THE UNIVERSITY *of* EDINBURGH
School of Physics
and Astronomy

# Modelling the role of anti-vaccination sentiment in disease outbreaks on a heterogeneous network

**MPhys Project Report**

**L. M. Wilkinson**

*Submitted for the 40pt MPhys Project course PHYS11016*
April 4, 2022

**Abstract**

It is widely understood that preventative treatments such as vaccination are vital tools in the control of disease outbreaks, and yet the effects of the spread of anti-vaccination opinions are mostly unstudied on networks. This report investigates the interplay of a disease outbreak and opinion exchange on a bespoke heterogeneous network informed by British population data. It is found that higher anti-vaccination populations can significantly increase the length and size of epidemics. The model is also able to replicate theoretically predicted dynamics on a network, while demonstrating more complex behaviour associated with the randomness of the system. These results validate the novel model used, providing scope for its further utilisation in the future.

Supervisors: Professor R. Kao, Dr. E. Colman

# Personal statement

Over the course of the project, I attended two types of alternating weekly meetings to discuss my progress: individual meetings with Professor Kao and Dr. Colman, and group meetings with the other MPhys students undertaking epidemiology projects and staff from the wider research group.

I spent the first 8 weeks familiarising myself with the field and studying relevant literature. As an experimental physicist, epidemiological modelling was a new frontier for me, and I needed to become familiar with the terminology and conventions of the field. In early November, I attended the "Modelling Behaviour to Inform Policy for Pandemics" event series, where I learned more about the roles of data analysts, epidemiologists and sociologists in modelling human behaviour during disease outbreaks. Attending presentations by these researchers also inspired some of the features of my own simulation, such as the age-based network structure and the inclusion of case severity as a factor in behavioural decision making.

At this stage, I began to build the network and disease simulation in Python. I was provided with a very basic disease model on a single-ring network by Dr. Colman, which I incrementally expanded to include more detail and functionality: incorporating small world links, adding two more rings to the overall shape, and linking the three rings together. I also modelled several methods of vaccination (including age-based waves similar to the UK's COVID-19 vaccination roll-out, and completely random blanket vaccination) and compared the effects of these on the outbreak of disease.

Over the Christmas break, I built a basic voter model to incorporate behaviour into the simulation and linked nodes' behavioural statuses with the vaccination mechanism to allow vaccine refusal. I also used data to differentiate the three rings' populations: changing the relative ring sizes to match British age-stratified population data and processing the POLYMOD contact matrix data to obtain accurate edge densities. I also started to write this report and work on the presentation.

At the start of Semester 2, I contracted COVID-19 and was unable to work on the project for two weeks due to illness. When I returned to work, I designed and implemented the case severity system, allowing the disease model to affect the behaviour model and fully linking the two systems. This completed the model and allowed me to start data collection. For the next four weeks, I constrained most of the many parameters within the simulation, using both real-world data and preliminary results of the simulation to settle on sensible values for $R_0$, initial anti-vaccination proportion, and case severity distributions. At this time I also became aware of the Papst *et al.* paper, and tried to visualise the theoretical results using my simulation.

After Flexible Learning Week, I finished with data collection and began to focus on the final report, presentation and public summary. I spent the final weeks of the semester writing about my work, as well as creating a version of my code that could be used by a general audience to explore my work.

# Acknowledgments

My utmost thanks must go to Dr. Ewan Colman and Professor Rowland Kao for all of their time, support, and useful answers to my many lengthy emails. I would also like to thank Professor Graeme Ackland, Dr. Chris Banks, Dr. Anthony Wood and Aeron Sanchez of the extended Epidemiology MPhys group for their support and insights at our fortnightly meetings. Finally, I must thank my flatmates for supplying enough hugs, kind words, cups of tea and snacks to emotionally and nutritionally support a small nation.

# Contents

# 1 Terminology of this report

This project is rooted in the field of epidemiology, and as such this report will use field-specific terminology which may be unfamiliar to those from other disciplines. For the purposes of this project, key terms are defined as follows:

- **Node**: A point on a network representing one person. Each node exists in a combination of disease and behaviour states (e.g. infected and pro-vaccination), which are subject to change over the course of a simulated outbreak.
- **Link/edge**: A connection between two nodes, along which transmissible states can propagate in one or both directions.
- **Network**: A collection of nodes connected by edges. It is not necessary for every node to be reachable from any other node along edges.
- **Outbreak**: A scenario where, as time progresses, transmissible states propagate across a network. An outbreak ends when disease is no longer present in the network, or when disease is deemed to have become endemic (persistent within the population over many years).
- **Model**: A scheme designed to quantify and replicate a real-world phenomenon such as an outbreak, either theoretically or within a simulation.

Unless stated otherwise, these definitions also apply to discussion of previous work.

# 2 Introduction

Epidemiology is a curious scientific field in that conducting experiments to verify theories is highly discouraged and usually very unethical. Instead, researchers observe real-world disease outbreaks and build models to replicate the dynamics of known epidemics. These models provide a flexible tool with which to study a wide range of scenarios in a controlled – albeit idealised – environment. Though real-world outbreaks are far too unpredictable to foresee in perfect detail, epidemiological models offer valuable general descriptions of possible outcomes.

The ability to anticipate the spread of disease is vital to inform decision making and control real-world outbreaks when they arise. Modelling can be used for strategic decision making before an outbreak manifests itself: theoretical epidemics can be studied to identify the best course of action if such an outbreak occurs, allowing decision-makers time to prepare [1]. During an ongoing outbreak, researchers can provide a model with existing data about the situation in order to compare response methods and make tactical decisions [2].

Many outbreak responses require the population to behave in a specific way to curb the spread of disease: for example, during the COVID-19 pandemic, governments advised people to work from home, socially distance from one another, wear face masks and become vaccinated when possible. While many of these behaviours were mandated by law in various countries around the world, vaccination has mostly remained optional [3, 4]. This means that is vital that the population are willing to receive the vaccine in order to halt the outbreak as effectively as possible.

Unfortunately, it is extremely rare that an entire population will follow the advice of scientists. As well as general complacency or a lack of access to healthcare services, misinformation "outbreaks" can spread rapidly between citizens, leading to distrust in vaccination [5]. This leads to a reduced level of vaccination uptake, which in turn limits the impact on the spread of disease. Simulating the interplay between this behaviour and disease propagation can help us to understand the dynamics of an epidemic, as well as revealing which factors are most pivotal in tackling outbreaks.

Extensive mathematical work has been conducted to model the interaction between vaccination uptake and the spread of disease. Various factors have been used to drive opinions on vaccination, including social pressure [6] and the relative perceived risks of disease versus vaccination (in cases where vaccination carries a risk of illness or death, or some other "cost" such as time taken out of work to travel to a vaccination centre) [7, 8]. Both of these methods independently highlight different facets of human decision making: the former models the well-observed phenomenon of social influence, while the latter allows the disease outbreak to directly affect behaviour. The optimum model may require a combination of these methods, but as of yet there is a lack of work in this area.

While numerical modelling offers useful insights into outbreak dynamics, the method has historically been limited by unrealistic assumptions made when converting disease models into mathematical equations, such as the need for a homogeneous network where all nodes have the same number of connections [9]. The relatively recent development of computational network simulations has allowed progress to be made in this area, and work has been undertaken to model disease-behaviour interplay on more complex simulated networks, with a wide range of focus behaviours such as social mixing [10] and use of hand sanitiser [11]. However, there has been very little use of networks to study the effects of vaccine refusal on the spread of disease. A recent paper investigated the interaction between anti-vaccination sentiment and disease on a variety of common network shapes [9], but used an unconventional method of opinion "diffusion" across the network, where anti-vaccination sentiment was transmitted in the same way as disease. While this is an established model for the spread of information [12], it is less suitable for opinion dynamics, as it models the pro-vaccination opinion as a "passive" state which cannot be transmitted and is always overridden by an incoming anti-vaccination wave.

Another advantage of complex network simulation is the ease with which the model can be informed by and compared to real-world data. Simulated network topology can trivially be customised to mimic or recreate complex observed networks such as brain networks [13]. In terms of epidemic modelling, rich data exists that describes population demographics

[14] and the degree to which different subpopulations interact with each other (or "social mixing") [15]. The COVID-19 pandemic has also provided a wealth of data which can inform dynamic network design, such as the COMIX study [16], which has measured the weekly evolution of social mixing over the course of the pandemic.

This project aimed to construct a novel network shape inspired by data about the British population and their social mixing patterns, and use this to explore the interplay between vaccine refusal and the spread of disease. Existing disease and opinion propagation models were adapted such that they could be coupled to one another, allowing interplay between the parallel outbreaks over the course of an outbreak. Several additional features were included to lend further realism to the simulation, and tested as a "proof of concept" for future simulation work. Details of the network construction and its features can be found in Methods.

It was found that a larger anti-vaccination subpopulation can cause a significant increase in both the length and size of outbreaks, but this group also disproportionately affected by disease throughout the course of an epidemic. When using a parameter space that consistently produces endemic disease conditions (outbreaks lasting longer than five years), periodic dynamics were observed between the amount of active cases and the amount of immune nodes in the network, supporting previous mathematical modelling work [7]. Further details can be found in Results, as well as more in-depth discussion of the findings in Discussion.

# 3 Background

## 3.1 Networks

Networks are an invaluable tool in epidemiological modelling: in the real world, outbreaks spread between contacts (represented by nodes) which have access to one another (via links). Network theory is a natural framework with which to model these connections between members of a population, and thus understand how real contagions spread.

### 3.1.1 Topology

Networks can be classified as either simple or complex according to their topology. A simple network is built using a consistent algorithm, such as a square lattice. A commonly used simple network in epidemiology is the ring network, where nodes are arranged in a circle and connected to their neighbours, as shown in Figure 1(a). Simple network shapes such as a ring are ideal for mathematical modelling due to their homogeneity, but are often too over-simplified to produce realistic outcomes.

Complex networks, on the other hand, include heterogeneous shape features and irregular structures. Most real-world networks, including contact networks across which epidemics

take place, are complex networks. One typical aspect of complex networks is the small world phenomenon: most nodes are connected to most other nodes by a relatively small number of edges, regardless of their position and immediate neighbours [17, 18]. In network topology, a regular ring network can be modified into a "small world network" by including edges across the ring (an example of this is shown in Figure 1(b)). Using data about the inter-connectivity of real-world networks, small world links can be included as a simple way to increase network complexity.

