Investigating the relationship between network properties and the spread of disease

Yujie Dai

200010781

Supervisor: Professor Simon Dobson

University of St Andrews

16 August 2022

Abstract

Disease modelling has taken on increasing importance in the recent outbreak of COVID-19. In order to simulate the propagation process of a disease among a population, networks of contacts are inseparable from it, combining both to form the network epidemiology.

Biology theory suggests that increasing connectivity of the human population tends to select pathogens with higher infectivity. Inspired by this, we aim to investigate how the size and connectivity of networks affect the transmission process of diseases. We create the disease models based on the compartmented SIR (Susceptible-Infected-Removed) model and construct two types of networks - the modular network and the core-periphery network. For each network, it has a self-defined metric to describe its connectivity. In our experiments, we simulate diseases spreading over both networks, and observe the simulation process when changing disease infectivity, network sizes, and network connectivity. The results show that network properties have impacts on the spread of diseases. A disease can infect more people in the larger and more connected networks, and diseases with higher infectivity behave similarly tendency.

Acknowledgement

First of all, I would like to thank my supervisor, Professor Simon Dobson, who patiently provides me with guidance and advice during the project. I am appreciated having this project at the end of my master's period.

Secondly, I would like to thank my parents and friends, who always give me spiritual encouragement and make my time being St Andrews wonderful and impressive.

Declaration of authorship

I declare that the material submitted for assessment is my own work except where credit is explicitly given to others by citation or acknowledgement. This work was performed during the current academic year except where otherwise stated.

The main text of this project report is 10, 238 words long. In submitting this project report to the University of St Andrews, I give permission for it to be made available for use in accordance with the regulations of the University Library. I also give permission for the title and abstract to be published and for copies of the report to be made and supplied at cost to any bona fide library or research worker, and to be made available on the World Wide Web. I retain the copyright in this work.

16th August 2022

Yujie Dai

Contents

1 Introduction

The importance of disease modelling is highlighted by the fact that coronavirus disease (COVID-19) causes pandemics in many countries around the world after 2019. Disease modelling can not only help us to prevent outbreaks before a pandemic occurs, for example, people vaccinate against a disease [1], but also inhibit transmission and spread after an outbreak has occurred, such as by wearing masks and controlling social distances [2, 3]. Diseases can be transmitted through a variety of channels, such as respiratory or sexual transmission, and the spread of such diseases in the population is usually due to interactions between susceptible and infected individuals. Thus, population networks of contacts are closely linked to the simulation of disease propagation. Nodes in a network represent individuals, and edges between nodes represent interactions between individuals. Many traditional theories of disease development ignore the influence of networks [4, 5], but recently more and more research has begun to understand their importance [6, 7, 8, 9, 10, 11], forming network epidemiology. Besides, this theory has been applied to other fields, like social sciences, where researchers study the spread of culture and knowledge [12].

One of the most popular and fundamental disease models is the susceptible-infected-removed (SIR) model. It classifies individuals into three groups, susceptible, infected, and removed groups. Susceptible individuals capture the virus from infected individuals and become a member of the infected group. Infected individuals move into the removed group if they recover or die. SIR model assumes that there is no secondary infection of the disease, which means that once an infected person has recovered, he/she will not be infected again. There are other derived models that change the assumptions for addressing different types of diseases, such as SIS(Susceptible-Infected-Susceptible), SIRS (Susceptible-Infected- Removed -Susceptible) [6], and SEIR (Susceptible-Exposed-Infected-Removed) [13]. In our project, we model diseases according to the SIR mode.

In recent years, many studies have studied network epidemiology based on SIR models. Some studies apply it to study various diseases. Dottori et al. study whooping

cough before the vaccine era with an SIR model on dynamical networks [14], and Kathakali et al. study Covid-19 with an SIR model on a Euclidean network [15]. There are some other studies focusing on the details of the process of general disease transmission. For example, researchers study the infection delay during disease spreading on complex networks [16, 17]. Different from these studies, in this project, we focus on the relationship between the properties of networks and the disease spread processes. The network properties here include the size and connectivity of networks.

