Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

Alexander Orvik

Network based pandemic modeling

Master's thesis in Artificial intelligence
Supervisor: Keith L. Downinng
June 2023

NTNU
Norwegian University of
Science and Technology

Alexander Orvik

Network based pandemic modeling

Master's thesis in Artificial intelligence
Supervisor: Keith L. Downinng
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Abstract

Computational epidemiology is a large field with many approaches and toolkits for modeling pandemics. These approaches ranges from statistical analysis and differential equations, agent-based modeling and even hybrid methods combining the aforementioned ones.

This thesis aims to expand the toolkit with a novel toolset that integrates graph theory and agent-based methods. This allows one to explore a wide range of settings and environments that changes the impact and development of pathogen activity in a given population. Depending on the environment provided, the pathogen will attempt to optimize itself based performance function weights provided as an input to the system through the use of a variation of the genetic algorithm.

The flexible nature of the system gives rise to numerous different scenarios that can be tested. The system variables range from the social structure of the population through different network topologies, simulated immune responses with static and dynamic populations, sickness isolation policy and more.

Preface

I have always enjoyed creating systems. Practical applications of knowledge. This thesis explores a process of creating a modeling system for pathogens built from the ground up.

Plans panning out is a rare occurrence, and the work laid down in this thesis does not serve as an exception to the rule. But nonetheless, now that I am here. At the end of the road. I can still be content with the final result, and the effort that went into it.

Although I have reached road's end, I never would have managed this without proper guidance, or "veivisning" if you will. I extend my thanks to Keith Downing, for supervising and guiding the direction of this thesis.

Alexander Orvik
Trondheim, June 11, 2023

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Disease has always been an undesired, but unavoidable chapter, in the history book of humankind. Be it the black plague in medieval times, the Spanish flu or most recently, Covid-19 it will always be an ever looming potential threat to the human population. Epidemiology, the study of diseases and how they interact with a population, has understandably had a surge of interest in the last couple of years.

1.1 Background and Motivation

Covid-19 made it very clear the social, economic and health impact a world-wide epidemic could have on society. It also made it very apparent how difficult it is to accurately predict the outcome of an epidemic. The non-static nature of human behavior, and unpredictable evolution of disease adds up to an extremely complex modeling task. Taking everything into account gives rise to a computational problem we do not have the resources for, meaning we often have to reduce problems into less resource intensive problems to solve.

We still have to, with a certain degree of accuracy, be able to model some of these human-disease interactions. Disease requires carriers, a medium to be spread, but we want to be able to structure the relationship between carriers in some manner. A natural way to portray social relationships are networks, or graphs as often referred to within the field of mathematics. Graphs allow us to tie together human interactions into direct social relationships. Modeling populations using graph theory gives us a strong tool to be able to trace how diseases manage to traverse from individual to individual, the drawback then being the large computational cost of managing these relationships when they grow in order of magnitude to realistic proportions.

Agent based modeling has the advantage of being able to model behavior in closed systems on a more granular level. Individual behavior and traits can drastically alter how pathogens interact with a population. If a single individual contained data such as immune system traits, social behavioral traits and a graph of their relationships, this could help to more realistically reproduce pathogen behavior. This however, similarly with social graphs, comes at the cost of computational complexity.

With today's state of computation it is not possible to try and simulate every single individual in a given population and all their traits, interactions and how they come together with the policies of the society they live in. But there should still be interesting behavior to see by implementing a choice few of these different approaches to modeling.

1.2 Goals and Research Questions

Goal *Develop a tool leveraging a variation of network structures and low-level agent-based interactions to synthesize complex epidemic events*

Epidemiology is a broad field where networks have been used as a means to an end to model social interactions. Many different efforts have been made to try and capture human social behavior using different network topologies, equation based epidemic models with difference equations or other agent based approaches.

The infection is often considered to be static where the variables changing are social policies of the network or to some extent social behavior of nodes in the system. But creating a realistic social contact network with a coexistent agent-based model with a synthetic population and a pathogen able to exhibit agent-like behavior requires some work

Research question 1 *How does an adaptable pathogen handle policies to mitigate spread and damage?*

Research question 2 *How does the network topology of a social contact network affect the impact of an agent-based pathogen?*

Research question 3 *How does population diversity affect affect evolution and impact of a pathogen*

1.3 Research Method

The methodology used is a hybrid approach reading up on literature within epidemiology, network theory, behavioral psychology, artificial life and agent-based modeling. Based on initial findings I will design a model for the different agents and their inherent mechanisms. This will then be implemented in a system where its possible to run trials where parameters can be adjusted to see how these change the outcome of the trials.

Chapter 2

Background Theory and Motivation

2.1 Epidemiology

2.1.1 The SIR model

The SIR model is considered to be the start of mathematical epidemiology. SIR is an acronym for Susceptible, Infected and Recovered, and is a differential equation model for the spread of disease first introduced by Kermack and McKendrick in their 1927 publication Kermack and McKendrick [1927]. This model is also commonly referred to as a compartmental model in the sense that any individual in this model can only exist in one of the states at any point in time. The constraints of the model is a homogeneous population where the population size remains constant. The time-span of being disease-ridden is considerably smaller than that of the human lifespan. It is also also only possible to be attacked once, where you then are moved to the Infected state and then later permanently to Recovered which either implies death or recovery with no chance of re-infection.

This model has a multitude of versions in which the mechanisms of state transitions and the states available have been modified in some manner. Most modifications introduce new states such as Exposed, implying infected but not aware (asymptomatic). Vaccinated and deceased which could often be considered a replacement for the Recovered state which often is either dead or recovered and cannot be re-infected.

2.2 Graph theory and networks

Network topologies largely depend on the context of their creation. In order to model the spread of pathogens what is interesting will differ from delivery route optimization. In the case of disease spread we want to find topologies which might apply to social relationships. Creating a specific network topology means we need to describe the process of its creation with some type of algorithm; a systematic way to create the relationships between nodes, where the underlying structure can be associated with a trait we are interested in.

2.2.1 The random network

The most naive way to generate a network is by random attachment. We start with a single node and then as long as we have not reached a desirable size for the given network we keep adding nodes. Although it is hard to make use of the random network in the context of social relationships, there are two properties of the random network which are worth taking note of. The degree of clustering in a network is a way to describe whether or not nodes have a tendency to group together. Another property is average path length. A path between two nodes is a sequence of edges that connect the two nodes. The average path length of a network is found by finding the shortest path between all unique node pairs and then taking the average value of this.
The random network is often recognized by its tendency to have a low degree of clustering, there is no inherent logic behind how nodes are connected so there is no reason for nodes to be grouped for any other reason than random chance. The average path length in a random network tends to be low, meaning that the number of "jumps" we have to make from one node to another is on average lower than most other network topologies.

2.2.2 Scale free networks

The scale free network is a topology that follows a a power-law distribution using a preferential-attachment expansion algorithm. Preferential attachment is the process of determining which node a new node will form a connection to. What are the preferences of a new node connecting to an existing network? The proposed model follows a probability based on the number of connections a node already has, where the probability of being chosen for a new connection increases with the number of existing edges.
This results in a network with a high degree of clustering since nodes will favor nodes that are already popular. This however often leads to a longer average path length because traversal from one node to another will often require to use a highly interconnected node as an intermediary step. Nodes that have a large portion of connections making up the entire network are also referred to as hubs.