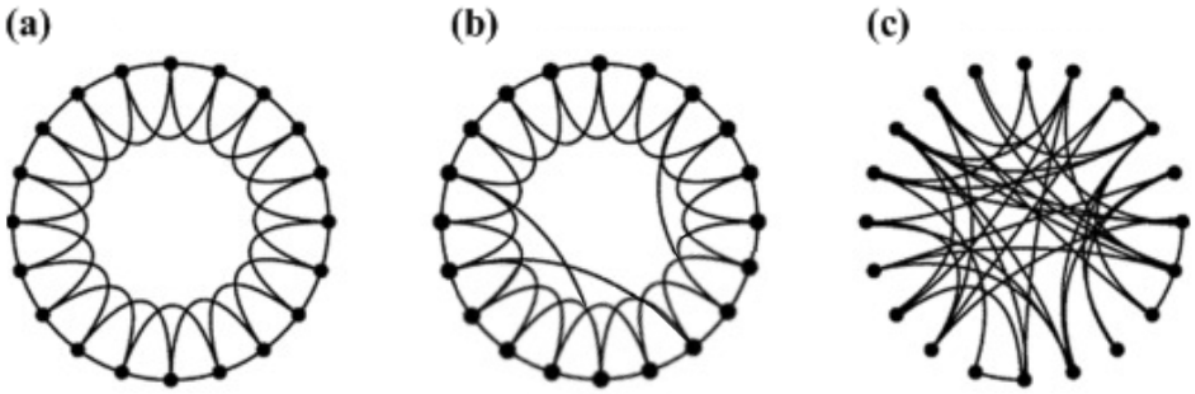


Figure 1. Three ring-based network shapes: (a) A basic ring network, where nodes are connected to their two closest neighbours on each side. (b) A Newman-Watts small world network, where random links are added across the ring without removing or rewiring existing rink links. (c) A random network, where all links are randomly generated. Adapted from [19].

### 3.1.2 Incorporating age assortativity

Historically, the network shapes used to model disease epidemics have faced several challenges due to their over-simplicity, as outlined by Pellis *et al.* in a 2015 paper [20]. One feature of real social networks that is often not captured is age assortativity: people of the same age have a much higher tendency to interact with one another than they do with members of other age groups. The POLYMOD study quantified this phenomenon in detail in 2008, surveying participants from several European countries about their daily contact patterns [15]. As shown in Figure 2, there are heterogeneous levels of both overall contact (shown in Row A) and physical contact only (shown in Row B) between age groups. As this study was specifically conducted to gather contact information for use in respiratory disease outbreak modelling, physical contact was defined as any skin-to-skin contact such as a kiss or handshake.

The average number of daily physical and social contacts for each age group within each nation is also provided in the form of a "contact matrix". The British contact matrices are included in the Appendix for reference. By assigning age values to nodes, this data can be

applied directly to networks by adjusting the interconnectivity of nodes from different age groups. It should be noted that these matrices are not perfectly symmetric - this is because of the different sizes of the age groups involved with the study.

Pellis *et al.* recommended the use of contact matrix data to improve the output of disease network models, and subsequent work has indeed proven that networks informed by POLYMOD data produce more realistic outbreak dynamics [21]. Several studies have incorporated the data into their network models to study the spread of various diseases [22, 23], but there is little existing work that simulates both disease and opinion dynamics using POLYMOD contact data.

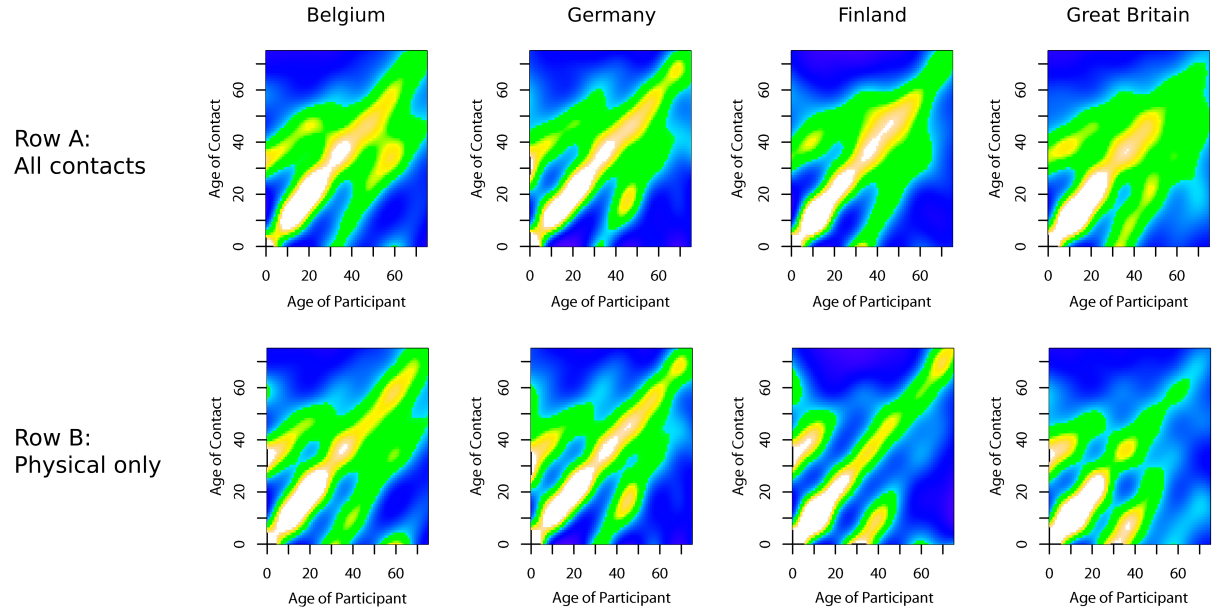


Figure 2. Heat maps visualising the levels of daily contact between different age groups in Belgium, Germany, Finland and Great Britain. White indicates the highest contact rates, green intermediate contact rates, and blue the lowest contact rates, relative to each country's contact intensity [15]

## 3.2 Disease modelling

Once a network configuration has been chosen, it is possible to introduce a disease outbreak model and study its propagation through the system. While different diseases have varying intrinsic properties such as method of transmission, infectivity and length of infectious period, the underlying mechanics of transmission are usually modelled in the same way, using compartmental models.

### 3.2.1 Compartmental models

Disease outbreaks are generally modelled by categorising the population into compartments, which are groups with different statuses with respect to the disease. The most common compartmental model is the SIR model [24], where nodes cycle through the following compartments:

- **Susceptible** - individuals are unprotected from disease, and may become infected if exposed to another infected individual.
- **Infected** - individuals with an active case of disease, who can transmit the infection to any susceptible neighbours they may have.
- **Recovered** - individuals who have recently been infected but are no longer contagious. These individuals are afforded post-infection immunity and cannot be infected again for a period of time.

The SIR model has inspired a wide range of variations, such as the SIRS model (nodes can loop back to the susceptible state after they recover), the SIRV model (nodes can be in a "vaccinated" state to protect them from disease), the SEIR model (nodes incubate the disease while in an "exposed" state before becoming infectious) and many others [25]. All main variations of the SIR model have been extensively mathematically modelled on homogeneous networks, but the irregularity of complex networks means that this is generally not possible [25]. However, computational simulations allow us to study the dynamics of SIR-like models observationally, without the need for explicit mathematical treatment.

### 3.2.2 Quantifying the spread of disease

Several parameters must be supplied to any disease model in order to quantify the way that infection spreads through the population.

One key variable that describes the ability of a disease to spread is its basic reproduction number $R_0$: the average number of new cases that an infected person will create by infecting its neighbours, in a population that is entirely susceptible to infection. By definition, if $R_0 < 1$, the number of active cases will decrease over time and the disease will eventually die out. If $R_0 > 1$, the number of active cases will rise, and if $R_0 = 1$, the outbreak will maintain its size [26].

However, if the population includes immune nodes $R_0$ will not accurately predict the growth of an outbreak, but will instead provide an upper limit for the transmission rate. In this case, the effective reproduction number $R_e$ (also referred to as "the $R$ number") describes how many new cases are *actually* being generated by each infected node [27]. In a SIR-like model where nodes become immune after infection, $R_e$ therefore provides a better description of the status of the disease over the course of an outbreak.

In a stochastic disease model, whether an infected node infects each of its neighbours is decided by probability and random number generation. At any point, the probability that a node will infect each of its neighbours is given by

$$\beta = \frac{R_0}{\langle N_{neighbours} \rangle}$$

where $\langle N_{neighbours} \rangle$ is the average number of physical neighbours any node has [26]. This probability will translate into a lower $R_e$ when some neighbours are immune.

Another aspect of transmission that varies between disease is the amount of time that a node spends in each compartment of the disease model. For instance, for seasonal influenza it has been observed that the typical infectious period is ∼2 days and the recovered (immune) period is ∼4-6 years [24], whereas measles has an infectious period of ∼8 days and affords permanent immunity after infection [28].

### 3.2.3 Vaccination schemes

In real outbreaks, populations are usually inoculated using a targeted vaccination scheme, where individuals are prioritised according to a specific epidemiological trait. For example, the primary method for inoculation against smallpox is ring vaccination, where all contacts of a confirmed patient are vaccinated, as well as all contacts of those contacts [29]. Another example is the roll-out of the UK's COVID-19 vaccination program, where the most vulnerable subgroups of the population (in this case, the oldest) were vaccinated first [30].

When computationally modelling disease outbreaks, the implementation of these vaccination schemes is simple as the systematic methods can easily be converted into algorithms. However, network shape has a major impact on the spread of disease: for instance, introducing small world links to a systematic network changes the initial growth of the epidemic from exponential to power law [31]. Therefore, the network shape affects which vaccination strategies will be most successful: the vaccination method which is most successful in a real outbreak may perform poorly on a simulated network if the network shape used is not similar enough to the real network [32].

Existing work has been undertaken to compare vaccination methods on various network configurations. In one paper, Bansal *et al.* used real-world network data about the population of Vancouver, Canada, and found that ring vaccination and targeted vaccination for at-risk groups were the most effective methods of protecting the population as a whole [33]. In general, it has been shown that targeted vaccination is always more effective than random vaccination [32].

## 3.3 Behaviour modelling

There are three broad mechanisms used by opinion dynamics models to simulate the way that group members change their opinions over time: game theory models (where members have perfect knowledge of their environment and use this to act in their best personal interest), peer influence models (where members have imperfect knowledge of their environment, so rely on the opinions of their neighbours to make decisions) and personal experience models (where members have imperfect knowledge of their environment, but use their limited knowledge to draw their own conclusions about what is in their best personal interest).

Though it is likely that most real-world opinion dynamics require a mix of these techniques, many vaccination opinion simulations use only one of these established mechanisms to model behaviour in a system [7, 8, 9]. However, each scheme has its own pitfalls that can limit its ability to model human behaviour, as will be described in the following sections.

### 3.3.1 Game theory and free-loading

In perfect-information game theory models, members of a group have access to complete knowledge about their environment: they are aware of their neighbours' decisions and how their own decisions can affect their personal risk level [34]. As time progresses, the system evolves towards the Nash equilibrium, where all members of the group reach a conclusion about what strategy is the most personally beneficial given the decisions of the rest of the group, and the system reaches equilibrium.

As well as a Nash equilibrium, where all members of the system are *individually* best-served given their surroundings, every system will have at least one Pareto optimum, which is the overall best outcome for the group *as a whole*. These two configurations may not be identical: at the Pareto optimum, some members may have to use a strategy that is not the most beneficial move for them personally, but better serves the group overall [34].

In the case of vaccination opinion modelling, game theory generates a problematic paradigm known as "free-loading". As each node chooses the most personally beneficial strategy, any unvaccinated nodes with many vaccinated neighbours are aware that they are not likely to contract disease, so if there is any vaccination cost at all, it is preferable for them to remain unvaccinated, leading to a Nash equilibrium which is far less effective than the Pareto optimum. While free-loading has been shown to exist in real populations' decision making on vaccination [35], this leads to an unviable long-term strategy for dynamical study [36].

### 3.3.2 Peer influence models and divorce from reality

In a peer influence system, group members rely only on the opinions of their neighbours to inform their own opinion.

