Benedictow, Gino Fornaciari, and Valentina Giuffra. Quinto tiberio angelerio and new measures for controlling plague in 16th-century alghero, sardinia. *Emerging Infectious Diseases*, 19(9):1478–1483, 2013. doi:10.3201/eid1909.130311.

[3] Kenrad E. Nelson and Carolyn M. Williams. *Infectious Disease epidemiology: Theory and Practice*. Jones & Bartlett Learning, Burlington, 3rd edition, 2014.

[4] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, and Wei et al. Zhang. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798):270–273, 2020. doi:10.1038/s41586-020-2012-7.

[5] Alessandra Zappa, Antonella Amendola, Luisa Romanò, and Alessandro Zanetti. Emerging and re-emerging viruses in the era of globalisation. *Blood Transfus*, 7(3): 167–171, 2009. doi:10.2450/2009.0076-08.

[6] Badar Nadeem Ashraf. Economic impact of government interventions during the covid-19 pandemic: International evidence from financial markets. *Journal of Behavioral and Experimental Finance*, 27:100371, 2020. doi:10.1016/j.jbef.2020.100371.

[7] Anant Kumar and K. Rajasekharan Nayar. Covid 19 and its mental health consequences. *Journal of Mental Health*, 30(1):1–2, 2021. doi:10.1080/09638237.2020.1757052.

[8] Camila Saggioro De Figueiredo, Poliana Capucho Sandre, Liana Catarina Lima Portugal, Thalita Mázala-De-Oliveira, Luana Da Silva Chagas, and Ícaro et al. Raony. Covid-19 pandemic impact on children and adolescents' mental health: Biological, environmental, and social factors. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 106:110171, 2021. doi:10.1016/j.pnpbp.2020.110171.

[9] The World Health Organisation (WHO). Covid-19 dashboard, April 2022. [Internet]. Retrieved 17.04.2022 from `https://covid19.who.int/`.

[10] Xiaoyue Liu, Jianping Huang, Changyu Li, Yingjie Zhao, Danfeng Wang,

Zhongwei Huang, and Kehu Yang. The role of seasonality in the spread of covid-19 pandemic. *Environmental Research*, 195:110874, 2021. doi:https://doi.org/10.1016/j.envres.2021.110874.

[11] Faïçal Ndaïrou, Iván Area, Juan J. Nieto, and Delfim F.M. Torres. Mathematical modeling of covid-19 transmission dynamics with a case study of wuhan. *Chaos, Solitons & Fractals*, 135:109846, 2020. doi:10.1016/j.chaos.2020.109846.

[12] André Voigt, Nikolay Martyushenko, Emil Karlsen, Martina Hall, Kristen Nyhamar, Stig William Omholt, and Eivind Almaas. Containing pandemics through targeted testing of households. *BMC Infectious Diseases*, 21(1), 2021. doi:10.1186/s12879-021-06256-8.

[13] Erik Cuevas. An agent-based model to evaluate the covid-19 transmission risks in facilities. *Computers in Biology and Medicine*, 121:103827, 2020. doi:10.1016/j.compbiomed.2020.103827.

[14] Cécile Viboud, Pierre-Yves Boëlle, Simon Cauchemez, Audrey Lavenu, Alain-Jacques Valleron, Antoine Flahault, and Fabrice Carrat. Risk factors of influenza transmission in households. *British Journal of General Practice*, 54(506):684–689, 2004.

[15] Lin T Brandal, Trine S Ofitserova, Hinta Meijerink, Rikard Rykkvin, and et al. Minimal transmission of sars-cov-2 from paediatric covid-19 cases in primary schools, norway, august to november 2020. *Eurosurveillance*, 26(1), 2021. doi:10.2807/1560-7917.es.2020.26.1.2002011.

[16] Wei Xu, Xue Li, Marshall Dozier, Yazhou He, Amir Kirolos, Zhongyu Lang, Catherine Mathews, Nandi Siegfried, and Evropi Theodoratou. What is the evidence for transmission of covid-19 by children in schools? a living systematic review. *Journal of Global Health*, 10(2), 2020. doi:10.7189/jogh.10.021104.

[17] Edward Goldstein, Marc Lipsitch, and Muge Cevik. On the effect of age on the transmission of sars-cov-2 in households, schools, and the community. *The Journal of Infectious Diseases*, 223(3):362–369, 2021. doi:10.1093/infdis/jiaa691.

[18] Yen-Chin Liu, Rei-Lin Kuo, and Shin-Ru Shih. Covid-19: The first documented coronavirus pandemic in history. *Biomedical Journal*, 43(4):328–333, 2020. doi:10.1016/j.bj.2020.04.007.

[19] Andrew G. Harrison, Tao Lin, and Penghua Wang. Mechanisms of sars-cov-2 transmission and pathogenesis. *Trends in Immunology*, 41(12):1100–1115, 2020. doi:10.1016/j.it.2020.10.004.

[20] S. A. Meo, A.M Alhowikan, T. Al-Khlaiwi, I.M Meo, D.M Halepoto, M. Iqbal, A.M Usmani, and N. Ahmed. Novel coronavirus 2019-ncov: prevalence, biological and clinical characteristics comparison with sars-cov and mers-cov. *Eur Rev Med Pharmacol Sci*, 24(4):2012–1019, 2020. doi:10.26355/eurrev˙202002˙20379.

[21] L. F. Wang and B. T. Eaton. *In: Childs, J.E., Mackenzie, J.S., Richt, J.A. (eds) Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*, volume 315, book section Bats, Civets and the Emergence of SARS, pages 325–344. Springer, Berlin, Heidelberg, 2007. doi:10.1007/978-3-540-70962-6˙13.

[22] Aleksandra Synowiec, Artur Szczepański, Emilia Barreto-Duran, Laurensius Kevin Lie, and Krzysztof Pyrc. Severe acute respiratory syndrome coronavirus 2 (sars-cov-

2): a systemic infection. *Clinical Microbiology Reviews*, 34(2):e00133–20, 2021. doi:10.1128/CMR.00133-20.

[23] Fei Xiao, Meiwen Tang, Xiaobin Zheng, Ye Liu, Xiaofeng Li, and Hong Shan. Evidence for gastrointestinal infection of sars-cov-2. *Gastroenterology*, 158(6):1831–1833.e3, 2020. doi:10.1053/j.gastro.2020.02.055.

[24] Ewen Callaway. The coronavirus is mutating - does it matter? *Nature*, 585:174–177, 2020. doi:10.1038/d41586-020-02544-6.

[25] Elisabeth Mahase. Covid-19: What do we know about the delta omicron recombinant variant? *BMJ*, page o792, 2022. doi:10.1136/bmj.o792.

[26] Koronakommisjonen. Norges offentlige utredninger (nou): Myndighetenes håndtering av koronapandemien, April 2021. [Internet]. Retrieved 02.05.2022 from `https://www.regjeringen.no/no/dokumenter/nou-2021-6/id2844388/`.

