NTNU

Norwegian University of
Science and Technology

Computational SIS Modeling of the Spread of Antibiotic Resistance within Bacterial Metapopulation Networks

Nora Rosvoll Finstad

Biotechnology
Submission date: May 2018
Supervisor: Eivind Almaas, IBT
Co-supervisor: Pål Røynestad, IBT

Norwegian University of Science and Technology
Department of Biotechnology and Food Science

Acknowledgements

The work of this thesis was conducted at the Faculty of Natural Science at the Norwegian University of Science and Technology (NTNU), as a part of the 2-year Master of Science program in Biotechnology.

First, I would like to thank my supervisor Professor Eivind Almaas for helpful guidance and support throughout the work of my thesis. I am grateful for being educationally introduced to this, for me, somewhat unfamiliar and exiting field of biotechnology.

I would also like to express my gratitude to my co-supervisor PhD Candidate Pål Røynestad for providing invaluable help regarding the development of the network model tool used for simulations. I highly appreciate his drive and effort concerning the work of my thesis, and his availability around-the-clock.

Finally, I would like to thank my family and friends for moral support and encouragement during these last two years. Thank you all.

Trondheim, May 15th, 2018

Nora Rosvoll Finstad

Abstract

The world faces a continually ongoing emergence of antibiotic resistance, and an almost non-existent development of new antibiotics to fight resistant bacteria. This leads to a substantial need for new and effective methods to stagnate the spread of antibiotic resistance. There exists a diversity of approaches for modeling spreading through a system. Each approach with their own strengths and weaknesses. For this thesis, a combination of a compartmental susceptible-infected-susceptible (SIS) modeling approach and a metapopulation modeling approach was chosen. The choice was based upon available data, and the general aim of the thesis. Various existing software were carefully evaluated and considered, but none of the studied software seemed to satisfy the initial needs and desire of the tool.

Initially, the aim was to develop a simple tool to study the effect of different parameters and network topologies in spreading situations. The desire was to show how a tool of this kind could be of use to predict the spread of infection; in this case antibiotic resistance, and how the tool could possibly help us establish measures to be taken in order to cease a spread – or at least control the spreading activity. During the work of this thesis, a tool was developed, and several simulations were run in order to illustrate the utility of such a remedy as a help to stagnate the ongoing development of antibiotic resistance. Results from these simulation runs proved the importance of transit nodes for a spreading to progress, and that different network types display different spreading patterns. Results also demonstrated how parameters such as the frequency of cleaning, the growth rate of bacteria, and the fitness cost related to the acquisition of resistance genes, might affect systems. Among other things, results indicated an optimal cleaning interval of 6-9 hours, and a proportionality between increased growth rate and spreading activity. To illustrate how the tool may be utilized in real-life events, a semi-realistic simulation of a virtual hospital was conducted. Last but not least, limitations to the tool were established; limitations such as degree of realism.

This thesis presents a review concerning a selection of already existing research within the field of computational epidemiological modeling, as well a presentation of the developed network model tool and some of its functions.

Sammendrag

Verden står overfor en kontinuerlig pågående utvikling av antibiotikaresistens, og en tilnærmet ikke-eksisterende utvikling av nye antibiotika for å bekjempe resistente bakterier. Dette fører til et betydelig behov for nye og effektive metoder for å stagnere spredningen av antibiotikaresistens. Det finnes et mangfold av tilnærminger for å modellerere spredning gjennom et system. Hver tilnærming med sine egne styrker og svakheter. For denne avhandlingen ble det valgt en kombinasjon av sensitiv-infisert-sensitiv (SIS)-modellering og metapopulasjonsmodellering. Valget baserte seg på tilgjengelige data, og det generelle målet med avhandlingen. Ulike allerede eksisterende programvarer ble nøye evaluert og vurdert, men ingen av de aktuelle programvarene syntes å tilfredsstille de primære ønskene og kravene for verktøyet.

I utgangspunktet var målet å utvikle et enkelt verktøy for å studere effekten av ulike parametere og nettverkstopologier i spredningssituasjoner. Målet var å vise hvordan et slikt verktøy kan være til nytte for å forutsi infeksjonsspredning; i dette tilfellet antibiotikaresistens, og hvordan verktøyet muligens kan hjelpe oss med å etablere tiltak med hensikt å stoppe spredningen – eller i det minste kontrollere spredningsaktiviteten. Under arbeidet med denne oppgaven ble det utviklet et verktøy, og flere simuleringer ble gjennomført for å illustrere bruken av et slikt verktøy for å stagnere den pågående utviklingen av antibiotikaresistens. Resultatene fra disse simuleringene påviste viktigheten av transittknutepunkter for spredning, og at forskjellige nettverkstyper viser forskjellige spredningsmønstre. Resultatene viste også hvordan parametere som rengjøringsfrekvens, bakteriell vekstrate og den metabolske kostnaden knyttet til å erverve resistensgener, kan påvirke systemer. Resultatene indikerte blant annet et optimalt rengjøringsintervall på 6-9 timer, og en proporsjonalitet mellom økt vekstrate og spredningsaktivitet. For å illustrere hvordan verktøyet kan utnyttes i det virkelige liv, ble det gjennomført en halvrealistisk simulering av et virtuelt sykehus. Sist men ikke minst ble noen begrensninger til verktøyet etablert; begrensninger slik som grad av realisme.

Denne avhandlingen presenterer en gjennomgang av et utvalg av forskning innen beregningsmessig epidemiologisk modellering, samt en presentasjon av det utviklede modelleringsverktøyet og noen av dets funksjoner.

Table of contents

1 Introduction

Since the introduction of antibiotic use in medical therapy, antibiotics have been one of the most important and successful remedies of medical therapy (Lin et al., 2015).

There are a multitude of different antibiotics available on the market today. Antibiotics are classified on the basis of their structure, characteristics and functions. They may be classified based on their specificity towards different types of bacteria; broad spectrum antibiotics or narrow spectrum antibiotics, or whether they kill bacteria (bactericidal antibiotics) or inhibit bacterial cells from reproducing (bacteriostatic antibiotics). Most commonly, antibiotics are classified based on how they affect bacterial cells; by inhibiting cell wall synthesis, protein synthesis, DNA synthesis, RNA synthesis, or folic acid synthesis (Kohanski, Dwyer, & Collins, 2010).

Due to different factors, both evolutionary and human-induced, bacteria have developed resistance against antibiotics. The consequences of antibiotic resistance are serious, and antibiotic resistance has already given rise to increased numbers of fatal outcomes of bacterial infections (Laxminarayan et al., 2013). The Centers for Disease Control and Prevention (CDC) has for the United Stated alone, estimated an approximate number of 80 000 severe cases of MRSA infections annually, of which over 11 000 cases are associated with death (Frieden, 2013, p. 16).

Antibiotics are essential in treatment of infections. In addition, they are necessary components in medical procedures such as major surgeries, organ transplantations, and cancer chemotherapy. Without effective antibiotic remedies, these procedures will be nearly impossible due to high infection risks. Not only will antibiotic resistance affect the health of individuals, it will also have a considerable impact on the economy of health care (WHO, 2018a).

As previously mentioned, the development of antibiotic resistance is a consequence of many different factors. The most important factor today is human behavior. Earlier, the treatment of antibiotics was under little control, which had a huge impact on the development of antibiotic resistant bacteria. Today, the guidelines of antibiotic use are a lot more restrictive, but we still have a long way to go. In countries of low to medium income (LMICs), the restriction of antibiotic use is not as adequate as in highly developed countries. Resistance against antibiotics has developed in bacteria as a consequence of mutations in the genome or as an

obtainment of resistance-coding genes via horizontal gene transfer (Lázár et al., 2013). Due to selective pressure, a higher incidence of antibiotic resistance development is seen in environments of extensive use of antibiotics. This phenomenon is seen in hospitals, agriculture- and aquaculture environments, to mention some (Gullberg et al., 2011; Levy, 2002; Laxminarayan et al. 2013).

Statistics show that a large quantity of the cases with antibiotic resistant infections are infections acquired in hospitals (Ventola, 2015). World Health Organization (WHO) has estimated the number of deaths due to antibiotic resistant infections acquired in hospitals in Europe; out of 400 000 registered cases of antibiotic resistant infections, WHO estimate a number of 25 000 deaths (WHO, 2018b). This number will increase drastically if antibiotic resistance keeps developing and spreading.

Though it is a huge field of research, the search for new antibiotics and other mechanisms to overcome bacterial infections shows little progress. We therefore must think of other ways to stagnate the spread of antibiotic resistance. A helpful tool to get a better understanding of the extent antibiotic resistance through horizontal gene transfer, is computational modeling. Computational modeling gives good insight to how and why spreading occurs. However, more important is that computational modeling may help develop and test predictions for how one may prevent further development and spreading of resistance. Some experiments are high-cost and require lots of planning, time, and equipment. Computational modeling is a valuable tool in simulating experiments that, due to different reasons, are hard to execute in practice. It may also be a great preparatory work for future experiments by "drawing a picture" of what will be the most successful way to conduct experiments. Several studies using computational modeling on this topic, have been conducted previously, but these efforts may not include all the necessary factors to be directly applicable in real life.

Existing computational models of antibiotic resistance usually include microbiological factors such as genetic evolution and acquisition of new resistance mechanisms, individual factors such as intra-strain growth competition, and population factors such as transmission of pathogens between populations. Some models focus on the bacterial dynamics in individuals, and other focus on bacterial dynamics in populations. While the models may be great at certain areas, they lack information to be fit for use at other areas. Though the existing models may be improved, they already provide important contributions to the research of antibiotic resistance (Opatowski, Guillemot, Boëlle, & Temime, 2011).

This thesis presents a review concerning a selection of already existing research within the field of computational epidemiological modeling, as well as a network model tool developed in relation to the thesis.

1.1 Thesis objectives

The first aim of this thesis was to conduct a literature survey of relevant publications, and to identify possible improvements to computational methods that already exists. The second aim was the development and application of a new approach for modeling antibiotic resistance spreading through horizontal gene transfer in bacterial populations.

The new approach was based upon computational modeling in the shape of a network model. The model was intended to simulate the spread of resistance genes due to horizontal and vertical gene transfer among populations of susceptible individuals. The model was designed with an ability to simulate spreading processes on multiple different networks with different structures and topologies. The possibility of making simple changes to the parameters provides a chance of customizing the network model to fit different approaches. Not only are the network parameters themselves flexible for adjustment, the model may also be adjusted according to type of microbes and antibiotics. The network model may be customized on the basis of each individual desire for simulation.

I wished to develop a network model tool that was able to visualize the effect of different parameters, in order to possibly be beneficial in predicting spread prevention measures. Another aspect that seemed interesting in relation to this thesis was to study how simulations with the same parameter settings would appear when run on types of networks with different properties; how does model type affect spreading patterns?

1.1.2 Deciding model approach

Before starting the process of designing the tool, extensive literature search was conducted. The aim was to see if similar simulations had previously been done. If so, the idea was to employ existing tools for the simulation.

Generalized epidemic mean-field (GEMF) modeling

Gehring et al. (2010) discussed a similar approach to the problem. They intended to develop a mathematical compartmental model that combined with pharmacokinetics could be used to visualize the optimal treatment plans with antibiotics, that has the lowest impact on resistance development. They intended to combine molecular bacteriology and pharmacology to illustrate horizontal transfer of resistance genes among bacteria in a quantitative manner (Gehring et al., 2010).

I reached out to the authors of the article (Gehring et al., 2010) and requested the availability of the tool. It turned out that the tool of interest was no longer available. However, a modified version of the tool was provided, with consent of further use. This modified tool was called GEMF (Generalized epidemic mean-field) (Sahneh, Scoglio, & Mieghem, 2013). GEMF has shown to be a suitable tool in the simulation of spreading processes in complex networks. In its simplest form, it simulates the spreading process with N nodes that may or may not be in M compartments. The GEMF model is able to simulate multi-layer networks by implementing edges between the nodes on different levels.

The concept was based upon the nodes in a network being able to transit through different states. The transition equations of this software were designed to follow a stochastic process called the Markov process. After careful consideration I decided that the modified tool was not as good fit as the one presented by Gehring et al. (2010). The structure of the model seemed as a good fit at first glance, but I soon discovered that each node was designed to represent single individuals, not subpopulations, as intended for this thesis. GEMF had the possibility of making multilayer networks, but it lacked the ability of considering metapopulations. Thus, I concluded that designing the tool from scratch was the better option.

Airline transportation network modeling

The network model of this thesis was inspired by the airline transportation network model by Colizza et al. (2006). The airline transportation network model is a stochastic model with the intention to forecast global epidemics based on the combination of worldwide airline transportation infrastructure and registered population data. Their main aim was to study how the airline transportation infrastructure influenced global epidemics. Additionally, they aimed to assess the authenticity of disease outbreak predictions and scenarios (Colizza et al., 2006).

As the airline transportation network model, the model of this thesis is based on compartmental modeling with differential equations. We have, like the team behind the airline transportation network model, developed a metapopulation model where different compartments represent subpopulations. The contact pattern of the airline transportation model is dependent of the data obtained from sources such as the International Air Transport Association (IATA) database. For the model presented in this thesis, data may be varied and obtained according to the aim of simulation.

Population based modeling

The network model was designed to follow a population-based approach. A population-based approach allowed us to investigate the dynamics of susceptible individuals and infected individuals over a period of time, based on the population composition. An agent-based modeling approach would also allow for the possibility to investigate dynamic changes to a population. However, the dynamics were in this thesis studied at an individual level, not by the characteristics of the population as a whole. The network model was primarily designed for bacteria; individuals that are short-lived, and usually appear in large quantities. Following each individual in an agent-based manner would thus not provide an outcome to fulfill our requisites for the network model.

Staphylococcus aureus

The main simulation executed to present as an example of the model in this thesis, is the spread of *Staphylococcus aureus* at Orkdal hospital. This is a hypothetical example with the sole purpose of presenting the potential of the tool. *Staphylococcus aureus* was selected as the microbe of interest for the simulation, due to it being one of the most common sources of hospital infections. Additionally, *Staphylococcus aureus* is one of the most common bacteria to acquire antibiotic resistance (Murray et al., 2012). *Staphylococcus aureus* was thus found as an interesting and important microbe to study.

The fact that *Staphylococcus aureus* is a rather well-studied microbe, opened for the possibility to include real biological data in the simulations of this thesis. This includes molecular data of the microbe, approximate bacterial numbers, and antibiotic treatment dosages. Data is retrieved from available molecular data, previous published studies and a

bachelor thesis written in 2015 (Brunes & Finstad, 2015). Although the simulation is intended as a hypothetical example, it was of great interest to make it as realistic as possible. Analyzing how network topology and different parameters influence spreading patterns, may be possible by running several parallel simulations on the same network structure with varying input parameters.

1.1.3 Hypotheses

In order to illustrate the potential and possible applications for such a tool as the one developed in relation to this thesis, several hypotheses were formulated. The hypotheses are based upon published studies and research, and general theory.

Hypothesis 1:

"Transit nodes are important for the spreading through the network."

Hypothesis 2:

"Simulations on different model types; Erdős-Rényi and Barabási-Albert will display different spreading patterns. The ER model will show a uniform spreading pattern, while the BA model will show an uneven spreading pattern."

Hypothesis 3:

"The length of the cleaning intervals clearly affects the spread of *Staphylococcus aureus* through the system."

Hypothesis 4:

"Reduced growth rate increases susceptibility to antibiotics, while increased growth rate affects the spreading in total to progress more rapidly."

Hypothesis 5:

"Increased fitness cost will affect the spread of antibiotic resistance in a negative manner; the higher the fitness cost, the lower the spreading activity."

In addition to the hypothesis presented above, a hospital network was simulated. This simulation was mainly intended as an example of how our tool may be employed in a real-life scenario.

2 Background and theory

The start of the antibiotic era is associated with the discovery of the antibiotic penicillin. Penicillin was discovered by Alexander Fleming in 1928, when he by chance observed a petri dish with mold that inhibited the growth of a particular bacterium. Fleming and his team started investigating the mold and found it to be the strain called *Penicillum notatum*. During the next few years, much research in conjunction to Fleming's findings was carried out. This resulted in information on which bacterial strains were affected by the *Penicillum*, and how it could be purified and quantified. Howard Florey and Ernest Chain published in 1940 a paper on the latter (Chain & Florey, 1940), which in turn lead to the mass introduction of Penicillin for therapeutic use in 1945 (Aminov, 2010).

A great success in finding effective antimicrobials followed. However, a downside to the frequent and increasing usage of antibiotics was quickly discovered. This was due to the discovery of bacteria being resistant to antibiotics. Ability to withstand the antibiotic penicillin was the first well known mechanism for antibiotic resistance. It was learned that this type of resistance mechanism was due to the enzyme penicillinase. Penicillinase hydrolyzes and breaks the beta-lactam ring of beta-lactam antibiotics (Kong, Schneper, & Mathee, 2010). The beta-lactam ring structure constitutes a part of the core structure of beta-lactam antibiotics, and thus when degraded by penicillinase, the antibiotic will not be effective in destroying certain bacteria. Beta-lactam antibiotics work by inhibiting the synthesis of the bacterial cell wall. They do so by interfering with enzymes necessary in synthesizing the peptidoglycan layer (Tenover, 2006).

The first identified penicillinase was discovered as early as in 1940 (Davies & Davies, 2010). During the years following these findings, several other resistance mechanisms were discovered. Furthermore, the number of bacteria resistant to treatment soared significantly. While the beta-lactam antibiotics works by interfering with the synthesis of the bacterial cell wall, there also exists other common mechanisms for antibiotics to eradicate bacteria. Such mechanisms are inhibition of the protein synthesis, inhibition of different metabolic pathways, or interference in the synthesis of nucleic acids. Figure 1 is included to give an impression of the rapid increase of resistance development. This figure shows a timeline of the evolution of antibiotic resistance for various *Staphylococcus aureus*; community-acquired MRSA, livestock-acquired MRSA, VISA (vancomycin intermediate *Staphylococcus aureus*), and

VRSA (vancomycin resistant *Staphylococcus aureus*). Additionally, the introduction of some important antibiotics is included in the timeline.

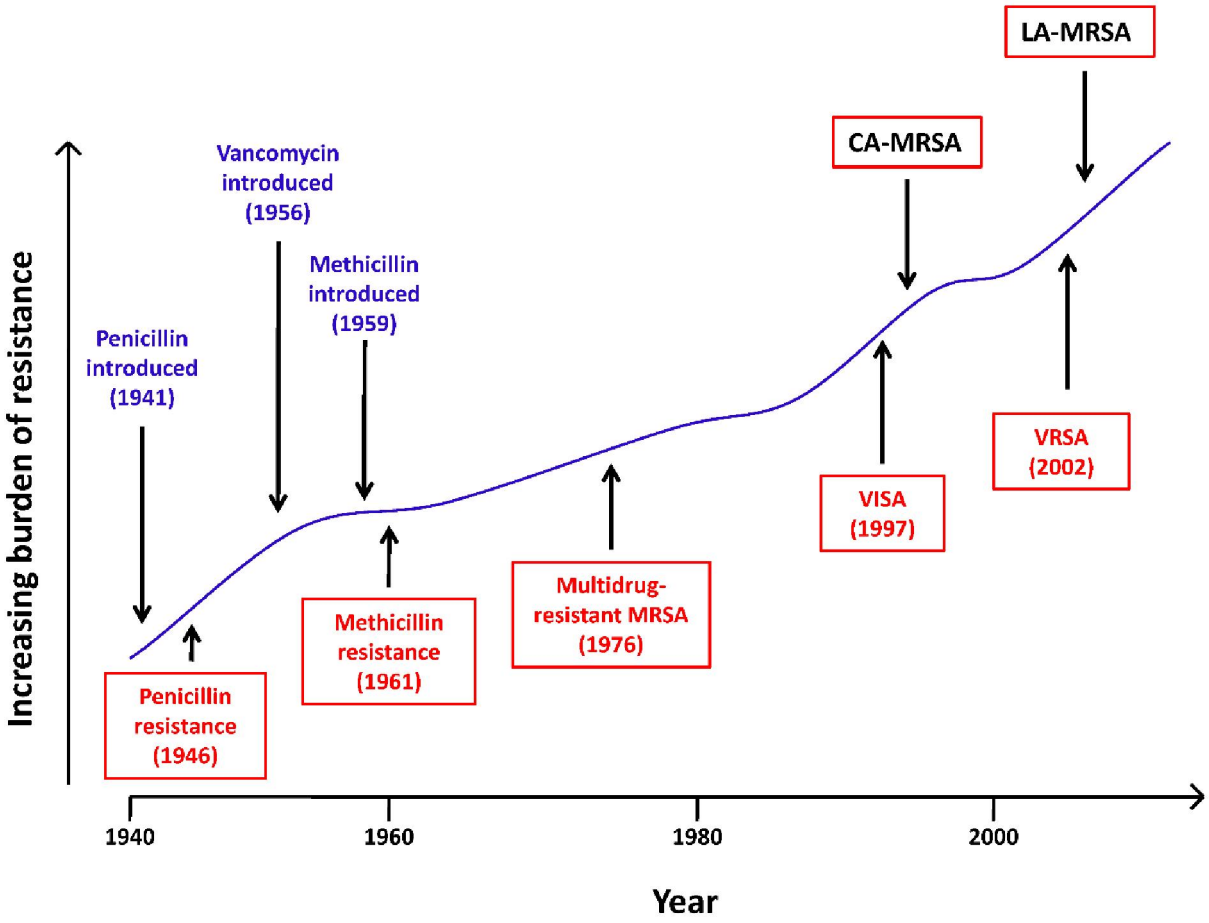


Figure 1: The evolution of antibiotic resistance for various Staphylococcus aureus; community-acquired MRSA, livestock-acquired MRSA, VISA (vancomycin intermediate *Staphylococcus aureus*), and VRSA (vancomycin resistant *Staphylococcus aureus*) (Schmidt et al., 2015).

2.1 Antibiotic resistance

Resistance mechanisms

Just as the different antibiotics work by a variety of mechanisms, there exists a multitude of resistance mechanisms for bacteria. Bacteria may possess genes encoding certain enzymes that destroy the antibiotics before they have the chance to constitute any effect. Bacteria may also possess efflux pumps that prevents antibiotics to reach their target site inside the bacterial cell. Consequently, an antibiotic will have no effect on bacteria. A third mechanism is that some bacteria possess traits that make them able to alter the bacterial cell walls and its

binding sites. This mechanism prevents the antibiotic to bind the cell walls to exercise effect. Bacteria may acquire mutations that hinders effect of antibiotics (Tenover, 2006). Bacteria have several mechanisms of acquiring resistance towards antibiotics. Incidents where spontaneous mutations cause changes in the bacterial genome is one way of acquiring resistance. Another way for bacteria to acquire resistance towards antibiotics is due to a mechanism called horizontal gene transfer, which is the mechanism of focus in this thesis.

Horizontal gene transfer

Horizontal gene transfer is the transmission of genetic material between genomes of different organisms. Transmission of genetic material due to horizontal gene transfer is possible between organisms that are not related. This differs from vertical gene transfer, where the requisition of genetic material to offspring is from parents. Transmission of genes may be mediated by mobile genetic elements such as plasmids, transposons, or bacteriophages. These are elements that frequently carry resistance genes. Transmission of the genetic material from one bacteria to another is possible through different mechanisms; transformation, conjugation, and transduction. Transformation is a mechanism whereas bacteria acquire new genetic material by the uptake of extracellular DNA from surrounding environment. Conjugation differs from this mechanism in several ways. The most important differentiating factor is the necessity for cell-to-cell contact. The genetic material is in this manner transferred through channels of the cell membranes. The latter mechanism, transduction, is the transmission of genetic material between cells with the help of bacteriophages.

The most common mechanism of bacteria to receive genetic material carrying antibiotic resistance from other bacteria has shown to be conjugation (Dzidic & Bedeković, 2003). Once a recipient cell has acquired new DNA, it may be incorporated into the cell genome. The newly acquired genomic material may be included in the host genome by homologous recombination or additive integration. Homologous recombination integrates sequences of high similarity. The DNA is integrated into the host genome by exchange. Size and functionality will in most cases be maintained due to the exchange mechanism. Sequences with little or no similarity may be inserted into the host genome by substitution or addition of DNA. This mechanism opens for the possibility of horizontal gene transfer between different species and organisms with highly dissimilar genomes.