The basic voter model evolves as follows [37]: a random node is selected at a random time, and a random neighbour of this node is selected. The node then inherits this neighbour's state. As this system evolves, clusters with diffuse edges form as shown in Figure 3(a). This is not dissimilar to the effect of the Ising model (shown in Figure 3(b)), a two-state system from statistical mechanics where a node's state is changed according to the overall state distribution of its neighbours, as opposed to one of the neighbours at random [38].

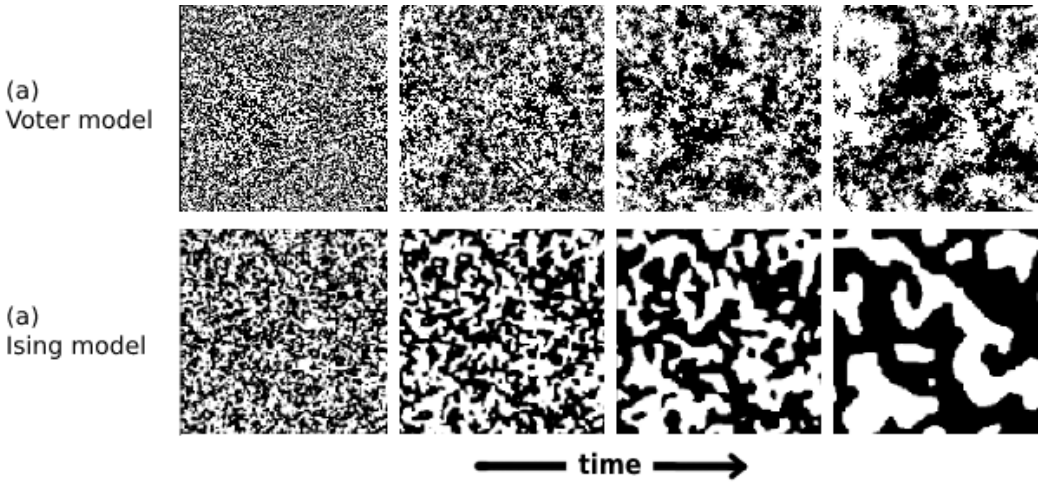


Figure 3. Side-by-side comparison of the time evolution of (a) the voter model and (b) the Ising model, both initialised on random lattices. Adapted from [38].

One limitation of this model is that it is completely ignorant of the nodes' environment. In terms of vaccination sentiment, a group member may easily become anti-vaccination even when they are at extreme risk of infection, as risk is not considered. While this may be suitable for situations where group members have no information about their risk level, it is not realistic when they can observe the world around them and use agency to make decisions.

### 3.3.3 Personal experience models and lack of social pressure

An alternative to peer-influence models is to change opinion based on nodes' personal experiences. For instance, previous work has used perceived risk to inform vaccination opinion [7, 8]. To do this, each node uses one strategy (becomes vaccinated or does not), and then observes how successful that strategy is: if they became vaccinated but experienced a breakthrough infection, this would be deemed unsuccessful, as would remaining unvaccinated and then contracting disease. If the current strategy proves unsuccessful, they switch methods.

This system creates a strong link between the environment and decision making, leading to mostly "sensible" decisions over time. However, this mechanism completely removes the aspect of social pressure, which is observed to play a key role in real-world decision making [39]. In order to avoid this, this project uses a combination of techniques from personal experience modelling and peer influence modelling, as described in Methods.

# 4 Methods

## 4.1 Network design

To incorporate POLYMOD data into the model, a custom network shape was designed: multiple rings were included to represent different age groups, each with their own internal contact density (representing contact between members of the same age group) and density of links to the other rings (representing how connected age groups are to one another).

For simplicity, three rings were simulated representing the three broad age categories: children (under 16s), adults (aged 16-64), and the elderly (aged 65+). The populations of the three rings were weighted 19%, 62.5% and 18.5% respectively in order to reflect the approximate age distribution of the United Kingdom [14]. The child and adult groups were represented by two Newman-Watts small world rings, as were shown in Figure 1(b), and the elderly group was represented as a random network, as was shown in Figure 1(c). For the child and adult rings, ring links were created for neighbours of distance 1 (nearest neighbours only). As will be discussed in Section 4.2.1, low contact rates among elderly people meant that no ring links were included for the elderly ring, hence the random network shape with no systematic ring links. Each of these rings was then connected to the others using more random links. The bespoke network structure is shown in full in Figure 4.

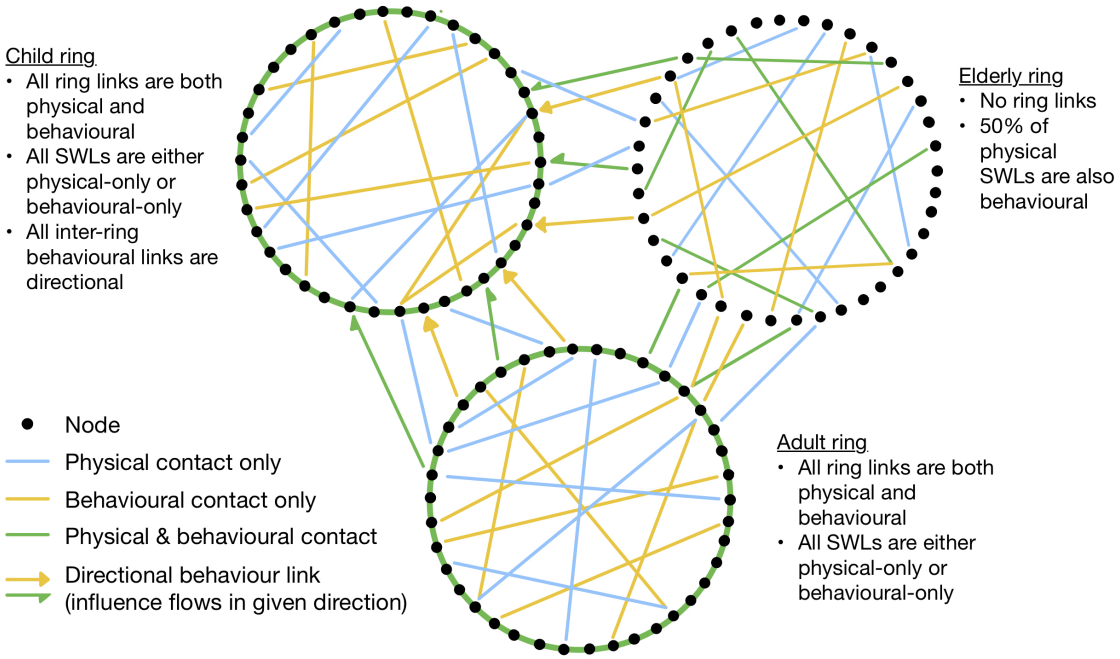


Figure 4. Simplified visual representation of the network edges used in the simulation. Note that the number of nodes and links in this figure are chosen for visual clarity and are not accurately generated.

As shown in Figure 4, the behaviour and disease models propagate across their own sets of links, in order to generate different social and physical contact densities as described by the POLYMOD data. The neighbours of a node were deemed to fall into one of three categories: physical contact only (such as passing contact with a worker in a shop), behavioural contact only (such as contact made over the internet), or physical and behavioural contact (such as close family members). Links to neighbours falling into the third category were shared by both the disease network and the behavioural network, with all other links being exclusive to their respective network. In addition, the behavioural network included directional links: this was incorporated to prevent the opinions of children from influencing adults and the elderly, while older age groups could still influence children. This is a novel feature which is not commonly modelled in the study of the spread of disease, but aims to add realistic detail to the opinion dynamics.

This network design also addresses several of the issues of conventional networks which were highlighted by Pellis *et al.* [20]: as well as the age-specific contact rates creating heterogeneity in contact densities across the network, the use of random links (and of different levels of randomness across the three rings) creates clustering heterogeneity: an irregular mix of well-connected and less connected nodes. The use of directional behaviour influence also adds realistic complexity to the system. In these respects, the network design takes several steps towards a more realistic generic shape for outbreak modelling.

## 4.2 Processing the POLYMOD data

The three age groups were chosen according to data from the Office of National Statistics, which reports the population breakdown in terms of children (aged 0-15), adults (aged 16-64) and the elderly (aged 65+) [14]. POLYMOD data splits the population into five-year brackets, e.g. 0-4, 5-9, 10-14, and so on. As a result, the POLYMOD data was collated into the age groups 0-14, 15-64 and 65+. As there was a slight discrepancy in the age bands for children and adults, the resulting contact densities were rounded to one decimal place.

The collation process was as follows:

1. Contacts of the participants (matrix rows) were combined by summing over the rows to obtain a total number of contacts for each of the three broader age ranges.
2. Matrix columns (the ages of participants) were combined by obtaining a weighted mean, using the number of participants in each POLYMOD age group (this data was provided by the study).

### 4.2.1 Physical contact densities

As the network is static (connections between nodes do not change over time), each node must have access to the total number of neighbours that they could contact during their

entire infectious period. As will be discussed in Section 4.4, the simulation uses a log-normal distribution to determine the time between a node becoming infected itself and infecting its neighbours, with a mean of 5 days and a dispersion of 1.3 days. For the purposes of the following discussion, we will consider the overall infectious period to be around one week.

The POLYMOD contact matrices only provide average daily contact information. However, multiplying these numbers by seven would likely be unrepresentative of reality, as many real-world situations such as workplaces and schools cause the same people to mix repeatedly over several days. There is a lack of age-stratified information about repeated contact habits, so the extent to which daily contacts reoccur is unknown.

As such, a POLYMOD "scale factor" was chosen experimentally by simulating one hundred outbreaks with $R_0 = 1.2$ for each integer scale factor from 1-7, and examining the variance of the results. Scaling the daily data by a factor of 3 produced the most consistent outbreaks, with the least observed variance in outbreak length and size. As such, this scale factor was selected for the disease network contact density, producing the following contact matrix:

| Age group | Child | Adult | Elderly |
|---|---|---|---|
| Child | 11.7 | 3.6 | 1.8 |
| Adult | 10.2 | 10.5 | 4.2 |
| Elderly | 0.3 | 0.9 | 2.7 |

Table 1. Scaled contact matrix showing the average weekly physical contacts between the three age groups. Columns headers are the ages of participants, rows are the ages of contacts.

Note the apparent asymmetry in these results: this is due to the variation in the population sizes for each of the three groups. For instance, there are approximately three times as many adults as there are children, which accounts for how, on average, children contact 10.2 adults each while adults contact 3.6 children each.

It should also be noted that initial testing of the system took place before the scale factor was selected. At this point, the average number of elderly-elderly contacts was 0.6, which was too low to include systematic ring links (providing a baseline of 2 elderly contacts to each node) in the structure of the elderly ring. Though it was later decided to scale this value to 2.7, the original network design was maintained to allow for flexibility in any future use of the simulation, which may require a different scale factor.

### 4.2.2 Behavioural contact densities

Similarly, the static network requires nodes to have contact with a larger number of social contacts than the daily figures. In this case, there are three complicating factors to consider.