2.2.3 Small world networks

Watts and Strogatz Watts and Strogatz [1998] introduce the concept of small world networks. They unified two extremes of network topologies from fully ordered to completely random, arguing that social networks exhibited traits from both ends of the spectrum. What came to be was the small world network which had high degree of clustering while also maintaining a low average path length much like a randomly generated network.
The clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. It is defined as the ratio of the number of edges that are present between the neighbors of a node to the maximum number of edges that could potentially exist between them. This measure can be used to assess the local connectivity of a node and its neighbors (the local clustering coefficient), or the overall connectivity of the entire graph (the global clustering coefficient). The concept of the clustering coefficient was introduced by Holland and Leinhard Holland and Leinhardt [1971], and was further developed by Watts and Strogatz Watts and Strogatz [1998]. Newman YANG [2013] provides a more detailed treatment of the local and global clustering coefficients and their applications.
Consider a network where nodes are spatially ordered in a ring, and each node is connected to its immediate two neighbors. This forms a lattice network with a high degree of clustering but

also a long average path length. Watts and Strogatz describes this as the polar opposite to the random network which has a low degree of clustering but low average path length.
Social relationships are often forged within a given context which gives rise to varying degrees of clustering. This is then a valuable metric when trying to generate a network that should have some resemblance to a social network.
To create a small-world network, you start with a regular lattice of nodes, where each vertex is connected to its neighboring nodes. This creates a highly clustered network with a relatively long average distance between nodes. To create the "small-world" effect, a few long-distance connections, or "shortcuts," are added between randomly selected nodes. This reduces the average distance between nodes and increases the network's overall connectivity. The result is a network that has a high degree of local clustering, like a regular lattice, but also has a few long-distance connections that create a short average distance between nodes. This process can be repeated multiple times to create networks with different levels of clustering and average distances.

2.2.4 The strength of weak ties

Granovetter [1973] argues that weak ties (or casual acquaintances) play a crucial role in the formation and maintenance of social networks. According to Granovetter, strong ties (such as close friendships and family relationships) are important for providing emotional support and a sense of belonging, but weak ties are equally important for providing access to new information and opportunities.
Granovetter suggests that because weak ties are less emotionally intense, they are more likely to connect individuals who are from different social groups or networks. This allows for the flow of information and resources between different groups, which can be beneficial for all involved. For example, a job opening at a company might be advertised through a weak tie (such as a friend of a friend), allowing an individual who might not have otherwise heard about the opportunity to apply and potentially get the job.
He also talks about the concept of *triadic closure* which is the tendency for people to become connected through intermediaries. It's based on the idea that people are more likely to form connections with others who are already connected to their existing friends and acquaintances. This can create a situation where three people who may not have otherwise been connected end up becoming friends through the connections of others.
Overall, Granovetter's ideas highlight the importance of both strong and weak ties in the construction of social networks, and suggest that a balance of both is necessary for a healthy and functional network.

2.2.5 The ego network

Freeman [1982] presents the concept of the ego network through centered networks. Zhou argues that our social relationships are organized in a hierarchial structureZhou et al. [2005] taking inspiration from the work of Dunbar and Dunbar's number. Dunbar's number is a concept that is closely related to the social brain hypothesis [Citation needed]. It is a suggested cognitive limit to the number of people with whom one can maintain stable social relationships. This number is thought to be around 150, and it is based on the idea that the human brain has a limited capacity for social connections. The social brain hypothesis suggests that the size and complexity of the human brain evolved in part to support the cognitive demands of living in a complex social environment, and Dunbar's number provides a way to measure the extent to which this is true.

2.3 Genetic algorithm

The genetic algorithm is a search heuristic that draws inspiration from the evolutionary process in nature. First introduced by Holland Holland [1975], it takes the concept of survival of the fittest by randomly generating a pool of possible solutions. This pool, being the first generation is scored based on some fitness function which determines how good of a solution it is relative to the others. Subsequently, the best solutions (the 'fittest') are selected to create offspring for the next generation. This is typically achieved through the processes of crossover and mutation. Crossover combines parts of two parent solutions to create new offspring, while mutation introduces random changes. This evolution of the population ideally leads to the emergence of a solution that is superior to the initial ones. Over successive generations, the algorithm navigates the search space, progressively refining the population towards optimal or near-optimal solutions, depending on the complexity of the problem and the fitness landscape. It's this capability that has made genetic algorithms a popular choice for tackling a wide range of optimization and search problems in various fields.

Chapter 3

Related work

There is a lot of interesting work done in the world of epidemiology. This project touches a lot of different subject areas which means that related work is spread across many interdisciplinary fields. There is network theory and creating realistic social network topologies based on research on human capacity for maintaining social relationships. Given its realistic nature those networks can be used as a point of entry for contact based spread in an epidemic. How networks have been applied in epidemic models vary greatly, but what they have in common is that they attempt to encapsulate how disease manages to spread from person to person.

3.1 Social networks

3.1.1 A model to represent human social relationships

Building on the concepts of the ego network Conti, Passarella and Pezzoni Marco Conti [2012] describe an algorithm for creating an ego network consisting of locally connected ego networks by creating a union of their single-ego model and a multi-ego model.
The single-ego model is from the perspective of a single individual, an ego. It then proceeds to expand the network of this ego on the given premise that it has a finite number of relationships it can establish. An important distinction is made when establishing relationships in an ego network, and that is the type of relationship it is. We divide them into three different layers, which are called "support clique", "sympathy group" and "active network". Their respective averages sizes are 4.6, 14.3 and 132.5. The ego network is expanded using density functions until the target size/budget for a given layer is exceeded and the algorithm then moves on to an outer layer.
The single-ego model does not provide a lot of utility when looking at a complete social graph of a population. This motivates the design of the multi-ego model. The main difference between the single- and multi-ego model is that the multi-ego model takes into consideration every single ego in the network. When adding a link or another ego it takes into consideration both ends of the connection. Every single ego in an ego-network is a part of another ego-network. When a single ego's social budget has been expended we can traverse the network and continue expanding those networks recursively.
The expansion of these networks makes use of Granovetter's Granovetter [1973] *triadic closure* and *weak social ties*. *Triadic closure* can be used to expand the innermost social layer "support clique" whereas weaker ties of acquaintances to create bridges/expand an egos "active network" layer.

3.1.2 Triadic closure drive scaling laws

Klimek and Thurner Klimek and Thurner [2013] consider triadic closure to be a driving force behind formation of networks used for communication, trading and friendships. They use a model to show patterns in a non-growing network that explain the scaling laws related to scaling of the probability of a node acquiring a new link, the scaling of the probability of finding a node with degree k and the scaling of the clustering coefficient as a function of node degree.
An interesting issue they bring up is that the the concept of preferential attachment which is used in expansions of scale-free networks is that it requires global knowledge of the network which is not realistic in ordinary social networks. Instead they use it as extensions of local network formation rules. An integral part of this is triadic closure since the chance of closing a triad is higher than forming a new connection between two arbitrary nodes.
The model begins with a static network of N nodes, each linked to a random node. It iteratively picks a random node i, adds connections based on its current number of links and a probability r, and occasionally removes a random node while introducing a new one. The new node connects to m randomly selected nodes, thus continually altering the network's structure.

Initialize a static network of N nodes with each node having one link to a random node.
while True **do**
  Pick a random node i
  if node i has less than two links **then**
    Connect it to another random node j
    Go to Step 7
  else
    Pick one of the node's neighbors at random
  end if
  Given probability r
  if r is True **then**
    Create a link between j and another random neighbor of i
  else
    With probability $1 - r$, pick its new link to be to some random node from the entire network
  end if
  With a probability p, remove a random node from the network and sever all its links
  Introduce a new node to the network connecting to m randomly chosen nodes, where m is a chosen parameter
end while

They state that their findings show strong evidence that triadic closure plays a significant role in the formation of social networks. It appears that three different scaling laws are influenced by the properties of triadic closure. The model however shows some weakness because triadic closure is not well suited for networks describing hostile sentiment. The enemy of my enemy is in most cases not my friend.