Horizontal gene transfer is an important factor in evolution. Since genetic changes are necessary for bacteria to, for example, withstand changes in the environment. We are therefore likely to find an increasing number of bacteria possessing traits giving advantages during such changes; survival of the fittest! Though horizontal gene transfer may result in great changes in genotype and phenotype, it will not necessarily always cause significant changes. Despite horizontal gene transfer being necessary for bacterial survival, the consequences are not always convenient. An example of this is its essential role in the spread of antibiotic resistance (Thomas & Nielsen, 2005).

Bacteria may possess natural resistance towards certain antibiotics, a phenomenon called intrinsic resistance. These bacteria are often sensitive towards other types of antibiotics. However, if bacteria carrying natural resistance additionally acquires resistance towards other types of antibiotics, consequences such as treatment difficulties, may arise. Antibiotics that would normally eradicate certain bacteria are no longer effective (Olivares et al., 2013). The frequency of these events varies. Transmission rate is dependent on several factors, and may vary significantly between different species. It may also depend on environment, on what resistance mechanism is being transmitted, on growth rate of the bacteria, and so on (Dzidic & Bedeković, 2003).

Growth rate and resistance development

Unrestrained growth of bacteria in favorable conditions typically follow an exponential growth curve. The growth is usually divided into four different phases; lag phase, exponential phase, stationary phase, and death phase. During lag phase, directly following inoculation of cells to growth medium, there is no proliferation of bacterial cells. Nevertheless, cells may possess metabolic activity during the lag phase. When cells start to proliferate, they enter the next phase; the exponential phase. Bacterial growth rate during the exponential phase is constant. The growth rate of the exponential phase is often denominated as the generation time or the doubling time; the average time it takes to double the size of a population. The generation times of well-studied bacteria are known. The times differ a lot from species to species; *Mycobacterium tuberculosis* has a generation time of 792-932 minutes, while *Staphylococcus aureus* has a generation time of about 30 minutes (Todar, 2006, pp. 68-69). When the growth reaches a certain limit, for example when all necessary nutrients are consumed, the growth curve flats out and enter the stationary phase. During the stationary

phase there is no population growth. Some cells may divide or die, but the population size remains constant. As in the lag phase, the bacterial cells may be metabolically active. Cells may for example produce secondary metabolites during the stationary phase. The last phase – death phase – is reached when a decrease in cell population occur due to cell death.

The duration of each phase largely depends on various factors such as available nutrients and other growth conditions. In nature the growth of bacteria may proceed differently than growth experiments conducted in a laboratory. In the laboratory, environment factors affecting growth may be controlled in a manner that is not achievable in nature (Monod, 1949).

The rate of which genes are transferred due to horizontal gene transfer is significantly lower than the rate of reproduction. Consequently, an acquired trait is spreading much faster vertically than horizontally. This leads to the assumption that if a trait is transmitted by horizontal gene transfer, affected bacteria would be outcompeted before the trait would manifest in a population. In theory, natural processes will prevent changes in a population as a result of horizontal gene transfer. However, studies show that this is not a reflection of the reality. Horizontal gene transfer has shown to constitute a relatively big part of the evolution of genomes. How can this be? A possible reason to this is the phenomenon of migration. Migration introduces foreign genotypes into the population. Natural extinction processes may be disturbed, and the horizontally transferred traits may endure in the population (Niehus, Mitri, Fletcher, & Foster, 2015).

Efforts to stagnate the development of antibiotic resistance

Simultaneous with the increase in antibiotic resistance, the development of new antibiotics stagnated. Regulations on the use of antibiotics for disease treatment was previously primarily based upon the outcomes of individual patients. Since the severe increase in antibiotic resistance, factors affecting the potential for further resistance development were to a higher degree included in these regulations (Gehring, Schumm, Youssef, & Scoglio, 2010). Changes in the restrictions and regulations are supremely due to research suggesting that the frequency of treatments with antibiotics, and the dosages used, have an impact on the development of antibiotic resistance. Research on this area is broad, and there are several different approaches and methods in use. Computational modeling is one such approach (Roberts, 2004). By considering different parameters that have an impact on antibiotic resistance in microbes, it is

possible to make simulations of how development will proceed (Arepyeva et al., 2017). There are several different modeling options for simulations of this kind.

2.2 Compartmental modeling of the spread of antibiotic resistance

Simple deterministic differential equations may be helpful tools in giving a rough overview of systems, and to determine relationships within systems. These equations may be used to design epidemic models called compartmental models. Compartmental models are based upon the idea that the individuals in a system are divided into compartments (figures 2, 3 and 4). Common compartmental models used in epidemiology are the SI (susceptible-infected), the SIS (susceptible-infected-susceptible) models, and the SIR (susceptible-infected-recovered). Individuals in these models move between the compartments of the model depending on their state; whether they are susceptible, infected, or recovered. The SI-model is the simplest of these models, where one assumes that the infection will spread until all individuals are infected. The infected state is thus a permanent state, until death possibly occurs. The SIS-model assumes that the infected individuals may return to the state of being susceptible. The individuals recover from the infection, and return to the state of being susceptible. Hence, not recovered and immune, as in the SIR-model. The SIR-model assumes that when an individual has recovered from the infected state, it is immune to infection of the same kind (Barabási, 2016).

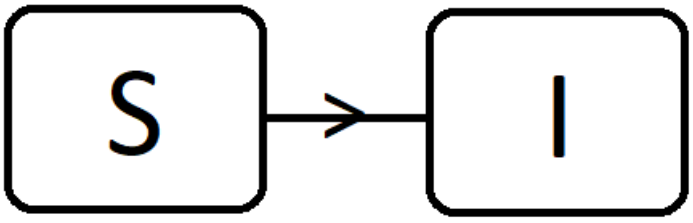


Figure 2: Flow chart of the susceptible-infected (SI) model. S = susceptible bacteria, I = infected bacteria.

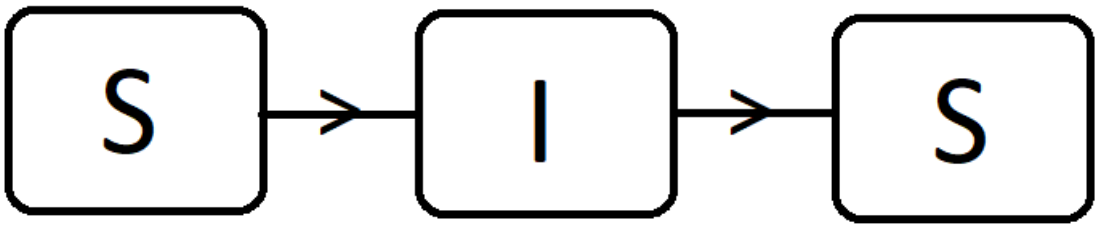


Figure 3: Flow chart of the susceptible-infected-susceptible (SIS) model. S = susceptible bacteria, I = infected bacteria.

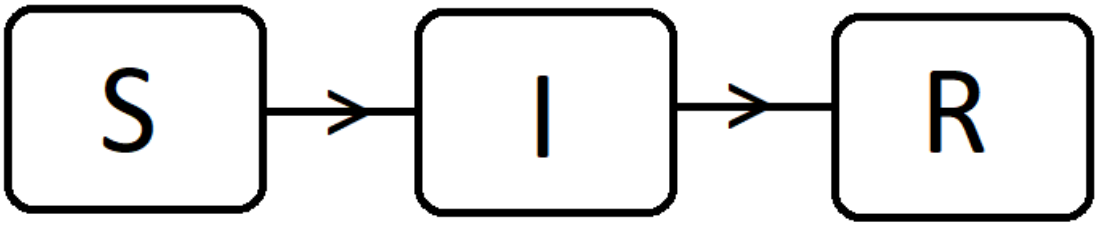


Figure 4: Flow chart of the susceptible-infected-recovered (SIR) model. S = susceptible bacteria, I = infected bacteria, R = recovered bacteria.

The SI, SIS and SIR models are built on the basis of differential equations, where in the simplest forms, the rate of infection and recovery is the main determining property.

We can define the change in number of susceptible individuals with the SI model as:

$$\frac{dS}{dt} = -\beta \cdot S \cdot I, \qquad (1)$$

$$\frac{dI}{dt} = \beta \cdot S \cdot I.$$

Here, β is the rate/probability of getting infected, S is the number of susceptible individuals, and I is the number of infected individuals.

We can define the change in number of susceptible individuals with the SIS model as:

$$\frac{dS}{dt} = -\beta \cdot S \cdot I + \alpha\ I \cdot, \qquad (2)$$

$$\frac{dI}{dt} = \beta \cdot S \cdot I - \alpha \cdot I.$$

As previously, β is the rate/probability of getting infected, S is the number of susceptible individuals, I is the number of infected individuals. Additionally, α is the rate/probability of the infected returning to the susceptible state (Barabási, 2016).

For the SIR model, we can define the change in number of susceptible individuals as:

$$\frac{dS}{dt} = -\beta \cdot S \cdot I,$$

$$\frac{dI}{dt} = \beta \cdot S \cdot I - \alpha \cdot I, \qquad (3)$$

$$\frac{dR}{dt} = \alpha \cdot I.$$

The parameters in equation 3 have the same meaning as in equation 2.

These models in their simplest form do not consider the influence of other factors, such as the localization of the individuals or randomness. Such models are referred to as deterministic models. Deterministic models are based upon knowing all the information of the parameters before calculations are conducted. Models that include randomness and the aspect of chance, are called stochastic models. Hence, if you run several simulations of the same stochastic model, the outcomes will vary depending on the sequence of random numbers calculated. While with deterministic models the outcomes will be the same for every run.

Say you have a network consisting of different compartments. Individuals in these compartments may be in different states at different time points. At each time point the individuals may with some probability either transit to another state or remain in their current state. The state of the individuals is determined by a process referred to as a Markov process. Markov chain models are probabilistic models, they do not follow a linear growth. Instead they have stochastic features. Each step of the simulation is independent from the previous steps, and the probability of change of state is always between 0 and 1 (Voit, 2012).

2.3 Epidemiological modeling

Biological networks almost always contain some form of variability and uncertainty, hence using deterministic models simulating development of for example disease may be unfavorable because it does not reflect the complete complexity of real-life situations.

Variations may occur depending on factors like climate, daily variations of contact patterns, the genetic properties of individuals, and previous infections. It is however a helpful tool in doing rough calculations of epidemiology. While stochastic models are suggested to be better fit for realistic probability simulations, they do not work well with large numbers of compartments and individuals. Deterministic models allow for higher complexity, and is commonly used for epidemic modeling (Voit, 2012; Andersson & Britton, 2000).

Metapopulation networks

In simple compartmental models, each compartment commonly represents different states; susceptible, infected or recovered/removed. The compartments may however represent something else, such as for this thesis; subpopulations. Within each subpopulation individuals are either susceptible, infected or recovered/removed. These types of models are referred to as metapopulation models. Metapopulation network models are networks of multiple interacting networks, and may present a more holistic view. They may allow a broader approach, where population dynamics are not only simulated in isolated populations, but in complete communities. Some form of immigration or interaction between the subpopulations of the community is required for the model to reflect a metapopulation network where spreading may occur.

The metapopulation network approach makes it possible to simulate population dynamics in subpopulations depending on both the natural dynamics within the population, and on the influence from the other populations in the community. Populations are rarely completely isolated from external influence in real life, and thus basing a simulation on isolated network models may be seen as quite naïve (Pastor-Satorras, Castellano, Van Mieghem, & Vespignani, 2015).

Geographical epidemiology

The geographical spreading of epidemics due to traveling has been a field of considerable focus the last few years. Detailed information about world-wide air-transport, for example obtained by the International Air Transport Association (IATA) database, is combined with known molecular data of pathogens. Researchers believe this approach may provide efficient

tools for developing successful strategies to prevent for example pandemic outbreaks (Hufnagel, Brockmann, & Geisel, 2004).

Outbreaks of diseases like, amongst others, influenza (Longini, 1988), and SARS (Hufnagel, 2004) have previously been modeled. The development of mathematical models for simulating the geographical spreading of diseases can be dated back several decades (Sattenspiel, 1990). It has long been of great interest knowing how pandemic outbreaks may be prevented. Although research during recent years has contributed to increased knowledge, this area of research is more of focus now than ever. Models are constantly tried, validated, and improved. Increasing knowledge and detailed information is an indispensable factor for the development of effective and reliable models. An applicable example of such models is the airline transportation network model by Colizza et al. (2006) Their contribution is a stochastic model that intends to forecast global epidemics based on the combination of worldwide airline transportation infrastructure, and registered population data. Their main aim of the model is to study how the airline transportation infrastructure influences global epidemics, and how spreading may be prevented; or at least inhibited to some extent. In addition, they aim to assess the authenticity of disease outbreak predictions and scenarios (Colizza, Barrat, Barthélemy, & Vespignani, 2006).

As with the air-transport models, detailed information such as registered travel patterns give a much higher degree of realism. Due to the development and the increased use of technology, information about people's whereabouts is a lot easier to obtain. For example, every time we use our cell phones we leave so-called digital footprints. This information may provide detailed information about our whereabouts, as well as how much time we spend each place and who we interact with (Frias-Martinez, Williamson, & Frias-Martinez, 2011).

Susceptible-infected-susceptible (SIS)

The SIS model is well suited for simulating the spread of antibiotic resistance. Bacteria is susceptible to receiving genes for antibiotic resistance, and will enter the infected state when having received these traits. While being infected, the bacteria is immune to receiving the same trait over again. Bacteria may also lose the received trait and thereby revert to the state of being susceptible. There is generally always a fitness cost related to receiving new traits. This cost is of great influence in the development of for example antibiotic resistance. A trait of no value in a certain environment, combined with high fitness costs is not desirable for the

bacteria. The newly retained trait is therefore likely to be lost within a period of time. The rate of receiving traits is a lot higher than losing them. In most cases new traits will impose changes in the genome of the bacteria which will be maintained throughout its life. Thus, loss of acquired traits is rarely seen at the individual level. The change may rather be seen in the population dynamics; the best fitted bacteria survive, and therefore ends up dominating the population. Despite the acquisition of a resistance traits usually coming with a fitness cost, the bacterial genome often accumulates compensatory mutations which in turn cancels the fitness costs. This is part of the explanation to why acquired traits tend to stay in the bacterial populations despite fitness cost in the acquisition process (Andersson & Huges, 2010; Dzidic & Bedeković, 2003). The degree of fitness cost varies depending on, among other things; type of species, type of drug, type of mutation, previous mutations or acquired traits, epigenetics, and growth environment (Melnyk, Wong, & Kassen, 2015).

This model in its simplest form does not take into account the population dynamics that is due to births and deaths. Including these parameters will possibly make the simulations more trustworthy for real-life situations. Considering population dynamics in the SIS model may be very complex. Yet making some assumptions may simplify the process. Such assumptions may be assuming a constant population size. The most elementary deterministic epidemic models are based upon the assumption that the individuals that are removed are immediately replaced by incoming individuals, and thus hold a constant population size (Greenhalgh, Liang, & Mao, 2016). On the other hand, making assumption may result in simulations being less reliable for real world situations. Assuming a constant population size could be somewhat approximate for short-time acting, rapidly spreading diseases. These are diseases that are not associated with death as a common outcome, or diseases where the disease-related deaths have no significance on the total population size. Diseases like smallpox and the Black Death are contrary to this. Smallpox and the Black Death had a crucial impact on the population size during the time they reigned the most. During this time, the death rate of the infected individuals was of a higher rate than the birth rate of susceptible individuals (Greenhalgh et al., 2016; Lahrouz & Settati, 2013). In slow spreading diseases, the population size will vary to some extent, but the population loss due to deaths will be somewhat cancelled out by the population gain due to births.

Susceptible-infected-recovered (SIR)

The SIR model may also be used for the simulation of spreading processes. Due to bacteria usually having a short life period, the bacteria will be removed from the network rapidly after being introduced to the new trait. The SIR model is similar to the SIS model, but different in that the infected individuals will not return to a susceptible state. The infected individuals will rather transit into a state called recovered or removed. Recovered means either that the individual has recovered from the infection, and is now immune towards being infected by the same agent again, or that the individual is removed from the network. The general SIR models with the assumption of a constant population size would be well suited for simulating the spread of diseases like influenza. Influenza spreads rapidly though the population and gives immunity to surviving individuals (Newman, 2002).

As mentioned, the SIR model includes the stage whereas individuals have recovered from infection, and are immune. While this is a suitable model for multiple spreading simulations, it is not optimal for the spread of antibiotic resistance in a population. This assumption is made on the basis of the fact that "recovering" from a horizontal gene transfer rarely happens. When bacteria acquire new traits due to horizontal gene transfer, as in this case a gene for antibiotic resistance, it will become a part of the bacterial genome. The bacteria will presumably remain infected throughout its life. This is due to the short lifespan of bacteria, as mentioned earlier. If, on the contrary to these presumptions, the antibiotic resistance trait is lost, the bacteria are not left immune towards gaining it again. Thus, SIR may not meet the criteria required for a model simulating the spread of antibiotic resistance.

Susceptible-infected(SI)

The last of the compartmental population dynamics models mentioned above; the SI model, is similar to the SIS model. However, this model differs in that the infected individuals has no point of return to the susceptible state. A SI model will reach an endpoint where all individuals in the system are infected. A spreading situation will thus persist indefinitely in the population unless infected individuals are removed in some other way.

Mass-action principle

A common assumption in population dynamics models is homogeneous mixing. These models follow a law called the mass-action principle. The mass-action principle states that in a population with homogeneous mixing, the rate of transition from susceptible to infected is proportional to the sizes of the two subpopulations. It states that despite a reduced population size, the number of contacts between individuals in the population is still, on average, the same (Anderson & May, 1992; Keeling & Rohani, 2008).

The process in which susceptible individuals transit into an infected state is referred to as the infection process. This transition is a function of the number of neighbors that are already infected, and it depends on the interaction between the different individuals in the network. Transition from susceptible to infected is therefore defined as an edge-based transition. The process in which the infected individuals returns to the susceptible state is referred to as the recovery process. This process is not affected by the surrounding individuals and is therefore defined as a node-based transition or nodal transition (Sahneh, Scoglio, & Mieghem, 2013).

Basic reproductive number, R0

In order for a spreading to occur in a population, certain requirements must be fulfilled. An example is the requirement of a certain threshold number of individuals within a population carrying the trait of interest. This requirement is commonly referred to as the basic reproductive number, R0. The basic reproductive number represents how infectious a spreading is. It is defined as the expected number of susceptible bacteria, from a completely susceptible population, that will get infected during a typical spreading situation. This particular parameter is highly relevant in a lot of disease research. It may be used to indicate the severity of a spreading. R0 may also be helpful in developing vaccine programs and procedures. It may be used to find estimates for how vaccination should be carried out to be as beneficial for the eradication of disease as possible. The basic reproductive number must be higher than unity for a spreading of a disease to be able to proceed in a population. If this requirement is not met, the spreading activity will cease. A high reproductive number is likely to result in a disease taking over a population (Dietz, 1993; Keeling & Rohani, 2008).

2.4 Network modeling

Population dynamics are conveniently visualized as networks. There exist countless different network models in use. Network structure and network properties may be customized to suit the problem at hand.

Continuous and discrete time modeling

Network models may follow what is referred to as a continuous time approach, or a discrete time approach. Whereas continuous models are suitable for continuously evolving systems, discrete time models are suitable for systems with fixed intervals (Pastor-Satorras et al. 2015). For population dynamics modeling based upon differential equations, a continuous time approach is more common. As for population dynamics, births and deaths commonly occur continuously and not in fixed time intervals. There are exceptions. Some insects have consistent breed patterns; they breed a certain number of times at certain times during the year. The number of times the organism naturally breeds per year is referred to as voltinism. Humans are semivoltine; they breed less than once per year. An organism that breeds only once a year is called univoltine. The population dynamics of univoltine organisms are thus better suited for discrete time models than the population dynamics of humans (Singh, 2004).

Data that depends on, and that are divided based on the time they were collected, are better suited for discrete modeling approaches (Iannelli & Pugliese, 2014). The population-based HUNT study (The Nord-Trøndelag Health Study) is an example. These studies are based on answers from questionnaires that carry on for a certain period of time. Data from the questionnaires are presented in time periods; HUNT 1 – from 1984-1986, HUNT 2 – from 1995-1997, HUNT 3 – from 2006-2008, and the latest and ongoing; HUNT 4. These questionnaires carry on for a certain period of time. Changes to the data may be seen from one HUNT-period to another, and thus discrete time approaches may be used (About HUNT, 2018). Nonetheless, one type of data set is not applicable for solely one type of model approach. Data may be used in different ways. The suitability of the model depends on, among other things, the aim of the modeling. One must decide which data are desired as results. Are the changes over time the main interest, or is the main aim to study the correlation between different factors? Such questions should be asked and answered in the process of selecting model approaches.

Erdős-Rényi and Barabási-Albert network models

One may be able to simulate how spreading patterns are affected by network structure. Valuable information may be obtained by observing differences upon comparison of results. Different network types display different network properties. Erdős-Rényi (ER) models and Barabási-Albert (BA) models for example, are network models of different network properties. The ER model is a random network model in which all nodes existing in the network has a probability *p* of receiving an edge. The degree distribution of ER models follows a Poisson distribution; the majority of the nodes have approximately the same number of edges, and there are no highly connected nodes.

While Erdős-Rényi models have a fixed number of nodes from the start, Barabási-Albert models continually grows by the addition of nodes with *m* edges. Addition of nodes occurs in a process called preferential attachment. Preferential attachment means that the new nodes prefer to connect to already well-connected nodes. The degree distribution of the BA model thus follows a power law; the majority of the nodes are sparsely connected, while a few of the nodes are highly connected; often referred to as hubs. BA models are referred to as scale-free networks. As a result of the degree distribution of ER models one would possibly expect a uniform spreading throughout the network. How aggressive a spreading will run through a network depends on, among other things, the characteristics of the spreading agent. Spreading patterns of scale-free networks may be tough to predict considering the existence of hubs in the networks. Spreading activity may be low for some time before it suddenly bursts when a spreading agent reaches a hub. ER models are less prone to model disease dynamics due to the absolute randomness (Barabási, 2016).

One way to cease a spreading event is by inflicting some sort of damage to the network that will hinder the access to uninfected nodes. Network damage is commonly referred to as the removal of nodes or edges. In relation to this thesis, targeted network damage is more relevant than random network damage. By targeting and inflicting network damage to the Achilles heel of a network; the most vulnerable point of the network, one may break the network into pieces. An active spreading situation will not be able to persist in a broken network. Targeting hubs is thus an effective way to eliminate spreading through a system. BA networks are relatively vulnerable to targeted attacks due to their highly connected hubs. BA networks are on the other hand more resistant to random attacks due to hubs only composing a small

fraction of the total number of nodes; the probability of affecting the highly connected hubs is less than affecting the sparsely connected nodes. The ability of networks to withstand pressure and damage is referred to as robustness. Network robustness largely depend on the size giant component; the largest connected node cluster of nodes, and the average shortest path; the distance between two nodes that includes the least number of edges (Barabási, 2016).

Approaches to modeling disease spreading

Previously, the compartmental models SI, SIS and SIR has been mentioned in relation to epidemiological network modeling. Other network models commonly used in this field of research are contract networks, multilayer networks, and agent-based networks.

While the SI, SIS and SIR models assumes homogenous mixing, the contact networks, the multi-layer networks, and the agent-based models contain more specific information about interaction and movement within the networks. Contact networks use information about social contacts to predict how infecting agents will spread through a network (Seilheimer, 2008). Multilayer networks contain multiple layers of nodes interacting. This means that one layer of the network containing nodes connected by edges, are in some way also connected to other layers of nodes connected by edges. Multilayer network allows for more complex network modeling (Kivelä et al., 2014). Agent-based modeling is a modeling approach where the individual is of focus. Agent-based modeling gives the opportunity to study how individual agents behave, interact, and affect a system (Bonabeau, 2002). This is contrary to population-based modeling, where the dynamics are observed in a population as a whole.

A population-based modeling approach is often preferred upon modeling a large number of individuals. Nevertheless, it is not said that a population-based modeling approach is correct for any type of population dynamics modeling. The correct approach may vary, and it largely depends on the aim of each individual case (Frias-Martinez et al., 2011).