Firstly, the opinion model timescale is of the order of weeks (as will be discussed in Section 4.5), requiring an increase from the daily number of social contacts. Secondly, not all social contacts are equally capable of changing a person's opinion - a close friend or family member will have a much greater influence than a barista in cafe, for instance - requiring a decrease in the number of contacts considered. Finally, there is a fundamental issue with the data provided: the POLYMOD study only provides values for "physical contacts" (including physical-only contacts and physical-social contacts) and "all contacts" (which additionally includes social-only contacts). With no way to determine how many social contacts are included in the "physical contacts" data, combined with the two other factors, it is not possible to isolate accurate social contact numbers from the POLYMOD study. As such, this data was processed as follows:

1. The contact matrix for "all reported contacts" was collated in the same way as the physical contact matrix, as described in Section 4.2.1.
2. The scale factor of 3 was retained and applied to this new reduced matrix.
3. The amount of contacts that were *purely* behavioural was obtained by subtracting the physical contacts matrix (Table 1) from the "all reported contacts" matrix. This "behaviour only" contact matrix is shown in Table 2.

| Age group | Child | Adult | Elderly |
|---|---|---|---|
| Child | 8.7 | 1.8 | 0.9 |
| Adult | 6.9 | 17.1 | 12.9 |
| Elderly | 0.6 | 0.9 | 2.7 |

Table 2. Scaled contact matrix showing the social contacts for the three age groups. Columns headers are the ages of participants, rows are the ages of contacts.

As it was unknown how many of the physical contacts were also behavioural, an arbitrary number of additional links were chosen to be included in both the disease and behavioural networks: 50% of randomly assigned links (small world links and inter-ring links) were retained, and all systematic links were retained.

The implementation of these contact figures is discussed in Section 4.3.

## 4.3 Network construction

A total of 1000 nodes were simulated across the three rings. As discussed in Section 4.2, up to 50% of the disease network links were assumed to be shared by the behavioural network. As such, the disease network was generated first, and then used as a basis from which to generate the behavioural network.

### 4.3.1 Disease network

The three rings of nodes were generated, and systematic ring links were added to the child and adult populations. Small world links were then added within each ring until the required physical contact rates within each age group were met. These links were generated at random between any two previously unconnected nodes within the same ring.

Inter-ring links were then added to recreate the required physical contact rates between the three age groups. These were also generated in the same way as small world links, by connecting any two previously unconnected nodes, but this time the nodes belonged to separate rings.

### 4.3.2 Behaviour network

As discussed in Section 4.1, the behavioural network inherited some of the edges from the disease network to simulate contacts that were both physical and behavioural. Within the child and adult rings, the ring links were used as the inherited edges. As there were no ring links in the elderly ring, 50% of the small world links were copied over from the disease network. 50% of all inter-ring links were also inherited from the disease network.

Once the inherited edges were established, the behaviour-only links were generated in the same way as the small world and inter-ring links from the disease network (as described in Section 4.3.1).

A unique feature of the behaviour network mentioned in Section 4.1 was directionality: inter-ring edges involving children were modified such that child nodes could inherit behaviour from adults and the elderly, but could not transmit their behaviour to the older groups. All other links were bi-directional by default.

## 4.4 Disease propagation

A compartmental SIRS disease model was implemented, with the following sequence of events repeating for every infection of the next generation:

1. A node $j$ is infected at time $t$ by a previously infected physical neighbour $i$.
2. Node $j$ infects each of its susceptible neighbours with probability $\beta$. The generation time (an offset time from the infection of node $j$ to the infection of the neighbour $k$) is generated from a log-normal distribution with a mode of 5 days and a dispersion of 1.3 days (chosen as a generic infectious period for a respiratory infection).
3. Node $j$ is marked as immune to further infection.

4. Node $j$ becomes susceptible to further infection after a time generated from a log-normal distribution (by default, this had a mode of 20 days and a dispersion of 3 days).

At time $t = 0$ seconds, the simulation is initialised with five simultaneous infections of randomly chosen nodes in the network. The simulation was initialised with $R_0 = 1.3$.

## 4.5 Behaviour propagation

In order to combat the limits of individual behaviour schemes described in Section 3.3, behaviour propagation was modelled using a two-state voter model combined with a personal experience aspect.

Every node was considered to be in either a "pro-vaccination" or "anti-vaccination" state. The basic voter model mechanism functioned as follows:

1. A node was chosen at random, at a random time in the first year of simulation
2. A random behavioural neighbour of the node was chosen
3. The node inherited the behaviour state of its chosen neighbour (if the neighbour behaviour was the same as the node behaviour, there was no noticeable effect)
4. The next opinion-change event was scheduled for the node after a fixed time period (by default, this was 12 weeks).

Under this basic voter model, there will be fluctuations in the anti-vaccination and pro-vaccination numbers, but overall the two groups will not change size. However, there is emerging evidence that vaccine hesistancy has slowly decreased over the course of the COVID-19 pandemic [40]. In light of this information, the personal experience system was implemented to slightly decrease the size of the anti-vaccination population over time.

To do this, a system was introduced to represent perceived threat of illness, where every case of disease had an associated severity value $S$. This severity was drawn from a probability distribution according to the age group of the infected node, such that children were more likely to experience a low-severity case and elderly people were more likely to experience a severe case, in accordance with data from the COVID-19 pandemic [30]. Figure 5 illustrates the forms of the distributions used for the three age groups.

In the case that a pro-vaccination node was going to inherit anti-vaccination behaviour from a neighbour, the severity of cases experienced by the node's behavioural contacts and the node itself were first checked. If any of the node's neighbours had experienced a disease case with $S \geq 0.8$ (a value chosen to represent the most severe cases, approximately analogous to hospitalisation or death), the probability of the switch to anti-vaccination behaviour was

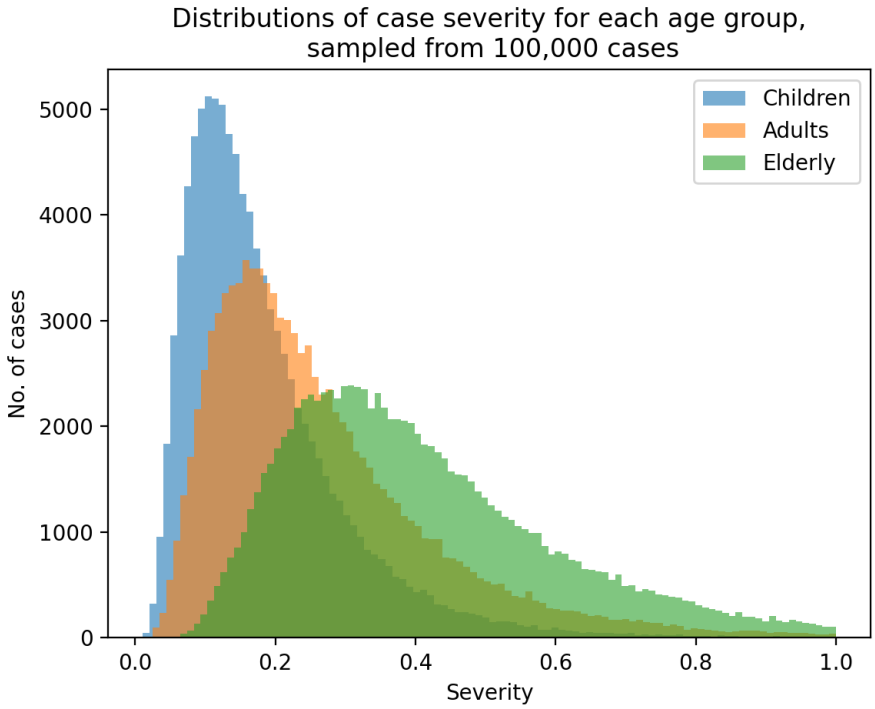


Figure 5. Case severity distributions for the three age groups. These distributions are drawn from log-normal distributions ($\mu_c = 0.2, \sigma_c = 0.6, \mu_a = 0.6, \sigma_a = 0.6, \mu_e = 1.1, \sigma_e = 0.5$), then divided by eight to attempt to normalise the value, but redrawn if the value is still greater than 1. As such, they are only intended to be relative to one another, rather than rooted in any real world expression.

reduced from 100% to 50%. If the node itself had experienced a case with $S \geq 0.8$, this would also reduce the probability by 50%, for a maximum reduction to 0%. Over time, this results in a slowly declining anti-vaccination population as the rate of change from pro-vaccination to anti-vaccination is reduced while the inverse remains unaffected. This system means that more cases of disease in the network lead to a more severe (though still fairly gradual) decline of the anti-vaccination population, as more nodes are likely to have experienced or witnessed a severe case of disease.

By default, the network was initialised with 75% of nodes in the pro-vaccination state and 25% in the anti-vaccination state, randomly assigned across all nodes with no age weighting.

## 4.6 Vaccination mechanism

As discussed in Section 3.2.3, there are many real-world vaccination schemes which can be implemented into a simulation with relative ease. Three vaccination methods were successfully implemented in this simulation: random vaccination (where every node is randomly selected and offered their first vaccination appointment at a random time), age-based waves

of vaccination (where nodes in the elderly ring are offered their first appointments first, then adults, then children), and a log-normal distribution method (where first appointments are offered to the entire populace using a log-normal time distribution, to simulate a mass-uptake scenario).

As previously established, the topology of the network can alter the effectiveness of different vaccination methods, and as such the results of all three schemes were compared to decide on the most suitable method for data collection. The random vaccination method was chosen, as it ensured that all nodes had been offered their first vaccination by the end of the first year of the simulation. In much of the parameter space considered, outbreaks would only last up to one year, so this method gave a clearer picture of vaccine uptake versus refusal.

The random vaccination model worked as follows:

1. Each node is offered a first vaccination at a random time within the first year of the simulation.
2. The node accepts or refuses the vaccination based on its opinion status at the time of the appointment.
3. Regardless of the node's choice, its next vaccination is scheduled for 365 days later.

Vaccination granted nodes 100% immunity from the moment of vaccination until the second that vaccination wore off, but no reduction in the severity of cases after vaccination. By default, the protective period lasted for a time drawn from a log-normal distribution with a mode of 90 days and a dispersion of 7.5 days. After this time, the node would return to a susceptible state and could henceforth become infected.

## 4.7 Default parameter values for data collection

For reference, the default parameters mentioned in this section are collected here. The following parameters were not altered in any of the results discussed in later sections:

- Network size: 1000 nodes (190 children, 625 adults, 185 elderly)
- Log-normal distribution for infectious period: mode of 5 days, dispersion of 1.3 days ($\mu_I \simeq 1.68, \sigma_I \simeq 0.26$)
- Amount of time between each node's opinion change events: 12 weeks
- $R_0 = 1.3$

Depending on the role of vaccination in the results obtained, two main parameter combinations were used to achieve specific dynamics for study, or to change the relative effect of vaccination over infection:

- Consistent endemic outbreaks: post-disease and immunity drawn from identical log-normal distributions with a mode of 20 days, dispersion of 3 days.
- Vaccination is protective for significantly longer than infection: post-disease mode of 20 days and dispersion of 3 days, post-vaccination mode of 90 days and dispersion of 7.5 days.

For each individual result in Results and Discussion, the combination of immunity timescales used will be mentioned and justified.

# 5 Results

Investigation of the outputs of the model followed three main lines of enquiry:

1. How does the size of the (initial) anti-vaccination population affect the characteristics of outbreaks?
2. Inversely, how does the disease outbreak affect the anti-vaccination population in comparison to the pro-vaccination population? Are anti-vaccination nodes more heavily exposed to disease?
3. What dynamics occur within individual outbreaks? How are these dynamics affected by changing key variables of the simulation?