3.2 Modern epidemic models

3.2.1 Epidemic spread in scale free networks

Scale free networks have been a way to model our highly interconnected computers on the world wide web, where most communication travel through highly connected hubsPastor-Satorras

and Vespignani [2001]. Making use of a SIS model where instead of moving from an infected to a recovered state each node is simply vulnerable to being infected once again. Each node is a computer connected to the network, and if infected transmits the virus to its connected counterparts at a given rate. A comparison can be drawn to what has been coined superspreaders in epidemics; individuals with a very large social circle and a high rate of contact with unique individuals.
If a very interconnected computer in a network contracts a virus and spreads it to all other connected computers implies a large security risk if such a hub is compromised. Interesting things to consider here are targeted policies for very exposed computers that may pose a high risk and also at what critical threshold the virus is self-reliant in the sense that it spreads at a higher rate than the system cleanses itself. This way the virus may remain in the system indefinitely in a type of equilibrium.

3.2.2 Epidemic spread in trajectory networks

Pechlivanoglou Pechlivanoglou et al. [2022] took an approach where they made use of GPS-coordinate data and trajectory networks to model the spread of Covid-19. The goal was to be able to capture individual variability based on their location data. A trajectory network is a term of the aggregated structure of temporal networks, which again is an extension of a proximity network. A proximity network is a structure in which the connection of nodes is inferred by their proximity. The issue with the proximity network is their static nature, introducing discrete time steps and movement patterns of individuals the resulting graph type is the temporal (or time-varying) network. The collection of temporal networks for a given time span $t \in [0, T]$ is the trajectory network.

3.2.3 Model infection with random graphs

Croccolo, Fabrizio and Eduardo introduce a network based epidemic model making use of percolation theory Croccolo and Roman [2020]. They constructed a two-dimensional square lattice network, essentially a two-dimensional grid where each square is a node and is connected to each neighboring square, nearest neighbor (NN). Each node has a state belonging to any compartment in a SIRD model, where D is Dormant and is a group of people isolated from the population with no way to interact with the population.
Their model sets four rules for the spread "*(1) The virus can cross a bond from an infected site to a susceptible one; (2) no virus transmission occurs otherwise; (3) infected sites heal after* $_H$ *days, becoming recovered sites, so that they can neither infect others nor become infected again (immune sites); and (4) dormant sites do not participate in the spreading process*". Links are also added between nodes (sites) dynamically to represent nodes moving around, which creates an effect similar to ordinary social movement in a population. They found that the ability of the infection to cross gaps between clusters of people, or lack thereof, simulates the effects of lockdowns on disease transmission.

3.2.4 Covid-19 models using networks of social interactions

Herrmann and Schwartz Herrmann and Schwartz [2020] criticize the state of current (2020) models for Covid-19 due to their tendency to assume random diffusion without taking an underlying social network into consideration when modeling the spread. They propose using a scale free network stating that human social networks are known to be scale free.
They constructed a scale-free network of 10 000 nodes based on BarabÃ¡si and Albert's algorithm. Using a cutoff of $n = 8$ for node connections with the exception of hubs which had connections

far greater than n. Nodes were initialized as a population of susceptible nodes according to a SIR model. Patient zero was then a random node in the network. An infected node was sick for two weeks with an infection rate depending on which day in the infection cycle the node was in. The highest infection rate being $4 \leq d \geq 6$. They use a classic SIR model in which the recovered state is irreversible and no longer interacts with the infected or susceptible part of the population.
As a type of containment policy they varied node to node interactions. First model being an unrestrained scale-free model, the other completely containing all hubs and the last one limiting hubs to only 8 of their existing links chosen at random, detaching all the others. They were called scale-free, mitigation hub and mitigation random respectively.
Their simulations showed that there could be some value in mitigating strategies, where node-to-node interactions are limited in order to "flatten the curve". When hubs remain uninfected the simulations showed that the infection died very early, or the peak of the infection was noticeably delayed until a hub was infected.

3.2.5 Hybrid epidemic models

Bobashev et al. Bobashev et al. [2007] made use of a hybrid model in order to balance the pros and cons of agent-based and equation-based modeling. The main benefit of the agent-based model is its ability to capture high quality details if designed to. Individual behavioral patterns of interaction based on scheduled events in local communities and personal preferences which might more accurately reflect how we behave in a community are possible to implement. The main problem is the raw computational demand of such models, requiring powerful machines to run. On the other hand equation-models making use of difference equations to model infections in a community are much less resource intensive and can be scaled up to much larger populations, then again at the cost of losing high quality data that agent-based models offer.
Their model introduced a hybrid based mechanism that stems from the premise that when the number of active agents reaches a certain threshold the laws of large numbers apply meaning we can aggregate the behavior of the agents in the system using mean-field approximations. They use a SEIR model in addition to a T denoting the total population. Only the interactions of exposed and infected individuals are of consequence. This allows us to only consider these agents at each individual time-step and keeping other agents inactive until they have made contact with an infected agent. This reduces the number of agents we need to consider at the beginning of a simulation.
The equation-based model consisted of a coupled system of 155 difference equations. They take into consideration problems such as chance of spread on contact, chance of travel between cities and the chance of starting an epidemic in a city when traveling.
The agent-based model uses five homogeneously mixed groups from the model used by Longini et al. Longini et al. [2004]. Households, playgroups, school, work and social groups. They placed age-based restrictions such that if an agent is too old they don't go to school anymore but work and vice versa.
At any point in time a certain part of the population of a given city has a chance to travel to another city. An uninfected city only has inactive agents, but could then be visited by an infected agent where there is a chance of contact and then contracting the pathogen. If so infected agents in the city are activated and depending on the spread it may die out or induce a city-wide epidemic.
Whether or not a city is using agent-based or equation-based modeling depends on whether or not they have reached the threshold switch. Different cities can find themselves in different states depending on the virus activity in the given city.

They used this model with different switch thresholds to see how the hybrid model changes depending on how it models its epidemic. The switch threshold would vary depending on the size of the city, requiring a dynamic value for this threshold. This is set when a simulation-based estimator of the transmission rate of the disease starts to stabilize.

The validity of the model is stated to require to be confirmed in further work.

3.2.6 SIR based epidemic model for Covid-19 in communities

Cooper et al. Cooper et al. [2020] discuss a traditional SIR model with the twist that the population is not defined or kept constant and that susceptible individuals do not decline at a monotonic rate. It can instead be increased in surges. Unlike traditional models that keep the total population constant, their model allows the number of people susceptible to the disease to change, better reflecting possible surges in susceptible individuals.

Their model was able to fit into published data of Covid-19 infection spread in different communities "reasonably well", with the exception of China where the infection was contained more than predicted but was thought to be due to successful implementation of policies by the Chinese government.

3.2.7 Covasim

Kerr et al. Kerr et al. [2021] designed a Covid-19 agent-based simulation (Covasim) tool in order to aid in exploration of intervention scenarios and estimation of resource needs.