2.5 Hospital networks

Nosocomial infections

Communities, as previously mentioned, may for example be hospitals. Outbreaks of antibiotic resistant bacteria in health institutions are of great focus today. This is especially due to the vulnerability of the patients and the increased chance of fatal outcomes. Infections acquired in hospitals are referred to as nosocomial infections. Infections may be caused by cross-contamination and cross-infection, or by endogenous infection. The latter being that infectious agents are present in patients at the time of their admission, but not yet having caused infection. An infection is more likely to manifest in a patient when the immune system is altered, as commonly occurs during hospital stays (Chartier et al., 2014).

Both healthy and sick people carry a considerable number of bacteria. These bacteria are defined as the normal bacterial flora, and it is vital for normal bodily functions of humans. Among other things, the normal flora is important for the natural defense against pathogens. These bacteria may however cause infections if they are introduced to environments they do not belong to. Bacteria will only cause infection if the number of bacteria exceed the critical number. The critical number is defined as "the lowest number of bacteria, viruses or fungi that cause the first clinical signs of infections in a healthy individual" (Chartier et al., 2014). People may also be carriers of pathogens without being infected themselves. Such symptomless carriers are hard to discover, and they may pose a threat to patients with compromised immune systems. Typically, carriers that poses the biggest threat to patients are the health workers at the hospitals. Health workers that have been in contact with patients with certain diseases are required to get tested before being allowed back into patient care. However, without symptoms of infection or other indications of presence of pathogen, testing is rarely conducted.

Selective pressure

In addition to hospitals being environments dominated by immunosuppressed patients, hospitals are dominated by extensive use of antibiotics. This may constitute a larger threat in terms of developing resistance due to selective pressure. Bacteria with the most optimal phenotype to a certain environment will survive and proliferate to a greater extent than bacteria of less optimal phenotype (Roberts, Kruger, Paterson, & Lipman, 2008). This

phenomenon does not exclusively apply to hospital environments, but to all environments of excessive presence of antibiotics. For instance, the extensive use of antibiotics in agriculture contributes a considerable amount to the overall spreading of antibiotic resistance. A significant amount of the antibiotics used in agriculture is secreted out into the free environment. In the environment, antibiotics are diluted to non-lethal concentrations, which in turn may contribute to selective pressure. The result may be alterations to the microbiological compositions with increasing ratios between resistant and susceptible bacteria (Levy, 2002; Gullberg et al., 2011).

2.6 Hospital hygiene

Cleaning

The primary strategy to keep hospital infections to a minimal is by isolation of infected patients, as well as cleaning. Eliminating all cross-contamination is rather impossible, but measures may be done to minimize cross-contamination within hospitals. Most hospitals have established standard procedures for handling contagious infections, and standard procedures for cleaning. Cleaning primarily include mechanical removal of visible dirt with the help of water, soap, and detergents. Proper cleaning has the potential of eliminating over 90% of microorganisms, even without the inclusion of antimicrobial products. In addition to removal of the microorganisms themselves, the bacterial breeding-ground may be eliminated. Improper and sloppy cleaning, on the other hand, may cause the opposite outcome. The use of contaminated equipment at new places may for example spreading contaminants to uncontaminated areas (Chartier et al., 2014).

Despite a common belief that cleaning is important for disease prevention, the importance of cleaning in hospitals, as a part of infection control, has received surprisingly little attention. Nevertheless, research has been carried out, with results indicating that cleaning is an important factor in lowering the risk for nosocomial infections. A selection of studies is listed below:

- Wojgani et al. (2012) conducted an investigation to study the impact of design, location, and the usage of door handles, on contamination of microbials. They found a clear correlation between microbial contamination on door handles and the degree of

use. The other factors, such as design; lever handles, pull handles, or push plates, and location were also proved to influence the degree of contamination. The authors concluded that neglecting door handles in hospital cleaning procedures, possibly would have serious impact on the transmission probability within a hospital (Wojgani et al., 2012).

- Marshall et al. (1998) discussed in a letter the discovery of MRSA in a hospital ward. The discovery highlighted the significance of proper cleaning. Samples from a new hospital ward were collected, both before and after opening. 13 locations were tested prior to opening, and 25 locations were tested weekly for six weeks. One of the 13 locations showed presence of *Staphylococcus aureus* prior to opening of the ward. The samples from within the first week of opening showed presence of *Staphylococcus aureus* at five locations, whereas one of the strains turned out to be MRSA. There was no knowledge of any MRSA-infected patients being admitted during the first week. The conclusion to this discovery was that the bacteria was introduced to the new ward through dust on old furniture that was moved into the ward prior to opening. In this letter, Marshall et al. (1998) therefore stressed the importance of proper cleaning procedures in infection control (Marshall, Sen, Chadwick, & Keaney, 1998).
- Krilov et al. (1996) conducted an experiment in a specialized preschool to determine the value of infection control programs. Cleaning constituted a significant part of this infection control program, and they made changes to their old cleaning routines. Changes included the order in which the rooms were cleaned, more frequent and consistent change of mop water, and proper rinsing of mops and buckets at the end of the day. Additionally, they established procedures for cleaning of the toys that were used by the children, and they distributed information on proper infection control. Their results showed a decrease in number of acquired infections at the preschool (Krilov et al., 1996).

Hospitals house a great diversity of microorganisms, both in patients and on surfaces. Infecting agents may remain in the environment after successful treatment of infected patients. The survivability of microorganisms varies from species to species and from strain to strain. A common feature is however that poor growth conditions are unfavorable. Making the environment less optimal for growth is thus a significant measure (Dancer, 1999). Some areas are more prone to serve as transmission “hot spots”. Door handles, as mentioned in the study above, may be examples of such “hot spots”. Door handles are frequently touched, and less

frequently cleaned and disinfected. Additionally, door handles may be touched by a large variety of people over a short period of time. Thus, bacteria may easily be transferred despite dry surfaces composing poor growth environments for most bacteria. Hospitals contain a multitude of health workers, patients, relatives, visitors, and other personnel. Some areas are strictly prohibited for others than authorized personnel, while other areas are accessible for all people; some areas are defined as clean – or even sterile – areas, while other areas are not given much extra attention when it comes to cleaning.

Sterilization

Sterilization is another method used for eliminating microorganisms. A sterile environment should in theory be 100% clear of microorganisms. Sterilization will however not be able to eradicate the absolute number of microorganisms, but by proper sterilization 99.9999% of the microorganisms could be eradicated. Sterilization should be combined with common cleaning. There are different methods of sterilization, both chemical; by gas treatment or use of disinfectants, and physical; by heat treatment, irradiation, or filtration. (Chartier et al, 2014). Increased use of detergents and disinfectants are not necessarily an advised action for hospitals. This is due to research suggesting that excessive use of certain substances may induce resistance to microorganisms. Not only is it suggested that the microorganisms may become resistant to the substances of use, it is also suggested that it will have a possible impact on the resistance towards antibiotics (Levy, 2000).

Biolfilms

Microorganisms have a remarkable ability to adapt to changes in their environment. An example of such adaption is their ability to form biofilm on surfaces. Biofilms consists of microorganisms that are attached to a surface, and often to each other. These cells become embedded in a matrix of their own production, a matrix of so-called extracellular polymeric substances. Microorganisms that constitute these biofilms behave somewhat different than free cells, planktonic cells. Due to for example limitation of nutrients, biofilm cells tend to have a slower growth rate than planktonic cells (McDonnell & Russell, 1999; Donlan & Costerton, 2002). Reduced growth rate may, in addition to the protective layer matrix, affect how susceptible the bacteria are to cleaning and antimicrobial agents. Tuomanen et al. (1986)

showed in their research that slow growth rate reduced the effect that penicillin and other β-lactam antibiotics had on the bacteria. Evans et al. (1990) conducted a similar study. They aim was to establish the effect growth-rate had on Gram-negative biofilms' resistance to the antiseptic cetrimide. Their results showed that in a biofilm of the Gram-negative rod bacteria *Escherichia coli*, the slowest growing bacteria was the least susceptible to cetrimide (Evans, Allison, Brown, & Gilbert, 1990).

The biofilm matrix functions as a protective shield for bacteria. It protects the bacteria from external stress such as drying, washing, and the use of disinfectants (Flemming & Wingender, 2010). Biofilms may thus constitute a relatively great threat to the challenges in eradicating bacteria from locations. Biofilms may consist of several bacterial species. The different species and different strains may exchange genetic material within the biofilm. In fact, it has been suggested that the exchange is more efficient between biofilm cells than between planktonic cells. How efficient the transfer of genetic material is within biofilms, and how phenotype and gene transfer is related, are questions discussed further in a review article by Molin et al. (2003). They suggest that efficient gene transfer is an important factor in adapting to environmental changes, and that it is both the motive and the consequence for the evolution of biofilms (Molin & Tolker-Nielsen, 2003).

A need for infection control strategies

With this knowledge, and the studies presented above, one could ask why hospital cleaning is not of greater focus today. There has been development towards the better, but there is still a considerable improvement potential in this regard (Mehta et al., 2014). We do not only require guidelines such as presented by Mehta et al. (2014), we require the guidelines to be followed, revised and improved at any given time.

Tools for analyzing possible outbreaks and their spreading pattern would be of high value in order to develop and evaluate strategies to eradicate potential hazards. It could, among other things, provide insight to the effect of different infection control strategies. It opens for the possibility of testing and evaluating different scenarios without conducting them in real life. A possibility that may save both time and money in the long run.

Staphylococcus aureus

One of the most common bacterial strains responsible for hospital infections is the bacteria *Staphylococcus aureus*, and its resistant strain methicillin-resistant *Staphylococcus aureus* (MRSA) (Horan et al., 1986). *Staphylococcus aureus* is one of two main constituents of the *Staphylococcus* family. The bacterium is a natural part of the human flora, primarily on skin and mucous membranes; mucosa. It is estimated that as many as 30-60% of personnel at hospitals are carriers of *Staphylococcus aureus* (Chartier et al., 2014). *Staphylococcus aureus* contains the enzyme catalyze, which catalyzes hydrogen peroxide (H_2O_2) into water (H_2O) and oxygen (O_2). This property makes the bacterium able to grow despite the lack of oxygen in surrounding environment. Due to its rigid, thick peptidoglycan layer, *Staphylococcus aureus* is able to survive at dry surfaces for relatively long (Murray, Rosenthal, & Pfaller, 2012).

Staphylococcus aureus may develop resistance towards the antibiotic penicillin. This resistance is due to the production of the enzyme beta-lactamase (penicillinase), as previously explained. This is a common resistance in *Staphylococcus* strains. In fact, less than 10% of *Staphylococcus* strains are susceptible to beta-lactam antibiotics at current time (Murray et al., 2012; Solberg, 2000). Over the years certain strains of *Staphylococcus aureus* have developed resistance against several other, later developed, semi-synthetic penicillin. The multi-resistant *Staphylococcus* called methicillin-resistant *Staphylococcus aureus* is resistant against all beta-lactamase antibiotics, as well as methicillin, hence the name (Murray et al., 2012). The first documented case of MRSA is dated back to 1961, in the United Kingdom. Additional cases were reported from the United States in 1968 (Jevons, 1961).

The spread of MRSA is identical to the spread of *Staphylococcus aureus*; directly through contact between individuals, or indirectly through contact with for example door handles. Infection through indirect contact is relatively easy to prevent by common hygiene measures. This is partly due to the number of bacteria required for inducing infection; the infectious dose, is higher than by direct contact. However, patients in hospitals generally have a weaker immune system than normal and are more susceptible to infection despite thorough hygiene measures (Murray et al., 2012). Though the bacterium *Staphylococcus aureus* is common as normal human flora, it may cause infections. Wound infections are most commonly seen in conjunction with *Staphylococcus aureus*. If infections are caused by the antibiotic resistant strain, MRSA, successful treatment is more challenging. Despite *Staphylococcus* quite readily developing resistance towards several types of antibiotics, there exists antibiotics that will

eradicate the bacteria. The most common treatment of MRSA is the antibiotic vancomycin. Vancomycin is an antibiotic that blocks the formation of the cell wall in gram-positive bacteria. Today, *Staphylococcus* strains have developed resistance to vancomycin as well. These strains are called VRSA (Tenover, 2006). Due to the level of immunosuppression among the community in hospitals this is an alarming and critical development.

The increased world-wide development of antibiotic resistance clearly demonstrates the importance of efficient prevention plans. Such plans may include, among other things, more appropriate use of antibiotics, better cleaning routines and better infection control. Yet, we will not be able to stop all development of antibiotic resistance by the use of these methods. We must consider the natural aspect; antibiotic resistance is not human made, it is a natural consequence of evolution. And natural evolution is not eliminable. We are however able to slow down the process of further development (Spellberg, Bartlett, & Gilbert, 2013).

3 Material and methods

3.1 The model

The network model was based upon the compartmental susceptible-infected-susceptible (SIS) model, a system dynamics approach (Homer & Hirsch, 2006). Each compartment in the network constituted their own SIS model, based on simple differential equations:

I) $$\frac{dx1}{dt} = r1 \cdot x1 \cdot \left(1 - \left(\frac{x1 + x2}{x2max}\right)\right) - b \cdot x1 \cdot \frac{x2}{x1 + x2} + beta_2 \cdot x2 - d \cdot x1 \cdot del_A \cdot x3$$

II) $$\frac{dx2}{dt} = r2 \cdot x2 \cdot \left(1 - \left(\frac{x1 + x2}{x2max}\right)\right) + b \cdot x1 \cdot \frac{x2}{x1 + x2} - beta_2 \cdot x2$$

III) $$\frac{dx3}{dt} = ab_{supply} - x3 \cdot ab_{uptake}$$

Where:

r1 = the growth rate of S

r2 = the growth rate of I

x1 = S

x2 = I

x3 = the concentration of the antibiotic

b = beta – the rate of flow from S to I

$beta_2$ = beta2 – the rate of flow from I to S

x2max = bacterial_load – the maximal number of bacteria that may exist in a patient

del_A = delA – the effectiveness of the antibiotic

ab_{supply} = ab_suppply – the supply rate of antibiotics during treatment

ab_{uptake} = ab_uptake – the uptake rate of antibiotics during treatment

These differential equations were implemented in the model through an input file that upon simulation was run as a program file in a command prompt. In addition to the network input file, other input files containing information about node and edge parameters were included in the program file. In this way a network model was created.

Random network generator

If the aim is to simulate large, random networks (see Results and Discussion for examples), specified program files may be used to generate either Barabási-Albert (BA) networks or Erdős–Rényi (ER) networks. The number of nodes, edges and edge probability may be adjusted as desired. In relation to this thesis, seven different BA networks and one ER network were generated (table 1). *m* represents the number of edges to be attached from new nodes of the network to already existing nodes for BA networks and *p* represents the probability of an edge connecting each pair of nodes in the network for ER networks.

Table 1: Table displaying network structure for the Barabási-Albert or Erdős–Rényi networks generated in relation this thesis.

Network type	Associated with	Number of transit nodes	*m*	*p*
Barabási-Albert	Hypothesis 1	100	1	
Barabási-Albert	Hypothesis 2	1000	5	
Erdős–Rényi	Hypothesis 2	1000		0.01
Barabási-Albert	Hypothesis 3	200	5	
Barabási-Albert	Hypothesis 4, part 1	500	5	
Barabási-Albert	Hypothesis 4, part 2	500	5	
Barabási-Albert	Hypothesis 4, part 2	1000	5	
Barabási-Albert	Hypothesis 5	1000	5	

In cases where networks are small, and structure is known, it may not be desirable to generate random networks. The network model tool is therefore implemented with the possibility of creating custom made networks. This may be done by changing the parameter values of the transit nodes and the transit node network structure.

The simulation process

The program file that encodes the actual network simulation initiates a simulation and a sensitivity analysis of the network. Input files necessary for this command to work are files containing information about the edges of the network, and files containing information about the nodes of the network. Additionally, a file containing information about varying parameters is required. The output files from the simulation program file are not easy to interpret without further processing. The files are thus further used as input for other program files to retrieve more understandable and interpretable results. Information such as the total number of bacteria in the network, the number of nodes infected with S bacteria and I bacteria, the number of transit nodes infected with S and I, and the number of S bacteria and I bacteria in the nodes at each iteration, is information provided by these program files.

Network visualization

In order to achieve a more presentable visualization of the result, a program file generating files viable for presentation in Cytoscape, is run. Cytoscape is a software that, among other things, may be used to visualize interaction networks in a tidy and apparent manner. The software is implemented with a lot of various options which makes it convenient for analysis of data sets, or solely for visualization purposes (Shannon et al., 2003).

The simulation may additionally be visualized dynamically as a video file. See appendix 1 for more detailed description of the components of the tool.

Population dynamics

As previously mentioned, simple population dynamics models may not be exquisite for simulating real-world problems. However, by doing some modifications to the equations simulations can be made quite realistic. A common simplification for population dynamic models is the assumption that the population size is constant. In a spreading simulation of antibiotic resistance this assumption may be based on that the acquisition of a resistance trait is not a cause of death. It is merely a trait that may or may not come with a fitness cost. Bacteria with acquired resistance rarely return to the susceptible state within their lifetime, but the distress due to high fitness cost may lead to an early death. In the real world, births and deaths are significant properties of bacterial populations. This property is significant for the

simulation of the spread of antibiotic resistance because acquiring antibiotic resistance may be a result of cell division, as well as by direct horizontal gene transfer. Thus, by excluding the factor of birth the model will be incomplete.

Bacteria have a short life span and a high growth rate. If no other interfering factors are present, the number of births is estimated to be cancelled out by the number of deaths. This is due to the assumption that the birth rate and death rate is approximately equal, and thus the number of individuals will not fluctuate significantly. Therefore, one may assume a constant population size. The variations in such models will rather lie in the ratio of susceptible individuals and infected individuals.

What complicates the network model of this thesis, is the inclusion of interaction between the different compartments/subpopulations; inter-compartmental movement, as well as the aspect of immigration and emigration; the option of connecting/disconnecting patient nodes to/from transit nodes. The model is built to simulate an overview of a dynamically changing population, not a fixed population following only the initial individuals. The assumption is made that population dynamics is both a result of natural fluctuations; births and deaths, but also due to the aspect of immigration, emigration, and cleaning. These are parameters that may be adjusted based upon the intention of the simulation. The population size is directly related to the variations of these parameters. If the parameters are set to cancel each other out, the population size will remain constant. The last S (SIS) represents susceptible individuals that are new to the system, either due to birth, or due to immigration from other compartments or from the external environment.

Inter-compartmental movement

To make the simulations more life-like and probable I have designed a model with multiple compartments. The interest of this thesis is observing the dynamics of the populations in each location on the basis of which traits they hold; whether the individuals in the compartments are susceptible or infected. More importantly however, is observing how spreading proceeds through the complete network. A flow chart of the network model is shown in figure 5.

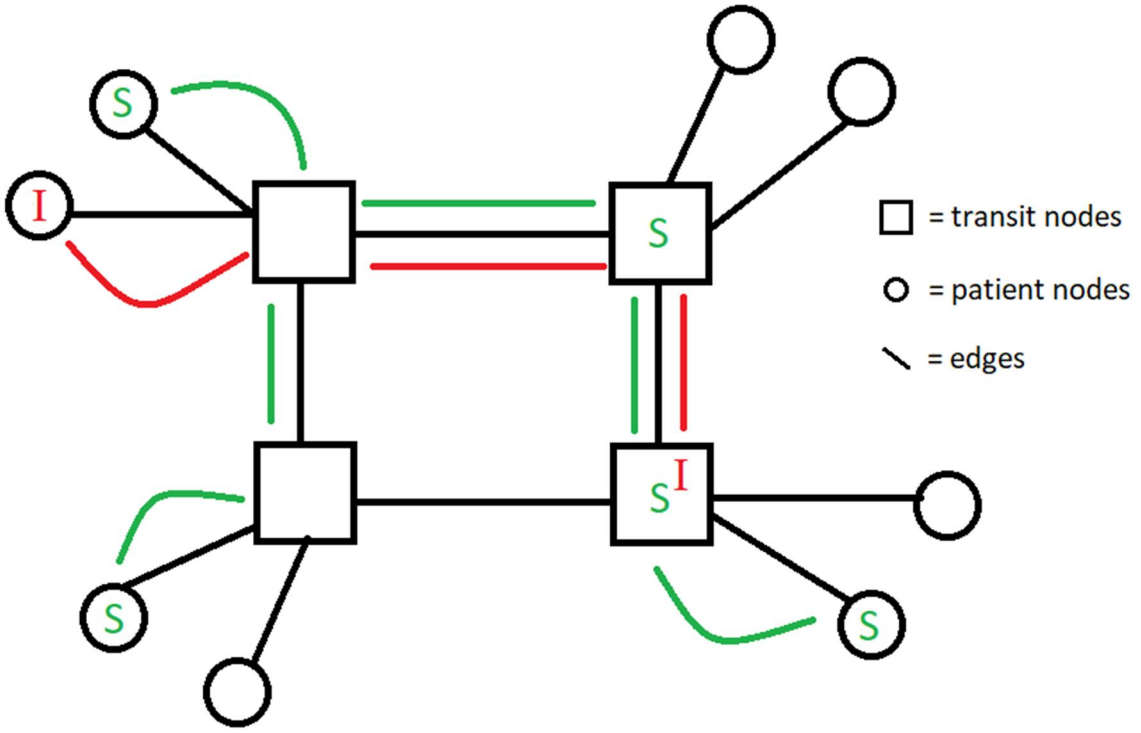


Figure 5: Schematic flow chart of the network model. Round nodes = patient nodes, square nodes = transit nodes, S = susceptible bacteria, I = infected bacteria, green lines = possible transfer route for S bacteria, red lines = possible transfer route for I bacteria.

Each compartment represents a location with a population. This is simulated by nodes being connected by edges. Edges represent the correlation amidst the connected nodes; the transmission flow of individuals from one compartment/node to another. The complete network of transmission possibilities constitutes a contact network. The actual movement of individuals from one compartment/node to another makes out a transmission tree. The transmission tree uncovers the route of transmission through the network – the spreading pattern. Mapping out a transmission route within the contact network is a complex process. It requires lots of detailed information. Information that one may not have. Thus, approximations and simplifications are often incorporated in epidemic dynamics models (Eames, Bansal, Frost, & Riley, 2015).

Assumptions, approximations, and simplifications are however not necessarily equivalent to unreliable and incorrect. As long as results are considered and interpreted carefully with a critical mind, they may be of high value. Results may serve as suggestions for possible outcomes. Suggestions being the operative word. We are not certain of the impact of any potentially missing information. We do not always know the essentiality of parameters to the networks and their behavior.

Results may also be compared to known biological data to determine validity. Simulations run on network models that are built identical to real networks gives the possibility of result comparison. Differences and similarities may be used to map out weaknesses and strengths, and possibly identify and separate essential and nonessential parameters.

Population interaction

Another important factor to include for population dynamic simulations to be realistic is population interaction. The simplest models assume that all individuals in an entire network are in contact with each other. This is however rarely the case with biological systems. Take a hospital as an example; assuming all bacteria in the entire hospital community interacts at the same level is not very realistic. Different bacteria may favor different environmental factors for survival, and they may reside at different locations. The inclusion of network topology in these models may make a big difference as to how realistic outcomes are. In our network model this is considered by the inclusion of different movement patterns between the different compartments. The edges are weighted. The aspect of homogenous mixing is however not eliminated entirely. Transition from the susceptible state to the infected state depends on the interaction with other individuals in the network. Due to lack of specific information about the interaction pattern within the bacterial populations being simulated, this model follows the most basic approach; assuming all individuals within the separate populations are in contact with each other. It is assumed that they interact at the same level; that each separate subpopulation follows the mass-action principle. The transition from the infected state to the susceptible state is however more complicated. For our network model this transition means the replacement of infected individuals with new susceptible individuals. The transition may be dependent on various factors which have to be defined for the specific model of use. Removal of individuals may for example be due to cleaning. The extent of the cleaning, the time intervals between each time the location is cleaned, and the effectiveness of the cleaning supplements being used, are among other things factors that may be defined. The sum of such factors will, together with natural death rate, constitute the rate of removal from the system (Pastor-Satorras et al., 2015).

Cleaning intervals

Our network model is built with the possibility to adjust time intervals for cleaning. Due to lack of detailed data, the effectiveness of the cleaning is in the model determined by a random probability. For each round of cleaning each transit node exhibits a certain probability of being cleaned. A transit node of high connectivity and throughput may or may not be cleaned during a round of cleaning, and it may or may not be cleaned during the next. These contingencies may affect the system, and the outcomes may be quite dissimilar for each completed round of cleaning. The reliability for a real-world simulation is thus decreased relative to a simulation based upon detailed and accurate data. On the other hand, it may be rather impossible to retrieve accurate detailed data of this sort. Nature plays a great part in such parameters; despite impeccable procedural training, humans are unable to behave identically from time to time. Variation and coincidence will always exist in such cases.