The following sections provide outputs from the simulation with which these questions can be addressed.

## 5.1 Effects of the size of the anti-vaccination subpopulation

Figure 6 shows how increasing the initialised anti-vaccination population size affects the outbreak size (total number of cases) after five years. One hundred outbreaks were performed for each anti-vaccination fraction, using the same randomly generated network in order to minimise additional random uncertainty. The "consistent endemic outbreaks" timescale combination was used, as described in Section 4.7. Outbreaks which did indeed become endemic (disease was still present in the network at five years, when data collection was stopped) are plotted in purple, and non-endemic outbreaks (which ended before five years) are plotted in grey.

This graph highlights two distinctive trends: the mean outbreak size increases with the initial anti-vaccination population size (shown by the positive trend in y values), and the proportion of outbreaks which lead to endemic disease increases with the initial anti-vaccination population size (less grey points are plotted as the x value increases).

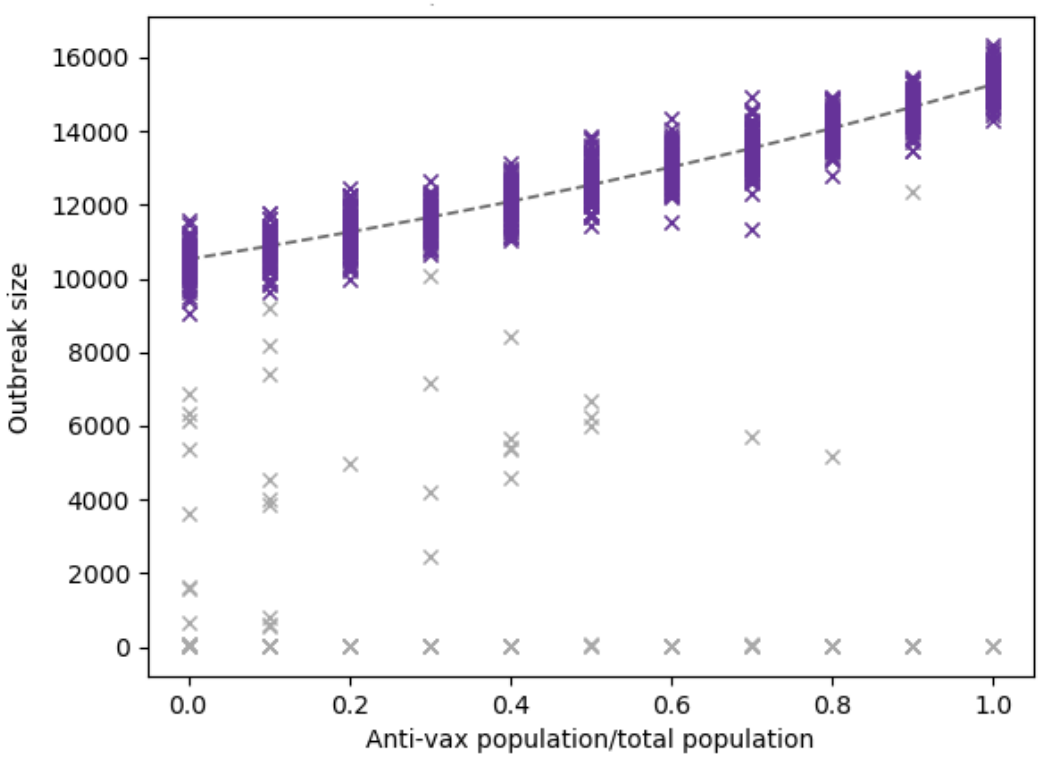


Figure 6. Initialised anti-vaccination subpopulation size vs. total cases after 5 years of simulation time. Purple plot points are outbreaks which became endemic (disease was still present in the network after 5 years)

To examine the outbreak size trend in closer detail, the mean value of each "cluster" of endemic outbreaks was plotted in Figure 7, with the errorbars representing the standard deviation of each cluster. The mean values follow an exponential trend ($y = ae^{-bx} + c$), with the parameters $a = 5751.49 \pm 1001.22$, $b = -0.60 \pm 0.08$ and $c = 4771.03 \pm 1030.05$.

The percentage of outbreaks which led to endemic disease was also plotted in Figure 8. A linear fit ($y = mx + c$) was obtained with parameters $m = 11.1 \pm 3.1$ and $c = 85.5 \pm 1.8$.

Due to the nature of the data collection, it was not possible to obtain an error value for these points. The 27.9% uncertainty in the fit gradient ($m$) suggests that numerical analysis of this trend line should be considered preliminary - it is recommended that further investigation of this trend should collect data for more than the 100 outbreaks per data point used here, and if possible should use more $x$ values to obtain more data points and constrain the fit further. Regardless, using a linear fit, it can be concluded from this data that an entirely anti-vaccination population will increase the likelihood of disease becoming endemic by at least 8.0% over an entirely pro-vaccination population.

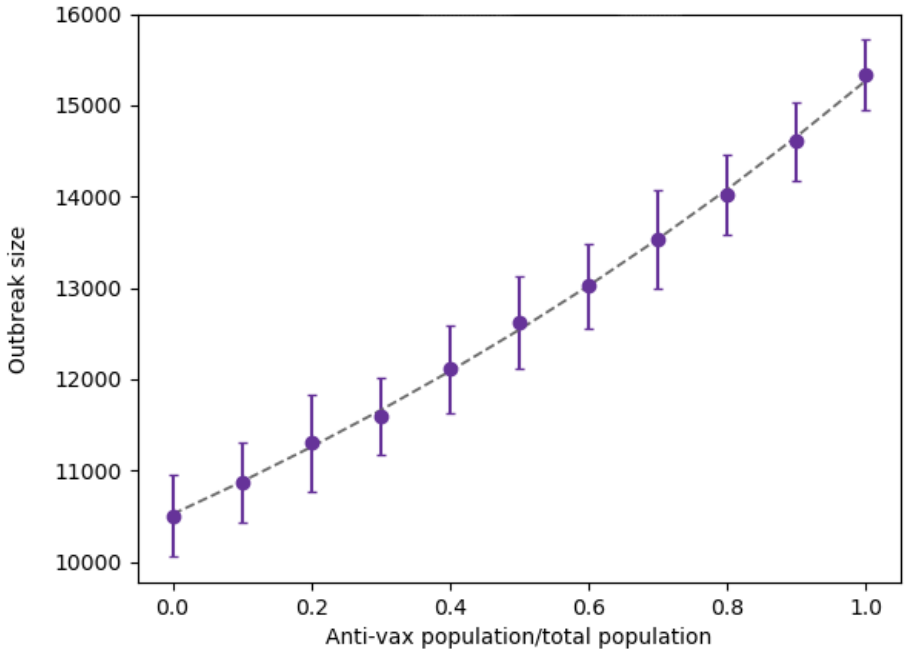


Figure 7. Initialised anti-vaccination group size vs. mean endemic outbreak size after five years of simulation time, showing an exponential trend.

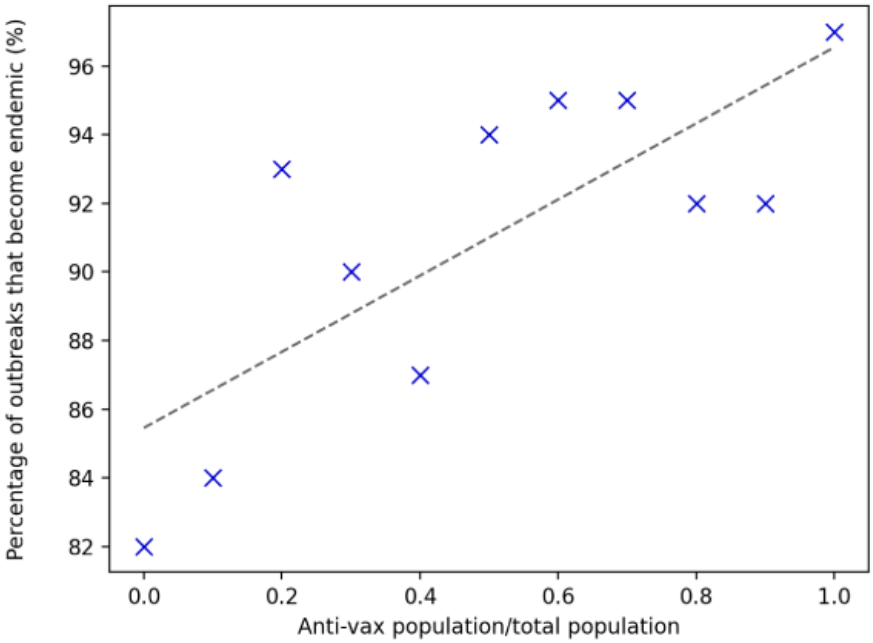


Figure 8. Initialised anti-vaccination subpopulation size vs. percentage of outbreaks which became endemic (lasted longer than five years).

## 5.2 Case burden on anti-vaccination nodes

To explore whether the disease outbreak disproportionately affects anti-vaccination nodes, the percentage of infected nodes in the subpopulation at any given time during an outbreak was investigated, as shown in Figure 9.

In order to examine whether anti-vaccination sentiment was more harmful for any specific age group, this data was plotted for all nodes and also for each age group separately.