Covasim incorporates country-specific demographic data, such as age structures and population sizes, to model realistic transmission networks across various social groups and communities. The system further integrates disease risk factors, dependent on the infected population's attributes, such as age-specific outcomes. Additionally, it provides an array of intervention measures to explore their effects on different outcomes

The model is an agent-based system with discrete time steps where state changes of any given agent are calculated through probabilistic means. It uses a modified SEIRD model where exposed (E) has a nested structure where there can be different individual states. An exposed agent can either be presymptomatic or asymptomatic. A presymptomatic agent then progresses into a state of either mild, severe or critical. All three sub-states can transition into a deceased state where only critical illness can possibly be fatal.

Their contact network design falls into three different types, a random network, SynthPop or hybrid. The random network is generated with the premise that any individual can come in contact with any person in the population. Each individual is designated daily contacts through a Poisson distribution and this number can also be sampled randomly for each day for an individual.

SynthPop is an open source tooling for generating synthetic population networks which create a high-resolution network that is able to capture age-specific distributions and then give individuals in the population context based social network layers for work, school, household and communities.

The hybrids approach is a middle ground of the two where one does not have to comply with the large input-data requirement of a pure SynthPop approach

The model also comes with intervention methods such as physical distancing, masks and hygiene, testing and diagnosis, contact tracing, isolation of positives and contact quarantine or vaccines and treatment.

3.2.8 Interaction based based approach to computational epidemiology

Barret et al. Barrett et al. [2008] talks about the general concept of interaction-based multi-agent approach towards epidemic modeling called *Simdemics*. The intent is that it should provide a detailed model that should aid decision makers in evaluating available information and potential outcomes of enforcing policies.
Simdemics borrows concepts such as multi-agent systems, social network analysis and Markov decision processes. It also further makes use of dynamic social networks and disease diffusion on such networks. The model is split into four distinct parts, first one being a way to create a synthetic population of individuals. The nest step is to form a type of interaction mechanism through a network in which this population may be in contact with each other. The third step is to have a system for simulating an epidemic process given step one and two. Lastly a system for applying and collecting data on intervention policies for a given simulation.

Chapter 4

Model architectures

The simulation architecture is built on leveraging a network graph to run experiments where a pathogen attempts to infect the entities in the network structure. During the life cycle of a pathogen during a simulation there are a number of different interactions which can either be enabled, disabled or changed as will be outlined in this chapter. The main models of the simulation engine are the *individual* and *pathogen* models and how they interact with each other. Individuals are interconnected using weighted relationships. We connect together individuals with specific network topologies where each individual is a node in the graph. A specific network topology follows a specific algorithm. A graph of individuals is then a *population*. Individual and node are used interchangeably in this thesis. In a single experiment we perform a search using a genetic algorithm to explore how a pathogen attempts to adapt to changes in its environment.

4.1 Pathogen

The attributes of the pathogen and its mechanics are the main drivers during a simulation run of a population. The pathogen model has five attributes, namely *infection rate*, *death rate*, *incubation time*, *infection time* and *attack vector*. Infection and death rate are given as a value $0 \leq r \leq 1$, whereas infection, and incubation time are given as integers $0 \leq d$. The attack vector is a binary string used to emulate an immune-system response from its infected host. This interaction is explained in more detail in section 4.4.
A pathogen is given a finite amount of points to spend on its own attributes, excluding the attack vector. The number of points spent is the normalized with respect to a minimum and maximum value for each field. Spending all points on a single attribute would result in that attribute being maximized and all others receiving their minimum value. As an example, we have a pathogen with minimum rates of 0.0 and maximum of 1.0. Minimum incubation and infection time is 0 and maximum 15. With a 100 points to spend an example of a skill-point distribution would be as shown in table 4.1. A thing to note here is the 5 points spent into incubation time resulting in 0. This by design so the pathogen cannot overspend its skill point allowance.

4.2 Individual

The individual acts as an entry point for the pathogen. An individual is connected to other individuals through *connections*, representing its social contact area. All populations are undirected weighted graphs. The weight of a connection is a value $0 < w \leq 1$ stating the probability

Attribute	Points	Normalized value
Infection rate	31	0.31
Death rate	16	0.16
Incubation time	5	0
Infection time	48	4

Table 4.1: Pathogen skill point distribution example

of two connected individuals making contact when cycling through active pathogens during a simulation loop. This is the *encounter rate* of the connection. The *immune system* attribute is a 10-length bit-string. The *immune system* is used to diminish the effect a pathogen has on a certain individual. This includes reducing its death and infection rate, depending on the overlap it has with the *attack vector* the infecting pathogen. At any given moment an individual exists in a state of either *susceptible*, *infected*, *recovered* or *deceased*. While infected there are two sub-states where the pathogen is incubating or capable of infecting others. An individual is susceptible if there is no active pathogen infecting it, and if previously infected the immunity duration has expired. While infected an individual can infect others, but this state has two sub-states which is the incubation period and infection period. These sub-states differ by the fact that during the infection period the host has chance of dying.

4.3 Network topology strategies

The network is the attack surface of the pathogen. We use specific strategies to synthesize desired network topologies.

Scale free network

The procedure begins by initializing the network by creating a connection between the first two nodes in the collection it has accepted as input. To create the remaining connections, the procedure employs a lottery-based approach. For each node in the collection, starting from the third one, the procedure selects a random node from the existing network to establish a new connection with. The selection is biased such that nodes with more existing connections have a higher probability of being chosen. This reinforces the scale-free property of the network by simulating a "rich-get-richer" effect. Each time a connection is formed, the procedure updates the total count of connections and the individual connection counts of the involved nodes. It maintains a record of the number of connections each node has in a data structure akin to a dictionary, using the node identifiers as keys.

Small world network

The procedure begins by constructing a ring lattice topology among the nodes. This is a cyclically connected network in which each node is connected to its k nearest neighbors, where k is a user-defined parameter. After forming the initial ring lattice, the procedure re-configures the network to introduce the small-world property. It does this by traversing each connection and, with a probability p, rewiring the connection to a randomly chosen node. p is another user-defined parameter which controls the randomness in the network. During this rewiring process, the procedure ensures that multiple connections between the same pair of nodes are avoided, and

that no node connects to itself. This is done by repeatedly selecting a new random node until a suitable one is found whenever the initially selected node is either the same as the source node or already connected to it.

Clustered small world networks

As an extension of the small world network we have the *clustered small world network* meaning that instead of having one large rewired ring lattice, we split them into smaller clusters of small world networks representing smaller social communities where there are some nodes acting as links between the communities.

Weighted networks

Weighted network versions follow the algorithms as described, but with a randomized *encounter rate* used as the connection weight. The weight is set to 1.0 for all connections in the network if it is not a weighted version.

4.4 Model interactions

Pathogen life cycle

The pathogen life cycle first determines whether or not it exists in a state where it can do anything actionable. There are three different states to be aware of. The incubation period, where the host can infect other individuals, but does not suffer from any symptoms. In this period the host cannot die. When the incubation period is done, the pathogen transitions into the infection period. While infectious the host can still transmit the pathogen, but at any given time-step the host can also die depending on the death rate of the active pathogen. If the pathogen specifies a number of immunity days, it delays self-termination until the total days of infection surpass the combined sum of the incubation, infection, and immunity periods. Consequently, this leads to a temporary period of immunity in the host. An undefined immunity period implies immunity throughout the duration of the simulation.

Immune response

The immune response is the interaction between a pathogen and its host. During an iteration of a simulation loop the pathogen will increment its state, and depending its internal state might attempt to infect another individual or kill its host. The simulation will calculate a mitigation factor $0 \leq m \leq 1$ that is multiplied with the relevant rate. The mitigation factor is calculated using a function based on overlapping binary values. The degree of overlap between the two results in reducing the death and infection rate for this given pathogen-individual interaction, by a factor as determined by the mitigation function. The mitigation function can either be a step-wise function is to have clearer levels of immunity instead of a linear relationship between the overlap and the mitigation factor or directly return the fraction of overlap as the factor.