Weighted edges

The edges in the network model are weighted. By adding weight to the edges, a rather important inaccuracy may be eliminated. If all edges in the network were equally weighted, spreading would in theory always persist even if only a few infected individuals were present. Not necessarily at a high rate, but it would still persist. Within the hospital network model different compartments represent bacterial populations at different locations. Different locations may be patient rooms, personnel rooms, operating rooms etc. The movement pattern between these locations may differ a lot, and it may constitute a relatively great difference in transmission probability. If the edges connecting these locations were equally weighted, there would be an equal chance for bacteria to be transferred to either of the other locations. This does not reflect the reality. In reality, certain locations have a higher or lower turnover rate, and edges connecting the locations have a higher or lower transmission probability. With weighted edges there would thus be a chance of infected individuals to be present without there being any actual spreading through the system. If infected individuals are present in populations at locations where transmission probability is low, a spread may cease naturally. Eliminating all infected individuals from a system like a hospital is virtually impossible, and it is not an applicable approach. This unrealistic factor is eliminated by adding edge weight as an adjustable parameter. It is assumed that all bacteria may be transferred in both directions.

Examples of other parameters included in the network model that may affect the simulation outcomes are: the number of iterations, the number of initial patients, the maximum number of patients, the growth rate of susceptible (S) individuals and infected (I) individuals, the number of patients entering and exiting the model during each iteration, the proportion of bacteria being transferred between nodes, the maximum number of bacteria in a patient, the proportion of patients infected with S bacteria and I bacteria, the time at which I individuals enter the network model, the start time of antibiotic treatment, the flow rate of S individuals becoming I individuals – and vice versa, the effectiveness of antibiotics, the supply rate and uptake rate of antibiotics, the frequency of cleaning, and the life-time of bacteria in the transit nodes. See appendix 2 for full overview and explanation of network parameters.

Parallel simulations

To determine which actions and approaches that are more favorable than others, several parallel simulations with different parameter settings may be conducted. By running several parallel simulations on identical networks, while varying only specific parameters, one might be able to determine their significance and function. Certain parameters are highly sensitive. Small changes to these parameters may result in large variations of the outcome. Other parameters may need considerable changes to cause noticeable changes of the outcome. Some parameters are necessary for a network model to work properly, but unimportant in terms of changes and variations of the outcome. Significance and function of parameters is not necessarily established beforehand. It is nevertheless information that is vital to our simulations, and our results. We are not able to interpret a result if we have no knowledge of how and why. Statistical significance and correlation was calculated for results from several of the parallel simulations by running regression analysis in Microsoft Excel (2016).

For this thesis, simulations were run on both large random networks, and on a smaller and more empirical network. Simulations were run with varying values of for example cleaning interval, growth rate, and fitness cost. Parameters were chosen on the basis of hypotheses that were outlined during the planning phase of the thesis. Simulations were run on both Erdős-Rényi (ER) and Barabási-Albert (BA) networks with identical parameter settings. The main purpose of performing identical simulations on different network types was to study the effect of network structure on spreading processes. The main example simulation was run on a hospital model with data included from a bachelor thesis from 2015 (Brunes & Finstad,

2015). The simulation was run on a partly empirical model, and it was thought to reflect the application value of the tool in a more realistic manner. The hospital model was designed as a network model with a number of 48 nodes and 199 edges.

Susceptible-infected-susceptible (SIS) model

The network model of this thesis is based upon the susceptible-infected-susceptible (SIS) model. At the initiation point of the simulation there are no infected individuals present in the system. The I (infected) is introduced due to immigration from the external environment. The model accounts for immigration and emigration, as well as internal dynamics such as movement, genetic exchange, births and deaths. The patient nodes, as you can see while running the simulations, may suddenly disappear. The patient nodes disappear simply due to emigration from the network. This is a simplification of the reality, where patients may also disappear form the network due to death, relocation, etc.

The network model consists of nodes representing subpopulations at specific locations, and edges representing the connections or flow of individual between these populations. The model simulates interaction at more than one level; within the subpopulations at each location, and between the different subpopulations. Thus, a network inside the network, as discussed in Background and theory.

Patient and transit nodes

The model is built somewhat different than the majority of other simple epidemic models. It includes both transit nodes; which are in a fixed structure, and patient nodes; which are in constant movement. The transit nodes are represented in the network as square nodes, and the patient nodes are represented as round nodes. Every node represents a bacterial population of constant evolvement. This evolvement is determined by differential equations in a SIS model. Figure 6 shows a simplified schematic of a network with transit nodes and patient nodes.

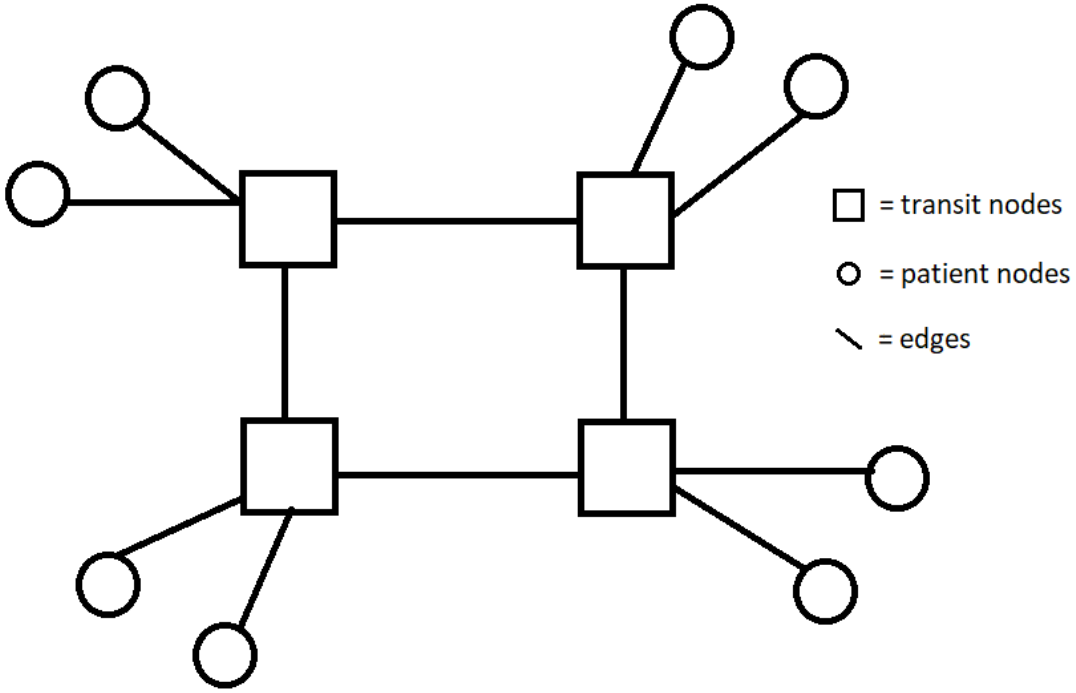


Figure 6: Simple schematic of a network with transit nodes and patient nodes.

Most transition from one state to another occurs in the round nodes. That is, in the nodes representing the carriers of bacteria that move through the system. This is due to bacterial growth and dynamics being higher in optimal environments, such as in the body of a patient, rather than at dry surfaces in the hospital. However, as previously mentioned, some bacteria have the ability to survive at dry surfaces. Thus, the bacterial populations at dry surfaces should not be neglected in such a simulation. These locations are represented by the transit nodes in our model. The transit nodes may cause outbreaks of disease despite efficient treatment of the patients infected with the bacteria. Examples are shown in Results and discussion. The transit nodes work as bacterial pools from which bacteria is obtained by patients/hospital workers, and others, and transferred to another patient or location. The patient nodes do not include hospital staff and visitors. Hospital staff and visitors are however not excluded from the network. Hospital staff, visitors, and other people that may for different reasons be relevant, are included in the network model through the edges. The transfer between the nodes are dependent on their movement, and thus constitutes the edge weight as explained previously.

Approximations, assumptions and simplifications

Simplifications of the reality are made at several places in our network model. It is important to stress that the network model is in no way a true reflection of the reality, and that results must not be interpreted as such. The network model is merely a simplified sketch of a virtual reality. It is intended and developed as a tool to help investigate the effect of different network structures and parameters on spreading pattern. Results should not be interpreted as the whole truth. Nevertheless, the results may be used as supplementary data, and they may be good indicators for further study and action. The intention of this thesis is not to invent a solution to all challenges of antibiotic resistance. The intention is rather to show that computational modeling may be highly valuable as an additional apparatus in the fight against the spread of antibiotic resistance. In other words, it constitutes only a small piece of a large puzzle.

Despite the improvements and progress seen the last few years in this area of research, computational modeling still faces challenges in making models that reflect the real world entirely.

3.2 Collection of material

Material for this thesis was collected in various ways. The first part of the thesis description proposes that extensive literature search should be conducted. Literature search was conducted to find theory for the thesis as well as researching already existing methods and models for modeling the spread of antibiotic resistance. Literature search was conducted both on the internet to find relevant research papers and review articles, and in the university library at the Norwegian University of Science and Technology (NTNU) in order to find books to cover basic theory.

Bachelor thesis

The main simulation for this thesis is a hypothetical simulation of a hospital. The hospital visualized is virtual, but some of the conditions for the simulation itself were obtained from a

bachelor thesis approved in May 2015, written by Tonje Brunes and Nora Rosvoll Finstad (Brunes & Finstad, 2015). The topic of this bachelor thesis was microbiological examination for the presence of antibiotic resistant bacteria on equipment and inventory at different departments at Orkdal hospital. The thesis was written during a period of which this topic was highly relevant. This was due to recent and more frequent discoveries of certain antibiotic resistant bacteria at health institutions, such as Orkdal hospital. Several cases of infections caused by vancomycin resistant enterococcus (VRE) were in 2014 registered at St. Olavs hospital. The spread was presumed to originate from St. Olavs hospital, department Orkdal hospital. The chief of infection control at St. Olavs hospital, Andreas Radtke suggested that the bacteria entered the hospital by unknowing carriers of the bacteria, or by insufficiently cleaned equipment (VRE sniker seg inn på norske sykehus, 2014).

A research collaboration between St. Olavs hospital and NTNU called The Operating Room of the Future (FOR) were the initiators of the bachelor thesis. The main ambition for FOR is conducting research that will affect patient care, logistics and architecture of operating departments, in a safer and more efficient manner (FOR – The Operation Room of the Future, 2017).

Focus for the bachelor thesis was extended from concerning only VRE to concerning all antibiotic resistant bacteria. Several microbiological samples were collected from equipment and inventory at Orkdal hospital. Collected samples were cultivated and further differentiated. The microbes of interest were typed, and resistance tested. The number of microbes at each test spot was indicated by the number of colonies on the growth medium after incubation. The microbial concentration in the samples were unknown. However, a common approach is to assume that it is required an amount of over 1 million bacteria to give a visible colony. For the model simulation of this thesis, this assumption is followed. Thus, the result was a semi-quantitative indication on the amount. Testing was conducted at several different locations and wards at the hospital, and the number of tests varied from place to place. The results were not qualified to give precise comparisons, and they are merely used as indications. For the simulation in this thesis a hypothetical hospital ward was visualized. The number of bacteria at the different locations was based on the bacterial findings in the bachelor thesis.

The hypothetical simulation conducted of the hospital only considered the common bacterium *Staphylococcus aureus* and the methicillin-resistant *Staphylococcus aureus*. During the testing at Orkdal hospital, methicillin-sensitive *Staphylococcus aureus* was found at several places. Both in areas defined as clean and unclean. *Staphylococcus aureus* is a part of the normal

flora of humans. *Staphylococcus aureus* is therefore expected to be detected even at locations without infected patients. Since the strain is capable of acquiring resistance, which is of great concern in hospitals, the simulation is based upon the spread of this bacterium.

3.3 Orkdal hospital

At the time of the bachelor thesis (Brunes & Finstad, 2015), Orkdal hospital was organized somewhat differently than today. Due to this master thesis being based upon the research conducted in 2015, the information presented in relation to the hospital will primarily be based upon the former data.

The four wards of focus in 2015, was A3, B2M, B3 and B4.

A3 is a 5-day unit, which is closed on weekends. Opening hours are from Monday at 7 am to Friday at 10 pm. Most procedures at this ward are elective and pre-planned. Patients that are not ready for discharge before closing time on Friday are moved to ward B3. Some of the nurses employed at this ward commonly have shifts at ward B3 every third weekend.

B2M is a general internal medicine unit. Acute cases are treated at this ward. This includes diagnoses such as heart disease, stroke and cerebral hemorrhage, and infection.

B3 is a surgical and orthopedic unit. The ward is a long-stay unit, and patients from other wards, such as A3, may be moved here.

B4 is a general internal medicine unit. Acute cases are treated at this ward. This includes diagnoses such as pulmonary- and heart disease, diabetes and cancer (Brunes & Finstad, 2015).

3.4 Scenario

To illustrate how a spreading situation could emerge, a scenario has been drawn out. Figure 7 shows the movement patterns of people included in the scenario, and the contact sites between carrier and spreading agent. Note that spreading situations like these are not possible to avoid entirely due to coincidences and natural variations. They are however, especially the more obvious and controllable situations, possible to reduce.

Step 1 – The patient:

At the beginning of the scenario there are no infected individuals, only carriers of MRSA. The scenario starts with a patient entering the hospital for a pre-planned operation. The patient filled out a form prior to scheduling the procedure, to see if he/she should get tested for MRSA. The patient answered "no" to all questions related to the risk of carrying MRSA, and thus was not tested. However, he/she unknowingly carried the bacteria.

Step 2 – The first nurse:

A nurse with employment at ward A3, which works weekend shifts every third weekend at ward B3, treats this patient. The nurse unknowingly picks up the bacteria from the patient. The nurse does not show any symptoms of infection; the bacteria have merely become a part of the normal flora in her anterior nasal cavity. Every time she blows her nose, or in other ways come in contact with the mucus membranes where the bacteria is colonized, she poses a risk of transferring the bacteria to surfaces or to patients. The risk of her directly infecting patients is low due to her being well trained on procedures of hand hygiene. The risk of her spreading the bacteria indirectly is however higher. The nurse is bothered by a running nose due to allergies. In this scenario the nurse sneezes into her hands while she walks through the hospital corridor. She knows well enough to wash her hands in order to maintain a good hygiene. She spots a bathroom just in front of her and enters. On her way in she touches the door handle. She carries out proper hand wash before returning to her duties. The nurse no longer poses a threat of further transferring the bacteria from her hands.

Step 3 – The cleaning worker:

A cleaning worker on shift that day touches the door handle on his way into the bathroom to clean. Well inside, he puts on gloves to prevent contamination while he cleans. If he follows

protocol on how to use gloves, he will protect himself from fecal oral transmission of potential infectious agents in the bathroom. He will also reduce the risk of spreading these agents further. Bacteria from the door handle is nevertheless still on his hand inside the gloves. The cleaning worker finishes cleaning the bathroom like the cleaning procedure instructs him to, and he removes his gloves. He then rushes on to the next ward to continue his cleaning schedule. He starts in the corridor. All the cleaning cloths on his cart is used, so he needs to pick up new ones before starting at the current ward. He approaches a cupboard containing cleaning cloths and clean sheets for the patient rooms down the hall. The cleaning worker touches the cupboard handle in order to pick up new cloths, and by this action the MRSA bacterium from ward A3 is now transferred to ward B2M. This cupboard is used by both cleaners and nurses.

Step 4 – The second nurse:

A nurse in ward B2M picks up new sheets for a bed in a patient room. Ward B2M deals with a lot of post-operative cases, and the next task for the nurse is to redress a post-operative wound. Open wounds are especially prone to infections, and the patient ends up acquiring a post-operative MRSA infection.

The nurse that primarily started the spread within the hospital works shifts at ward B3 every third weekend. This further increases the risk of transfers across wards.

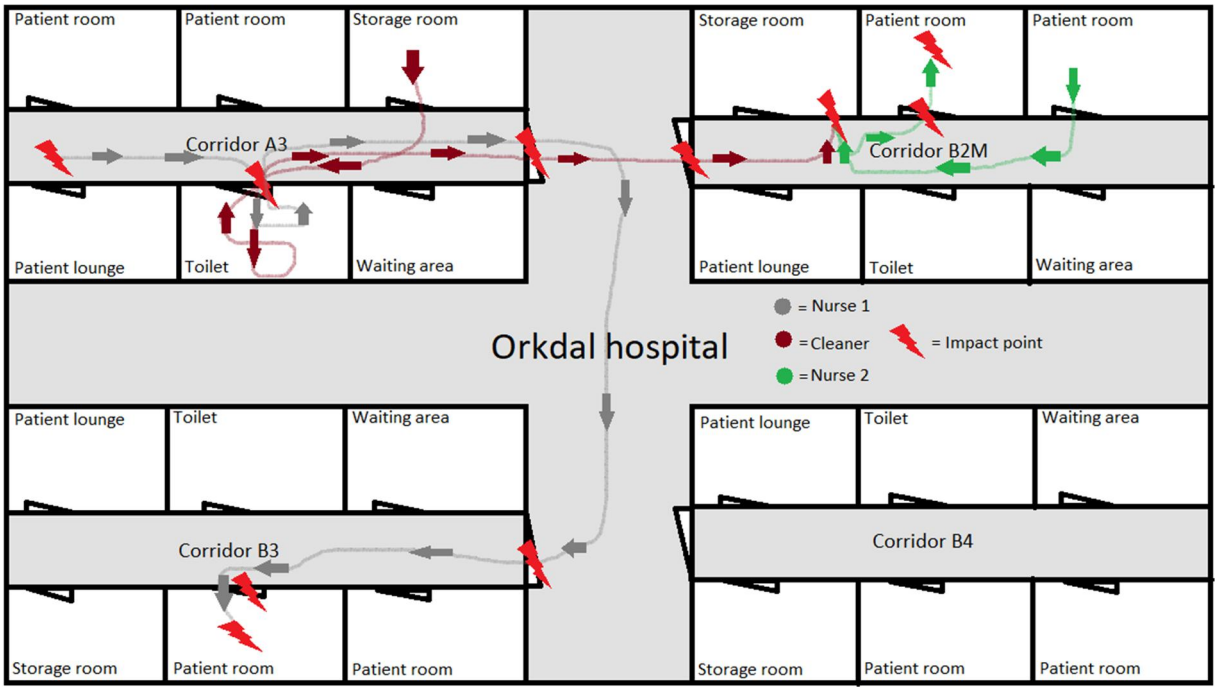


Figure 7: An illustration of the spreading scenario described previously. The direction of the arrows indicates the direction of movement, and the different colors indicates who represents the movement pattern (grey = nurse 1 – the source of infection, purple = the cleaner, green = nurse 2). The red lightening indicates the impact point of bacteria transmission.

This is a pretty naïve example of a possible outbreak situation. Still, it gives an impression of how easily a spreading situation may occur, and how important proper procedures are in infection prevention. Despite procedures in this scenario allegedly being followed, the MRSA bacteria managed to spread from one ward to another. Chances are the bacteria would not survive this transfer and to manifest as an active spreading due to few microbes and poor living environment. Nevertheless, it is not impossible. Which is precisely what makes prevention so important. If one is able to map out various scenarios, both reasonable and unreasonable, one is also better equipped to develop prevention plans. Imagine the outcome of the same scenario if infection control procedures were not established and followed!

3.5 Designing the hospital network

A virtual illustration of Orkdal hospital was used for the main example simulation of this thesis. Locations at the hospital were represented by transit nodes, and they constituted equipment inside the rooms within the four hospital wards. Transit nodes may contain all kinds of different bacteria. The overall composition of bacteria was considered irrelevant for

the sake of this thesis, and bacteria were thus not identified. As a result, transit nodes may in theory contain *Staphylococcus aureus* even prior to the bacteria being simulated to enter the system.

Staphylococcus aureus is a part of the normal bacterial flora for a portion of the population, and the species is generally not seen as a pathogen unless it is the cause of infection. For that reason, the S bacteria present in the hospital network are meant to reflect S bacteria that are cause of infection. The simulation only considers a pathogenic *Staphylococcus aureus*. Note that this assumption is made for simplicity, and it does not fully reflect the reality of microbes as such. A *Staphylococcus aureus*-infected patient entering the hospital may or may not further transfer the infecting agent to transit nodes. The following events after *Staphylococcus aureus* is introduced to transit nodes vary. The bacteria may either be removed by cleaning or natural death, or persist in the system. The latter may have unfavorable consequences; *Staphylococcus aureus* may be transferred to another transit node, or worst-case scenario; be transferred to another patient.

Bacterial population numbers were chosen randomly from the sample collection on the basis of seeming presentable for the simulation. The rooms within the four wards were further divided into transit nodes representing equipment found inside. Two locations inside each room were defined by a number of bacteria. Thus, the total number of transit nodes in the hospital model was 48. An overview of the transit nodes for one of four wards are listed in table 2.

Table 2: An overview of locations selected for transit nodes, and their approximate bacterial load.

Room	Equipment	Number of colonies	Ward for collection
Corridor			
	Cupboard handle	Approx. 6 colonies	B4
	ECG-machine	Approx. 22 colonies	B2
		In total: Approx. 52 colonies	
Storage room			
	Blood pressure monitor	Approx. 25 colonies	A3
	Stethoscope	Approx. 8 colonies	B4
		In total: Approx. 30 colonies	
Toilet			
	Toilet handrail	Approx. 22 colonies	A3
	Cupboard handle	Approx. 8 colonies	B2
		In total: Approx. 30 colonies	
Patient lounge			
	Remote control	Approx. 47 colonies	B2
	Sofa	Approx. 68 colonies	B2
		In total: Approx. 115 colonies	
Waiting area			
	Chair	Approx. 30 colonies	B2
	Table	Approx. 16 colonies	B2
		In total: Approx. 46 colonies	
Patient room			
	Night table	Approx. 19 colonies	B3
	Bed lamp	Approx. 30 colonies	B3
		In total: Approx. 49 colonies	

Table 3: An overview of locations with *Staphylococcus aureus* at Orkdal hospital, and their approximate bacterial load.

Bacteria	Equipment	Number of colonies	Ward for collection
Staphylococcus aureus	Blood pressure monitor	Approx. 8 colonies	A3
Staphylococcus aureus	Remote control	Approx. 22 colonies	B2
Staphylococcus aureus	Bed frame	Approx. 30 colonies	B3
Staphylococcus aureus	Cupboard handle	Approx. 3 colonies	B4

A precise arrangement of the hospital is not included in the network simulation, and the connections in between different locations are randomly distributed. An option to adjust for more precise and realistic structural characterizations may be implemented by modifications to the network model. I wished not to overcomplicate the network model at this time.

Some of the transit nodes were set to be unable to have any contact with patient nodes. This was to represent locations such as storage rooms, where solely hospital workers were allowed. Transmission via these nodes was caused by other people transferring them; nurses or cleaning crew for example.

Presence of *Staphylococcus aureus*

In our hospital simulation *Staphylococcus aureus* is present at some locations/transit nodes prior to patients infected with the bacteria is introduced to the system (table 3). This is based upon the findings in the bachelor thesis (Brunes & Finstad, 2015). During the research at Orkdal hospital four samples were positive for *Staphylococcus aureus*. The strains were detected at a blood pressure instrument, a remote control, a bed frame, and a cupboard handle; one sample at each of the wards tested. The *Staphylococcus aureus* strains were typed and found to be of different origin, which was interpreted as an indication that there was not a case of internal spreading. For this thesis, the numbers from the bachelor thesis are included as the maximal number of bacteria in the transit nodes. Additionally, the transit nodes corresponding to the locations where *Staphylococcus aureus* strains were detected are set to have presence of S bacteria in the transit nodes from the simulation start. These values are set

to be optional for this network model tool, due to different bacteria being of different nature, and to make the tool more flexible for adjustment.

For simplicity, the four different wards simulated in the model contained the same information. All four wards consisted of the same locations and are based upon the same numbers. The transit nodes of detected *Staphylococcus aureus* were only included once; one transit node for each of the wards. Consequently, they were introduced to the hospital network model as unique transit nodes. The total number of transit nodes were however not changed. The survival frequencies were chosen at random, following an exponential distribution.