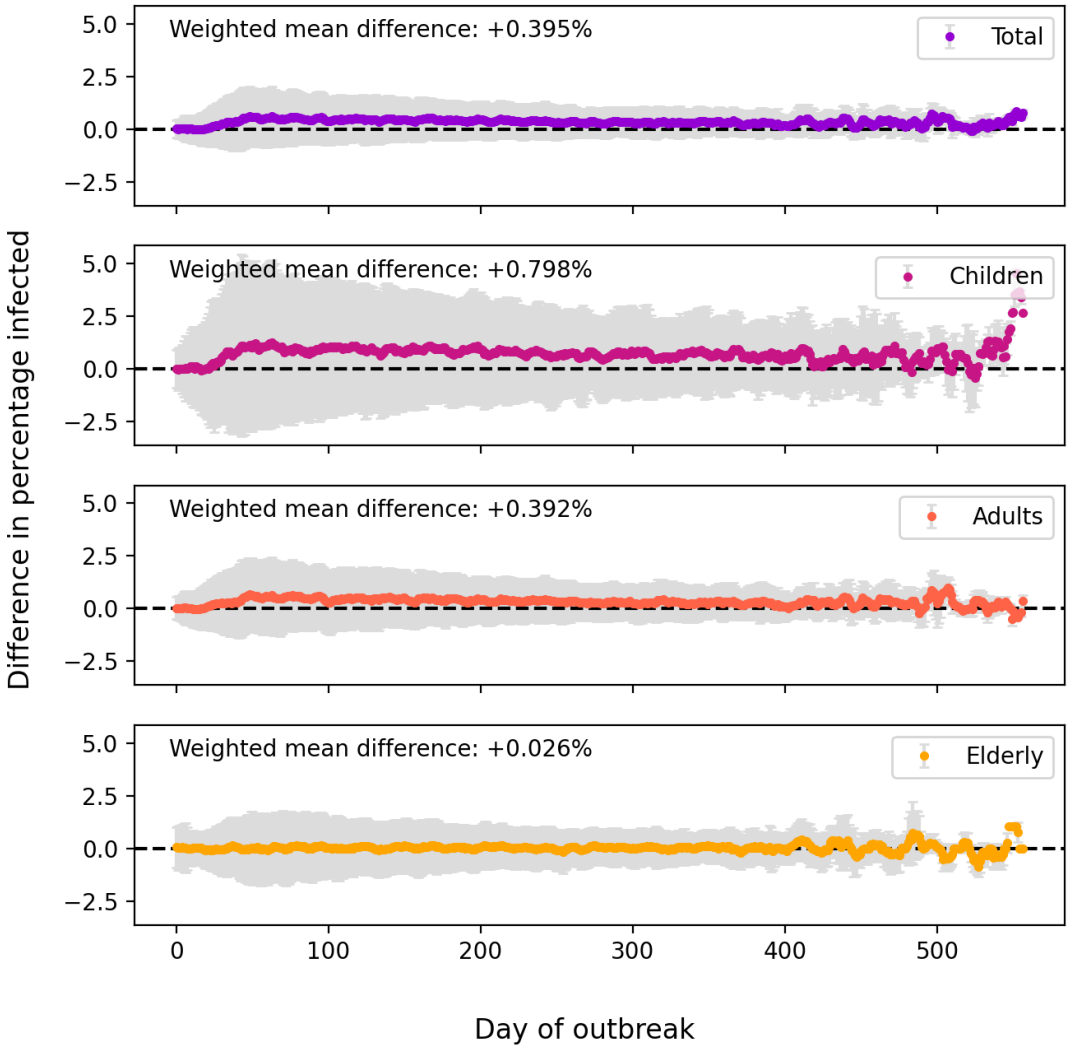


Figure 9. Difference in the percentage of the pro- and anti-vaccination subpopulations that were in the infected state, averaged over 500 outbreaks. A positive value represents a day where a higher proportion of the anti-vaccination population were infected than in the pro-vaccination population. Overall difference is plotted in purple, as well as a breakdown by age in pink, orange and yellow.

To collect this data, 500 outbreaks were simulated, each on a newly generated network. This maximised the variation in the results and ensured that no systematic bias was introduced by the unwitting use of a network with any peculiar randomly-generated features. The timescale combination was used where vaccination protection lasted significantly longer than post-infection immunity (90 days and 20 days respectively), as described in Section 4.7, to create a functional difference between those who were vaccinated and those whose immunity came from previous infection. The initial fraction of anti-vaccination nodes was also set to 50%, to provide the clearest information about the difference between the two populations (as their sizes were approximately equal).

At many points over the course of each outbreak, data was collected on the current number of pro- and anti-vaccination nodes and the current number of active cases within each subgroup. This data was used to create a snapshots of the percentages of each subgroup that were infected at a given time, and thus the difference in these percentages (where a positive difference signified that a higher proportion of anti-vaccination nodes were infected than pro-vaccination nodes). The time of the collection of each data point was then rounded to the nearest day, to create a discrete number of plot points with error bars showing the standard deviation across all outbreaks, as shown in Figure 9. Due to randomness, there was a wide range in the length of outbreaks, with a mean outbreak length of $289 \pm 99$ days. However, a few outbreaks persisted for over 400 days, providing data with increased uncertainty given the lack of corroborating data points (though the standard deviation error bars appear small, fluctuations in the mean caused by uncertainty can be visually observed in all plots in Figure 9).

Therefore, to obtain an overall measure of the difference between the two populations across the entirety of outbreaks, a weighted mean of all data points was performed. This ensured that the high uncertainty data points did not distort the final values.

It was found that the mean difference between the active case rates for the two subpopulations was +0.395% overall. For children, the difference was +0.798%. For adults, the difference was +0.392%. For the elderly, the difference was +0.026%.

## 5.3 Dynamics of individual outbreaks

In order to compare the results of the model to previous work in the field, the dynamics of individual outbreaks were also explored.

In their mathematical work, Papst *et al.* [7] describe a trend where the number of active cases increases at the start of an outbreak, leading to a surge in vaccinations. This in turn drives cases down, but when vaccination wears off, the number of cases surges again, and so on. If one were to plot active cases as a function of vaccinated nodes, this would create a periodic plot with a circular trend line. This part of the project aimed to determine whether this trend could be observed on a network, as suggested by Papst *et al.* as an avenue for further work.

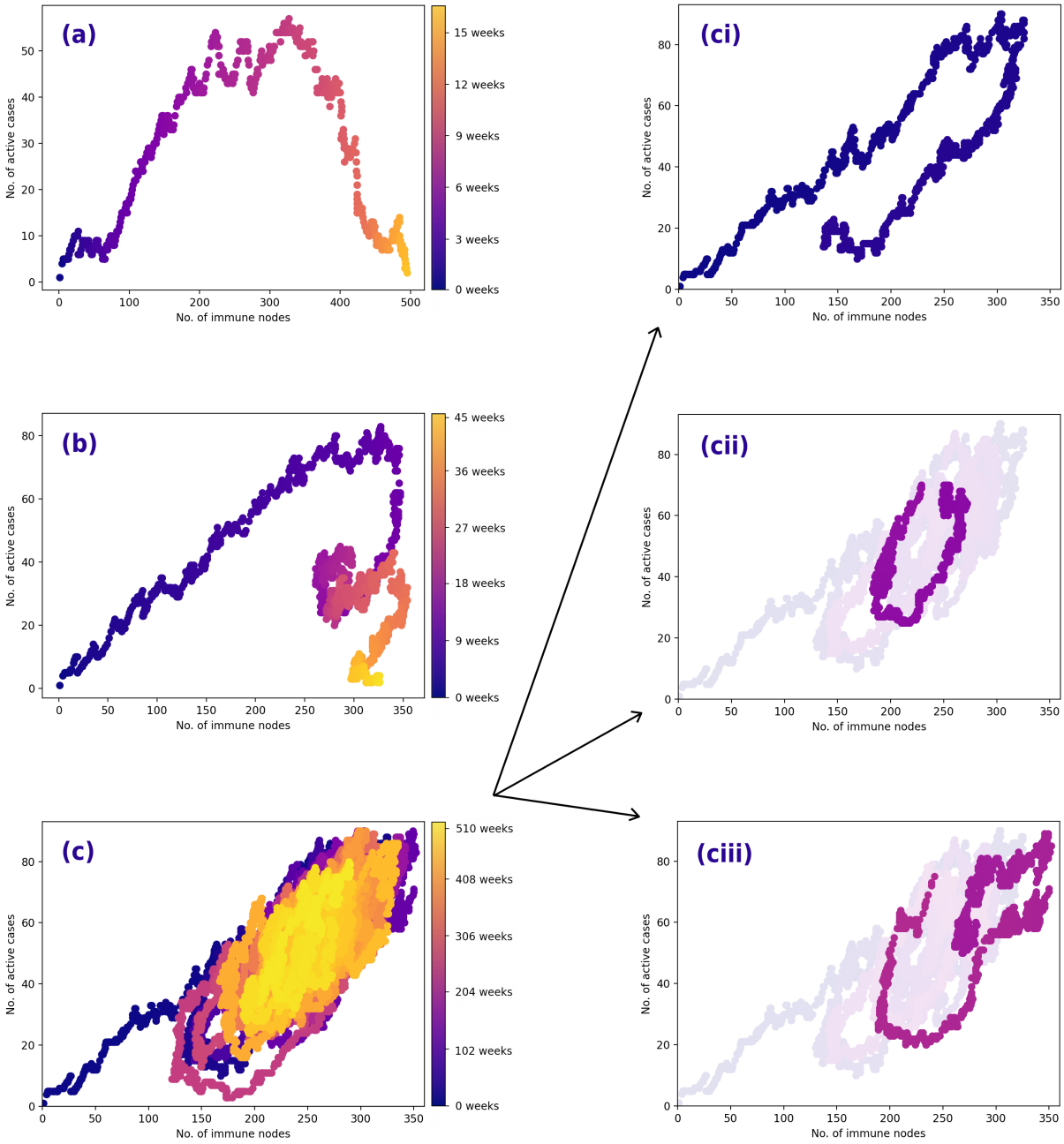


Figure 10. Various plots of the interplay between active cases and active immunity: (a) when both vaccination and post-infection immunity last for 90 days, (b) when post-infection immunity lasts for 20 days and vaccination lasts for 90 days (the "significant vaccination" regime), and (c) when both post-infection immunity and vaccination last 20 days (the consistent endemic regime). The plots on the right show specific dynamics within the endemic outbreak: (ci) showing the first 120 days, (cii) showing a "loop" between 565 and 620 days, and (ciii) showing a breakdown in the trend followed by recovery, between 680 and 790 days.

The aforementioned paper considered the spread of seasonal influenza, which functionally provides very little post-infection immunity due to the large number of new immunity-resistant variants that emerge over time. As a result, due to the substantial role that post-infection immunity played in the network simulation, it was necessary to plot the total number of immune nodes (including actively vaccinated and recovered nodes) as opposed to the total that were actively vaccinated.

Several plots were made using different timescale combinations for post-infection immunity and post-vaccination immunity, which can be seen in Figures 10(a)-(c).

Figure 10(a) shows the trend for an outbreak with post-infection immunity and post-vaccination immunity timescales of 90 days. This outbreak lasted approximately 16 weeks. The number of immune nodes rose throughout (due to the relatively long period of immunity granted by each case/vaccination), leading to an early extinction of the disease outbreak.

Figure 10(b) shows the trend for an outbreak with post-infection immunity of 20 days and post-vaccination immunity of 90 days. This outbreak lasted approximately 46 weeks. Both active cases and vaccinations rose steadily for the first $\sim$10 weeks of the outbreak. Active cases then fell rapidly, while the level of immunity fell slightly before rising again, causing a spiral trend until the outbreak died out.

Figure 10(c) shows the trend for an outbreak with post-infection immunity and post-vaccination immunity timescales of 20 days. This outbreak was endemic, and the first ten years (520 weeks) are plotted. This regime produced the clockwise cyclic trend described by Papst *et al.*, though caution should be taken when comparing the results directly to this work due to the inclusion of post-infection immunity in the $x$ plot.

The cyclic trend is dissected and examined in closer detail in Figures 10(ci)-(ciii). Figure 10(ci) shows the first 120 days of the outbreak, where the numbers of active cases and immune nodes both rise steadily, before both falling at similar rates. After this, the cyclic trend begins to appear: the trend line draws several elliptical tracks, generally decreasing in radius, and culminating in the cycle shown in 10(cii) after 620 days. After this, the system experiences a period of instability shown in 10(ciii), where the clockwise cycles stop and an anticlockwise loop is drawn, before returning to the clockwise trend. This breakdown in the usual trend leads to a larger ellipse shape, which "resets" the trend and allows the loops to begin to shrink once more.

This pattern is observed throughout the 10 years of the epidemic: the radius of the loops decreases slowly to a minimum, at which point a period of instability is triggered which generates another large loop.

As it was not possible to clearly define the start and end of these loops, it was not possible to obtain clear information about the duration of the loops, but the general trend showed that, as the radius of the loops decreased, so too did the time taken to complete a full cycle.

# 6 Discussion

When considering all results of this model, it should be reiterated that the parameters used are not representative of any one infectious disease. While the use of the POLYMOD contact rates makes the model most suitable for the study of human respiratory disease (as this is how physical contact was defined in the study), and parameters such as $R_0 = 1.3$ and the compartmental timescales may be similar to diseases such as COVID-19 and seasonal flu, they were not implemented based on data and should not be interpreted as conclusions about real diseases. However, the flexibility of the model would allow for specific parameter sets to be used in future work, in order to apply the simulation to real world diseases and investigate whether the conclusions of this work hold for known illnesses.