4.5 Experiment design

The experiment requires bootstrapping the network topology and the first generation of pathogens. The network structure is kept constant for all generations as to give the pathogen the same conditions in all simulation cases. This means replicating the whole population, including connections

between individuals and their respective weights. The topology is generated according to a specific network strategy, this includes some degree of randomness. Specific network topologies follow their own strategy for generating the relationship between nodes as described in 4.3

4.5.1 Patient zero

Patient zero is the first infected node in a simulation. There are three options to chose, worst-case, best-case and random. The strategies are from the perspective of the pathogen, meaning that in the best-case patient zero strategy we retrieve the node with the highest number of connections in the network. In the worst-case scenario we do the opposite, we retrieve random node selected among those with the smallest number of connections. A limitation in the implementation is that we do not consider the chosen node's distance from other highly-connected nodes. The less available nodes with a high social contact area are, the harder it is for the pathogen to spread.

4.5.2 Main infection loop

The main infection loop it the simulation logic that happens every time-step in a simulation. For any given time-step, we retrieve all pathogens in the population and update their local state. Updating the local state means we increase the counter for how many days it has been infecting the host. After this we check if it capable of infecting others and if it can kill the host. If applicable we retrieve the host's connections and then if an encounter is triggered attempt to infect the connection based on the infection rate and the mitigation factor as determined by the immune response, if enabled. The main algorithm is outlined in the appendix **??**

4.5.3 Experiment-level variables

An experiment-level variable is a variable/setting that is defined for all simulations run in that specific experiment. These vary in type between being able to be turned on or off, or changing its value.

Infection isolation

Infection isolation determines if an infected node self-isolates when experiencing symptoms. The state of experiencing symptoms is defined as its infection phase, in which the host is susceptible to death. If enabled the infection-period is less valuable to a pathogen, since it will stop it from further spread.

Immune system response

Immune system response determines if the mitigation factor of a pathogen-individual interaction should be taken into account. If enabled all attempts to either infect or terminate an individual will include another calculation which will reduce the impact of the pathogen based on attack vector and immune system string overlap

Mitigation factor function

If immune system response is enable we can also chose between two methods of calculating the mitigation factor, one being a linear relationship between overlap of the two strings, and step-wise requiring a minimum of overlap at discrete steps.

Population heterogeneity

The immune system string can either be set to be the same for the whole population in an experiment, giving a homogeneous population. The other option is randomly generating this string for all individuals in the population, giving a heterogeneous population.

Immunity duration

The state of immunity means a host cannot be infected, this happens after both incubation and infection periods have respectively passed. Immunity can either exist as an integer $I \geq 0$, or be undefined which implies infinite immunity. A host can never be re-infected. If defined, and the number of days spent in immunity surpasses the immunity duration the population will count it as a recovery event and the pathogen will self-terminate in the host. A recovery event is not triggered if the host is given infinite immunity.

4.5.4 Genetic algorithm

We employ a variation of the genetic algorithm where each generation ranked by a specific performance function and then either transferred without alterations (selection with elitism), transferred with minor re-allocation of attributes or complete replacement.

Performance

Performance metrics are scored as a linear combination of deaths d, infections i and recoveries r. Their weights are defined respectively as the variables w_d, w_i, w_r. The value of each weight is determined as an experiment-level variable.

$$f(d, i, r) = w_d d + w_i i + w_r r$$

Selection

There are two selection pools to be aware of, the elite sample which are the pathogens that are unaltered and the breeding pool. If we have a pathogen pool P, elite sample E, breeding pool B and remaining pool R we split them accordingly

$$P - E = B + R$$

The size of the breeding and remaining pool are equal $B = R$ so we subtract the total pool with the number of elite samples and divide this by two. All these splits are done on an ordered list of pathogens ordered by the performance function as described in 4.5.4. This way we keep current optimal solutions while exploring the solution space by altering existing solutions and generating new ones.

Offspring and skill re-allocation

The process of skill re-allocation is performed on the pathogens in the breeding pool B. It involves randomly choosing two traits or attributes of a pathogen. The process also involves picking a random value from a predefined list of trait values (5, 10). If the first randomly chosen trait has a value greater than this random value, the value is subtracted from the first trait and added to the second trait. However, if the first trait's value is less than the random value, the process is reversed: the random value is added to the first trait and subtracted from the second trait. The skill re-allocation process ensures that the total number of trait points remains constant.

4.6 Summary

The simulation model is highly dynamic, replete with numerous adjustable features. Our unique system allows for modifications spanning from the fundamental network structure to detailed interaction parameters, including the ability to define encounter rates with weighted networks, and even simulate individual immune responses to pathogen interaction. This system's flexibility provides a comprehensive tool-set that affords complete control over epidemic modeling conditions. Furthermore, it allows the users to set goals for the pathogen, which are evaluated based on custom-weighted scoring criteria.

Chapter 5

Experiments and Results

The experimental model is flexible and can be finely tuned to suit a variety of scenarios. Several variables can be adjusted or toggled on and off, providing substantial control over the simulation conditions. In a single experiment, we define a range of variables. These include the population size, which determines the number of nodes in the system, and the network topology, which defines how these nodes are interconnected. The duration of the simulation is determined by the number of time steps, while the size of each pathogen population and the number of these populations - or generations - are also variables within our control. We can even adjust whether or not to enable an immune response, which, if enabled, introduces a mitigation factor for all interactions between pathogens and individuals. The overlap thresholds used when calculating this mitigation factor, the duration of post-recovery immunity, and the minimum and maximum thresholds for translating pathogen skill points, can all be tailored to the needs of the experiment. Lastly, the model can also account for different scenarios for patient zero, ranging from best-case, worst-case situations and random selection

5.1 Experimental Plan

The measurable component of an experiment is the pathogen performance score. The performance score is determined by its performance function as explained in section 4.5.4. A key point of interest is the impact of pathogen conditions on its performance. The weights within the performance function identify the optimal attributes for maximizing its score, yet determining the maximum attainable performance in each scenario is another compelling facet to examine. This sequence of experiments will explore the effects of modifying these conditions and compare the results across different scenarios. We will be looking at isolated settings, meaning that most of the different mechanisms will be disabled in order to reduce the amount of noise in the resulting data set.

5.2 Setup

All the experiments share a common set of base settings. Each experiment spans 50 time steps, and each pathogen has 100 skill points available for allocation among four attributes: infection rate, death rate, incubation time, and infection time. The infection and death rates are fractional values ranging from a minimum of 0.0 to a maximum of 1.0, which translate to a probability range of 0-100%. The time variables, incubation time and infection time, have a minimum value

of 0 and a maximum of 15. Any additional variables that differ across experiments are not part of these base settings. Such variables are individually detailed within the specific sections dedicated to their respective experiments.

5.2.1 Pathogen performance in varying topologies

A vital component for epidemics is to have a medium to infect through. A fully connected network is more convenient than a ring lattice where each node is connected to only two other nodes. A scale-free network has more high-value targets with highly connected central nodes working that can heavily impact the spread of a pathogen. In small world and clustered small world networks one can also find nodes that are more connected than the average node but this is not to the extent as with scale free networks.