[27] Utdanningsdirektoratet. Utdanningsspeilet 2020: Hvordan preget koronautbruddet våren 2020 barnehager, skoler og fagopplæring?, 2021. [Internet]. Retrieved 02.05.2022 from `https://www.udir.no/tall-og-forskning/publikasjoner/utdanningsspeilet/utdanningsspeilet-2020/del-2/`.

[28] Utdanningsdirektoratet. Beredskap: veileder om smittevern for skoletrinn 1–7, April 2022. [Internet]. Retrieved 17.04.2022 from `https://www.udir.no/kvalitet-og-kompetanse/sikkerhet-og-beredskap/informasjon-om-koronaviruset/smittevernveileder/veileder-om-smittevern-for-skoletrinn-17/smitteforebyggende-tiltak/`.

[29] Tim Peto, Dominic Affron, Babak Afrough, Anita Agasu, Mark Ainsworth, Alison Allanson, Katherine Allen, Collette Allen, Lorraine Archer, Natasha Ashbridge, et al. Covid-19: Rapid antigen detection for sars-cov-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. *EClinicalMedicine*, 36:100924, 2021. doi:10.1016/j.eclinm.2021.100924.

[30] Wei Duan, Zongchen Fan, Peng Zhang, Gang Guo, and Xiaogang Qiu. Mathematical and computational approaches to epidemic modeling: a comprehensive review. *Frontiers of Computer Science*, 9(5):806–826, 2015. doi:10.1007/s11704-014-3369-2.

[31] Helen Abbey. An examination of the reed-frost theory of epidemics. *Human Biology*, 24(3):201–233, 1952.

[32] Jonathan L. Gross, Jay Yellen, and Mark Anderson. *Graph Theory and Its Applications*, pages 2–50. Chapman and Hall/CRC, Boca Raton, Florida, 3rd edition, 2018. doi:10.1201/9780429425134.

[33] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016. URL `http://barabasi.com/networksciencebook/`.

[34] D. Auber, Y. Chiricota, F. Jourdan, and G. Melancon. Multiscale visualization of small world networks. In *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714)*, pages 75–81, 2003. doi:10.1109/INFVIS.2003.1249011.

[35] Rinku Jacob, K. P. Harikrishnan, R. Misra, and G. Ambika. Measure for degree heterogeneity in complex networks and its application to recurrence network analysis. *Royal Society Open Science*, 4(1):160757, 2017. doi:10.1098/rsos.160757.

[36] Rogier Noldus and Piet Van Mieghem. Assortativity in complex networks. *Journal of Complex Networks*, 3(4):507–542, 2015. doi:10.1093/comnet/cnv005.

[37] Francisco Aparecido Rodrigues. *Network Centrality: An Introduction*, pages 177–196. Springer International Publishing, 2019. doi:10.1007/978-3-319-78512-7˙10.

[38] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006. doi:10.1073/pnas.0601602103.

[39] E. Ronald Walpole, Raymond H. Myers, , Sharon L. Myers, , and Keying Ye. *Probability & statistics for engineers & scientists*. Pearson, Harlow, England, 9th edition, 2016.

[40] Virinchi Srinivas and Pabitra Mitra. *Link prediction in social networks: role of power law distribution*. Springer, Manhattan, New York City, 2016. doi:10.1007/978-3-319-28922-9.

[41] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002. doi:10.1103/revmodphys.74.47.

[42] C. M. Macal and M. J. North. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162, 2010. doi:10.1057/jos.2010.3.

[43] W. John Edmunds, C. J. O'Callaghan, and D. J. Nokes. Who mixes with whom? a method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society B: Biological Sciences*, 264 (1384):949–957, 1997. doi:10.1098/rspb.1997.0131.

[44] Il-Chul Moon and Kathleen M. Carley. Modeling and simulating terrorist networks in social and geospatial dimensions. *IEEE Intelligent Systems*, 22(5):40–49, 2007. doi:10.1109/mis.2007.4338493.

[45] Uri Wilensky and William Rand. *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. MIT Press, Cambridge, Massachusetts, 2015.

[46] Sosiopatterns. Organisation, 2022. [Internet]. Retrieved 05.04.2022 from `http://www.sociopatterns.org/`.

[47] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009. ISBN 1441412697.

[48] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003. doi:10.1101/gr.1239303.

[49] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, 2020. doi:10.1126/science.abb6936.

[50] Zoltán Dezső and Albert-László Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65(5), 2002. doi:10.1103/physreve.65.055103.

[51] Valéria Romano, Mengyu Shen, Jérôme Pansanel, Andrew J. J. Macintosh, and Cédric Sueur. Social transmission in networks: global efficiency peaks with intermediate levels of modularity. *Behavioral Ecology and Sociobiology*, 72(9), 2018. doi:10.1007/s00265-018-2564-9.

[52] Utdanningsdirektoratet. Fag- og timefordeling og tilbudsstruktur for kunnskapsløftet udir-1-2021, June 2021. [Internet]. Retrieved 01.05.2022 from `https://www.udir.no/regelverkstolkninger/opplaring/innhold-i-opplaringen/udir-1-2021/`.

[53] Folkehelseinstituttet. Definisjoner: Definisjon av mistenkt og bekreftet tilfelle av covid-19, av nærkontakt, grunnvaksinert og av risikogruppe, 2022. [Internet]. Retrieved 12.05.2022 from `https://www.fhi.no/nettpub/coronavirus/testing/definisjoner/`.

[54] A. Meiksin. Dynamics of covid-19 transmission including indirect transmission mechanisms: a mathematical analysis. *Epidemiology and Infection*, 148:1–30, 2020. doi:10.1017/s0950268820002563.

[55] Brita Askeland Winje, Trine Skogset Ofitserova, Ola Brønstad Brynildsrud, and et al. Comprehensive contact tracing, testing and sequencing show limited transmission of sars-cov-2 between children in schools in norway, august 2020 to may 2021. *Microorganisms*, 9(12):2587, 2021. doi:10.3390/microorganisms9122587.

Appendix A

Methods supplement

This section will go into more detail on the methodology of this thesis. It contains an overview of python modules used (section A.1), class descriptions (section A.2) and calculating expected r_{00} and r_{0N} (section A.3).

A.1 Python modules

An overview of the models used throughout this thesis is given in Table A.1.

Table A.1: Overview of Python modules used in this thesis. The modules are separated into built-in modules and installed modules [47].

Module	Brief description	Version
Itertools	Helps iterate over iterables	–
Math	Provides mathematical functions	–
Pickle	Allows for binary storage of objects	–
Random	Random generator	–
Os	Helps running operating system commands	–
Enum	Stores symbolic names to values	–
MatPlotLib	Data visualisation module for plotting	3.4.3
NetworkX	Allows for creation and analysis of networks with nodes and edges	2.6.3
NumPy	Useful for working with arrays	1.21.2
Pandas	Dataframe library	1.3.3
Sklearn	Contains supervised and unsupervised learning algorithms	1.0
SciPy	Contains algorithms of scientific functions	1.7.1
Seaborn	Data visualisation, based on matplotlib	0.11.2

A.2 Class description

The module contains the following classes: Person class (subsection A.2.1), Interaction (subsection A.2.2), Network class (subsection A.2.3), Analysis class (subsection A.2.4) and Disease_transmission (subsection A.2.5).