3.6 Deciding parameter values for the hospital simulation

In the process of developing the main simulation to present the tool of this thesis, deciding parameter settings constituted a big part. As previously stated, these values are not fixed, and may be changed according to the aim of the study. Despite the main simulation being a hypothetical simulation, my aim was to present a simulation as realistic as possible. Some numbers were collected from the bachelor thesis (Brunes & Finstad, 2015), while other supplementary parameter values were inspired by research and known microbiological data. Accurate data proved hard to find due to the lack of detail in the data from the bachelor thesis. The remaining parameters are therefore approximate. Results may not be treated as valid scientific results. Reasoning behind some of the parameter settings are listed underneath. A complete overview and explanation of parameters is found in appendix 2.

- num_iter: *the number of iterations; how long the simulation runs for. Each iteration in this model runs for 30 minutes.* The simulation shows the development during a period of approximately 12.5 months, meaning the value of num_iter is set to 18000.
- initial_patients: *how many non-infected patients there are at time = 0.* The number of patients does not increase considerable, but it varies to some extent. The simulation starts at a random point, which means that a number of patients are already in the system. The number of initial patients is set to an arbitrary value of 200.
- max_patients: *maximal number of patients.* The maximal number of patients is not a crucial value, although it is useful to make sure the activity does not get out of hand. The value is set to 300.

- s_growthrate: μS*, the growth rate of S individuals*. The parameter value is estimated by following the calculated generation time of *Staphylococcus aureus*; 30 minutes under optimal growth conditions. One iteration in our network model lasts for 30 minutes, thus the bacteria should in theory divide once every iteration. To find the specific growth rate the formula: ln2/generation time. The growth rate for *Staphylococcus aureus* is calculated to be ln2/30=0.02310490601. Although the calculated growth rate probably is somewhat inaccurate due to the hospital environment not being equal to a laboratory environment, the parameter value is set to 0.02.
- i_fitness_cost: *how much slower I individuals grow.* As previously discussed, acquiring new traits usually carry a fitness cost. To what degree the acquisition of new traits affects the growth rate vary considerably, and it may be hard to predict. If i_fitness_cost is set to for example 0.8, it means that the I bacteria grow at 20% of the rate of S bacteria. The parameter value is set to 0.3.
- bacterial_load: *the maximal number of bacteria that may exist in a patient.* This parameter is included to prevent an improbable overgrowth of bacteria. The parameter value is set to 90000000. This is the highest detected number of microbes gathered from data from Orkdal hospital (Brunes & Finstad, 2015).
- s_frac: *the proportion of new patients that is already infected with S bacteria.* Since *Staphylococcus aureus* is a common part of the normal flora of humans, this the fraction of people already carrying S bacteria should be high. The number of patient infected with S bacteria upon admission to the hospital should be significant. However, due to the simulation only considering infected patients, not the carriers, the value is set to 0.02.
- i_frac: *the proportion of new patients that is already infected with I bacteria.* The fraction of patients already infected with I bacteria should according to known microbiological data be lower than the fraction of patients that are already infected with S bacteria. Studies show that about 1-5% of the population carry MRSA, and that as many as 30-60% of the population carry MSSA (MRSA Carrier: Questions and Answers, n.d.). The exact ratio between s_frac and i_fraq is unknown. The parameter value is set to 0.002. Note that the parameter value reflects the number of patients with MRSA infection, not the number of patients carrying MRSA.
- i_delay: *decides when patients carrying I bacteria starts to enter the model.* The parameter value is given in number of hours, and is set to 168 hours; a week.

- tresh: *decides when antibiotic treatment starts. This is a number between 0 and 1. For example, if the parameter is set for 0.5, the treatment starts when a patient has 50% of bacterial_load as S + I.* Treatment with antibiotics should only be given to patients that are infected with the bacteria, not the patients that solely carry the bacteria as their normal flora. To prevent antibiotics treatment being carried out for all patients with the common bacteria, the value is set relatively high. The number of I bacteria in a carrier will not be substantial compared to the infected patients, hence carriers will not be treated with antibiotics. The value is set to 0.2.
- beta: *the rate of flow from S to I when S and I come into contact with each other.* The transmission efficiency is dependent on several factors, and for our data the transmission efficiency is unknown. Nevertheless, transmission is known to happen slowly. More so in a natural environment than in a laboratory test tube. The value is arbitrarily set to 0.05.
- beta2: *the rate of flow from I to S.* This transition occurs independently of contact between I bacteria and S bacteria. The value is set to 0.06.
- delA: *tells how effective the antibiotic kills bacteria. This is a kinetic parameter which models the meeting between antibiotics and bacteria as in a chemical reaction; A (antibiotic) + B (bacteria) -> A with a rate of delA.* The parameter value is set to 0.9.
- ab_supply: *the supply rate of antibiotic during treatment, which means the amount of antibiotics supplied.* This parameter value is based upon *Staphylococcus aureus* treatment data (Rayner & Munckhof, 2005). Recommended treatment dosage of intravenous treatment with Dicloxacillin is according to Rayner and Munckhof (2005) 0.5-2 g per 6^{th} hour. The dosage was chosen to be 1 gram per 6^{th} hour: 0.167 g/h.
- ab_uptake: *the uptake rate of antibiotics during treatment, which means how fast it is removed from the system.* The half-life of Dicloxacillin is approximately 0.7 hours (Nauta & Mattie, 1976). The value is set to 0.06 g/h.
- clean_interval: *how often transit nodes are cleaned. The value corresponds to hours.* This parameter value was set to different values for parallel simulations to study its overall effect. For the main simulation it is set to 6 due to it generating interesting results with this value.
- halflife: *the half-life of cells in transit nodes. It is implemented to be a radioactive decay. By radioactive decay it means that the population in the transit nodes decreases exponentially with a half-life time of halflife. Staphylococcus aureus* is able to survive for quite some time on surfaces, and the parameter value is set to 150.

- decay: *decides if cells in transit nodes dies or not (0 or 1).* This parameter value is set to 1.

4 Results and discussion

In relation to this thesis, several simulations were run with varying parameters. The aim was to confirm or deny hypotheses made prior to simulation start. Five hypotheses were examined, in addition to a supplementary case; a simulation of a virtual hospital network. Time steps for all simulations are presented in iterations, whereas one iteration equals 30 minutes. Results are presented below in order.

4.1 Hypothesis 1: The importance of transit nodes for spreading activity

The bacterium *Staphylococcus aureus* is able to survive at surfaces for some time, despite sub-optimal growth conditions. Transit nodes in the network represent locations within the hospital that patients, hospital staff and visitors commonly come in contact with; door handles, remote controls, bed frames etc. The most effective way for an infection to be transmitted from patient to patient is through direct contract. This is due to the infectious dose being lower for direct contact; a lower number of bacteria is required for an infection to manifest in the next patient (Murray et al., 2012). Nevertheless, infection transmission is not impossible by indirect contact. Patients with known infections are often isolated from other patients, and thus transmission of the infectious agent by direct contact may actually be excluded. Additionally, patients are eventually removed from the network. Patients may however encounter the same equipment and inventory. Bacteria may in this way be transferred from patient to a surface, and then to the next patient. In the hospital network of this thesis, some of the transit nodes are disconnected from all patient contact. One would possibly assume that these nodes would be negligible from a potential spreading situation. The hypothesis of this thesis states the contrary, that these nodes are essential for disease spreading to persist in the network. It states that transit nodes are important for the spreading through the network.

To examine this hypothesis, a simulation was run on a network of 100 transit nodes, where solely two of the transit nodes (node 0 and 32) were connected to patients. The simulation was run for 1000 iterations; approximately 20.8 days. In order to maintain an active spreading through the network to be able to visualize for the sake of testing the hypothesis, cleaning

frequency was decreased considerably, transfer probability was increased, and transit nodes were set to have no death of bacteria. No nodes were assigned bacteria, neither S or I, before simulation start.

From this simulation network we can see that transmission still occurred across the edges despite a lack of patients being introduced to the network (figure 8). The spreading activity was considerably lower than by the inclusion of patient contact at the same parameter settings, but the spreading was still present in the network.

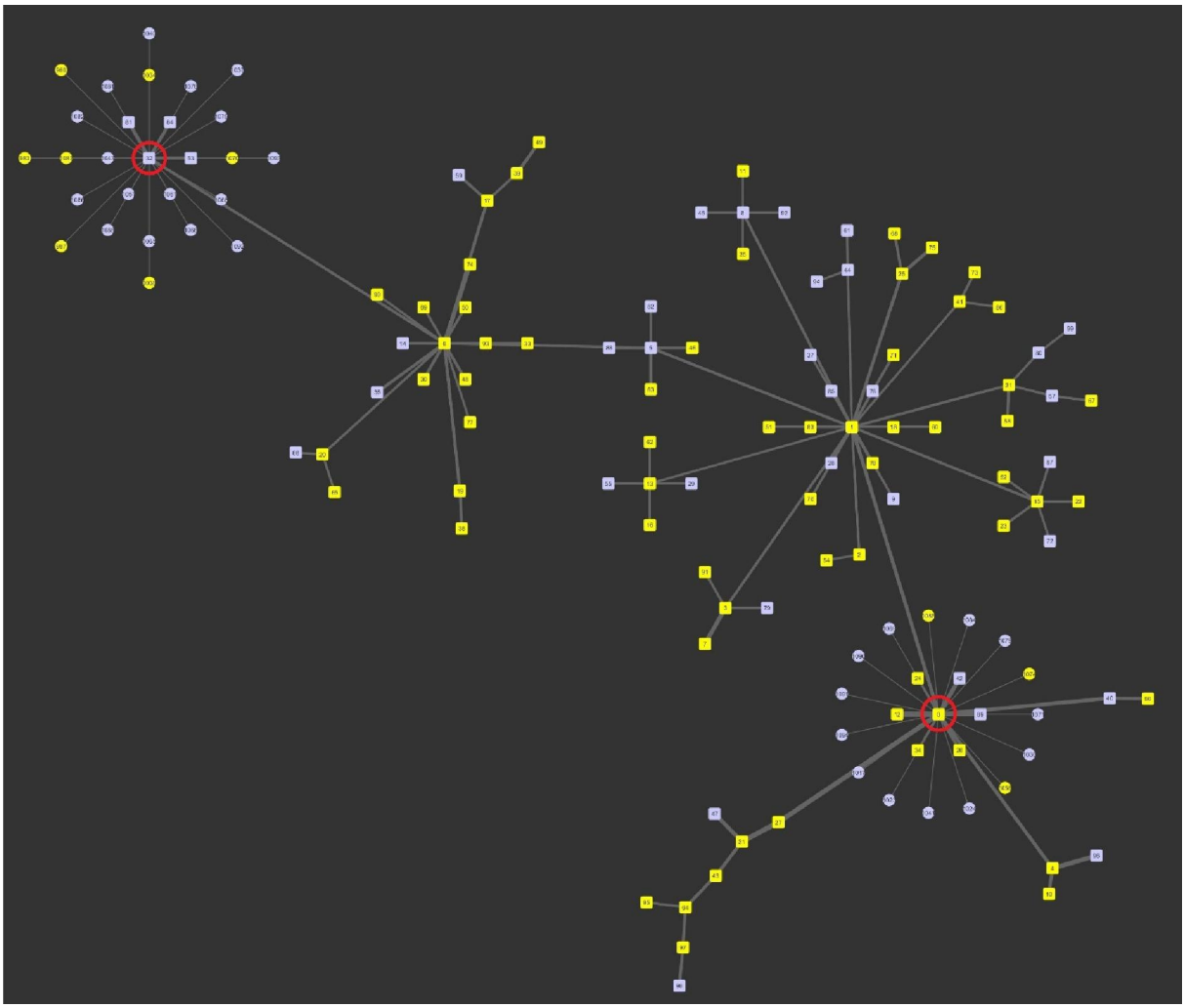


Figure 8: Cytoscape network screenshot showing the spreading situation of susceptible (S) bacteria at iteration 1000; day 20.83. The yellow nodes represent nodes containing S bacteria. At simulation start, no nodes contained S bacteria. Square nodes represent transit nodes; locations, and round nodes represent patient nodes; incoming patients. The red rings circle the transit nodes with patient contact (node 0 and 32).

Figure 8 shows that several of the transit nodes multiple steps away from patient nodes contained S bacteria. The thickness of the edges represents the fraction of movement across the edges from the start until iteration 1000, the “fraction_of_flow”. Several of the edges of

highest edge traffic are connected to node 0. Node 0 was one of the two transit nodes set to be connected to patients. When excluding patient nodes, node 0 displayed a degree of 10. Transit node number 32 displayed a degree of 4, and lower values for traffic across the edges in contact to the node.

As expected, the simulation results indicated that the transit nodes were essential for a spreading process to occur within the network. The two transit nodes connected to patients are marked in the visualization. From the network visualization, one can see that the spreading agent reached transit nodes several steps away from the patient nodes. This indicates that the spreading through the network depended on other factors than solely the inclusion of patients. Since no transit nodes were infected upon simulation start, it is obvious that some kind of movement between the transit nodes took place, despite no contact with patients. The cleaning interval was drastically decreased to better visualize the results. One may argue that this factor provokes an unnatural spreading through the network. However, this hypothesis simply states that transit nodes are important for spreading processes. Further details are not stated or requested. Thus, the results were considered credible.

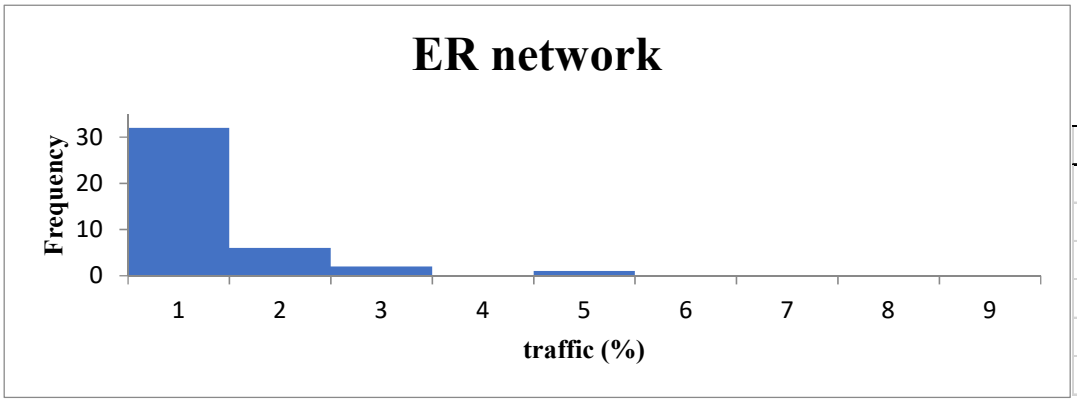
4.2 Hypothesis 2: The spreading patterns of Erdős-Rényi (ER) and Barabási-Albert (BA) networks

Spreading patterns are strongly affected by properties such as degree distribution and edge weights. The Erdős-Rényi (ER) model has a Poisson degree distribution, while the Barabási-Albert (BA) model has a scale-free degree distribution. In ER networks, close to all nodes have the same degree of connectivity. For that reason, a spreading progress is usually seen to occur in an exponential manner. BA networks consist of a few nodes with high connectivity (hubs), and the rest of low connectivity. Chances are the spreading activity will burst when the infecting agent reaches a hub (Barabási, 2016). The hypothesis thus states that simulations on different model types; Erdős-Rényi and Barabási-Albert will display different spreading patterns. The ER model will show a uniform spreading pattern, while the BA model will show an uneven spreading pattern.

To examine this hypothesis, simulations were run on both Erdős-Rényi and Barabási-Albert networks. The BA network was generated with 1000 transit nodes and $m = 5$. Here, m

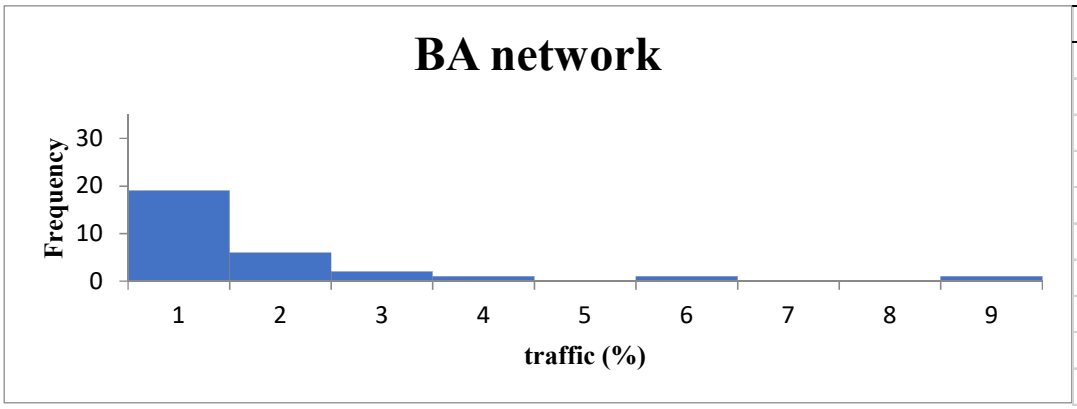
represents the number of edges to be attached from new nodes of the network to already existing nodes. The ER network was generated with 1000 transit nodes and p = 0.01. Here, p represents the probability of an edge connecting each pair of nodes in the network. To enhance coincidence, several simulations were run for both network types, with varying values for the parameters “seed” and “np.seed”. These are parameters that generate random numbers for the simulation. Other parameters were held constant for all parallel simulations. The average values for traffic across the edges, fraction_of_flow, was then calculated, and distributed in histograms (figures 9 and 10).

The phrase “fraction_of_flow” represents an edge property that represents the total fraction of total traffic across the edge of interest, from the start of the simulation until the iteration of choice. Hence, the property gives a good indication of how a spreading progress in a network. “fraction_of_flow” is from this point forward referred to as traffic. We can see from the histograms of the traffic values over 1% (figures 9 and 10) that the ER network displayed a more uniform spreading pattern than the BA network. The ER network contained a total of 41 edges with traffic of 1% or more. The majority of the values did correspond to the lower part of the scale, with the highest value of 5%. The BA network contained a total of 30 edges with traffic of 1% or more. The distribution was different from the ER network. In comparison to the distribution of the ER network the histogram showed a greater proportion of the values corresponding to higher values. This This indicates that a few nodes of the network are more important than others with respect to spreading dynamics.



%	Freq
1	32
2	6
3	2
4	0
5	1
total:	41

Figure 9: Histogram showing the average distribution of traffic (%) for the ER network.



%	Freq
1	19
2	6
3	2
4	1
5	0
6	1
7	0
8	0
9	1
total:	30

Figure 10: Histogram showing the average distribution of traffic (%) for the BA network.

An assumption was made that transit nodes of networks with a scale-free degree distribution, such as Barabási-Albert networks, would be associated with the edges of highest proportion of traffic. The theory was tested by comparing the busiest edges of the network; with highest traffic, against the most connected nodes. The comparison was done to solely one of the parallel simulations for each of the network types. Results are listed in the table below (table 4).

Table 4: Table showing the association between the edges of the Barabási-Albert and the Erdős-Rényi networks with the highest traffic, and the nodes of highest degree. Edges associated with the top 15 transit nodes of highest degree are highlighted in bold in the table.

Edge rating	Edge (BA-network)	traffic (%)	Edge (ER-network)	traffic (%)
1	180 (interacts with) 892	11 %	330 (interacts with) 454	11 %
2	**17 (interacts with) 721**	7 %	572 (interacts with) 902	4 %
3	**34 (interacts with) 164**	7 %	311 (interacts with) 642	4 %
4	137 (interacts with) 602	6 %	281 (interacts with) 642	2 %
5	70 (interacts with) 183	4 %	**14 (interacts with) 642**	2 %
6	21 (interacts with) 183	4 %	355 (interacts with) 620	2 %
7	21 (interacts with) 849	4 %	145 (interacts with) 781	2 %
8	**15 (interacts with) 849**	4 %	454 (interacts with) 781	2 %
9	**15 (interacts with) 892**	4 %	572 (interacts with) 620	2 %
10	**17 (interacts with) 180**	4 %	64 (interacts with) 572	2 %
11	**34 (interacts with) 182**	4 %	31 (interacts with) 203	2 %
12	**34 (interacts with) 189**	3 %	83 (interacts with) 203	2 %
13	**4 (interacts with) 205**	3 %	83 (interacts with) 642	2 %
14	**4 (interacts with) 189**	3 %	64 (interacts with) 516	2 %
15	137 (interacts with) 205	3 %	**514 (interacts with) 902**	2 %

Results obtained from the simulations showed a dissimilarity in traffic values; the fraction of traffic across edges of the Erdős-Rényi network and the Barabási-Albert network. The values did correspond to the original statements of the hypothesis, that the spreading patterns of Erdős-Rényi and Barabási-Albert networks are different. The Erdős-Rényi network displayed

a distribution of traffic that is likely to result in a uniform spreading pattern; several of the edges constitutes more or less equal proportions of traffic. Additionally, the number of values >1% was higher for the Erdös Rényi network than for the Barabási-Albert network. The Barabási-Albert network also displayed a more scattered distribution of traffic. This may indicate that the majority of traffic occurs across specific edges of the network. The differences are not immense, and the results should possibly not be treated as conclusive. However, they may be considered indicative, and when combining these indicative results with theory, the results seem rather reasonable.

The histograms related to this hypothesis (figures 9 and 10) were generated with the average results of multiple parallel simulations. Parallel simulations with the same parameters were conducted to enhance coincidence, and to possibly enhance the creditability of the results. Solely traffic values >1% were included in the histograms. The exclusion of traffic values <1% was a deliberate action to achieve more presentable results. Consequently, some of the edges were excluded despite displaying edge traffic. Perhaps would the differences be even more apparent with the inclusion of all traffic values.

Edges with the highest proportion of traffic; edges displaying highest traffic, in each of the networks, were compared to the transit nodes of highest degree. We expected to detect some of the top connected transit nodes of the Barabási-Albert network among the nodes included in the edges of the top 15 listing of traffic values. This observation was not necessarily expected for the Erdős-Rényi network, due to the uniform degree distribution typically displayed by such networks. Results shown in table 4 supported the predetermined assumption that several hubs of the Barabási-Albert network would be included in a majority of the traffic within the network. The 15 edges presented in table 4 equals 0.3% of the total number of edges in the system. Of these 0.3% edges, a total of 60% of the edges did correspond to the transit nodes of highest connectivity for the BA network. The same applied for the ER network with a percentage of 13.

The uniform degree distribution seen in Erdős-Rényi networks implied that most nodes were of approximately the same connectivity. It was thus not expected that a majority of the top connected transit nodes of this network would be associated with edges of highest traffic. Nevertheless, some of the transit nodes were associated with the edges of highest traffic. Despite most of the transit nodes being of similar degree in ER networks, some are commonly slightly more connected than others. The transit nodes associated with the edges highlighted

in table 4, were among the topmost connected ones. The correlation between node degree and traffic in the Erdős-Rényi network was of this reason not considered a coincidence.

4.3 Hypothesis 3: The effect of changes to cleaning frequency

Poor hygiene is generally associated with high bacterial numbers. There is however a lack of thorough research on the matter of cleaning within hospitals (Dancer, 1999). A hypothesis was made, stating that the frequency of cleaning has a great impact on the spread of bacteria through a system. To study this hypothesis, several different parameter values for cleaning interval was run on the same network. The aim was, in addition to see the impact of cleaning on spreading pattern, to find a reasonable time interval for cleaning. A time interval frequent enough to eliminate bacteria effectively, but not unnecessarily frequent. A too frequent cleaning schedule will possibly be a waste of resources.

The simulation was run on a Barabási-Albert (BA) network of 200 nodes and $m = 5$, with 6 different parameter values; 3, 6, 9, 12, 18, and 24 hours cleaning intervals. This means that cleaning was scheduled to occur every 3, 6, 9, 12, 18, and 24 hours. The simulation was run for 200 iterations, which corresponds to approximately 4 days. Figures 11-13 show the resulting networks. For each network, the iteration chosen was the iteration number closest to the last planned cleaning before simulation termination. This selection was conducted in order to achieve as similar terms as possible.