Experiment-setting	**Value**
Population size	2000
Number of generations	20
Generation size	60
Patient zero	Best-case
Weighted network	False
Immunity duration	Infinite
Isolate if sick	False
Immune response	False
Performance function death weight	4.0
Performance function infection weight	1.0
Performance function recovery weight	-1.0

Table 5.1: Varying topologies settings

The variable we will be changing there is the underlying network structure. We will run three simulations covering the three network structures covered in 4.3, namely scale free, small world and clustered small world networks.

5.2.2 Population diversity

When enabling the immune response we add another interaction layer in the simulation. The immune, from the perspective of the population acts as a line of defense that mitigates the impact of a pathogen spreading through the population. An interesting scenario to then examine is population diversity, meaning how a homogeneous population handles an epidemic in comparison with a heterogeneous one.

A homogeneous population will mean that we will not dynamically create the immune system string for each individual, but instead this will be a shared string. The heterogeneous population will instead have a randomly generated string for each node. We chose the scale free network topology using a best-case scenario. By initiating from the most connected node in a scale-free network, we can more readily assess whether a diverse population exhibits greater resilience by handling these starting conditions more effectively, especially considering these conditions are heavily tilted in the pathogen's favor.

Experiment-setting	Value
Network topology	Scale free
Population size	1000
Number of generations	20
Generation size	30
Patient zero	Best-case
Weighted network	False
Immunity duration	5 days
Isolate if sick	False
Immune response	True
Mitigation factor function	Step-wise
Performance function death weight	4.0
Performance function infection weight	1.0
Performance function recovery weight	-1.0

Table 5.2: Population diversity settings

5.2.3 Weighted vs. unweighted network

Network weights are interpreted as *chance of encounter* or *encounter rate*. Imagine three nodes, A, B and C. Connection AB has weight 0.9 and BC 0.2. If B is infected we will traverse all its connections, in this case AB and BC. An encounter between A and B will then have a 90% chance of being triggered. If triggered an infection event will start where transmission depends on the infection rate of the pathogen.

Experiment-setting	Value
Population size	1000
Generation size	40
Number of generations	15
Patient zero	Best-case
Weighted network	True
Immunity duration	5 days
Isolate if sick	False
Immune response	False
Performance function death weight	4.0
Performance function infection weight	1.0
Performance function recovery weight	-1.0

Table 5.3: Weighted vs. unweighted networks settings

5.2.4 Isolate if sick

Sickness isolation puts more weight on the pathogen investing in its incubation period. The difficult trade-off to be made here is when pathogen deadliness is the desirable metric. A pathogen can only invoke its death-event in its infection period, but investing in this while maintaining a

high death rate gives a poor prospect for doing well since it might easily kill off its host before managing to spread to a significant number of other individuals.

Experiment-setting	**Value**
Network topology	Scale free
Population size	1000
Generation size	40
Number of generations	20
Patient zero	Best-case
Immunity duration	5 days
Immune response	False
Performance function death weight	4.0
Performance function infection weight	1.0
Performance function recovery weight	-1.0

Table 5.4: Isolate if sick settings

Lethality vs. survival

A lethal pathogen does not really make sense from a evolutionary point of view. The performance weights dictate the metrics an experiment looks at when evaluating what pathogen does well and not. An interesting aspect to look at is how we can influence the skill distribution of a pathogen based on adjusting its performance function weights.

Experiment-setting	**Value**
Population size	1000
Network topology	Clustered small world
Number of generations	15
Generation size	30
Patient zero	Best case
Immunity duration	5

Table 5.5: Lethality vs. survival settings

In the experiments so far we encourage deadliness by scoring deaths with a weight of 4, infections 1 and recoveries 1. By adjusting these weights to -2, 2 and -1 respectively we punish deaths since it reduces its ability to spread. A dead host has no value if the objective is to multiply.

5.3 Experimental Results

5.3.1 Network topologies

An analogy to network topology is a variation of social structures. Scale free structures are more common when talking about client-server structures in computer networking, but can also be used when considering super-spreaders. There is a clear performance difference between the three different network topologies, as illustrated in figure 5.1. The more notable difference is

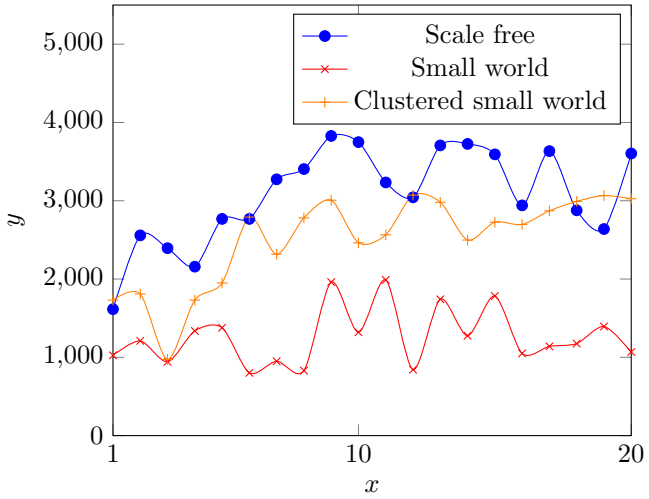


Figure 5.1: Network topology plot

the poor pathogen performance in the small-world network, whereas the other two plots have a larger degree of overlap.

The normalized skill point distribution of the pathogens provides some insight into the performance. Both small world and clustered-small world have a near identical skill distribution, with the small world network having a much higher infection rate. The performance function favors lethality to infection by a ratio of 1:4, which could explain its poor performance.

Attribute	**Scale free**	**Clustered**	**Small world**
Infection rate	0.32	0.32	0.56
Death rate	0.11	0.12	0.04
Incubation time	1	1	0
Infection time	4	3	4

Table 5.6: Network topology experiment skill distribution

5.3.2 Population diversity

Population diversity appears to have some impact on pathogen performance, but the difference is not as large as could be expected. In a homogeneous population the pathogen can find a combination of skill points and an attack vector that gives it a clear advantage since it will experience no mitigation. This would then be applicable to every interaction in the whole population. In comparison in a heterogeneous population this interaction will always differ from one individual to another.
Despite their similar performance plots the skill point distribution is not the same for their respective best performing pathogens. There is however a notable spike where the homogeneous population reaches a global maximum. The plot is also more volatile than the heterogeneous one. This helps substantiate the assumption that the heterogeneous population is more resilient.

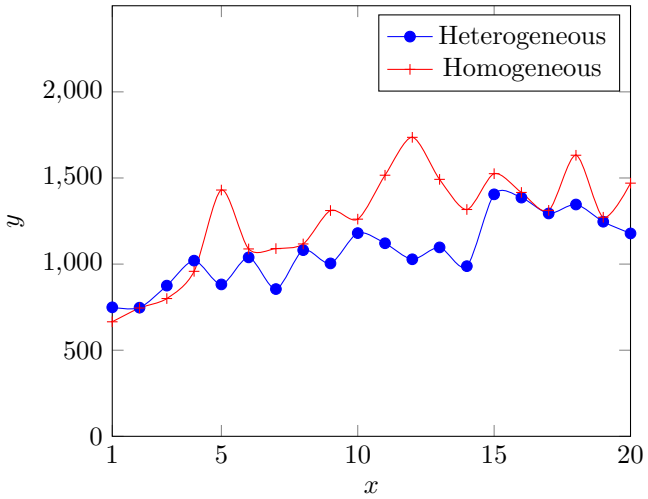


Figure 5.2: Population diversity plot

Attribute	**Heterogeneous**	**Homogeneous**
Infection rate	0.286	0.23
Death rate	0.114	0.18
Incubation time	2	0
Infection time	3	5