A.2.1 The Person class

The Person class has all the attributes explained in Table A.2. In network terminology, the Person object thus are the nodes, and any interactions between two Person objects are registered as edges. The attributes associated with each Person object both assist in network generation and disease transmission.

Table A.2: Type, description and examples of the attributes associated with Person objects

Attribute	Type	Description	Example
ID	int	A unique number attached to the person object ranging from 0 to the max capacity at each school	0
Sex	str	Denotes the sex of the person object. ”F” for female or ”M” for male	”F”
State	str	Disease state of individual. Possible states: S, E, Ia, Ip, Is, R	”S”
Grade	int	Number denoting the grade of the person object. Is limited by grades present at the school	1
Class_group	str	Describes the exact class a Person object is registered in. Could for instance be ”A” or ”B”	”A”
Age	int	The individual’s age	5
lunch_group	bool	True if the person object is in a grade lower than 4. Otherwise it is False. Can be changed based on schools and which grades interact most with each other	True
interactions	dict	Keeps a dictionary of each Interaction object that occur between this Person object and another. The key is the ID of Person object interacted with whilst the values are the Interaction object between the two Person objects (here denoted as their weight)	{4 : 1, 6 : 9, 17 : 14}
const_bias	float	Each Person object is initiated with a float value that denotes their bias for interacting with any other Person object	6.028
bias_grade	float	The bias this Person object has for interacting with other Person objects within the same grade	11.028
bias_class	float	The bias this Person object has for interacting with other Person objects within the same class	101.40
p-vector	dict	Dictionary that keeps track of the possibility that one Person object has to interact with another Person object. The key is the other Person object (here described by their ID), whilst the value is the p	{1 : 83.29 2 : 5.00 3 : 1.80}
Status	Disease_State	Denotes the disease state of an individual ranging from susceptible to infected and recovered	Disease_State.S
tested	bool	Returns true if an individual has tested positive for COVID-19	True
states	dict	Dict where the keys are Disease_States and values are the number of days an individual is in that state	{Disease_State.S : 1}
cohort	str	String that determines the cohort an individual belongs in	”1A1”
infected_on_day	int	Integer that denotes the day an individual was infected	3
recovered_on_day	int	Integer that denotes the day an individual was recovered	10
infected_by	Person	The Person object that infected this individual	Person p1

A.2.2 The Interaction class

Table A.3 describes all attributes within the Interaction class. The interaction class keeps track of interactions occurring between two Person objects and stores the number of interactions they have had with one another as a weight attribute.

Table A.3: Type, description and examples of the attributes associated with Interaction objects.

Attribute	Type	Description	Example
p1	Person object	Source Person object that is a part of the interaction	Person 0
p2	Person object	Target Person object that is a part of the interaction	Person 1
count	int	Keeps track of how many times two individuals have interacted	20

A.2.3 The Network class

Table A.4 describes all attributes within the Network class. The Network class is generated by students, which is a list of Person objects that attend a certain school, as well as all Interaction objects that exists between them.

Table A.4: Type, description and examples of the attributes associated with Network objects

Attribute	Type	Description	Example
parameter_list	list	List of parameters used in Person class to set the p_vector	[2, 0.04, 0.01, 0.01, 2, 0.4, 0.2, 11.5]
weights	dict	Dict with list of weights for setting threshold during interaction generation. Is only used when traffic light model is run	{'None': [100, 80, 60], 'G': [100, 80, 60], 'Y': [40, 30, 60], 'YC': [60, 40, 20], 'R': [20, 15, 10], 'RC': [50, 30,10]}
students	list	A list that contains all the Person objects that attend a certain school	[0, 1, 2.., 335, 336]
d	float	A number that helps scale the weights	0.01
available_grades	list	List of available grades in the school	[1, 2, 3, 4, 5]
available_classes	list	List of available classes in the school	[A, B]
graph	nx.Graph	A nx.Graph object that describes interactions in the network	nx.Graph with 336 nodes and 5500 edges
iteration_list	list	A list of daily nx.Graph objects	[graph_day1, graph_day2..]

A.2.4 The Analysis class

The Analysis class only has one attribute; network. This is a Network object containing, amongst other things, a nx.Graph representation of interactions. The analysis class consists of multiple methods for the analysis of the simulated model, and this is the class in which most results from section 4.2 and section 4.3 are generated.

A.2.5 The Disease_transmission class

In addition, the class Disease_transmission helps simulate transmission of disease on a Network object.

Table A.5: Type, description and examples of the attributes associated with Disease_transmission objects

Attribute	Type	Description	Example
network	Network	Network object containing a nx.Graph and students present in the network	Network(236,5,2)
stoplight	Traffic_light	Contains a Traffic_light entry from the traffic light model (green (G), yellow (Y), red(R)) or None	Traffic_light.G
graph	nx.Graph	nx.Graph object containing interactions between Person objects	nx.Graph with 236 nodes and 5500 edges
students	list	list of all person objects that are a part of Network	[Person 0, Person 1..., Person 236]
patient_zero	Person	The first person to introduce disease into the Network	Person 0
day_no	int	Keeps track of the current day Disease_transmission is run on	0
days	list	list of graphs generated for all days Disease_transmission is run on	[graph_day1, graph_day2]
p_0	float	Estimated Disease_transmission parameter	0.001
infectious_rates	dict	Dict containing Disease_states as keys and their relative infectiousness (in relation to the symptomaticstate) as values	{Disease_states.IAS: 0.1 , Disease_states.IP: 1.3 , Disease_states.IS: 1 }
Ias	float	Describes the percentage of individuals having a asymptomatic disease course	0.4
Ip	float	Describes the percentage of individuals having a presymptomatic/symptomatic disease course	0.6
positions	nx.spring_layout	A dictionary of positions keyed by node	dict over positions of nodes in graph according to spring layout

A.3 Calculating expected R_0

The expected value of R_0 used throughout this thesis was calculated using Equation A.1, where $P(r_0 = x)$ is the fraction of times r_0 has been x.

$$\sum_{x=0}^{N}(x * P(r_{00} = x)) \tag{A.1}$$

A.3.1 r_{00} of traffic light transmission

The r_{00} was measured for each traffic light level as shown in section 3.4.5. The following subsection will describe the calculations and present the data from which the expected R_{00}

was calculated.