## 6.1 Effects of the anti-vaccination group on outbreaks

When considering outbreaks with parameters that allow for disease to become endemic, it was shown that an increase in the anti-vaccination population has two main effects: the average number of cases is increased (Figure 7), and the likelihood that the disease will become endemic is increased (Figure 8).

The exponential trend of the increase in cases over five years shows that any small increase in the size of the anti-vaccination population is more impactful when the existing population is already large. For instance, the trend shows that an increase from 10% holding anti-vaccination opinions to 20% causes an additional 379 cases over five years (a 3.48% increase), whereas increasing from 80% to 90% anti-vaccination opinions causes an increase of 578 cases (a 4.11% increase). To control endemic disease on a larger scale, it is therefore imperative to keep the overall number of anti-vaccination members of the population as low as possible, to protect against the knock-on effects of any large fluctuation in opinion, such as a prominent piece of misinformation being shared through the social network, or any other development that could make the population lose trust in the effectiveness of vaccines en masse.

The likelihood of endemic disease seems to linearly increase with the size of the anti-vaccination population, though the shape of the best fit is uncertain due to the low-quality data. As mentioned in Results, this result could easily be constrained further by additional simulation and more detailed results. However, the initial finding is a compelling statement on the role of anti-vaccination sentiment in prolonging outbreaks: compared to an entirely pro-vaccination population, an entirely anti-vaccination group are 11% more likely to experience endemic disease. For the parameter set used, this also means that an entirely anti-vaccination population has only a 3.6% chance of preventing the disease from becoming endemic. Naturally, this figure is likely to change if the parameters used are accurate for a real-world disease, but this outcome highlights how crucial the anti-vaccination population can be in keeping an outbreak under control and bringing it to a halt.

## 6.2 Effects of outbreaks on the anti-vaccination group

In Figure 9, it was shown that anti-vaccination nodes suffer from a higher infected percentage than their pro-vaccination counterparts throughout outbreaks. This effect was most pronounced for anti-vaccination children, who had a mean burden of +0.798%, followed by adults with a mean burden of +0.392%, and was least pronounced for the elderly, who experienced a mean burden of +0.026%.

As described in Results, several steps were taken to ensure that these results were as representative as possible of the overall outcomes of anti-vaccination nodes. It is acknowledged that this method is not ideal, but given the need for an average result of many repeated outbreaks, the results presented provide a good picture of how the risk level of anti-vaccination nodes changes over the course of the "average" outbreak.

After stratifying the data by age, an interesting trend emerged: a pronounced difference in risk for anti-vaccination children as opposed to adults and elderly people. While the increased risk for anti-vaccination elderly people was virtually negligible at +0.026%, the increase for anti-vaccination children was approximately 31 times higher at +0.798%. These figures approximately align with the proportions of physical contacts for the three groups: the risk proportions for the age groups are 31:15:1, while the physical contact proportions are 32:22:8. This would suggest that the case burden associated with being anti-vaccination is roughly proportional to the number of physical contacts a node has: a common-sense result.

However, there is a significant discrepancy for the elderly population, for whom it seems that an anti-vaccination opinion is inherently less of a risk factor than for the other groups. The reasons for this are unknown, but it may be that they gain additional protection from the lack of ring links in their subgroup, leading to a less effective spread of disease and a lesser effect of vaccination in blocking the path of infection. Further work is required to determine whether this trend holds when all three groups have the same topology within the network.

It should be noted, however, that this effect was relatively small for all subpopulations. Only 190 children were simulated in each outbreak, and approximately half of these were anti-vaccination at $t = 0$ days, for a total of 95 children. An increase of +0.798% therefore only equated to an extra $\sim 0.75 \simeq 1$ anti-vaccination child being infected at any given time. Given the high level of randomness in the simulation, this is a negligible result. However, due to the persistence of the positive trend across 500 outbreaks, it is still noteworthy. Further work might repeat this work using a larger network, to see if this emerges as a statistically significant result when there are more active cases and more nodes in the system.

## 6.3 Cyclic dynamics in sustained outbreaks

In Figure 10, various outcomes of interaction between active cases and immunities were presented. It was shown that, in sustained outbreaks, periodic behaviour similar to that predicted by Papst *et al.* can be observed.

One interesting dynamic highlighted by these graphs is the way in which vaccination dramatically prevents the number of active cases from recovery after a "first wave". As shown in Figure 10(b), the active cases begins to rise again, but the number of immune nodes has not fallen enough to allow the cyclic behaviour of 10(c) to persist. Further work could be undertaken to locate the parameter combinations that trigger phase transitions between the three types of behaviour shown in Figures 10(a), 10(b) and 10(c), as well as locating any other regimes that were not identified here.

As described, a recurring trend could be observed for endemic outbreaks: the trend line would draw clockwise loops of decreasing radius, until a breakdown in the pattern and a period of instability, followed by the generation of a new large loops from which the pattern could begin again. The behaviour predicted by Papst *et al.* was much more consistent and did not feature the periodic "reset" behaviour shown in these results.

The gradual decrease of the size of the loops demonstrates that the system is attempting to oscillate around some equilibrium point. However, due to the level of randomness in the simulation, once it gets too close to a static state, any small random change will considerably perturb the system, as shown in Figure 10(ciii). In this example, an unexpected increase in cases triggers a subsequent increase in the number of immune nodes, causing the trend to move in the opposite direction (anticlockwise) for a short time. Eventually, there is an opportunity for clockwise motion to resume, but the perturbation has been substantial enough to create a larger loop than the one that preceded it.

In summary, it is likely that the unforeseen patterns in these cycles are mostly dictated by the stochastic nature of the network simulation, which is not accounted for in the mathematical model. Other influences such as the heterogeneous network shape will likely affect the shape and dynamics of this trend, though the exact effects of this are unknown and could be investigated further by future work. It is also recommended that a more thorough analysis of this output is performed with a broader combination of timescales for the post-infection immunity and post-vaccination immunity.

# 7 Conclusion

Disease outbreaks are impacted by a myriad of human behaviours, which must be understood in order to accurately model epidemics and understand the spread of illness. In particular, vaccine uptake is a key factor in controlling outbreaks, yet this behaviour has remained largely unstudied on network simulations.

Previous network models have also faced many limitations due to oversimplified topology and unreasonable assumptions in their design. While some studies have incorporated age-based contact information such as the POLYMOD data into network shapes, this work has mostly focused on disease transmission without the additional dimension of opinion dynamics.

This project aimed to apply leading features of several areas of disease and behaviour simulation on networks to model the interplay of vaccination opinion and disease propagation.

A bespoke network shape was devised to tackle some of the main limitations of disease networks as highlighted by Pellis *et al.* [20], and custom modifications were made to the existing voter model to introduce age-based directionality and allow the spread of disease to interact with decision-making. Through these features, the project aimed to propose a more realistic method of modelling outbreak dynamics, and examine the results of this simulation.

It was found that the size of the anti-vaccination population has a significant impact on the outcomes for long-term outbreaks: the number of cases over a five-year period increased exponentially with an increase in anti-vaccination nodes, and the probability of disease becoming endemic in the system increased by up to 11%. This group also faced a systematically higher level of cases over the course of an outbreak, though the magnitude of this increase was not functionally significant due to the level of randomness in the simulation. The additional risk associated with the anti-vaccination stance was found to be approximately proportional to the number of physical contacts of each node.

The simulation was also able to recreate dynamics predicted theoretically by Papst *et al.* [7]: in a particular parameter space, the relationship between the number of active cases and the number of immune nodes oscillated over time. Due to its stochasticity, the model also generated complex additional features within oscillations, which merit further study.

Other than the recommendations made in previous sections for further investigation of the presented results, there are many avenues for improvement upon the model itself. One significant limiting factor in this model is that it is static: the edges of the network do not change over time, and no nodes are added or removed. A commonly integrated dynamic feature is to change the contact densities at weekends as opposed to during the week. Another interesting implementation would introduce birth (additions to the child ring), age progression (migrating nodes to "older" rings) and death (removing nodes from the elderly ring).

Another application for this model would be to focus on the investigation of a real-world disease, such as COVID-19 or seasonal influenza. Choosing an illness to focus on would allow other disease-specific factors to be implemented, such as the age-based waves of vaccination mentioned in Section 4.6, or additional compartments for the disease model, such as an exposed state (where nodes incubate the disease for a short time before becoming infectious). Choosing a focal disease would also allow data-driven parameter selection, as well as allowing the model to be tested and compared to our existing knowledge of real-world disease outbreaks.

To summarise, this project has successfully proposed and constructed a novel network shape for the study of disease outbreaks, as well as combining previous methods of behaviour modelling to create a more realistic model of opinion transmission. Results from the simulation have shown that anti-vaccination sentiment can have a profound impact on the outcomes of disease outbreaks, and have corroborated previous theoretical work on the interplay between disease and vaccination uptake. While further work is recommended to constrain these results and improve the simulation as a whole, this project has provided an interesting avenue for the study of disease-behaviour interplay on networks.

# References


[1] G. Iacobucci. UK was advised to stockpile PPE and screen travellers in 2016 after coronavirus modelling. *British Medical Journal*, 2021.

[2] K. Rice et al. Effect of school closures on mortality from coronavirus disease 2019: old and new predictions. *British Medical Journal*, 371, 2020.

[3] H. Ritchie et al. Policy Responses to the Coronavirus Pandemic. *Our World in Data*, 2020.

[4] O. Mumcuoglu et al. Factbox: Countries making COVID-19 vaccines mandatory. *Reuters*, 2021.

[5] M. Mills et al. COVID-19 vaccine deployment: Behaviour, ethics, misinformation and policy strategies. *The Royal Society & The British Academy*, 2020.

[6] G. Dash et al. Effects of Vaccination Decisions and Peer influence on Epidemic Dynamics: A Network Perspective. In *2021 IEEE 18th India Council International Conference (INDICON)*, pages 1–6, 2021.

[7] I. Papst et al. Modeling the interplay between seasonal flu outcomes and individual vaccination decisions. *Bulletin of Mathematical Biology*, 84(3):1–17, 2022.

[8] M. Voinson et al. Beyond rational decision-making: modelling the influence of cognitive biases on the dynamics of vaccination coverage. *PloS ONE*, 10(11):e0142990, 2015.

[9] R. Prieto Curiel and H. González Ramírez. Vaccination strategies against COVID-19 and the diffusion of anti-vaccination views. *Scientific Reports*, 11(1):1–13, 2021.

[10] I. Z. Kiss et al. The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. *Journal of the Royal Society Interface*, 5(24):791–799, 2008.

[11] C. E. Walters et al. Modelling infectious disease transmission potential as a function of human behaviour. *medRxiv pre-print*, 2021.

[12] L. Burbach et al. Opinion Formation on the Internet: The Influence of Personality, Network Structure, and Content on Sharing Messages Online. *Frontiers in Artificial Intelligence*, 3:45, 2020.

[13] D. S. Bassett and E. D. Bullmore. Small-world brain networks. *The Neuroscientist*, 12(6):512–523, 2006.

[14] Office for National Statistics. Overview of the UK population: January 2021, 2021. Date retrieved: 28 January 2022.