Inspired by the research of Michael et al. who propose the biological theory that the increasing connectivity of the human population tends to select viruses with higher infectivity [18, 19], we create two different networks and change their connectivity and sizes to observe how different diseases propagate on them over time. The first network is a modular network, including a centre big cluster surrounded by several small clusters. It can be used to simulate a real-life town population, where a big town locates centrally, and several villages surround it. A metric $P_{modularity}$ is utilized to describe the connectivity between surrounding clusters and the central cluster in a modular network. The second network is a core-periphery network to simulate the central cluster. It includes a dense core cluster and a sparse periphery cluster to mimic the real-life population distribution. People who live in the city centre can meet more people than those who live in the suburbs. In core-periphery networks, a metric $P_{connectivity}$ is used to represent the connectivity between the core cluster and the periphery cluster.

To compare how diseases transmit on networks of different sizes, we create large-scale networks and ten times smaller networks. We also create diseases with various properties to discover how they have distinct transmission processes from each other.

Generally, our works includes:

(1) We create two kinds of networks, a modular network, and a core-periphery network, each of which has its own connectivity metric.

(2) We build an SIR model based on these two networks and simulate how different diseases spread on them.

(3) We simulate how diseases spread over these networks of different sizes.

(4) We simulate how diseases spread over these networks of different connectivity.

(5) We simulate how disease with different infectivity epidemic size in a population.

The outline of this paper is introduced as follows. In Section 2, we introduce the background of this project and show the related works that have been done by other researchers. The design and implementation are described in Section 3, while experiment results are discussed in Section 4. Finally, in Section 5, we draw our conclusions and point out our future works.

2 Context Survey

In this chapter, we illustrate the motivation for our research and some background knowledge in this area. Besides, we also introduce some relevant research and software packages we used in this section.

2.1 Motivation

In biology, diseases have their own properties, such as infectivity and transmissibility. The traditional theory holds that diseases have their own patterns of development. However, disease evolvement is not independent and can be influenced by external conditions. There are biological theories that suggest that a population with higher connectivity preserves those viruses that possess higher infectivity. Michael et al. use a lattice model and find lower virulence of diseases in more local infections [19]. Several years later, he did an experiment in a host-pathogen model system, and the results show that the infectivity of the virus decreases in a more viscous population [18]. We assume a similar situation may exist for the spread of disease in human societies. The higher connectivity between human society affects the evolvement of viruses. Therefore, inspired by his studies, we aim at finding out how the properties of networks affect disease transmission. The properties here include the size and the connectivity of networks. Based on the SIR models, we simulate different diseases spreading over two kinds of networks, and then change the connectivity and sizes of two networks to discover how they affect the propagation processes of diseases.

2.2 Background

Our project simulates the spread of diseases with different reproduction numbers $\mathcal{R}$ over two types of networks based on the SIR model. We will introduce some background knowledge about disease modelling and networks in this section.

2.2.1 Disease Modelling

Diseases are caused by pathogens. Each disease has its own specific pathogen, such as a certain virus or bacteria. The infection of disease is due to the presence of its pathogen in the body of the infected person. As each disease is transmitted in a different way, there are various channels of infection. Some diseases are airborne, where a susceptible person inhales the pathogen and becomes infected with the disease. Some diseases are transmitted through direct physical contact with an infected person, causing a susceptible individual infected. To help control the spread of disease and reduce the harm caused by pathogens on large populations, epidemiology science provides a basis for analysis and prediction by modelling the spread of disease. Next, we will present several necessary things that are important for building disease models.