The expected value, R_{00}, for **green light** is 2.36 and was calculated as follows after 1000 iterations as follows:

$$\sum_{x=1}^{N}(P(r_{00}=x)*x) = 0*\frac{6}{100}+1*\frac{26}{100}+2*\frac{24}{100}+3*\frac{19}{100}+4*\frac{20}{100}+5*\frac{5}{100} = \underline{\underline{2.36}}$$

The expected value, R_{00}, for **yellow light** is 1.9 and was calculated as follows after 1000 iterations as follows:

$$\begin{aligned}\sum_{x=1}^{N}(P(r_{00}=x)*x) = 0*\frac{16}{100}+1*\frac{23}{100}+2*\frac{30}{100}+3*\frac{22}{100}+4*\frac{6}{100}\\ +5*\frac{1}{100}+6*\frac{2}{100}\\ = \underline{\underline{1.9}}\end{aligned}$$

The expected value, R_{00}, for **red light** is 1.89 and was calculated as follows after 1000 iterations as follows:

$$\begin{aligned}\sum_{x=1}^{N}(P(r_{00}=x)*x) = 0*\frac{16}{100}+1*\frac{30}{100}+2*\frac{30}{100}+3*\frac{10}{100}+4*\frac{6}{100}\\ +5*\frac{4}{100}+6*\frac{3}{100}+7*\frac{1}{100}\\ = \underline{\underline{1.89}}\end{aligned}$$

A.3.2 r_{0N} of transmission with no testing, weekly testing, and biweekly testing

This subsection will describe the calculations necessary to find the expected value of r_{0N} for not tested, weekly tested and biweekly tested transmission. The approach for generating the different testing scenarios is described in section 3.4.5.

The expected value of disease transmission, R_{0N}, run on a **not tested** population is 0.99, when taking into account the number of infected individual each 767 recovered transmitted the disease to.

$$\begin{aligned}\sum_{x=1}^{N}(P(r_{0N}=x)*x) = 0*\frac{5793}{11572}+1*\frac{2769}{11572}+2*\frac{1520}{11572}+3*\frac{837}{11572}\\ +4*\frac{398}{11572}+5*\frac{157}{11572}+6*\frac{65}{11572}+7*\frac{34}{11572}\\ = \underline{\underline{0.9787}}\end{aligned}$$

The expected value of disease transmission, R_{0N}, run on a **biweekly tested** population is 0.97, when taking into account the number of infected individual each 279 recovered transmitted the disease to.

$$\begin{aligned}\sum_{x=1}^{N}(P(r_{0N}=x)*x) &= 0*\frac{5447}{10911}+1*\frac{2649}{10911}+2*\frac{1430}{10911}+3*\frac{796}{10911}\\ &+4*\frac{322}{10911}+5*\frac{151}{10911}+6*\frac{70}{10911}+7*\frac{47}{10911}\\ &=\underline{\underline{0.9796}}\end{aligned}$$

The expected value of disease transmission, R_{0N}, run on a **weekly tested** population is 0.95, when taking into account the number of infected individual each 364 recovered transmitted the disease to.

$$\begin{aligned}\sum_{x=1}^{N}(P(r_{0N}=x)*x) &= 0*\frac{4343}{8465}+1*\frac{1985}{8465}+2*\frac{1055}{8465}+3*\frac{568}{8465}\\ &+4*\frac{282}{8465}+5*\frac{136}{8465}+6*\frac{55}{8465}+7*\frac{42}{8465}\\ &=\underline{\underline{0.9723}}\end{aligned}$$

A.3.3 r_{00} of asymptomatic and symptomatic transmission

This section will detail the calculation of the expected r_{00} value of both symptomatic and asymptomatic disease transmission, which was run as described in subsection 3.4.6.

The expected R_{00} of **symptomatic** disease transmission was calculated to 3.16 with 100 iterations.

$$\begin{aligned}\sum_{x=0}^{N}(x*P(r_{00}=x)) &= 0*0.07+1*0.16+2*0.17+3*0.21+4*0.16\\ &+5*0.08+6*0.08+7*0.05+8*0.02\\ &=\underline{\underline{3.16}}\end{aligned}$$

The expected R_{00} of **asymptomatic** disease transmission was calculated to 1.01 with 100 iterations.

$$\sum_{x=1}^{N}(P(r_{00}=x)*x)=0*0.4+1*0.38+2*0.11+3*0.06+4*0.04+7*0.01=\underline{\underline{1.01}}$$

Ratio between symptomatic and asymptomatic disease is then $\frac{1.01}{3.16}=0.320$.

In order to compare this value with a predicted ratio, the following steps were calculated. For low values of $w_{ij}*I_p$, a first order approximation of the probability of transmission is as follows:

$$p_{inf} = 1 - (1 - 0.001)^{w_{ij} * I_p} \approx 0.001 * w_{ij} * I_p$$

Following this approximation, for each r_00 calculated per patient zero i with N neighbours j, the expected number of infected neighbours each day is $\sum_j^N (p_{inf})$. $days$ denotes the number of days in a particular disease state. From there on, the predicted r_00 for patient zero i is:

$$\begin{aligned} r_{00} &= days * \sum_j^N (p_{inf}) \\ &= days * \sum_j^N (0.001 * w_{ij} * I_p) \\ &= 0.001 * I_p * days * \sum_j^N (w_{ij}) \end{aligned}$$

Furthermore, the ratio between the R_00^A and R_00^P of the model is:

$$\begin{aligned} \frac{R_0^A}{R_0^P} &= \frac{0.001 * I_{p_a} * days_a * \sum_j^N (w_{ij})}{0.001 * I_{p_p} * days_p * \sum_j^N (w_{ij})} \\ &= \frac{I_{p_a} * days_a}{I_{p_p} * days_p} \end{aligned}$$

Thus, using the disease parameters from Ferretti et al. that are presented in Table 3.6, the predicted ratio between the R_0 of the two disease states can be calculated as follows:

$$\frac{R_0^A}{R_0^{PS}} = \frac{I_A * days_a}{I_P * days_p}$$

Where the R_0^A is the expected R_0 of only asymptomatic infections while R_0^{PS} is the expected R_0 of only presymptomatic infection. I_A describes the infectiousness of asymptomatic individuals in relation to symptomatic, while I_P describes the infectiousness of presymptomatic in relation to symptomatic. Furthermore, $days_A$ denotes the days an individual is in the state asymptomatic, while $days_P$ is the days an individual is presymptomatic.

$$\frac{0.1 * 6}{1.3 * 3} = \frac{0.6}{3.9} = \underline{\underline{0.154}}$$

The expected ratio of r_0 of asymptomatic and symptomatic infection for the generated model to 0.320, whilst the expected ratio based on probability and transmission was calculated to 0.154.

Appendix B

Result and analysis supplement

This chapter will provide more details and additional figures for the result and analysis chapter of this thesis. It will detail a further exploration of the empiric dataset in section B.1 and more detailed information on disease transmission as provided in section B.3.

B.1 Further exploration of empiric dataset

Analysis of the empiric dataset extended beyond the results presented in chapter 4, Results and analysis. The following section will describe the investigation of the distribution of weighted links for single nodes in subsection B.1.1 and address temporal aspects of the model in Figure B.1.