Table 5.7: Network topology experiment skill distribution

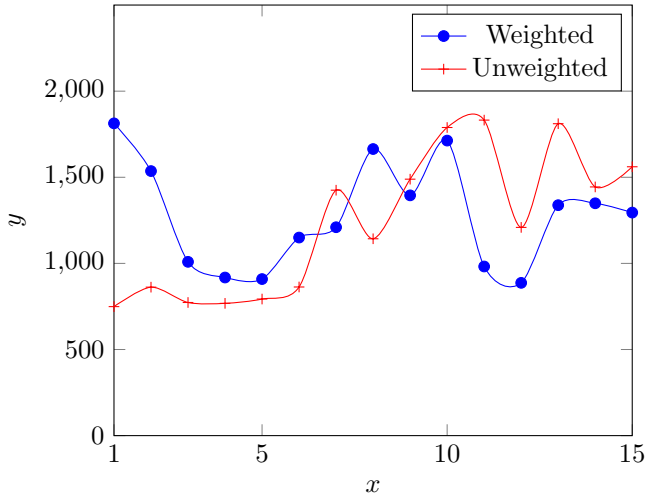


Figure 5.3: Network weight plots

5.3.3 Encounter rate

Initially the performance is completely inverted from what one would expect. Unweighted network is essentially a weighted network where the encounter rate is set to 100%. This could be attributed to the initial generation being lucky with its initial set of parameters. This however seems to switch as towards its final generation the pathogen performs the best in the unweighted network.

5.3.4 Sickness isolation

Sickness isolation has a clear impact on pathogen performance. The isolation plot also appears to exhibit a similar resilience as the heterogeneous population shown in figure 5.2. The difficulty for the pathogen comes from the fact that when isolating in order to achieve any significant level of lethality it has to be able to spread enough in its incubation phase, and then either have a very high death rate or enough infection days that compensates for a lower death rate. All this while having a satisfactory infection rate. Since this used a scale free network topology with the best-case patient zero strategy it could probably compensate for a lower infection rate, but in the isolation case it did not seem like this managed to change anything.

5.3.5 Performance score

The performance function is what dictates how the pathogen attempts to optimize itself. By changing this to heavily favor reproduction over lethality it appeared to affect the skill point distribution of the top-performing pathogens, as seen in table 5.8

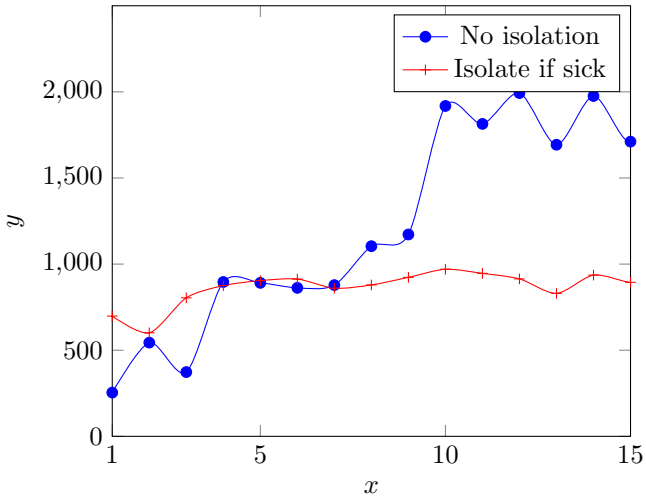


Figure 5.4: Sick isolation plot

Attribute	**Lethal**	**Infectious**
Infection rate points	32	58
Death rate points	20	12
Incubation time points	10	20
Infection time points	37	10

Table 5.8: Lethality vs. infectious skill point distribution

Chapter 6

Discussion and Final Remarks

The experimental design is intended to showcase the flexibility and range of the simulation toolkit. We cover a sequence of different variables to test and explore the capabilities of the system and in this chapter we discuss the findings in more detail, as well as limitations to the experimental approach and improvements we could make to the system in future work.

6.1 Discussion

As one would expect, how readily available susceptible hosts are is an important part of a pathogens ability to spread itself in a given population. A more surprising aspect of the results is the fact that the clustered small world network is closer to the scale free model than the small world network in terms of performance. The resemblance between the network structures is greater when comparing the small world network and its clustered variant. One could argue that the clustered variant is closer to having a super-spreader as the highly connected nodes in the scale free model has, whereas in the small world network its most notable attribute is that the rewiring process reduces the average path length of the graph. In the clustered small world we have a series of small world networks that are interconnected, meaning that when infiltrating these smaller communities available susceptible hosts notably increase. Giving a similar super-spreader effect as with the scale free network.

Population diversity allows us to observe emergent behavior of a low-level interaction between the pathogen and a host agent. Although the difference in the score itself was less than expected, the difference in volatility could imply some effect on the resiliency of a population. A problem with the design of this mechanism is that it introduces noise for the genetic algorithm that is difficult to take into account. When this feature is enabled the system now has to optimize the attack vector, as well as exploring the solution space space with its skill point distribution. There also is no mechanism for it to mutate this value either.

The encounter rate is a very under-utilized feature, mostly due to its binary implementation. The data from its experiment is not very conclusive either. The most probable reason for this is the fact that the weighted implementation of a network does not really have a logical structure to it. In these trials we simply have two scale free networks, where one of them for every edge has a random value between 0 and 1. The network structure in the unweighted graph is also not re-used for the weighted network. They only use the same algorithm to synthesize the network. The fundamental building block itself however is very powerful. It enabled us to construct net-

works with multiple levels of complexity, the structure could reflect social structures that account for the frequency of encounters in relationships between individuals.

Sickness isolation is something we would coin a policy-level interaction. By this we mean a specific targeted behavior, like isolating one-self when feeling sick. In the resulting performance plot for the isolation experiment there is a clear spike in the ninth generation for the experiment without an isolation policy. The other plot seems to be very stable and the pathogen did not manage to find an obvious solution during its search.

The different performance weights is what aids the pathogen in adapting. Changes in these weights balance between lethality of the pathogen and its ability to infect, two goals which are difficult to attain especially when operating with limited resources. The experiment managed to develop pathogens with different skill-distributions, but it begs the question. Is the skill point system too one-dimensional? There is not much complexity in how the distribution system works.

6.2 Main issues

A recurring problem with the implementation of the model is the one mentioned when discussing the encounter rate. There is no way to systematically re-use generated network topologies, they are always re-generated for each trial run. The only experiment which does not suffer from this is 5.2.1 which uses topology as its main variable. The rest of the trials fail to truly keep this experimental variable static throughout trials for a given experiment. An important thing to note is also the scale of the experiments run. At most the trials had 2000 nodes, generation size of 60 and 20 generations. The trials showcase the flexibility of the system, but never really attempts to run any of them a larger scale. When we use a genetic algorithm approach, it is possible that the provided parameters did not throw a large enough net to explore optimal solutions.

The skill point system is an interesting approach but it proved difficult to implement meaningful mutation mechanisms for the genetic algorithm. Using a binary representation of the normalized values was considered, where a mutation would be a simple inversion of a binary value along this string. This however was not compatible with the restrictive skill point system because it could translate to a total skill point total that exceeded its skill point allowance.

The patient zero strategy has a very shallow implementation. Worst-cases and best-cases are not linear relationships with the number of edges a node has. In a scale free model, the worst case is the node with the fewest other connections and is the furthest away from a highly connected node.
In general the system design suffers from a lack of scope. Although it manages to implement many different mechanisms and allow these to interact with each other, many of them feel unpolished.