Reproduction Number $\mathcal{R}$

Some essential elements of disease are needed to build a disease spread model. One of the most important elements is the reproduction number $\mathcal{R}$, which refers to the average number of secondary infections that arise from every new infection [9]. It describes how quickly a disease can transmit over a population. Higher R indicates a faster spread of disease. If $\mathcal{R} > 1$, the disease will grow dramatically and infect more and more people. If $\mathcal{R} < 1$, the disease will finally die out. If $\mathcal{R} = 1$, the disease will keep its originally infected number without changing. The value of $\mathcal{R}$ is not always constant during an epidemic. It may decrease as the body develops immunity or countermeasures taken by humans.

$\mathcal{R}$ not only provides a means of determining the rate of spread of the disease, but it can also be used as a reference for taking control measures [20]. Researchers propose

many ways to formulate $\mathcal{R}$ [20, 21]. In our project, we calculate $\mathcal{R}$ by the ratio of infectivity and remove probability. The details will be introduced in Section 3.

SIR Model

In order to construct a disease model, in addition to the reproduction number $\mathcal{R}$ that characterises a disease, we also need to show the transmission process, such as how many people are not infected yet, how many people have been infected, and how many have recovered. Showing the dynamic process of the disease can tell us how virulent the disease is and how wide the disease is likely to spread in the population. One of the most popular models of disease is the SIR model [22], which is a compartmented model by grouping people into one of three compartments at any time. In epidemiology theory, individuals are usually grouped into three main types - people who are susceptible to, infected by, and recovered from a specific disease. Related to this, SIR uses three compartments to represent the disease dynamics - susceptible, infected, and removed, forming its name SIR(Susceptible-Infected-Removed). Susceptible represents the individuals who can be infected by the disease. Infected refers to the individuals who have already captured the pathogen and can pass it to susceptible individuals. Removed denotes the individuals who have recovered from an infection or died due to infection. People in a certain group are not always static, they can move between compartments through some processes: (1) People move from susceptible to infected when they are infected; (2) People move from infected to removed when they recover or die. In SIR, it assumes that there is no secondary infection. There are some other models allow that people move from infected to susceptible again, which means they can be reinfected again even though they are recovered, like the SIRS model and SIS model [22]. We use several graphs to help compare their differences clearly, as shown in Figure 1.

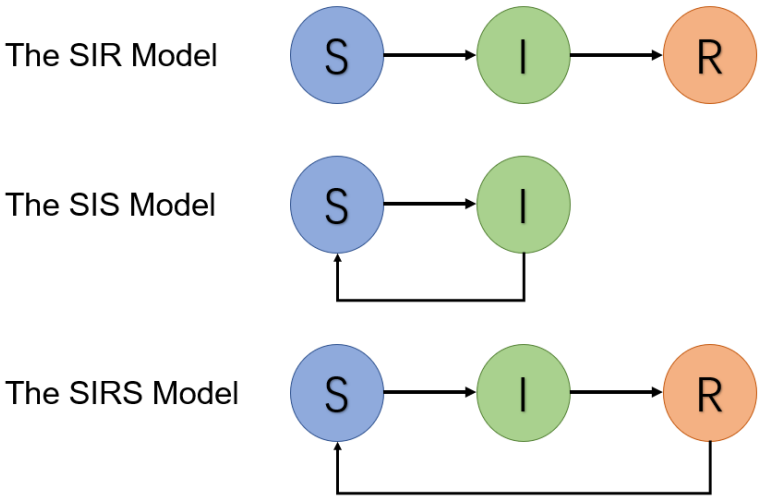

Figure 1 Different disease models

In our project, we focus on the most classic SIR model. In the SIR mode, consider there are N people in a population, in which S people are susceptible, I people are infected, and R people are in the removed group. It always follows the equation below