[15] J. Mossong et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine*, 5(3):e74, 2008.

[16] C. Jarvis et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Medicine*, 18(1):1–10, 2020.

[17] S. H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[18] S. Milgram. The small world problem. *Psychology Today*, 2(1):60–67, 1967.

[19] X. Liao et al. Small-world human brain networks: perspectives and challenges. *Neuroscience & Biobehavioral Reviews*, 77:286–300, 2017.

[20] L. Pellis et al. Eight challenges for network epidemic models. *Epidemics*, 10:58–62, 2015.

[21] S. Meyer and L. Held. Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18(2):338–351, 2017.

[22] M. Eichner et al. 4Flu - an individual based simulation tool to study the effects of quadrivalent vaccination on seasonal influenza in Germany. *BMC Infectious Diseases*, 14(1):1–21, 2014.

[23] A. D. C. Williams et al. An individual-based simulation of pneumonic plague transmission following an outbreak and the significance of intervention compliance. *Epidemics*, 3(2):95–102, 2011.

[24] S. Edlund et al. Comparing three basic models for seasonal influenza. *Epidemics*, 3(3-4):135–142, 2011.

[25] F. Brauer et al. *Mathematical Epidemiology*, volume 1945. Springer, 2008.

[26] J. H. Jones. Notes on R0. *Califonia: Department of Anthropological Sciences*, 323:1–19, 2007.

[27] K. M. Gostic et al. Practical considerations for measuring the effective reproductive number, R t. *PLoS computational biology*, 16(12):e1008409, 2020.

[28] P. Gastanaduy et al. Pinkbook: Measles. *Centers for Disease Control and Prevention*, 2019. Date retrieved: 15 March 2022.

[29] Centers for Disease Control and Prevention. Vaccination Strategies — Smallpox, 2019. Date retrieved: 15 March 2022.

[30] UK Health Security Agency. COVID-19: the Green Book, chapter 14a, 2022. Date retrieved: 3 March 2022.

[31] A. Vazquez. Spreading dynamics on small-world networks with connectivity fluctuations and correlations. *Physical Review E*, 74(5):056101, 2006.

[32] J. Ma et al. The importance of contact network topology for the success of vaccination strategies. *Journal of Theoretical Biology*, 325:12–21, 2013.

[33] S. Bansal et al. A comparative analysis of influenza vaccination programs. *PLoS Medicine*, 3(10):e387, 2006.

[34] R. Das et al. Relationship between Nash Equilibria and Pareto Optimal Solutions for Games of Pure Coordination. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–7, 2019.

[35] J. R. Meszaros et al. Cognitive processes and the decisions of some parents to forego pertussis vaccination for their children. *Journal of Clinical Epidemiology*, 49(6):697–703, 1996.

[36] M. Perc et al. Statistical physics of human cooperation. *Physics Reports*, 687:1–51, 2017.

[37] A. Sîrbu et al. Opinion dynamics: models, extensions and external effects. In *Participatory sensing, opinions and collective awareness*, pages 363–401. Springer, 2017.

[38] I. Dornic et al. Critical coarsening without surface tension: The universality class of the voter model. *Physical Review Letters*, 87(4):045701, 2001.

[39] A. Bish et al. Factors associated with uptake of vaccination against pandemic influenza: a systematic review. *Vaccine*, 29(38):6472–6484, 2011.

[40] Office for National Statistics. Coronavirus vaccine hesitancy falling across the regions and countries of Great Britain, 2021. Date retrieved: 28 January 2022.

# Appendices

*Note* — Appendices are provided for completeness only and any content included in them will be disregarded for the purposes of assessment.

# A POLYMOD contact matrices

## A.1 All reported contacts

| age of contact | age group of participant | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 00-04 | 05-09 | 10-14 | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70+ |
| 00-04 | 1.92 | 0.65 | 0.41 | 0.24 | 0.46 | 0.73 | 0.67 | 0.83 | 0.24 | 0.22 | 0.36 | 0.20 | 0.20 | 0.26 | 0.13 |
| 05-09 | 0.95 | 6.64 | 1.09 | 0.73 | 0.61 | 0.75 | 0.95 | 1.39 | 0.90 | 0.16 | 0.30 | 0.22 | 0.50 | 0.48 | 0.20 |
| 10-14 | 0.48 | 1.31 | 6.85 | 1.52 | 0.27 | 0.31 | 0.48 | 0.76 | 1.00 | 0.69 | 0.32 | 0.44 | 0.27 | 0.41 | 0.33 |
| 15-19 | 0.33 | 0.34 | 1.03 | 6.71 | 1.58 | 0.73 | 0.42 | 0.56 | 0.85 | 1.16 | 0.70 | 0.30 | 0.20 | 0.48 | 0.63 |
| 20-24 | 0.45 | 0.30 | 0.22 | 0.93 | 2.59 | 1.49 | 0.75 | 0.63 | 0.77 | 0.87 | 0.88 | 0.61 | 0.53 | 0.37 | 0.33 |
| 25-29 | 0.79 | 0.66 | 0.44 | 0.74 | 1.29 | 1.83 | 0.97 | 0.71 | 0.74 | 0.85 | 0.88 | 0.87 | 0.67 | 0.74 | 0.33 |
| 30-34 | 0.97 | 1.07 | 0.62 | 0.50 | 0.88 | 1.19 | 1.67 | 0.89 | 1.02 | 0.91 | 0.92 | 0.61 | 0.76 | 0.63 | 0.27 |
| 35-39 | 1.02 | 0.98 | 1.26 | 1.09 | 0.76 | 0.95 | 1.53 | 1.50 | 1.32 | 1.09 | 0.83 | 0.69 | 1.02 | 0.96 | 0.20 |
| 40-44 | 0.55 | 1.00 | 1.14 | 0.94 | 0.73 | 0.88 | 0.82 | 1.23 | 1.35 | 1.27 | 0.89 | 0.67 | 0.94 | 0.81 | 0.80 |
| 45-49 | 0.29 | 0.54 | 0.57 | 0.77 | 0.97 | 0.93 | 0.57 | 0.80 | 1.32 | 1.87 | 0.61 | 0.80 | 0.61 | 0.59 | 0.57 |
| 50-54 | 0.33 | 0.38 | 0.40 | 0.41 | 0.44 | 0.85 | 0.60 | 0.61 | 0.71 | 0.95 | 0.74 | 1.06 | 0.59 | 0.56 | 0.57 |
| 55-59 | 0.31 | 0.21 | 0.25 | 0.33 | 0.39 | 0.53 | 0.68 | 0.53 | 0.55 | 0.51 | 0.82 | 1.17 | 0.85 | 0.85 | 0.33 |
| 60-64 | 0.26 | 0.25 | 0.19 | 0.24 | 0.19 | 0.34 | 0.40 | 0.39 | 0.47 | 0.55 | 0.41 | 0.78 | 0.65 | 0.85 | 0.57 |
| 65-69 | 0.09 | 0.11 | 0.12 | 0.20 | 0.19 | 0.22 | 0.13 | 0.30 | 0.23 | 0.13 | 0.21 | 0.28 | 0.36 | 0.70 | 0.60 |
| 70+ | 0.14 | 0.15 | 0.21 | 0.10 | 0.24 | 0.17 | 0.15 | 0.41 | 0.50 | 0.71 | 0.53 | 0.76 | 0.47 | 0.74 | 1.47 |

Figure 11. United Kingdom contact matrix for all reported contacts, collected in 2007 for the POLYMOD study [15]

## A.2 Physical contacts only

| age of contact | age group of participant | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 00-04 | 05-09 | 10-14 | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 | 65-69 | 70+ |
| 00-04 | 1.49 | 0.59 | 0.25 | 0.18 | 0.42 | 0.61 | 0.57 | 0.74 | 0.18 | 0.20 | 0.36 | 0.15 | 0.18 | 0.26 | 0.07 |
| 05-09 | 0.74 | 3.82 | 0.53 | 0.44 | 0.37 | 0.42 | 0.68 | 0.99 | 0.66 | 0.15 | 0.17 | 0.19 | 0.41 | 0.30 | 0.07 |
| 10-14 | 0.36 | 0.73 | 3.19 | 0.79 | 0.17 | 0.14 | 0.32 | 0.51 | 0.69 | 0.27 | 0.20 | 0.19 | 0.17 | 0.26 | 0.17 |
| 15-19 | 0.26 | 0.22 | 0.52 | 3.10 | 0.85 | 0.12 | 0.17 | 0.29 | 0.48 | 0.51 | 0.27 | 0.13 | 0.09 | 0.26 | 0.37 |
| 20-24 | 0.39 | 0.19 | 0.09 | 0.38 | 1.32 | 0.49 | 0.28 | 0.16 | 0.23 | 0.44 | 0.20 | 0.28 | 0.11 | 0.15 | 0.00 |
| 25-29 | 0.53 | 0.37 | 0.11 | 0.20 | 0.58 | 0.64 | 0.35 | 0.21 | 0.23 | 0.25 | 0.30 | 0.30 | 0.20 | 0.04 | 0.03 |
| 30-34 | 0.77 | 0.72 | 0.29 | 0.17 | 0.31 | 0.42 | 0.80 | 0.21 | 0.27 | 0.38 | 0.23 | 0.20 | 0.23 | 0.15 | 0.07 |
| 35-39 | 0.73 | 0.65 | 0.61 | 0.35 | 0.20 | 0.24 | 0.47 | 0.76 | 0.47 | 0.29 | 0.18 | 0.11 | 0.32 | 0.04 | 0.03 |
| 40-44 | 0.38 | 0.59 | 0.58 | 0.47 | 0.27 | 0.29 | 0.25 | 0.49 | 0.45 | 0.40 | 0.12 | 0.19 | 0.23 | 0.15 | 0.30 |
| 45-49 | 0.22 | 0.23 | 0.22 | 0.34 | 0.46 | 0.29 | 0.17 | 0.17 | 0.35 | 0.58 | 0.15 | 0.19 | 0.14 | 0.11 | 0.20 |
| 50-54 | 0.26 | 0.21 | 0.14 | 0.17 | 0.22 | 0.17 | 0.15 | 0.14 | 0.19 | 0.27 | 0.33 | 0.35 | 0.20 | 0.04 | 0.07 |
| 55-59 | 0.22 | 0.12 | 0.10 | 0.10 | 0.14 | 0.22 | 0.13 | 0.21 | 0.05 | 0.25 | 0.17 | 0.52 | 0.32 | 0.44 | 0.10 |
| 60-64 | 0.22 | 0.15 | 0.10 | 0.10 | 0.10 | 0.14 | 0.15 | 0.20 | 0.19 | 0.13 | 0.09 | 0.24 | 0.27 | 0.30 | 0.13 |
| 65-69 | 0.05 | 0.08 | 0.07 | 0.08 | 0.07 | 0.12 | 0.08 | 0.16 | 0.11 | 0.04 | 0.05 | 0.07 | 0.17 | 0.37 | 0.43 |
| 70+ | 0.09 | 0.09 | 0.10 | 0.08 | 0.19 | 0.10 | 0.05 | 0.13 | 0.24 | 0.29 | 0.23 | 0.22 | 0.11 | 0.22 | 0.70 |

Figure 12. United Kingdom contact matrix for physical contacts, collected in 2007 for the POLYMOD study [15]