6.3 Future work

Amending the scope of the system is by no means a trivial task, but it is modular by design which makes it easier to work on. The network structure is the foundation, and should be extended to allow for re-use of existing network structures between experiments. It would also be interesting to design weighted network strategies that draws inspiration from real life social structures to

greater extent. This could perhaps lead to more meaningful results when there is a clear intent behind the structure of the network. The ego network as mentioned in section 2.2.5 would be an interesting network to implement and see the effect of.

The sickness isolation feature could also be extended to more general intervention tactics, instead of complete isolation an individual could restrict itself to a maximum number of other individuals. This could also be extended to be based on connection weight, where a higher weight implies a stronger emotional connection.

The scale of the system is on a community level. One could implement parallel communities which are connected together. Connection between communities could be restricted based on actionable policies or infection levels in communities. This pulls the epidemic from a local scale to a more global macro-level, a proper pandemic.

The main focus of the design was with the mechanisms of the pathogen and its traits. Other parts of the simulation feel more static, the only varying attribute in the individual is its immune system and connections. It would be interesting to add more detailed mechanisms, an example would be a happiness level that decreases as a function of active policies. An active policy could be restricting the number of social interactions, or travel that extends outside of the local community.

Bibliography

Barrett, C. L., Eubank, S., and Marathe, M. V. (2008). An Interaction-Based Approach to Computational Epidemiology. page 4.

Bobashev, G. V., Goedecke, D. M., Yu, F., and Epstein, J. M. (2007). A Hybrid Epidemic Model: Combining The Advantages Of Agent-Based And Equation-Based Approaches. In *2007 Winter Simulation Conference*, pages 1532–1537.

Cooper, I., Mondal, A., and Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139:110057.

Croccolo, F. and Roman, H. E. (2020). Spreading of infections on random graphs: A percolation-type model for COVID-19. *Chaos, Solitons, and Fractals*, 139:110077.

Freeman, L. C. (1982). Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291–304.

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 78(6):1360–1380.

Herrmann, H. A. and Schwartz, J.-M. (2020). Why COVID-19 models should incorporate the network of social interactions. *Physical Biology*, 17(6):065008.

Holland, J. H. (1975). *Adaptation in natural and artificial systems.* University of Michigan Press.

Holland, P. W. and Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*, 2(2):107–124.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics.

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., Núñez, R. C., Cohen, J. A., Selvaraj, P., Hagedorn, B., George, L., Jastrzebski, M., Izzo, A. S., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S. L., Bennette, C. S., Wagner, B. G., Chang, S. T., Oron, A. P., Wenger, E. A., Panovska-Griffiths, J., Famulare, M., and Klein, D. J. (2021). Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):e1009149.

Klimek, P. and Thurner, S. (2013). Triadic closure dynamics drives scaling laws in social multiplex networks. *New Journal of Physics*, 15(6):063008.

Longini, Jr., I. M., Halloran, M. E., Nizam, A., and Yang, Y. (2004). Containing Pandemic Influenza with Antiviral Agents. *American Journal of Epidemiology*, 159(7):623–633.

Marco Conti, A. P. . F. P. (2012). A Model to Represent Human Social Relationships in Social Network Graphs. pages 174–187.

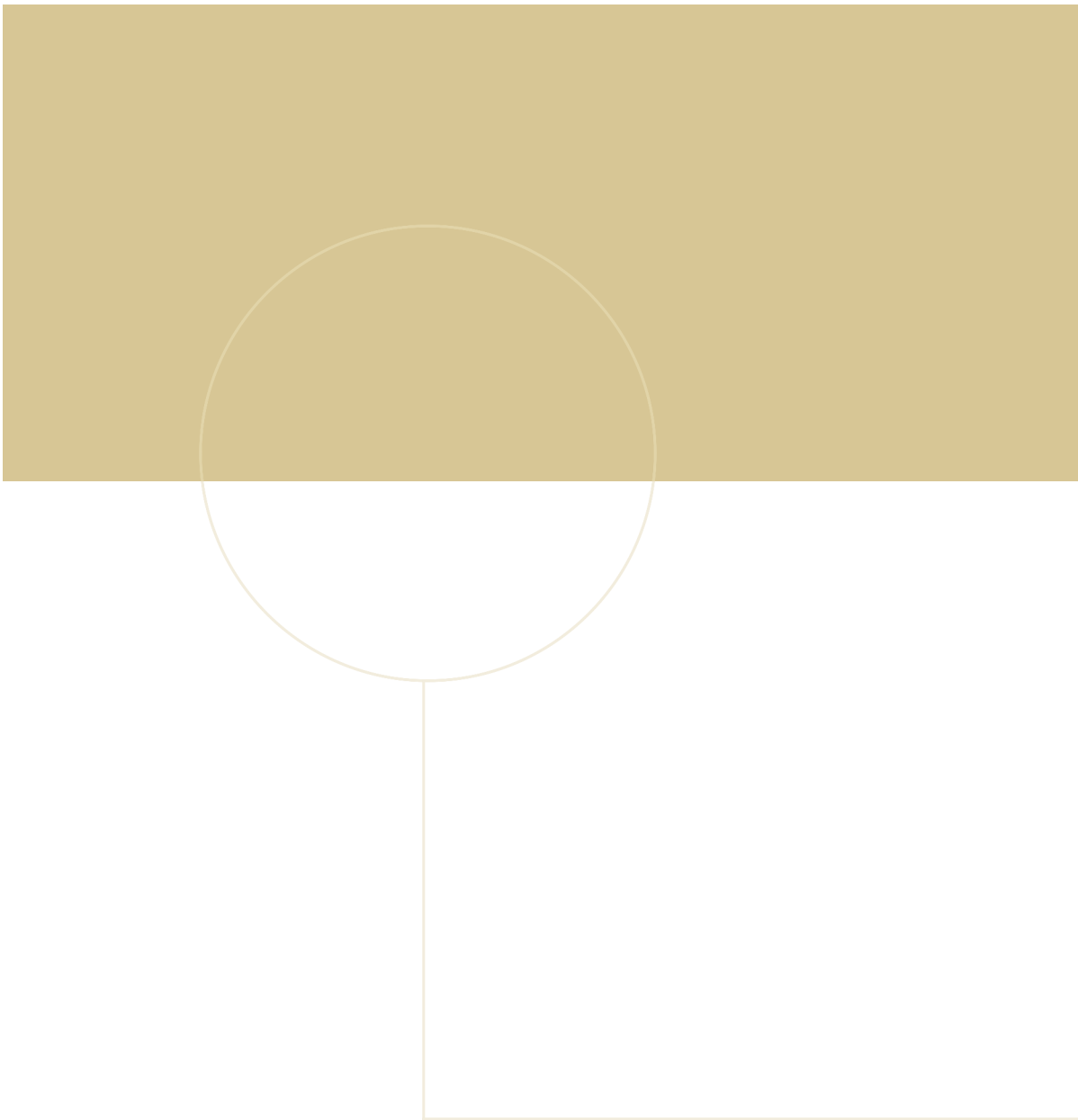
Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, 86(14):3200–3203.

Pechlivanoglou, T., Li, J., Sun, J., Heidari, F., and Papagelis, M. (2022). Epidemic Spreading in Trajectory Networks. *Big Data Research*, 27:100275.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

YANG, S. (2013). Networks: An introduction by m. e. j. newman. *The Journal of Mathematical Sociology*, 37(4):250–251.

Zhou, W.-X., Sornette, D., Hill, R. A., and Dunbar, R. I. M. (2005). Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 272(1561):439–444.



NTNU

Norwegian University of
Science and Technology