$$S + I + R = N$$

However, the values of S, I, and R are not static. They change with time. As time passes, susceptible individuals may get infected because of interactions with infected individuals, so those people move from compartment S to compartment I. Obviously, not all interactions between susceptible and infected individuals cause infections. Therefore, we use P_{infect} to describe the probability that an infected person will infect a susceptible person. Assuming that each person in the population has the same chance of meeting the others, there will be $S \times I = SI$ interactions between susceptible and infected individuals. Since there is a P_{infect} probability that interaction causes an infection, the number of susceptible people who may get infected after encountering an infected one is $SI \times P_{infect}$. This means as time passes, the number of individuals in the susceptible compartment will reduce by $SI \times P_{infect}$, and the number of people in the infected compartment will increase by the same number. But the number I not always increase. This is because infected individuals may recover from the disease or die at the end, which means people in the I compartment can move to the R compartment. We use P_{remove} to describe the probability that a person will recover

from the disease or die. Therefore, the value of I can reduce by $I \times P_{remove}$, while the value of R will increase for the same number. In summary, the number of people in the S, I, and R groups are not constant. Their changes can be represented as

$$\Delta S = -S \times I \times P_{infect} = -SIP_{infect}$$
$$\Delta I = (S \times I \times P_{infect}) - (R \times P_{remove}) = SIP_{infect} - RP_{remove}$$
$$\Delta R = I \times P_{remove} = IP_{remove}$$

P_{remove} and P_{infect} here are parameters of the SIR mode. Another parameter in this model is $P_{infected}$, which describes a fraction that the people get infected initially in the population.

Epidemic threshold

Another term related to disease modelling is the epidemic threshold. The epidemic threshold is used to describe whether a disease can become an epidemic or not. From the epidemic threshold point, more and more people will be infected and finally reaches an epidemic behaviour. It is a boundary to separate parameters that lead to an epidemic from those that do not [23]. If the parameters are above the threshold value, an epidemic occurs. Otherwise, diseases die out. The reproduction number $\mathcal{R}$ is an important value as a parameter of the epidemic threshold [24]. As described above, an epidemic occurs if $\mathcal{R}$ is greater than 1. Otherwise, no epidemic appears.

2.2.2 Networks

Diseases spread over populations. People capture a disease from an infected individual after interacting with him/her. However, we cannot meet everyone in society. A person living in Scotland has a small chance of being infected by someone living in the USA if they have not interacted with or met each other before. Furthermore, the number of people each person meets daily varies; some people meet more people because they have more friends, or because their work requires them to do so, such as a waiter or deliverer. Therefore, to model the spread of diseases, we should consider the population

structure. Networks are suitable tools to address this.

A network consists of a number of nodes connected through edges. It is a powerful tool that can describe patterns and interactions within connections. They have been applied in many fields, such as physical, biological, and social relationship sciences [22, 25]. A good example of a network is the social network, where nodes represent individuals and edges represent social relationships or contacts between them. It can be applied in the online environment, like friends on Facebook, or in real-life encounters, like meeting your families or friends. In epidemiology, network science also plays an important role. Nodes in a network are individuals, and edges are social contact that may carry the disease pathogen. We simulate diseases spreading over networks and analyse the process.

Topology

Networks have various structures. The structure of a network is known as topology. There are various types of topologies for networks. Some simple configurations include ring network topology or tree network topology, shown in Figure 2. More complex ones include modular networks and core-periphery networks, which we will introduce later in Section 3.

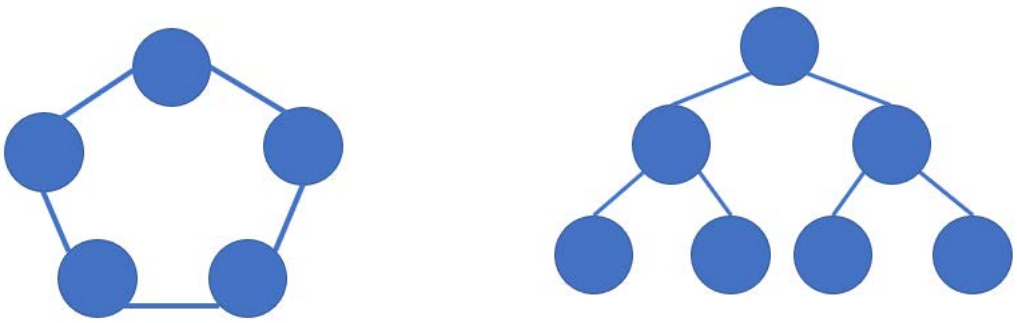

Figure 2 Examples of a ring network (left) and a tree network (right)