B.1.1 Single nodes and Power law

As the combined degree distribution of all nodes in a network did not yield a definitive distribution, it was decided to further investigate the distribution of each weighted edge per node. Furthermore, the edges, or interactions, were divided according to which interaction layer they belonged to. Each weighted edge was plotted for the whole network, off-diagonal, grade, or class interactions. Figure B.1 shows an overview of the first nine nodes to be investigated in the empiric network on day one. Of most interest is that each layer seems to follow a linear function on a log-log scale. This indicates a Power law distribution. Furthermore, the R^2 values given for the linear regression are relatively high for all layers investigated. Therefore, to some extent the single nodes have a Power law distribution of their weighted interactions with other nodes.

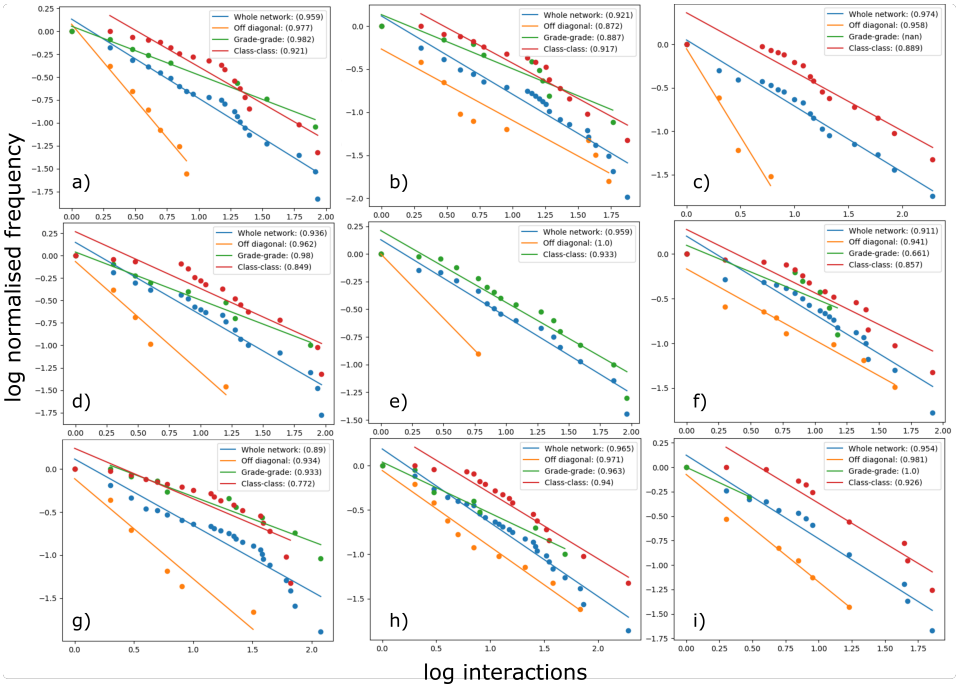


Figure B.1: The first nine nodes with log normalised frequency as a function of log degree, $\langle k \rangle$. The different colours indicates the layer the interaction occurred in with whole network (blue), off-diagonal (orange), grade (green) and class (orange). The R^2 value is given for each linear regression.

B.1.2 Hour by hour and lunch analysis

Figure B.2 shows the number of interactions, where each interaction is a 20-second interval between two individuals that are 1-1.5 meter from one another. Interactions are divided into those occurring within each grade from one to five. There are three significant spikes in the number of iterations throughout day one. Comparing the spikes with the schedule for the school seen in Figure 2.7, one can see that they occur during breaks. The two-hour lunch break in the middle of the day shows how much more grades one and two interact than grades three to five. This is most likely due to the older students going home to eat lunch during the longer break as opposed to the younger students. Most importantly, one can note that the interaction patterns change throughout the day.

Furthermore, Figure B.3 shows the distribution of interactions during lunch hours on both day one and day two. The division between grades one to three and four and five is evident on the first day. However, for the second day, it seems that class 1A has had lunch and/or break with grades two and three, while 1B has had lunch and/or break with grades four and five. Class 1A has few class interactions with each other, while 1B has more interactions. Similarly to day one, grades two and three interact the most with each other, while grades four and five interact with each other. Perhaps there was a deviance in the normal one to three and four and five division system on day two, where the first grade was split in two

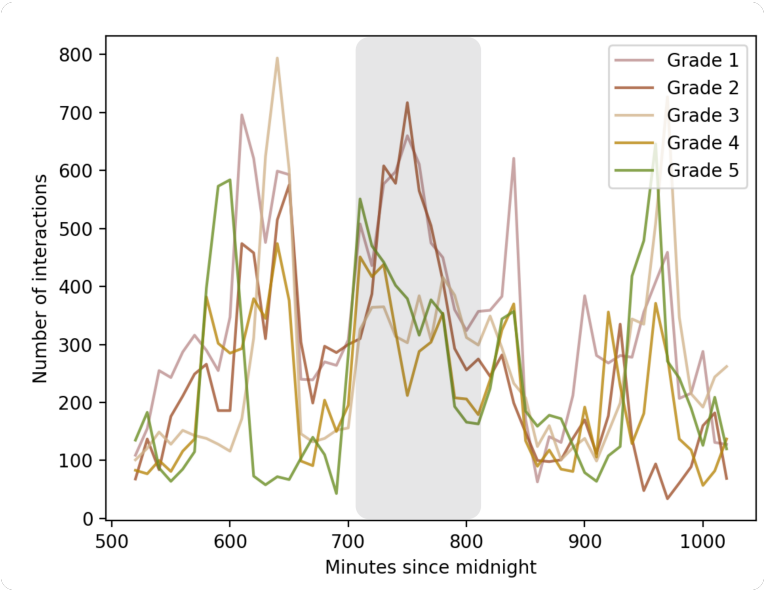


Figure B.2: Distribution of weighted interactions per grade occurring each minute on day one. A grey area highlights the time interval of the two hour lunch break.

for lunch and break time.

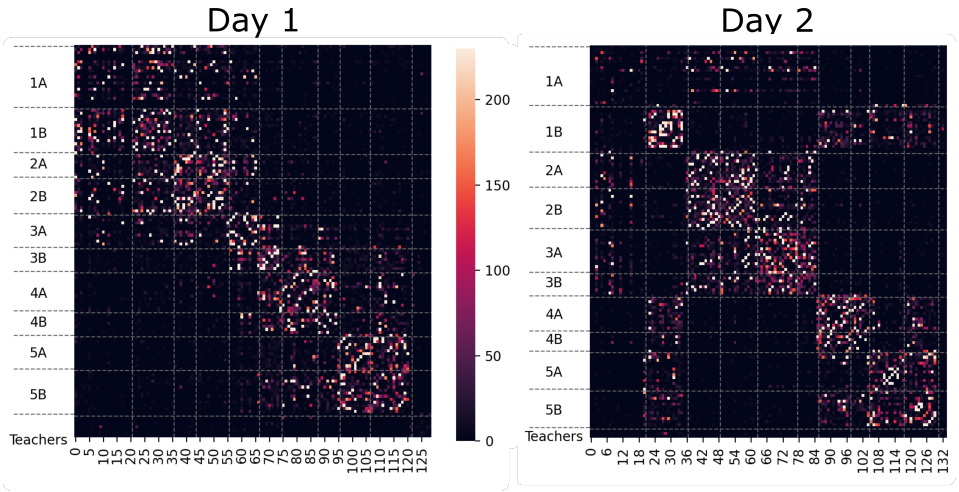


Figure B.3: Heatmap of interactions occurring during the lunch break. Helplines are drawn to show the class of each individual

Figure B.4 shows the degree distribution during lunch for the two days tracked by Barrat et al. [1]. The highest frequency of the degrees during lunch is around 100-300.