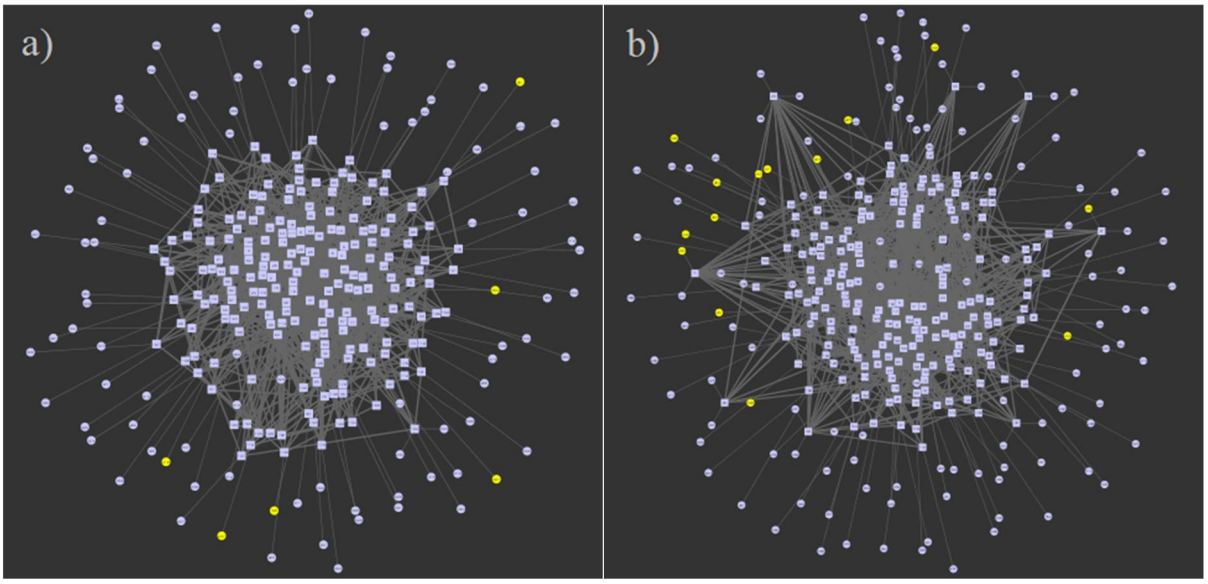


Figure 11: Cytoscape networks displaying S-infected nodes at iteration a) 196 with cleaning interval set to 3 hours and b) 196 with cleaning interval set to 6 hours. Yellow nodes = nodes containing susceptible (S) bacteria, square nodes = transit nodes, round nodes = patient nodes. Transit nodes are located centrally in the networks.

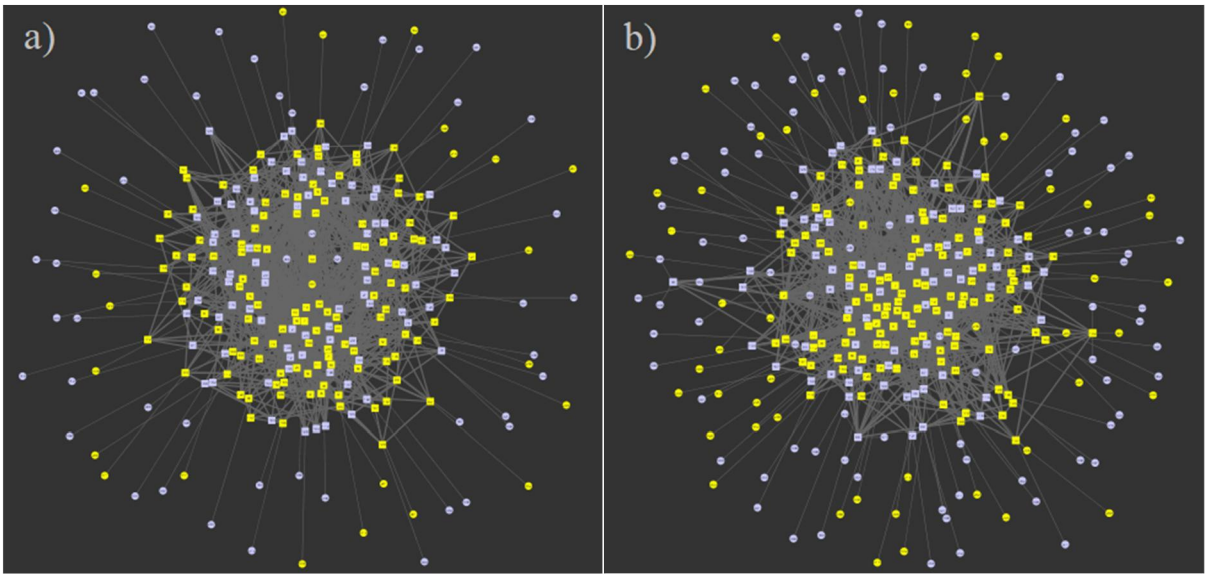


Figure 12: Cytoscape networks displaying S-infected nodes with same parameters as figure 11. Panel a) at iteration 196 with cleaning interval set to 9 hours, and b) at iteration 196 with cleaning interval set to 12 hours.

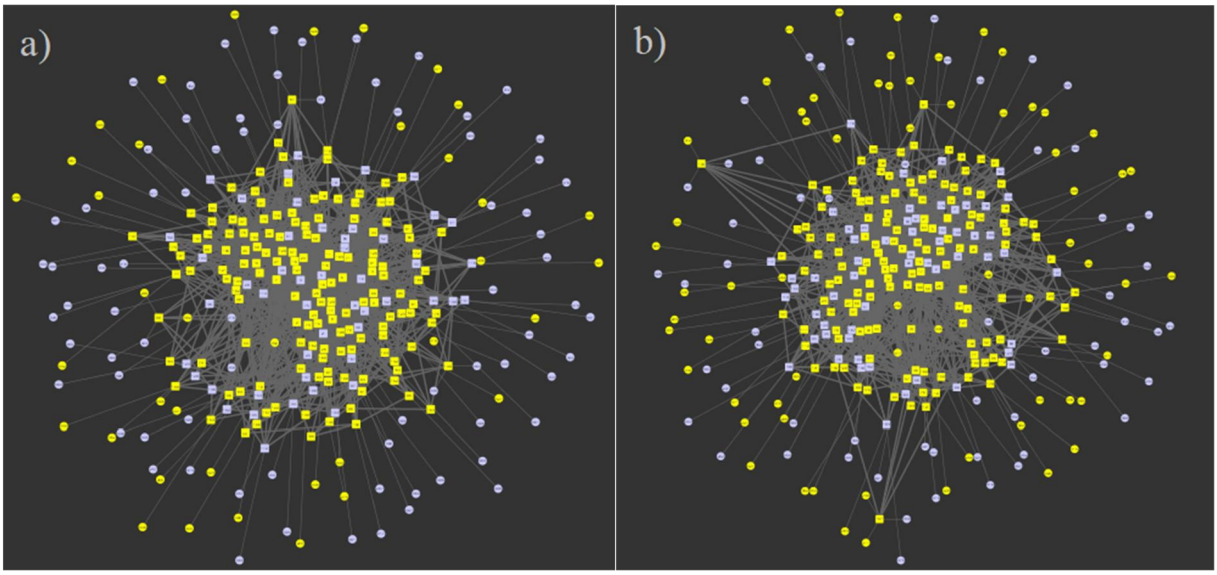


Figure 13: Cytoscape networks displaying S-infected nodes with same parameters as figure 11. Panel a) at iteration 198 with cleaning interval set to 18 hours, and b) at iteration 190 with cleaning interval set to 24 hours.

The Cytoscape networks (figures 11-13) showed an increasing number of S-infected nodes.

In an attempt to randomize data further before collecting detailed results, simulations were run some additional times with different parameter values for seeds; the parameters used to generate random numbers. Other parameters were held constant. Average data from the simulations showed that the network with cleaning interval of 3 hours contained 1 S-infected nodes, that the network with cleaning interval of 6 hours contained 35 S-infected nodes, that the network with cleaning interval of 9 hours contained 100 S-infected nodes, that the network with cleaning interval of 12 hours contained 163 S-infected nodes, that the network with cleaning interval of 18 hours contained 224 nodes, and that the network with cleaning interval of 24 hours contained 205 S-infected nodes. Of these, several were patient nodes.

In order to get a more precise view of the transit nodes infected, the patient nodes were removed from the network. S-infected nodes were once again selected. When solely looking at the transit nodes of the networks, results showed that the network with cleaning interval of 3 hours contained 0 S-infected nodes, that the network with cleaning interval of 6 hours contained 18 S-infected nodes, that the network with cleaning interval of 9 hours contained 37 S-infected nodes, that the network with cleaning interval of 12 hours contained 101 S-infected nodes, that the network with cleaning interval of 18 hours contained 142 nodes, and that the network with cleaning interval of 24 hours contained 124 S-infected nodes. See line chart (figure 14).

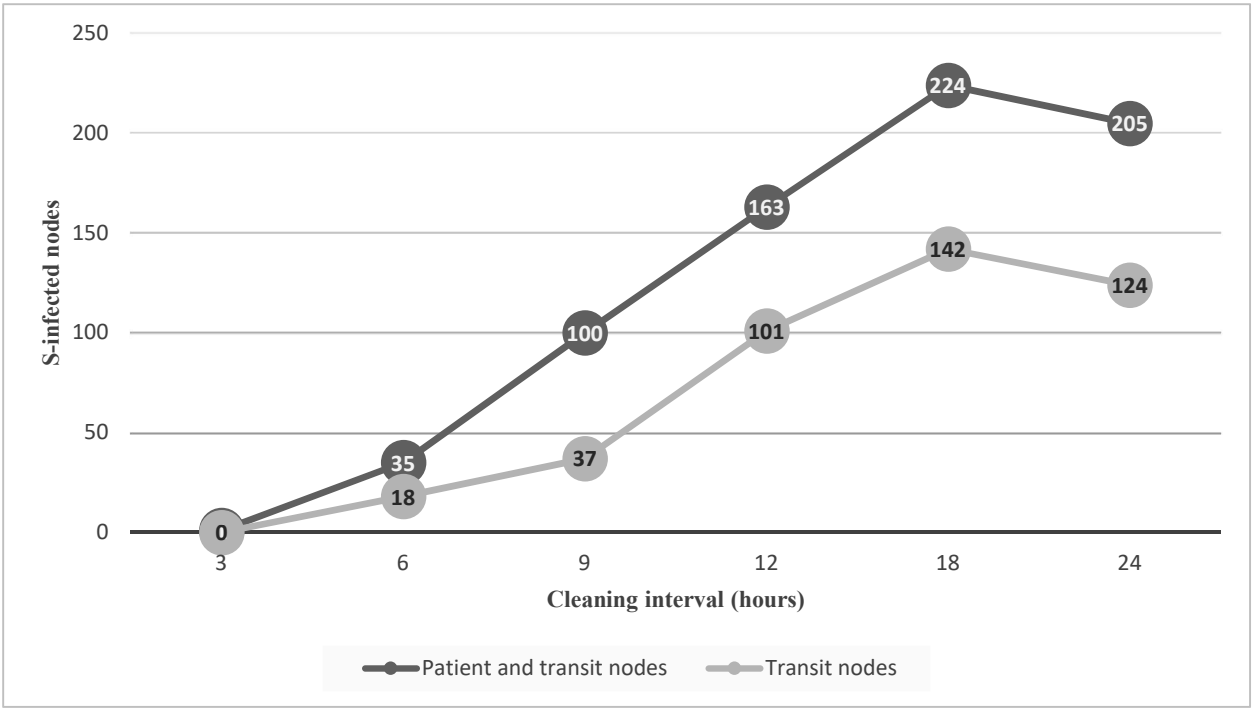


Figure 14: Line chart displaying the number of S-infected nodes; both with and without patient nodes after approximately 4 days with respective cleaning intervals.

Results showed that the increase in bacterial number was not absolute; the network with cleaning interval of 18 hours displayed a higher number of S-infected transit nodes than the network with cleaning interval of 24 hours. A likely explanation for this may be the parameter called dacay. This parameter allows for natural death in the transit nodes. Perhaps the bacterial population reached their peak before measurement, and then started do decrease exponentially due to natural fluctuations. To study this further, the numbers of bacteria in the top populated node of each of the networks were collected (table 5).

Table 5: Table showing the average number of bacteria in the top populated nodes; the nodes containing the highest number of bacteria, from each of the networks with different settings for cleaning interval.

Cleaning interval	Number of bacteria
3 hours	0
6 hours	11
9 hours	1728
12 hours	2616
18 hours	1055
24 hours	10722

Table 5 shows a distinct increase in number of bacteria in the top populated nodes of each of the networks corresponding to different intervals for cleaning, except for the network with cleaning interval of 18 hours. The point of measurement was, as mentioned, varied for the different simulations. If all simulations were to be studied at the same iteration, the networks would present their state at different times of their cleaning cycles. This could possibly lead to sources of error such as measuring bacteria at the population peak, versus at the population trough. One could thus argue that the number of S-infected would vary solely upon coincidence instead of the chosen time interval for cleaning. The decrease of bacteria in the top populated node did not correspond to the fact that the time of measurement was adjusted for each of the networks. This may indicate that the explanation of the discrepancy in table 5 corresponds to the theory proposed; that death within the transit nodes may be responsible for a decrease in population size. It may not be concluded with full confidence and certainty, but due to the time of measurement being at the same point of the cycle of each respective cleaning interval, it was seen as a relatively plausible explanation. Additionally, this model did not consider all transit nodes to be completely eradicated at each round of cleaning. Highly contaminated nodes that coincidently remain untouched during cleaning, may thus constitute a bigger probability for further spreading. Some simulations will of this reason possibly result in a more extensive spreading situation than others. Nonetheless, the overall trend shown in these simulations, was an increase of nodes containing S bacteria proportional to the frequency of cleaning.

From the results we can see that there is a more distinct increase of S-infected transit nodes from the network with cleaning interval of 9 hours to the network with cleaning interval of 12 hours, than for the networks with lower cleaning intervals. Thus, it seems like the realistic optimal time interval for cleaning would be no more than 6-9 hours. However, we have no information indicating that a burst will with certainty occur at the same time as shown in this example for other simulations with approximately the same parameter settings. BA networks are scale free networks; some of the nodes are highly connected (hubs), while most nodes are sparsely connected. When an infecting agent reaches such highly connected nodes, a spreading process is likely to escalate rapidly. Perhaps the infecting agent will reach a hub prior to measurement. Another reason for a potential sudden increase, may be that a higher number of patients get into the system at that specific time, and transfer more bacteria. That could possibly, despite removal of patient nodes before measurement, influence the results.

It is important to keep in mind that only a few parallel network simulations were presented in relation to this hypothesis, of which solely six different values for cleaning intervals were tested. Results from these individual simulations were not comprehensive enough to be used as templates for all similar situations. To achieve more reliable results, several more simulations should be run for longer periods of time, with an even higher number of different parameter values for cleaning interval. In addition, the simulations should be tried on several networks with different structures and parameter settings. If a considerable number of the outcomes are consistent and significantly correlated, a conclusion may be drawn with more certainty.

4.4 Hypothesis 4: The effect of changes to growth rate

The hypothesis regarding growth rate is twofold. It is partly formed on the basis of research suggesting that reduced growth rate leads to a reduced susceptibility to antibiotics and disinfectants (Evans et al., 1990). The other part of focus is the relationship between growth rate and spreading pattern. We know that a higher growth rate will result in a more rapid growth of the total population under the same conditions. We do however not know for certain how an increased growth rate will affect the spreading progress. An assumption is made that the number of bacteria in transit nodes proportionally relate to the number of bacteria transferred to the next node.

To examine this hypothesis several simulations were run with different parameter values for growth rate of susceptible (S) bacteria (μS). The simulations were split in order to obtain more reliable results for this twofold hypothesis; one round of simulations run with an infrequent cleaning interval, and one round of simulations run with a more optimal cleaning interval (based upon results from hypothesis 3). This was to ensure that results for the first part of the hypothesis was based solely upon removal of bacteria from the network due to natural death or death by antibiotics. Six different values for μS were chosen; 0.01, 0.02, 0.05, 0.08, 0.1. Other parameter values were held constant.

Under normal conditions we would not expect to see a remarkable spreading of resistance genes. Thus, to make sure a number of I individuals were introduced to the system during the simulation time, some of the parameters were given somewhat unnatural values; values that

would possibly provoke a more rapid transmission of resistance genes. These include parameters such as the transmission rate from S bacteria to I bacteria, and from I bacteria to S bacteria (beta and beta2). The transmission rate was set to be relatively high for these simulations. Resistance transmission is naturally a slow process, hence the simulations should, in addition to the adjustments made to the transmission rates, be run for a considerable amount of time.

4.4.1 Part 1: The effect of growth rate on susceptibility to antibiotics and disinfectants

Patient nodes were excluded from the network before results were studied. The network was generated as a Barabási-Albert (BA) network with 500 nodes and $m = 5$. Here, m represents the number of edges to be attached from new nodes of the network to already existing nodes. Growth conditions are more optimal in patients than on surfaces, and results including patient nodes would possibly cloud the results of interest. We were interested in observing the changes seen in transit nodes solely. Results showed that during the complete simulation run, the number of I-infected nodes did not differ much. At iteration 250; after approximately 5 days, the numbers of I-infected transit nodes were respectively 88, 96, 116, 129 and 136. The number of S-infected transit nodes also increased slightly proportionally to the growth rate. See the line chart below (figure 15).

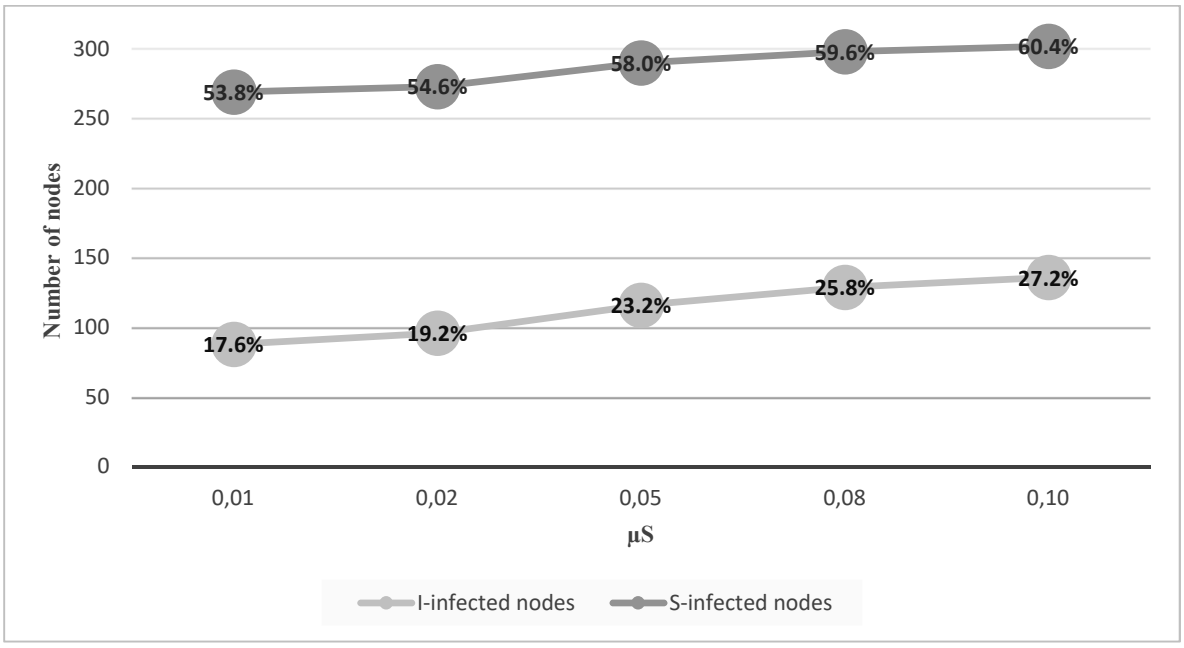


Figure 15: Line chart displaying the increasing number of nodes infected by S bacteria and I bacteria at iteration 250. Numbers displayed inside the data points correspond to the percentage of total number of transit nodes in the network.

Despite not expecting a huge difference of I bacteria due to the increase of growth rate, there was still expected some variation. This was due to the general growth rate being increased. The growth rate of I bacteria was the growth rate of S bacteria minus the fitness cost of I bacteria. The growth of I bacteria is thus assumed to follow the growth of S bacteria proportionally. Figure 15 shows a virtually identical increased number of nodes infected by S bacteria and I bacteria. A regression analysis was run to examine the actual correlation between the two datasets. The regression analysis calculated a correlation value of 0.997. It was concluded that the increase of I-infected nodes in this case is more dependent on the increase in S-infected nodes due to increased growth rate for S bacteria, than other parameters.

Changes in susceptibility to antibiotics may be of different reasons. It is commonly explained by slow growth giving more time for bacteria to acclimatize to potential environmental changes. Growth rate may affect fitness cost and transmission rate; other significant factors for susceptibility towards antibiotics. Fitness cost and transmission rate are adjustable parameters in this network model, but once set, they remain fixed throughout the simulation. The parameters will not be affected by the simulation itself, and valid results on the subject in question are thus not generated by this version of the tool. Additionally, the effectiveness for the antibiotics is set to 90% for these simulations. Research show that insufficient treatment dosages over lengthy periods of time may increase resistance development (Levy, 2002; Gullberg et al., 2011, Gehring et al., 2010). Antibiotic effectiveness was not taken into account while investigating the first part of the hypothesis. A reasonable conclusion to draw was that the simulation tool of this thesis was not, at this point, fit to perform analysis such as presented here.

It seems likely that the bacterial populations require some time to reach a steady level after cleaning. Selecting solely one iteration for presentation of results is possibly not an adequate course of action. The differences would possibly be more significant if run for a longer time. Nevertheless, no further investigation on this part of the hypothesis was conducted, due to the assumption that a realistic visualization of growth rate influencing susceptibility would not be possible with the tool at current time. Further investigation on part 1 of the hypothesis was discontinued.

4.4.2 Part 2: The effect of growth rate on spreading within a Barabási-Albert (BA) network

The second part of the analysis was run with a more optimal cleaning interval. Cleaning interval was set to 6 hours. This adjustment was done to possibly achieve a more realistic outcome; to visualize how growth rate might affect a real-life system. The network was generated as a Barabási-Albert network with 500 nodes and $m = 5$. The table below (table 6) displays the number of S-infected transit nodes at iterations 240-253. Cleaning was estimated to occur between iteration 240 and 241, and between iteration 252 and 253.

Table 6: Table displaying the number of S-infected transit nodes for each of the networks with different growth rate (μS), at iterations 240-253.

Iteration/ growth rate (μS)	0.01	0.02	0.05	0.08	0.1
240	52	54	60	74	79
Cleaning					
241	4	4	6	6	6
242	8	8	11	15	16
243	14	15	17	19	23
244	20	21	26	27	28
245	23	25	32	34	37
246	34	36	42	49	51
247	47	52	56	65	66
248	64	66	71	80	83
249	90	91	98	101	105
250	112	112	117	121	124
251	127	127	131	135	137
252	145	147	153	154	155
Cleaning					
253	16	16	16	17	19

From table 6 it can be seen that the number of S-infected transit nodes decreased quite drastically between iteration 240 and 241, and between iteration 252 and 253, which supports the prediction that cleaning had occurred. Cleaning was relatively sufficient, and number of S-infected transit nodes decreased to a fairly even level. The results also supported the hypothesis that spreading activity increases proportionally with increased growth rate. Frequent cleaning tends to keep a hold on the spreading process, which may be the reason that the differences shown in table 6 are not huge. The differences seen were however consistent,

and they showed that the growth rate did not only affect the growth in between the cleaning intervals, but also the total number of S-infected nodes subsequent to cleaning.

Values from the simulation at iteration 250 were collected in order to study the bacterial number at each of the top populated transit nodes. The values are presented in table 7.

Table 7: Table showing the number of bacteria in the top populated nodes; the nodes containing the highest number of bacteria, from each of the networks with different growth rate (μS). Values correspond to iteration 250, after approximately 5 days.

Node rating/growth rate (μS)	0.01	0.02	0.05	0.08	0.1
1	39388	39388	39388	39388	39388
2	3241	3241	3241	3241	3241
3	993	993	993	993	993
4	947	947	947	948	948
5	914	914	914	915	915
6	912	912	912	912	912
7	846	846	846	846	846
8	700	700	700	701	702
9	699	699	699	699	699
10	657	657	657	657	657
11	615	615	615	615	616
12	582	582	582	582	582
13	571	571	571	571	571
14	516	516	516	516	517
15	441	441	441	441	441

From table 7 it can be seen that the differences in bacterial numbers of the most populated nodes were negligible small – at this point of the simulation, they were even identical. Hence, it seemed likely that the differences seen in table 6 did not have the impact as first assumed. A regression analysis was run to calculate a p-value for the values of table 6, to control the actual significance of the differences displayed in table 6. Values corresponding to the lowest growth rate (μS) and the highest growth rate (μS) were compared. When running a regression analysis of five values of each column, the p-value was calculated to $1.4 \cdot 10^{-5}$.