Degree

In addition to determining the topology of the network, some other metrics can be used to help us to describe the properties within a network. Some networks have more edges, which means that the network is denser, while some networks have fewer edges, which means that the network is sparser. We use the degree of a node to describe the number of edges intersecting it. The degree of a node in a social network means the number of people you are connected to. In a network, each node has its own degree. To describe the density of a network, we can just sum up the degree of all nodes easily to get a number. However, this way leads to memory waste in a large network and is not realistic and useful in a complex network. Therefore, the mean degree is used to address the problem, which means the average degree for each node in a network. Denote the degree of node i as k_i, and the number of nodes as N. The mean degree $\bar{k}$ can be defined as

$$\bar{k} = \frac{\sum_N k_i}{N}$$

Some studies also use degree distribution to describe networks [9]. It shows the number of nodes with each degree in a network. If we use N_k to represent the number of nodes with degree k, we can define $\bar{k}$ in a different way as

$$\bar{k} = \frac{\sum_k N_k}{N}$$

One of the targets of our project is to observe the relationship between network connectivity and disease transmission. The network connectivity is essentially based on the degree of nodes, and we will introduce the concept of connectivity in Section 3.

Random Networks

One of the most famous network topologies is the Erdős-Rényi(ER) network, which is proposed in the late 1950s [26]. The idea of an ER network is that there exists an edge between any pair of nodes with a certain probability and the existence of any edge is independent of other edges. According to this, two parameters are needed to construct

an ER network: (1) the number of nodes, denoted as N; (2) the probability of having an edge between each pair of nodes in the network p_{edge}. We show four examples of ER networks constructed using python, as shown in Figure 3. As p_{edge} increases, higher probability stands for the existence of an edge, so the network becomes denser. In this project, we build two networks with different topologies based on ER network.