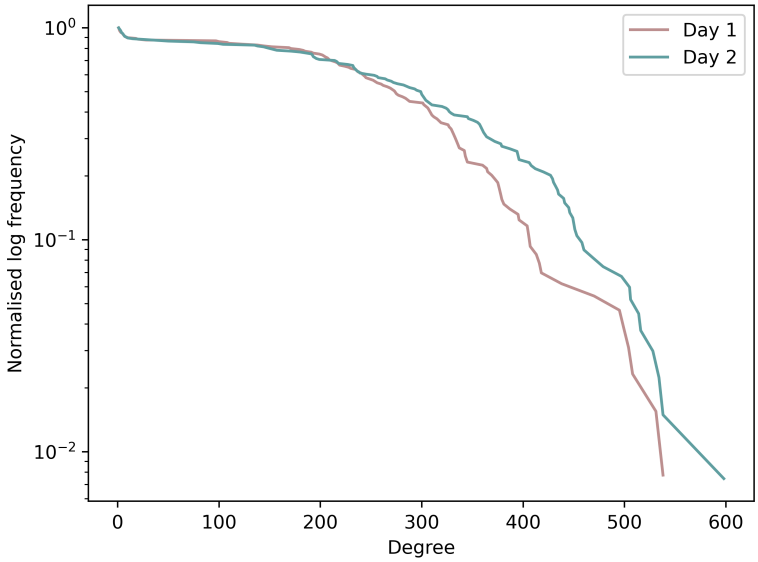


Figure B.4: Log cumulative frequency as a function of degree for day one and two during lunch

B.2 Analysis of model supplementary

fig. B.5 shows a drawing of the simulated model with edges drawn between nodes meaning an interaction occur between two individuals.

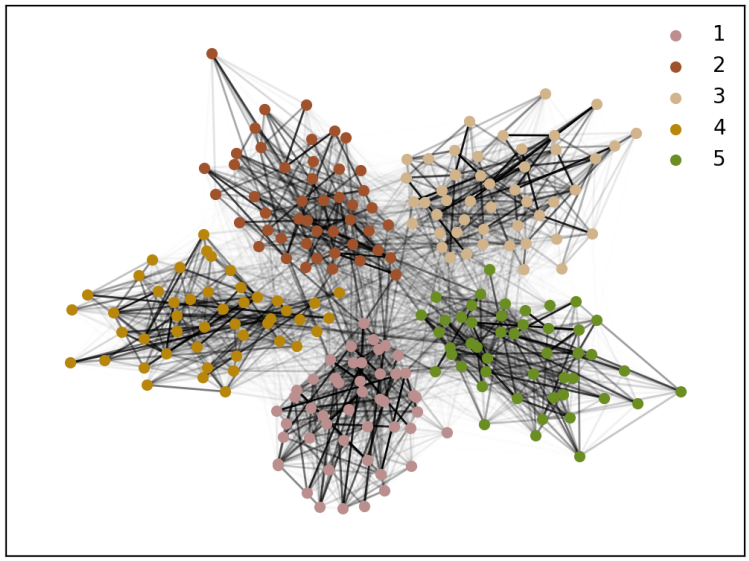


Figure B.5: Display of the simulated interaction model using the framework presented in this thesis. The colours of the nodes represent the grades the individuals belong in, whilst the strength of the colour on the edges increases with higher weight.

B.3 Disease transmission supplementary

More information and figures for disease transmission will be presented in this section. subsection B.3.1 describes how disease transmission looks on the empiric network and subsection B.3.4 gives more information on expected value of r_{0N} for different testing strategies.

B.3.1 Disease transmission

Figure B.6 describes a case where disease transmission has been run on the model. The nodes are coloured according to the state each individual is in. As seen, infection has a tendency to first spread within a grade, and from there on reach the rest of the school.

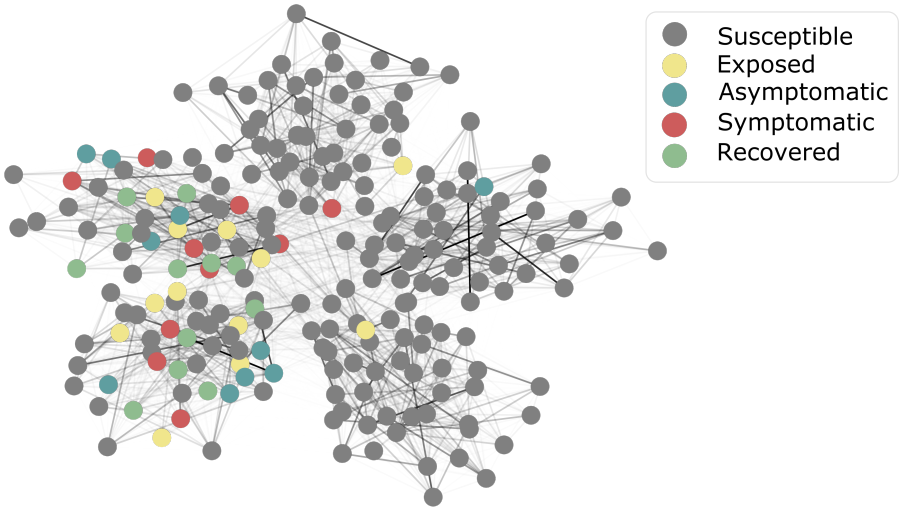


Figure B.6: Illustration of disease transmission run on 21 days. The colour of each node represents which state the individuals are in.

B.3.2 Traffic light

The error margin of the green traffic light (Figure B.7), yellow traffic light (Figure B.8) and red traffic light (Figure B.9) is illustrated by drawing 10 of the 100 iterations run alongside the average. As seen in the figures, many transmissions ends with few recovered, and the number increases for yellow and red traffic lights. The reasoning behind this, is that the randomly chosen patient zero most likely does not have many contacts with other individuals, and thus transmission is either greatly halted or completely stopped. Similarly, if a hub is chosen as patient zero, it may cause a significantly faster increase in the number of recovered and saturation is reached at a higher number of recovered. All data can be found in the GitHub described in section C.2 in Appendix C.