Despite the calculated p-value, it seemed that the differences shown in table 6 were deceptive: There is a possibility that the significance of the calculated growth rate is overestimated. Table 6 included all nodes containing >1 bacteria. We had no certain way of stating whether the additional number of nodes in table 6 were significant or not, due to the fact that slight

differences may have originated from nodes containing insignificant numbers of bacteria. Perhaps these very few bacteria were insignificant at this early point of measurement, but they could possibly become essential for changes occurring at a later point. Bacteria divide rapidly, and a node containing few bacteria may quickly become highly populated. Perhaps these very slim differences would contribute to more significant differences at a later point. To investigate if the results were consistent in the long run, a simulation was repeated with a prolonged time interval. Iteration number was set to 18000, which signifies approximately 12.5 months. The network for the second round of simulations for part 2 was generated as a Barabási-Albert network with 1000 nodes and $m = 5$. The other parameter values were kept identical to the previous run.

At iteration 18000; after approximately 12.5 months, a number of 107 transit nodes infected with S bacteria after simulation with growth rate 0.01 existed in the network, a number of 117 S-infected transit nodes after simulation with growth rate 0.02, a number of 146 S-infected transit nodes after simulation with growth rate 0.05, a number of 174 S-infected transit nodes after simulation with growth rate 0.08, and a number of 191 S-infected transit nodes after simulation with growth rate 0.1 (figure 16).

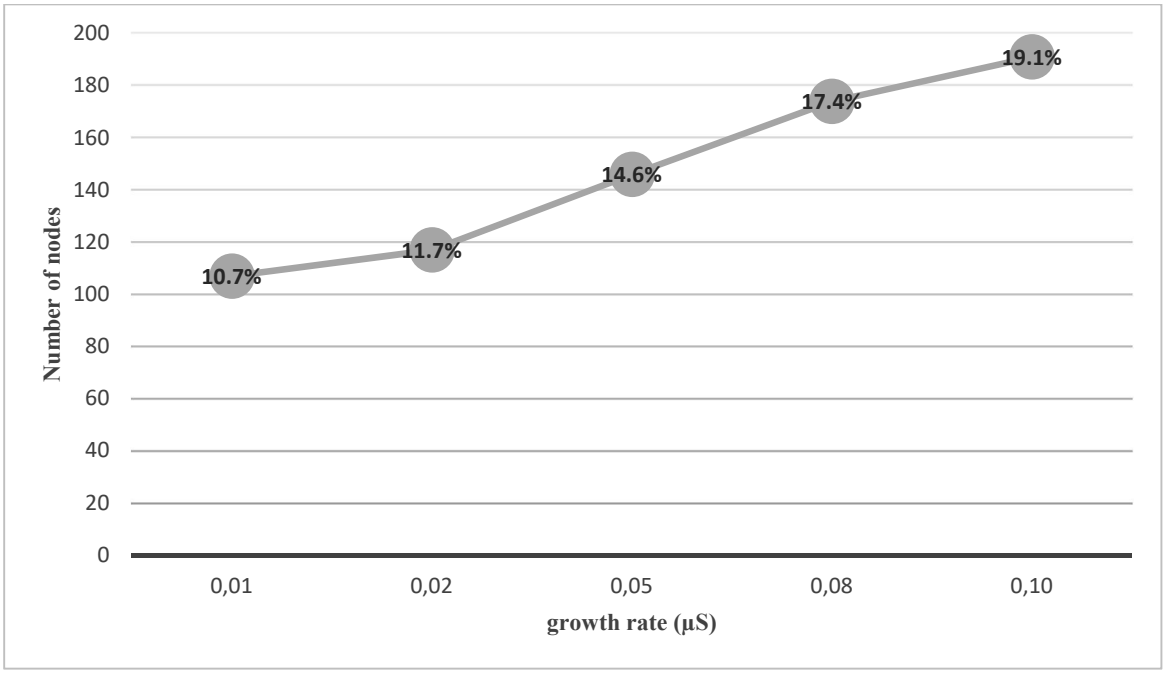


Figure 16: Line chart displaying the increasing number of nodes infected by S bacteria for each of the networks with different growth rate (µS). Values correspond to iteration 18000. Numbers displayed inside the data points correspond to the percentage of total number of nodes in the network.

A regression analysis was run to calculate a p-value for the new results. Values of the lowest μS (0.01) and the highest μS (0.1) were compared. The regression analysis calculated a p-value of $3.8 \cdot 10^{-9}$. The p-value corresponding to these values also indicated statistical significant differences. However, the p-value corresponding to the simulations run for approximately 12.5 months (18000 iterations) indicated more significant differences than the p-value corresponding to the simulations run for approximately 5 days (250 iterations). Results indicated that time was essential for distinct differences due to changes to μS to become apparent.

To further study the significance of the differences displayed in figure 16, the number of bacteria in the top populated transit nodes of each of the networks of different μS were collected. Table 8 shows an overview of the most populated nodes from the second round of simulations at iteration 18000.

Table 8: Table showing the number of bacteria in the top populated nodes; the nodes containing the highest number of bacteria, from each of the networks with different growth rate (μS). Values correspond to iteration 18000, after approximately 12.5 months.

Node rating/growth rate (μS)	0.01	0.02	0.05	0.08	0.1
1	258	304	487	618	888
2	137	160	293	570	724
3	121	151	251	317	371
4	88	103	162	204	239
5	83	98	157	199	233
6	61	72	117	154	199
7	53	63	103	132	189
8	42	50	81	127	167
9	27	33	65	125	156
10	25	29	45	61	75
11	23	27	44	57	67
12	20	23	38	48	67
13	19	22	35	47	63
14	18	22	33	46	60
15	15	18	29	45	60

None of the iterations for these simulations, neither the simulation run for 250 iterations nor the simulation run for 18000 iterations, displayed existence of I-infected transit nodes in the network. This may be interpreted as the fact that frequent cleaning, in this case, was substantial towards the eradication of I development in the network. The fitness cost

parameter was set to 0.3 for these simulations. A fitness cost of 0.3 will for this modeling tool signify the growth rate of I bacteria to be 70% of the growth rate of S bacteria. The fitness cost may, in addition to the cleaning, have constituted a significance to the results showing an extinction of I bacteria.

At current time, the simulation process runs a bit slow for large networks. This may be fixed in later versions of the tool by making some changes to how events are handled upon traffic through the network.

4.5 Hypothesis 5: The effect of changes to fitness cost

Acquisition of new traits is said to come with a fitness cost, which in turn may affect the spreading rate of that exact trait. A high fitness cost is assumed to affect the spreading through the network in a negative manner, due to it poses a disadvantage for growth and establishment of bacteria. I bacteria would thus be expected to have difficulties establishing in a population in the long run. An infrequent cleaning interval was used for these simulations in order to give a better opportunity for the I bacteria to proliferate, and possibly to more clearly observe differences.

The parameter fitness cost affects the rate of growth for I bacteria. If fitness cost is set to for example 0.2, it means that I bacteria grow at an 80% rate of S bacteria. Fitness cost is in other words a parameter that reflects the relationship between the growth rate for S bacteria and the growth rate for I bacteria. Simulations were run with fitness cost >0; a disadvantage for the growth of I bacteria, and with fitness cost <0; an advantage for growth of I bacteria. Several simulations were run on a Barabási-Albert (BA) network of 1000 nodes and $m = 5$, with varying parameter values for fitness cost; -0.5, -0.1, 0.1, 0.3, 0.5, and 0.8. The simulations were run for 18000 iterations, which constitutes approximately 12.5 months.

Table 9 shows a selection of the results generated upon simulation. The values collected in the table correspond to the number of I-infected transit nodes. The table displays results from iterations 17600 to 17610; the iterations around the time of the last scheduled cleaning. From table 9 we can see that there are slight differences between the values. Differences that did correspond to the original hypothesis that fitness cost affects the growth of I bacteria. The

significance of the differences was controlled by running a regression analysis to calculate a p-value. The regression analysis was based upon the values for highest fitness cost and the values for lowest fitness cost, finding the highly significant p-value of $1.2 \cdot 10^{-8}$. Additionally, a trend was observed with a difference proportional to increasing fitness cost.

Table 9: Table displaying the number of I-infected transit nodes for each of the networks with different fitness cost, at iterations 17600-17610.

Iteration/ fitness cost	-0.5	-0.1	0.1	0.3	0.5	0.8
17600	559	541	534	521	508	484
17601	60	58	54	52	50	47
17602	122	113	111	108	103	98
17603	170	159	156	153	146	138
17604	208	193	184	176	170	162
17605	260	241	228	213	204	196
17606	297	275	261	246	236	221
17607	328	294	279	270	255	234
17608	359	319	309	294	280	261
17609	367	334	323	306	291	275
17610	413	383	372	354	334	312

An assumption was made that time was essential for this parameter to cause any significant differences. Perhaps should simulations be run for several more iterations before more significant differences would become apparent in the results. The assumption was further investigated by comparing the values from table 9 to the corresponding values of iterations 110-115. All values were equal for all the different fitness cost values at this point of the simulation; the numbers of transit nodes infected with I bacteria were respectively 30, 39, 50, 57, 71, and 80. These results supported the assumption that time was essential for changes to become apparent. However, it did not seem likely that time alone would constitute a huge difference for this exact parameter.

As for hypothesis 4, an important factor that may have inhibited the results from showing more distinct differences could be fixed parameter values, such as growth rate, transmission rate, probability of transfer between nodes, and number of nodes transferred across edges. The fitness cost of this tool is for example not implemented in such a way that it may affect the transmission rate directly. Transmission rate is a fixed parameter value that may be customized before simulation start. Connecting these parameter values to vary depending on

each other to a higher degree may possibly give more realistic results. This option is not implemented in the simulation tool at this point, but it may be an idea for future versions. However, the tool is to some degree, able to simulate situations of which growth of I bacteria is constrained due to fitness cost alone.

Research suggests that the negative impact of fitness cost in bacteria may be cancelled out in time by compensatory mutations (Andersson & Huges, 2010; Dzidic & Bedeković, 2003). The network model tool of this thesis is not expected to reflect such high degree of detail, and highly credible results on details as these are not expected.

4.6 The virtual hospital simulation: Orkdal hospital

The hospital simulation was partly run on the basis of information collected from the bachelor thesis previously discussed (Brunes & Finstad, 2015), and partly run on the basis of approximations found in literature. The file including node information, and the file including edge information, were generated randomly as a Barabási-Albert (BA) network by running the random network generator. The files were further edited to represent the hospital network more accurately. The node properties were custom made for the hospital run. Customization involved inclusion of bacterial numbers collected from the bachelor thesis (Brunes & Finstad, 2015). See tables 2 and 3 for data. Certain transit nodes of the network were excluded from patient contact. This was done in order to mimic the locations at which patient contact is non-existent, such as for example storage rooms. The edge properties were adjusted to mimic the four wards of the virtual hospital. Some of the transit nodes, especially the nodes representing equipment of the corridors, were made the most well-connected nodes of the network; the nodes of highest degree. The highest level of interaction was intended within the wards, but several connections were made between transit nodes across the four wards as well.

The simulation was run for 18000 iterations; approximately 12.5 months. Other parameter values chosen for the simulation are described more in detail in Material and methods. Output files generated upon simulation were used to observe changes and resistance development continually from simulation start to end-point. I wished to study the dynamics of the spreading process; would the number of S and I bacteria peak at some point, would I bacteria

dominate the system in the long run, or would the spread of I bacteria eventually fade? These were some of the questions in mind before running simulation of the virtual hospital Orkdal.

The traffic; fraction of traffic across edges, was compared to the predetermined probability of transfer across edges. A prediction was made that the edges displaying high probability of transfer would correspond to the edges of highest traffic. The top 12 edges of transfer probability; the edges expected to partake in most of the transfer occurring during the simulation, were compared to the edges displaying traffic values>1%. 11 of the 12 edges of highest transfer probability were represented in the list of traffic values >1%. The top three edges of traffic values directly corresponded to the top three edges of transfer probability. In addition to comparing traffic against transfer probability, traffic was compared to the transit nodes of highest connectivity. As many as 82 out of the 92 edges with traffic >1% were associated with top ten connected nodes.

These results did not provide any remarkable new information, but the findings were concluded as relevant in order to visualize how parameter settings may directly influence results. Such knowledge opens for the possibility to customize the network model to fit each individual desire of use.

Population size in the four customized transit nodes from prior to simulation start was compared to the population size in the corresponding transit nodes at simulation end-point. Results are presented in the table below (table 10).

Table 10: Table containing information about the four customized transit nodes for the hospital simulation. The table contains information about the locations of the four transit nodes, the number of bacteria in each of the nodes at simulation start, the survival frequency of each of the nodes upon cleaning, and the number of bacteria in the nodes at simulation end.

Room, ward	Equipment/location	Transit node number	Number of S bacteria at simulation start	Number of S bacteria at simulation end	survival_freq
Storage room, A3	Blood pressure monitor	2	7547465	25691	0.0168
Patient lounge, B2	Remote control	18	21949238	103527	0.0354
Patient room, B3	Bed frame	35	25670474	58984	0.0083
Corridor, B4	Cupboard handle	36	2841516	426638	0.0006

Table 10 shows that transit node 2 contained 0.34% of the initial number of bacteria at simulation end, transit node 18 contained 0.47%, transit node 35 contained 0.22%, and transit node 36 contained as much as 15% of the initial bacterial number.

Additionally, bacterial numbers at the simulation end was compared to bacterial numbers in nodes that originally were unpopulated at simulation start. The numbers of bacteria in the four nodes represented in table 10, were subsequently compared with respect to connectivity and transfer probability. The top 10 most populated nodes at the last iteration were nodes 33, 34, 11, 28, 42, 47, 1, **36**, 29 and 4. Ergo, solely one of the four nodes from table 10 was represented among the top 10 populated nodes at end-point. From the result files related to the hospital simulation, it was shown that node 36 was included in 23 of the 199 edges of the network. Node 35 was included in six of the edges, node 18 in three, and node 2 in six. Node 36 was also included in the edge of second highest traffic.

The fact that only one of the four nodes from table 10 was represented among the top 10 populated nodes at end-point, seemed a bit strange considering the customization of node properties. One would possibly assume that the initially well populated nodes, such as node 35, would lie above node 36 in population size, since node 36 was given the lowest bacterial number of the four transit nodes of focus, and that the survival frequency was lower than for the other three nodes. The survival frequency of the four different nodes all did correspond to the fraction of bacteria that were existent in the respective nodes at simulation end versus simulation start, except transit node 36. If survival frequency was a critical parameter, node 36 should have been even less populated than the other nodes. A likely explanation to the fact that node 36 served as one of the most populated nodes at simulation end-point may be that node 36 had a higher connectivity, and that the node was included in a greater number of edges than the other three nodes. According to this discovery, it seemed likely that the connectivity and transfer probability were more decisive parameters than the probability of survival and initial population size. In addition to traffic, as discussed previously. However, connectivity and traffic seemed to be strongly correlated.

Further, to estimate the significance of patient contact, the transit nodes that were disconnected from simulation start were studied. The aim was to investigate hypothetically if a spreading situation could possibly be eliminated more rapidly if an increased number of locations were made inaccessible to patients. The numbers of bacteria in the specific transit nodes were compared to the numbers of bacteria in other transit nodes at the end-point of the simulation. In order to find the best comparable nodes for observation, node degree was

considered; nodes for comparison to the disconnected nodes were selected based upon similarity in node degree. There was no knowledge of the fact that node degree was with certainty the most viable parameter for selection, and there was no knowledge related to what other parameters should have been included. Node degree was nevertheless regarded to be the most relevant parameter in this case. The conclusion was based upon previous results in relation to the hospital simulation, showing that connectivity and traffic were important parameters when estimating node destiny.

The first results showed no nodes containing S bacteria at the time of simulation start, despite the node properties being edited in such manner. This discrepancy implied a source of error somewhere: The reason for the discrepancy between expectations and results turned out to be a scheduled cleaning at simulation start point. A new version of the simulation program file was made, in which cleaning was not scheduled to start until the iteration specified by the parameter file. The new version was not required for the simulations run in relation to the other hypothesis, due to none of the networks being customized to contain bacteria from simulation start. For the hospital network however, this edit was crucial.

The simulation was run again with the new version of the program file. The results generated by the second run seemed more reasonable, and the numbers were used further. From table 11 it can be seen that the number of bacteria in the transit nodes differ somewhat from the adjustments made prior to simulation start. The reason may be that some of the bacteria were being transferred across the edges already from simulation start point. The numbers did however not differ significantly, which seems obvious due to the time of measurement.

Table 11: Comparison of number of bacteria in transit nodes disconnected from/connected to patient nodes at simulation end. Nodes for comparison to the disconnected nodes are selected on the basis of node degree.

Transit node, disconnected	Number of bacteria	Transit node, connected	Number of bacteria
2	25691	47	694720
3	97261	44	56698
14	4328	23	21420
15	13737	29	349833
26	13688	16	254121
27	1075	6	64849
38	29884	45	23578
39	21285	7	14636

A regression analysis was run for the values of table 11 to estimate correlation between the two datasets, and to calculate a p-value. The regression analysis produced no significant results (p-value = 0.8). Further, with a correlation value of -0.106, the regression analysis disproved a correlation between the two sets of values. The results indicating no distinct trend for either of the two groups of nodes, were considered probable. Partly due to the arguments above, and partly due to the results from hypothesis 1, showing clearly that disconnected transit nodes are important for a spreading to progress through a network.

Through the hospital simulation there was also an aim to demonstrate how a potential development of I bacteria within a hospital could proceed in a semi-realistic manner. In comparison to the expectations of some of the previous hypotheses, in which parameters were set to speed up the process of I bacteria development, a remarkable manifestation and spreading activity of I bacteria was not expected in the hospital network.

The spread of I bacteria through the network was studied by following detected number of nodes in parallel to the iterations. I bacteria were programmed to have a delay before entering the network. According to the parameter-file, the delay for I bacteria to enter the system was set to 7 days; 336 iterations. However, from the results it was clear that the first patient node infected with I bacteria entering the system was at iteration 578. Further, we saw that I bacteria did not manifest in transit nodes until iteration 587. At iteration 588, a single node contained I bacteria. The same number was detected at iteration 588. The next iterations, as of iteration 589, did not show existence of I bacteria in any of the transit nodes. The next detection of I bacteria in the system was at iteration 651. Following iteration 651, transit

nodes containing I bacteria increased, until a new round of cleaning between iteration 660 and 661. Similar cycles were consistent for I bacteria in an increasing manner up until simulation stop; I-infected transit nodes fluctuated proportionally to cleaning interval. The number of S-infected transit nodes followed the same pattern. The fluctuations during the cycles were steady, and there was no sudden increase of infected nodes. At the last iteration, the network contained a total of 23 transit nodes infected with I bacteria.

The reason no I-infected transit nodes were detected in the network at the iteration scheduled for I bacteria to start entering the system, at iteration 336, may be that the cleaning interval of 6 hours lead to cleaning simultaneously. Another reason may be that no patient nodes infected with I bacteria entered the system until even more time had passed. The latter seemed to be the case of this simulation run. Results showed that I-infected patient nodes connected to the network at iteration 578, which was directly following scheduled cleaning. It thus seemed reasonable that I bacteria manifested in transit nodes at iteration 587, after some time of stabilization. The sudden disappearance of I-infected transit nodes after iteration 588, also seemed to be in line with programmed cleaning interval; cleaning was according to the parameter-file scheduled between iteration 588 and 589. The network tool did not neglect bacterial numbers <1, hence I bacteria could in principle have been existent in the network, despite not being evaluated. Half a bacterium is not viable; thus, bacterial numbers <1 were ignored.

Cleaning seemed to keep a hold on the spread of I bacteria within the hospital network for quite some time. Some increase in I-infected transit nodes was still observed at the time before the last measurement. Considering that cleaning was not programmed to eradicate the total amount of bacteria for each round of cleaning, this made sense. How drastically the number increased from time to time was to some degree coincidental. The individual chance of transit nodes being cleaned varied independently from the “importance” of the nodes. Additionally, we saw that an increasing number of patient nodes were connected to the network as time went by. Sometimes patient nodes connect to nodes of high importance related to spreading, such as hubs. This may constitute a greater significance for spreading through the network than if a patient node connects to a sparsely connected node.

We have yet not established a way to separate the I bacteria that appear as a result of horizontal gene transfer, vertical gene transfer, or immigration. This may be changed in later versions of the tool. By doing so, one would possibly have better chances of predicting and evaluating spreading activity.

One is able to observe the development of I bacteria at specific locations of the hospital by following specific nodes throughout the simulation. This act is only beneficial provided that all relevant information is known and accounted for. The hospital simulation of this thesis lacks detailed information related to for example the movement amidst the different locations and wards. Furthermore, edge transfer probability was generated randomly for the hospital simulation. This parameter was not customized specifically for the hospital network due to lack of information related to an actual movement pattern and transfer probability. This means that some of the locations of the hospital possibly were set to a higher/lower transfer probability than realistic for that specific location. Mapping out development at single locations is thus not expedient for the virtual hospital network. Lack of specific information was however not considered particularly problematic due to the aim of the simulation; the hospital simulation was anticipated as a hypothetical situation, not as a true copy of real life. Besides, a movement pattern is rarely fixed, and natural fluctuations will occur.

5 Conclusion and future perspectives

The ongoing emergence of antibiotic resistance worldwide raise the need for new and effective methods to stagnate the spread of antibiotic resistance. During the work of this thesis, a network model tool was developed with the intent to serve as a contribution to the fight against antibiotic resistance. Several hypotheses were formulated in order to illustrate the potential and possible applications of the tool, and these hypotheses were further examined. Through multiple simulations, it was demonstrated how parameters such as the frequency of cleaning, the bacterial growth rate, and the fitness cost related to the acquisition of resistance genes, may possibly influence the spreading through a network. Results proved that transit nodes were essential for spreading to progress, and that different types of networks display different spreading patterns. Results also indicated a seemingly optimal cleaning interval of 6-9 hours, and that increased growth rate was proportional to spreading activity in the network. Additionally, a semi-realistic simulation of a virtual hospital was conducted to present the usage of such tools in the study of real-life events.

Ahead in time, I see potential in further modification and development of the tool to better mirror real-life situations. By implementing a more specific framework, and by including more detailed information, I believe that the tool may serve as an effective and economical contribution to future work related to antibiotic resistance. Furthermore, I believe that the adaptability of the tool expands its potential. The tool is not limited to solely handling transmission of antibiotics genes within a hospital. It may be utilized as a helpful remedy in predicting spreading of infectious agents, in determining parameters and factors that are important and may affect various systems, and to determine how these parameters and factors affect the system. It may also be used to evaluate existing routines and procedures, such as hospital hygiene procedures and other infection control strategies, and as a means of establishing new routines and procedures.

References

About HUNT. (2018). Retrieved from https://www.ntnu.edu/hunt/about-hunt

Arepyeva, M. A., Kolbin, A. S., Sidorenko, S. V., Lawson, R., Kurylev, A. A., Balykina, Y. E., . . . Spiridonova, A. A. (2017). A mathematical model for predicting the development of bacterial resistance based on the relationship between the level of antimicrobial resistance and the volume of antibiotic consumption. *Journal of Global Antimicrobial Resistance, 8*, 148-156. https://doi.org/10.1016/j.jgar.2016.11.010

Aminov, R. I. (2010). A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Frontiers in Microbiology*, *1*, 134. https://doi.org/10.3389/fmicb.2010.00134

Anderson, R. M., & May, R. M. (1992). *Infectious Diseases of Humans: Dynamics and Control*. New York, NY: Oxford University Press.

Andersson, D. I., & Huges, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology, 8*, 260–271. https://doi.org/10.1038/nrmicro2319

Andersson, H., Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis.* New York, NY: Springer. https://doi.org/10.1007/978-1-4612-1158-7

Barabási, A-L. (2016). *Network Science*. Retrieved from https://barabasi.com/networksciencebook/

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences, 99*, 7280-7287. https://doi.org/10.1073/pnas.082080899

Brunes, T., & Finstad, N. R. (2015). *Microbiological examination for the presence of antibiotic resistant bacteria on equipment and inventory at different departments at Orkdal hospital.* (Unpublished bachelor thesis). NTNU, Trondheim, Norway.

Chain, E., Florey, H. W., Gardner, A. D., Heatley, N. G., Jennings, M. A., Orr-Ewing, J., & Sanders, A. G. (1940). Penicillin as a chemotherapeutic agent. *The Lancet, 236*(6104), 226-228. https://doi.org/10.1016/S0140-6736(01)08728-1

Chartier, Y., Emmanuel, J., Pieper, U., Prüss, A., Rushbrook, P., Stringer, R., . . . Zghondi, R. (2014). *Safe management of wastes from health-care activities*. World Health Organization.

Colizza, V., Barrat, A., Barthélemy, M., & Vespignani, A. (2006). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America, 103*(7), 2015-2020. https://doi.org/10.1073/pnas.0510525103

Dancer, S. J. (1999). Mopping up hospital infection. *Journal of Hospital Infection, 43*(2), 85-100. https://doi.org/10.1053/jhin.1999.0616

Davies, J., & Davies, D. (2010). Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews: MMBR*, *74*(3), 417-433. https://doi.org/10.1128/MMBR.00016-10

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research, 2*(1), 23-41. https://doi.org/10.1177/096228029300200103

Donlan, R. M., & Costerton, J. W. (2002). Biofilms: Survival Mechanisms of Clinically Relevant Microorganisms. *Clinical Microbiology Reviews, 15*(2), 167–193. https://doi.org/10.1128/CMR.15.2.167-193.2002

Dzidic, S., Bedeković, V. (2003). Horizontal gene transfer – emerging multidrug resistance in hospital bacteria. *Acta Pharmacologica Sinica, 24*(6), 519-26.

Evans, D. J., Allison, D. G., Brown, M. R. W., & Gilbert, P. (1990). Effect of growth-rate on resistance of Gram-negative biofilms to cetrimide. *Journal of Antimicrobial Chemotherapy, 26*(4), 473-478. https://doi.org/10.1093/jac/26.4.473

Eames, K., Bansal, S., Frost, S., & Riley, S. (2015). Six challenges in measuring contact networks for use in modelling. *Epidemics, 10*, 72-77. https://doi.org/10.1016/j.epidem.2014.08.006

Flemming, H.-C., & Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology, 8*, 623. https://doi.org/10.1038/nrmicro2415

FOR – The Operation Room of the Future. (2017, October 31). Retrieved from https://stolav.no/seksjon-engelsk/Sider/FOR.aspx

Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. (2011). An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. *IEEE.* https://doi.org/10.1109/PASSAT/SocialCom.2011.142

Frieden, T. (2013). Antibiotic Resistance Threats in the United States 2013. Retrieved from https://www.cdc.gov/drugresistance/threat-report-2013/index.html

Gehring, R., Schumm, P., Youssef, M., & Scoglio, C. (2010). A network-based approach for resistance transmission in bacterial populations. *Journal of Theoretical Biology, 262*(1), 97-106. https://doi.org/10.1016/j.jtbi.2009.09.002

Greenhalgh, D., Liang, Y., & Mao, X. (2016). SDE SIS epidemic model with demographic stochasticity and varying population size. *Applied Mathematics and Computation, 276*, 218-238. https://doi.org/10.1016/j.amc.2015.11.094

Gullberg, E., Cao, S., Berg, O. G., Ilbäck, C., Sandegren, L., Hughes, D., & Andersson, D. I. (2011). Selection of Resistant Bacteria at Very Low Antibiotic Concentrations. *PLoS Pathogens, 7*(7). http://doi.org/10.1371/journal.ppat.1002158

Homer, J. B., & Hirsch, G. B. (2006). System Dynamics Modeling for Public Health: Background and Opportunities. *American Journal of Public Health, 96(*3), 452-458. https://doi.org/10.2105/AJPH.2005.062059

Horan, T. C., White, J. W., Emori, T. G., Culver, D. H., Munn, V. P., Thornsberry, C., . . . Hughes, J. M. (1986). Nosocomial infection surveillance. *MMWR Surveillance Summaries, 35*(1), 17-29. Retrieved from: https://www.cdc.gov/mmwr/preview/mmwrhtml/00001772.htm

Hufnagel, L., Brockmann, D., & Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America, 101*(42), 15124-15129. https://doi.org/10.1073/pnas.0308344101

Iannelli, M., & Pugliese, A. (2014). *An Introduction to Mathematical Population Dynamics: Along the trail of Volterra and Lotka.* Springer International Publishing.

Jevons, M. P. (1961). "Celbenin" – resistant Staphylococci. *British Medical Journal, 1*(5219), 124–125.

Jiang, D., Yu, J., Ji, C., & Shi, N. (2011). Asymptotic behavior of global positive solution to a stochastic SIR model. *Physica A: Statistical Mechanics and its Applications, 54*, 221-232. https://doi.org/10.1016/j.physa.2017.04.100

Keeling, M. J., & Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. New Jersey, NJ: Princeton University Press.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks, 2*(3), 203-271. https://doi.org/10.1093/comnet/cnu016

Kohanski, M. A., Dwyer, D. J., & Collins, J. J. (2010). How antibiotics kill bacteria: from targets to networks. *Nature Reviews. Microbiology*, *8*(6), 423-435. https://doi.org/10.1038/nrmicro2333

Kong, K.-F., Schneper, L., & Mathee, K. (2010). Beta-lactam Antibiotics: From Antibiosis to Resistance and Bacteriology. *APMIS, 118*(1), 1-36. https://doi.org/10.1111/j.1600-0463.2009.02563.x

Krilov, L. R., Barone, S. R., Mandel, F. S., Cusack, T. M., Gaber, D. J., & Rubino, J. R. (1996). Impact of an infection control program in a specialized preschool. *American Journal of Infection Control, 24*(3), 167-173. https://doi.org/10.1016/S0196-6553(96)90008-5

Lahrouz, A., & Settati, A. (2013). Asymptotic properties of switching diffusion epidemic model with varying population size. *Applied Mathematics and Computation, 219*(24), 11134-11148. https://doi.org/10.1016/j.amc.2013.05.019

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A. K. M., Wertheim, H. F. L., Sumpradit, N., . . . Cars, O. (2013). Antibiotic resistance - the need for global solutions. *The lancet Infectious diseases*, *13*(12), 1057-1098. https://doi.org/10.1016/S1473-3099(13)70318-9

Lázár, V., Pal Singh, G., Spohn, R., Nagy, I., Horváth, B., Hrtyan, M., . . . Pál, C. (2013). Bacterial evolution of antibiotic hypersensitivity. *Molecular Systems Biology*, *9*, 700. https://doi.org/10.1038/msb.2013.57

Levy, S. B. (2000). Antibiotic and antiseptic resistance: impact on public health. *The Pediatric Infectious Disease Journal, 19(*10), 120-122.

Levy, S. B. (2002). Factors impacting on the problem of antibiotic resistance. *Journal of Antimicrobial Chemotherapy, 49*(1), 25-30. https://doi.org/10.1093/jac/49.1.25

Lin, J., Nishino, K., Roberts, M. C., Tolmasky, M., Aminov, R. I., & Zhang, L. (2015). Mechanisms of antibiotic resistance. *Frontiers in Microbiology*, *6*, 34. https://doi.org/10.3389/fmicb.2015.00034

Longini, I. M. (1988). A mathematical model for predicting the geographic spread of new infectious agents. *Mathematical Biosciences, 90*(1), 367-383. https://doi.org/10.1016/0025-5564(88)90075-2

Marshall, B., Sen, R. A., Chadwick, P. R., & Keaney, M. G. L. (1998). Environmental contamination of a new general surgical ward. *Journal of Hospital Infection, 39*(3), 242-243. https://doi.org/10.1016/S0195-6701(98)90265-1

McDonnell, G., & Russell, A. D. (1999). Antiseptics and Disinfectants: Activity, Action, and Resistance. *Clinical Microbiology Reviews, 12*(1), 147–179.

McManus, M. C. (1997). Mechanisms of bacterial resistance to antimicrobial agents. *American Journal of Health-System Pharmacy, 54*(12), 1420-1433.

Mehta, Y., Gupta, A., Todi, S., Myatra, S., Samaddar, D. P., Patil, V., . . . Ramasubban, S. (2014). Guidelines for prevention of hospital acquired infections. *Indian Journal of Critical Care Medicine: Peer-Reviewed, Official Publication of Indian Society of Critical Care Medicine, 18*(3), 149–163. https://doi.org/10.4103/0972-5229.128705

Melnyk, A. H., Wong, A., & Kassen, R. (2015). The fitness costs of antibiotic resistance mutations. *Evolutionary Applications, 8*(3), 273–283. https://doi.org/10.1111/eva.12196

Molin, S., & Tolker-Nielsen, T. (2003). Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Current Opinion in Biotechnology, 14*(3), 255-261. https://doi.org/10.1016/S0958-1669(03)00036-3

Monod, J. (1949). The Growth of Bacterial Cultures. *Annual Review of Microbiology, 3*(1), 371-394. https://doi.org/10.1146/annurev.mi.03.100149.002103

MRSA Carrier: Questions and Answers. Retrieved from https://www.staph-infection-resources.com/info/carrier/

Murray, P. R., Rosenthal, K. S., & Pfaller, M. A. (2012). *Medical Microbiology.* Elsevier Health Sciences. Retrieved from: https://elsevierelibrary.co.uk/pdfreader/medical-microbiology43045

Nauta, E. H., Mattie, H. (1976). Dicloxacillin and cloxacillin: Pharmacokinetics in healthy and hemodialysis subjects. *Clinical Pharmacology & Therapeutics, 20*(1), 98-108. https://doi.org/10.1002/cpt197620198

Newman, M. E. J. (2002). The spread of epidemic disease on networks. *Physical Review E, 66*(1). https://doi.org/10.1103/PhysRevE.66.016128

Niehus, R., Mitri, S., Fletcher, A. G., & Foster, K. R. (2015). Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nature Communications, 6*, 8924. https://doi.org/10.1038/ncomms9924

Olivares, J., Bernardini, A., Garcia-Leon, G., Corona, F., B. Sanchez, M., & Martinez, J. L. (2013). The intrinsic resistome of bacterial pathogens. *Frontiers in Microbiology*, *4*, 103. https://doi.org/10.3389/fmicb.2013.00103

Opatowski, L., Guillemot, D., Boëlle, P.Y., & Temime, L. (2011). Contribution of mathematical modeling to the fight against bacterial antibiotic resistance. *Current Opinion in Infectious Diseases, 24*(3), 279-287. https://doi.org/10.1097/QCO.0b013e3283462362

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics, 87*(3), 925-979. https://doi.org/10.1103/RevModPhys.87.925

Rayner, C. & Munckhof, W. J. (2005). Antibiotics currently used in the treatment of infections caused by Staphylococcus aureus. *Internal Medicine Journal*, *35*, 3-16. https://doi.org/10.1111/j.1444-0903.2005.00976.x

Roberts, F. S., (2004). Computational and Mathematical Epidemiology. Science. Retrieved from: https://www.sciencemag.org/careers/2004/02/computational-and-mathematical-epidemiology

Roberts, J. A., Kruger, P., Paterson, D. L., & Lipman, J. (2008). Antibiotic resistance – What's dosing got to do with it? *Critical Care Medicine, 36*(8), 2433-2440. https://doi.org/10.1097/CCM.0b013e318180fe62

Sahneh, F. D., Scoglio, C., & Mieghem, P. V. (2013). Generalized Epidemic Mean-Field Model for Spreading Processes Over Multilayer Complex Networks. *IEEE/ACM Transactions on Networking, 21*(5), 1609-1620. https://doi.org/10.1109/TNET.2013.2239658

Sattenspiel, L. (1990). Modeling the spread of infectious disease in human populations. *American Journal of Physical Anthropology, 33*, 245-276. https://doi.org/10.1002/ajpa.1330330511

Schmidt, T., Kock, M. M., & Ehlers, M. M. (2015). Antimicrobial Resistance - An Open Challenge. Retrieved from https://www.intechopen.com/books/antimicrobial-resistance-an-open-challenge/antimicrobial-resistance-in-staphylococci-at-the-human-animal-interface

Seilheimer, R. L. (2008). Contact Network Epidemiology: Mathematical Methods of Modeling a Mutating Pathogen on a Two-type Network. Retrieved from: https://repositories.lib.utexas.edu/handle/2152/13376

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Singh, T. (2004). *Principles and Techniques of Silkworm Seed Production.* New Delhi, ND: Discovery Publishing House.

Solberg, C. O. (2000). Spread of Staphylococcus aureus in Hospitals: Causes and Prevention. *Scandinavian Journal of Infectious Diseases, 32*(6), 587-595. https://doi.org/10.1080/003655400459478

Spellberg, B., Bartlett, J. G., & Gilbert, D. N. (2013). The Future of Antibiotics and Resistance. *The New England Journal of Medicine, 368(*4), 299–302. https://doi.org/10.1056/NEJMp1215093

Tenover, F. C. (2006). Mechanisms of Antimicrobial Resistance in Bacteria. *The American Journal of Medicine, 34*(5), 3-10. https://doi.org/10.1016/j.ajic.2006.05.219

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology, 3*, 711. https://doi.org/10.1038/nrmicro1234

Todar, K. (2006). *Todar's Online Textbook of Bacteriology.* University of Wisconsin-Madison Department of Bacteriology. Retrieved from https://www.textbookofbacteriology.net

Ventola, C. L. (2015). The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics, 40(*4), 277-283.

Voit, E. O. (2012). *A First Course in Systems Biology*. New York, NY: Garland Science.

VRE sniker seg inn på norske sykehus. (2014, March 11). Retrieved from https://sykepleien.no/2014/03/vre-sniker-seg-inn-pa-norske-sykehus

WHO. (2018a). Antimicrobial resistance. Retrieved from https://www.who.int/mediacentre/factsheets/fs194/en/

WHO. (2018b). Antibiotic resistance. Retrieved from https://www.euro.who.int/en/health-topics/disease-prevention/antimicrobial-resistance/antibiotic-resistance#

Wojgani, H., Kehsa, C., Cloutman-Green, E., Gray, C., Gant, V., & Klein, N. (2012). Hospital Door Handle Design and Their Contamination with Bacteria: A Real Life Observational Study. Are We Pulling against Closed Doors? *PLOS ONE, 7*(10). https://doi.org/10.1371/journal.pone.0040171

Appendices

Appendix 1: Description of the components of the tool

generate_barabasi_v2:

A program file that creates a random Barabási-Alberts network. Number of nodes and edges may be customized. Command line to create the network is “python generate_barabasi_v2.py test2.nx”.

The edges between transit nodes are divided into two different categories, category 1 and category 2.

- fraction_category_1 corresponds to the probability for an edge to be of category 1, and vice versa for fraction_category_2
- category_1_interval_from and category_1_interval_to correspond to the expected traffic across each edge per iteration. The values are given in a uniform distribution. The same goes for category_2_interval_from and category_2_interval_to, except for category 2 edges.
- transit_fixed_max is a parameter deciding variation between maximal number of bacteria in transit nodes. If the parameter is set to 1, all the transit nodes have the same max-population. If the value is set to 0, each transit node gets its own unique max-population. The values are generated in a normal distribution.
- minimum_survival_probability and maximum_survival_probability are parameters that determine the probabilities for transit nodes to survive cleaning. The values are generated in an exponential manner.
- probability_patient_attachment decides the probability for patients being connected to the transit nodes.

generate_erdos_v2:

A program file that creates a random Erdős–Rényi network. Number of nodes and edge probability may be customized. Command line to create the network is “python generate_erdos_v2.py test2.nx”.

The edges between transit nodes are divided into two different categories, category 1 and category 2.

- fraction_category_1 corresponds to the probability for an edge to be of category 1, and vice versa for fraction_category_2
- category_1_interval_from and category_1_interval_to correspond to the expected traffic across each edge per iteration. The values are given in a uniform distribution. The same goes for category_2_interval_from and category_2_interval_to, except for category 2 edges.
- transit_fixed_max is a parameter deciding variation between maximal number of bacteria in transit nodes. If the parameter is set to 1, all the transit nodes have the same max-population. If the value is set to 0, each transit node gets its own unique max-population. The values are generated in a normal distribution.
- minimum_survival_probability and maximum_survival_probability are parameters that determine the probabilities for transit nodes to survive cleaning. The values are generated in an exponential manner.
- probability_patient_attachment decides the probability for patients being connected to the transit nodes.

sens_simulate_v03:

A program file to simulate the created networks. The program is run by the command line "python sens_simulate_v03.py test.edgelist test2.nx parameters2.csv par.var 2". The command initiates a simulation and a sensitivity analysis of the network.

Input files necessary for the command to work properly are "test.edgelist"; contains information about the edges of the network, "test2.nx"; contains information about the nodes of the network, and "par.var"; contains information about varying parameters.

- test.edgelist: The first two columns refer to specific nodes that the edges connect. For example, "0 1" means the edge connects node 0 and 1. The last column refers to the expected value for transfer across the specified edge.
- test2.nx:
    - growth: decides if there is a possibility of growth in the transit nodes or not. This value is set to 0 at this time, to prevent there being growth in all nodes of the network.
    - S: refers to the number of S bacteria that are in the nodes at starting point. The I-column is the same, only for I bacteria.
    - max: refers to the maximal number of bacteria that may exist in the nodes

    - transit: a dead parameter. It has to be the value 0 for the program to work.
    - survival_freq: the probability that the bacteria in the transit node survives upon cleaning.
    - patient_attachment: decides if the transit node has any contact with patients. = is no contact, 1 is contact. A node referring to a location where solely staff is allowed will thus be set to 0.
- parameters2: the file containing all parameters for the network. See additional material x for more detailed information about parameters. These may be customized.
- par.var: the content of this file may be customized to apply to the parameters of interest. par.var is able to manage several varying parameters simultaneously; for example, s_growthrate, s_frac and clean_interval. It may also be customized to solely manage a single varying parameter.

The number at the end of the command; 2, indicates an intention of producing a file containing all traffic over transit-edges for every iteration. A datafile of each set of variable values in par.var is generated. They will be generated as
“test.edgelist_test2.nx_parameters2.csv_1_.txt”,
“test.edgelist_test2.nx_parameters2.csv_2_.txt”, etc. Edgeflow files will be created as well, containing information about the traffic over all edges in the network:
“test.edgelist_test2.nx_parameters2.csv_1_edgeflow”,
“test.edgelist_test2.nx_parameters2.csv_2_edgeflow”, etc.

A file called “test.edgelist_test2.nx_parameters2.csv_sens.csv”, will in addition be generated by running the command line. The file contains the fraction of S and I infected patients at the time of the last iteration.

scrapeFile_v201:

By running the command “python scrapeFile_v201.p test.edgelist_test2.nx_parameters2.csv_0_.txt 1000 1” the files “plot_data.csv” and “S_I_status.csv” is generated. The second last number represents the number of transit nodes in the network, and the last number corresponds to the minimum number of bacteria to be detected in each node.

- plot_data.csv: generates a file of the result from par.var iteration x. The example above will generate a result file for the first varying parameter in the par.var-file; iteration 0.
    - bac_cont: the total number of bacteria in the network.
    - inf_frac: the total number of nodes infected with I.
    - s_frac: the total number of nodes infected with S.
    - transit_inf_frac: the total number of transit nodes infected with I.
    - transit_s_frac: the total number of transit nodes infected with S.
- S_I_status.csv: the status of S and I from par.var iteration (in this case) 0; the number of S and I in each node at each iteration.

generate_network_state:

For the purpose of generating presentable files the command "python generate_network_state.py test.edgelist_test2.nx_parameters2.csv_0_.txt test.edgelist 500 test.edgelist_test2.nx_parameters2.csv_0_edgeflow.csv" is run.

- The first input is the result file from sens_simulate_v03. For example, if the first variant of parameter value is of focus – iteration 0: "test.edgelist_test2.nx_parameters2.csv_0_.txt" is used.
- The second input is the test.edgefile.
- The third input, the number (here 500), decides the iteration of focus.
- The fourth, and last input, is the edge flow file "test.edgelist_test2.nx_parameters2.csv_0_edgeflow.csv". This file contains information about the traffic over all edges.

The command generates files that are viable for presentation in Cytoscape. The files will be generated on the basis of the files generated by the "python sens_simulate_v03.py test.edgelist test2.nx parameters2.csv par.var 2" command line. The iteration number may be adjusted.

The two files generated are "cytoscape.edge" and "node_attribute.csv". These files are Cytoscape compatible, and they may be imported to Cytoscape to get a nice and presentable visualization of the network. The network may be shown in Cytoscape by importing the cytoscape.edge-file as a network, and the node_attribute-file as a table.

- cytoscape.edge:

    - source and target: the nodes that are connected by edges
    - S: the mean number of S that has been transferred across the edge during all iterations until the iteration that is specified in the generate_network_state-command.
    - I: the mean number of I that has been transferred across the edge during all iterations until the iteration that is specified in the generate_network_state-command.
    - combined: the S and I values combined.
    - fraction_of_flow: the fraction of total traffic in the network over the edge during all iterations until the iteration that is specified in the generate_network_state-command.
- node_attribute.csv:
    - Node: the node number
    - S: the number of S bacteria in the node
    - I: the number of I bacteria in the node
    - scale: a scale parameter; (S+I)/(max_bacteria in the node), multiplied with a scaling factor.
    - color: I/(S+I); the proportion of the total number of bacteria that is I.
    - transit: if the node is a transit node or not; 1 – transit node, 0 – not transit node

make_video_png:

A program file to create a video of the network simulations. The program is run by the command line “python make_video_png.py test.edgelist_test2.nx_parameters2.csv_1_.txt test.edgelist”. This will create a video from the entire simulation. Necessary input files for this command are “test.edgelist_test2.nx_parameters2.csv_1_.txt”; created by sens_simulate_v03, and test.edgelist.

Appendix 2: Explanation of parameters

num_iter: number of iterations; how long the simulation runs for. Each iteration in this model runs for 30 minutes

initial_patients: how many non-infected patients there are at time = 0

max_patients: maximal number of patients

s_growthrate: the growth rate of S individuals

new_pat_freq: expected value of number of new patients during one time-period/iteration

exit_pat_freq: expected value of number of exiting patients during one time-period/iteration

- new_pat_freq and exit_pat_freq follows a Poisson distribution with parameter lambda (λ)

transfer_mean/transfer_std: parameters deciding how big proportion of a population that will be transferred to a node when a transfer incident occurs. The distribution is half-normal, with mean = transfer_mean and sigma = transfer_std

transfer_prob: the probability for transfer from patient to transit node, or from transit node to patient

graphic: parameter for when putting together videos of figures. It affects the size of the nodes

i_fitness_cost: how much slower I individuals grow

bacterial_load: the maximal number of bacteria that may exist in a patient

s_frac: the proportion of new patients that is already infected with S

i_frac: the proportion of new patients that is already infected with I

i_delay: decides when patients infected with I starts to enter the model. The value corresponds to hours

tresh: decides when antibiotic treatment starts. This is a number between 0 and 1. For example, if the parameter is set for 0,5, the treatment starts when a patient has 50% of bacterial_load as S + I

beta: the rate of flow from S to I when S and I come in contact with each other

beta2: the rate of flow from I to S. This transition occurs independently of contact between I and S

stoptime: a time parameter. If decides how much happens during one iteration. 0,5 means 30 minutes

numpoints: decides how detailed the ODE solver is (the number of iterations)

delA: tells how effective the antibiotic kills bacteria. This is a kinetic parameter which models the meeting between antibiotics and bacteria as in a chemical reaction; A (antibiotic) + B (bacteria → A with rate of delA

ab_supply: the supply rate of antibiotic during treatment, which means the amount of antibiotics supplied

ab_uptake: the uptake rate of antibiotics during treatment, which means how fast it is removed from the system

clean_interval: how often transit_nodes are cleaned. The value corresponds to hours. For example, if clean_interval is set to 6, cleaning happens every sixth hour

seed/np.seed: what seeds are used to generate random numbers. The value is a random chosen number

halflife: the half-life of cells in transit nodes. It is implemented to be a radioactive decay. This means that the population in the transit nodes decreases exponentially with a half-life time of halflife

decay: decides if cells in transit nodes dies or not (0 or 1)

cores: how many cores that are used for calculations