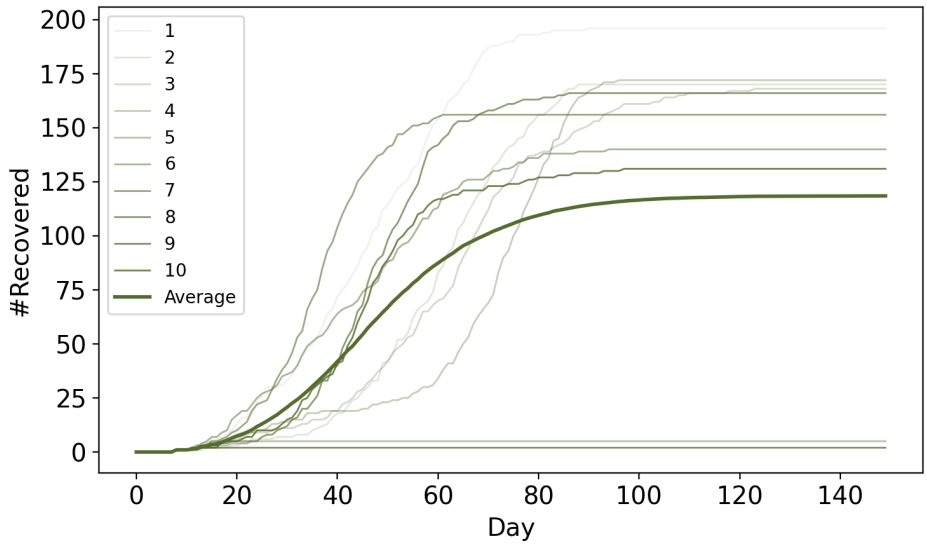


Figure B.7: Error margin of the green traffic light model with 10 randomly chosen iterations of transmission and the average value in bold

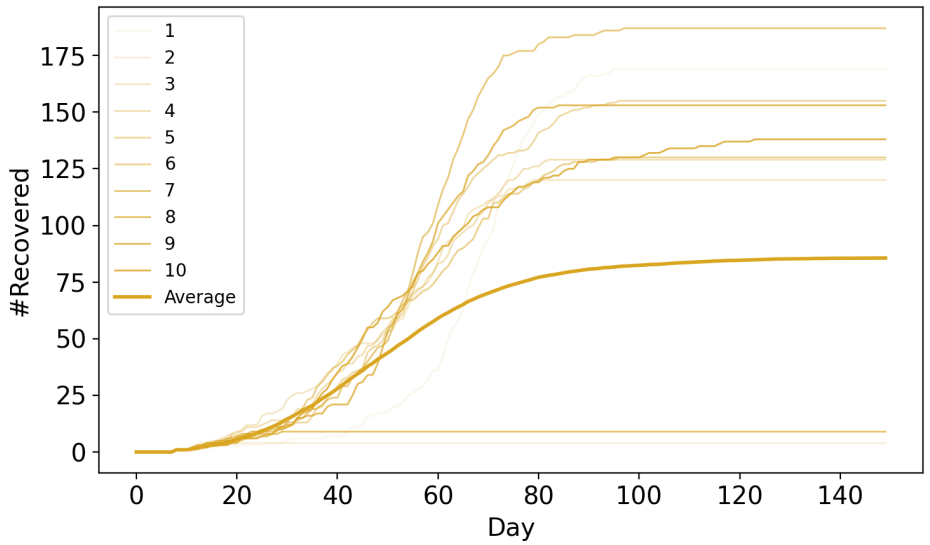


Figure B.8: Error margin of the yellow traffic light model with 10 randomly chosen iterations of transmission and the average value in bold

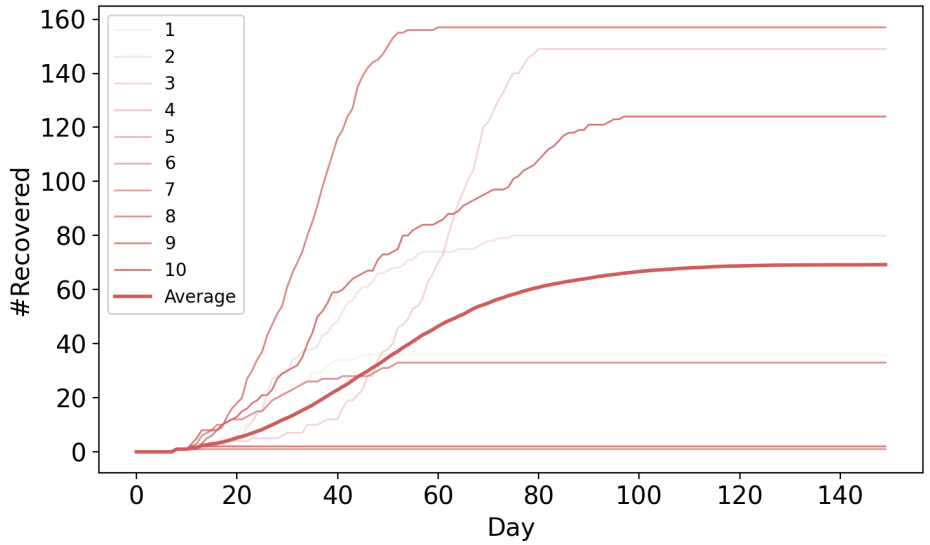


Figure B.9: Error margin of the red traffic light model with 10 randomly chosen iterations of transmission and the average value in bold

B.3.3 Recovered for different testing scenarios

The error margin of the no testing (Figure B.10), weekly tested (Figure B.11) and bi-weekly tested (Figure B.12) is illustrated by drawing 10 of the 100 iterations run alongside the average. Similarly to the traffic light error margin plots, there is great variation based on which individual was chosen to be patient zero. The randomly selected iterations of Figure B.12 shows this randomness, where the 10 selected iterations are higher than the average. For the data for all iterations, see the GitHub described in section C.2, in Appendix C.

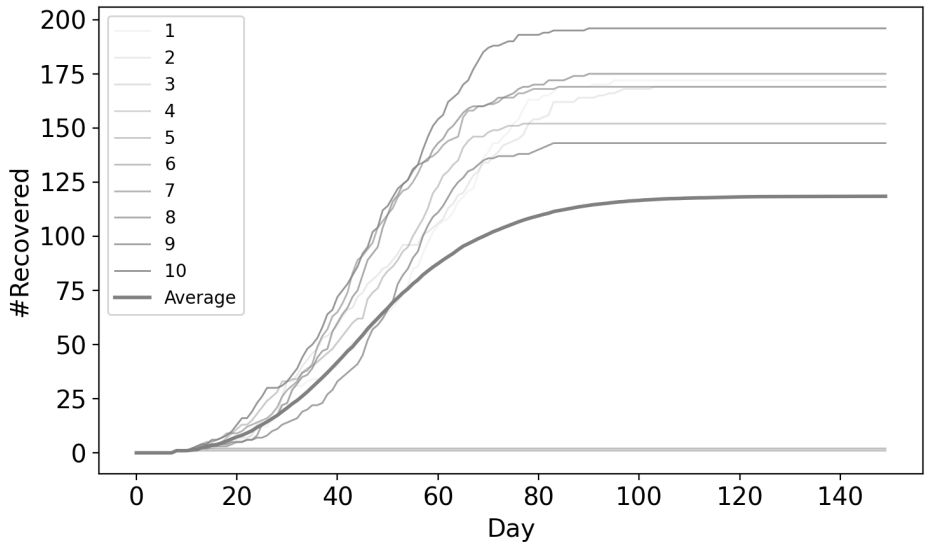


Figure B.10: Error margin for not tested disease transmission with 10 randomly chosen iterations of transmission and the average value in bold

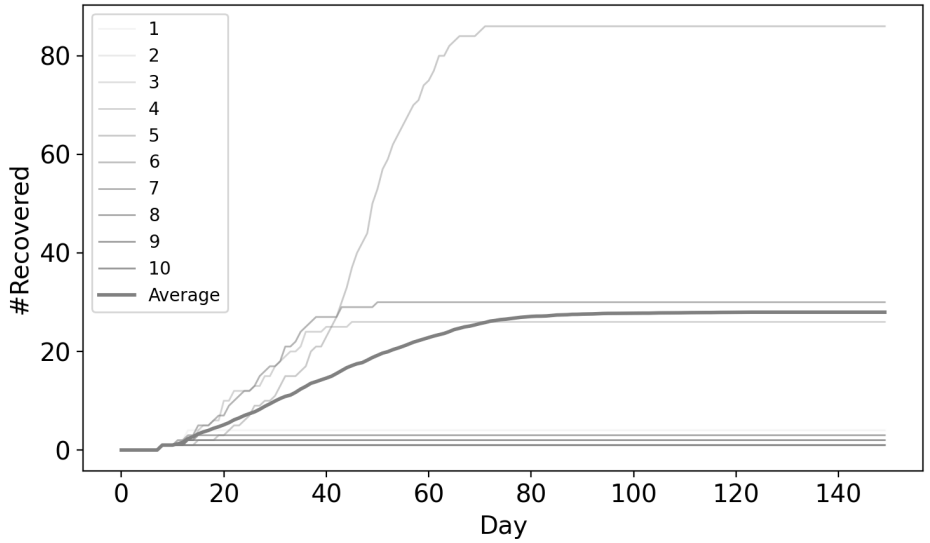


Figure B.11: Error margin for weekly tested disease transmission with 10 randomly chosen iterations of transmission and the average value in bold

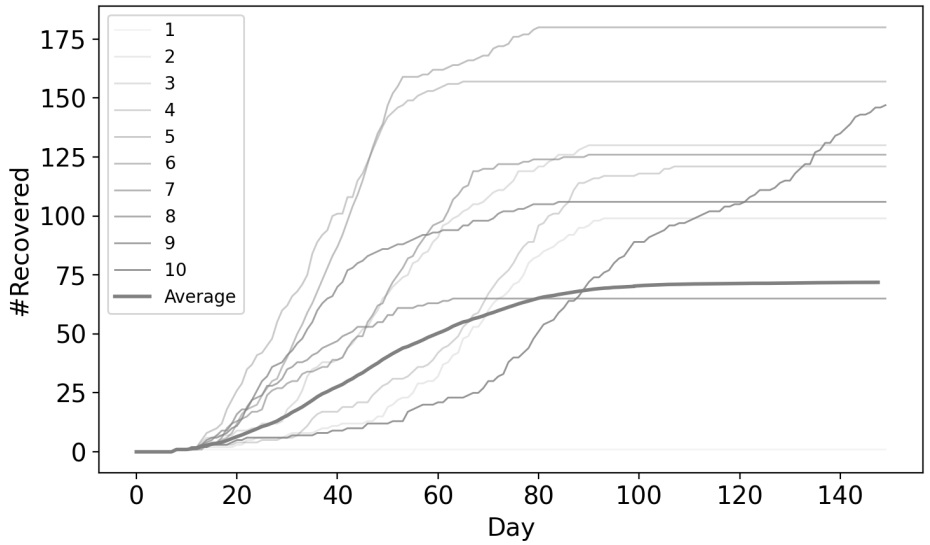


Figure B.12: Error margin for biweekly tested disease transmission with 10 randomly chosen iterations of transmission and the average value in bold

B.3.4 r_{0N} for different testing scenarios

Figure B.13 shows the frequency of different r_{0N} values as well as its expected reproductive number.

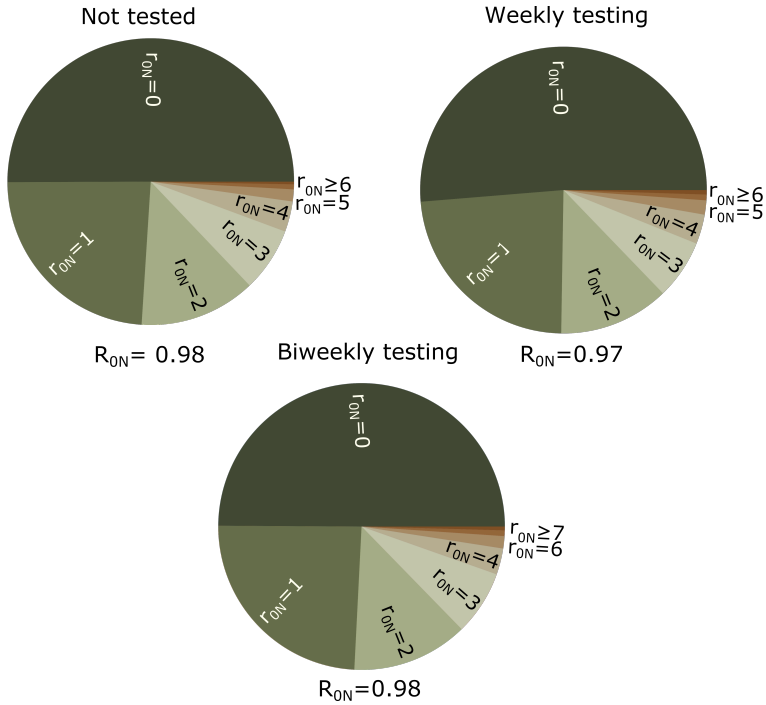


Figure B.13: Frequency of r_{0N} as well as expected value for three testing scenarios; no testing, weekly testing and biweekly testing

Appendix C

Github repositories

C.1 GitHub - Analysis of experimental data

The GitHub repository used to conduct all analysis described in section 4.1 can be found at `https://github.com/SaraAsche/PrimarySchool`. Furthermore, the overall description of the project can be read in the README.rmd file for the project. In addition, documentation for all functions and files are provided within their respective python files.

C.2 GitHub - The model

The code to run the primary school model proposed in this thesis can be found at `https://github.com/SaraAsche/PrimarySchoolSimulation`. The model was described in section 3.3, and disease transmission and different mitigation measures were described in section 3.4. Similar to section C.1, the overall description of the project can be read in the README.rmd file for the project. In addition, the documentation for each class is provided within their respective python files and a thorough description of all functions associated with the objects.

Sara Johanne Asche

Modelling the transmission of COVID-19 in primary schools

NTNU
Kunnskap for en bedre verden