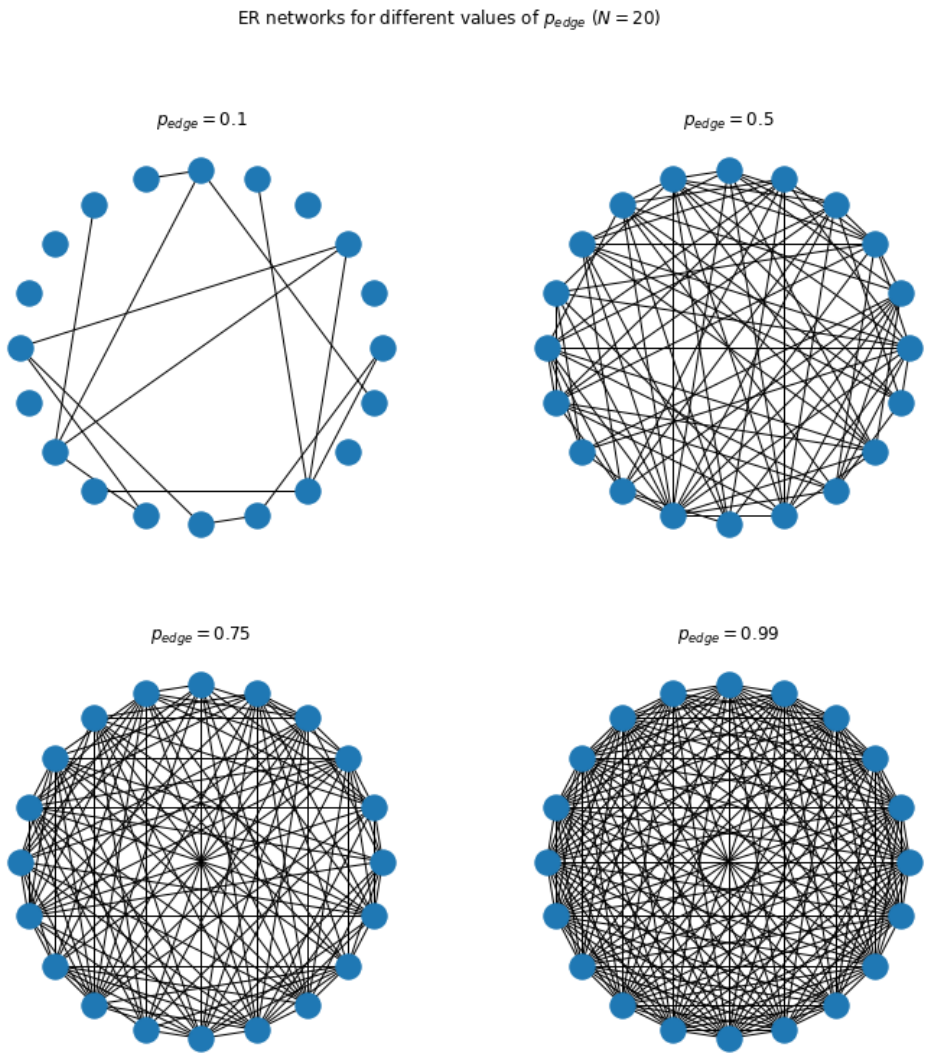

Figure 3 Examples of ER networks

2.3 Related Work

There are lots of researchers studied the spreading of diseases over networks. In 2002, Newman and his research group correlate the infectiveness time and probabilities together and set them dynamically to simulate a realistic case in human life [7]. Simon Dobson and his research team focus on the SIS model on coupled adaptive networks. They rewire edges to other uninfected nodes in the same cluster to simulate a situation where susceptible individuals avoid meeting infected people [10].

In real life, there can be more than one disease existing in populations. Some researchers focus on this and show their findings. Newman studied how two competitive diseases spread on the same network [8]. Simon Dobson studied how clustering affects the spread of sequential strains of a pathogen [11].

There are other studies that describe epidemiology in small-world networks. For example, Newman studies the disease spreading with various probabilities of transmission and infection on small-world networks and provides how to calculate the threshold in different cases [27].

A number of studies have broadly described the impact of contact networks on disease dynamics [28, 29]. In our project, we mainly focus on the effects of sizes and connectivity of networks on the spread of diseases. There are studies that have similar aims. Lopes et al. observe that the behavioural changes that cause connectivity varies in social contacts of mice can affect the process of disease spread [30]. But they use biology experiments to support their research. Our project uses computer science technologies instead, to simulate networks of various sizes and connectivity and disease spread over them.

Laurent uses smeared phase transitions in percolation on complex networks [31]. Two toy networks are used in his experiments, including modular networks and core-periphery networks. Inspired by his study, we decide to use these two kinds of networks as simulations of real-life population distribution. Differently, we add more metrics to describe the properties of networks. A self-defined modularity in modular networks to describe the connections between clusters and connectivity in core-periphery networks

to describe the connections between core and periphery clusters.

Researchers also studied strategies to prevent large infections. Taking the COVID-19 as an example, some studies show the importance of vaccination for controlling its spread [1], and some studies illustrate the effects of restricting social distances to Covid-19 [2, 3]. There are studies that build the COVID-19 model for researchers to analyse. For example, Taylor discusses the impact of the well-known Imperial College model of covid-19 on UK NHS bed availability [32].

2.4 Software Packages

2.4.1 Epydemic

Epydemic is a python library that focuses on simulating epidemic processes. It was built based on Epyc, which is an experiment management package that can run different simulations on a machine or in the cloud [33]. Epydemic provides different ways to simulate processes, including synchronous and stochastic (Gillespie) processes. It can handle various epidemic models, like SIR, SIS, and SEIR, and allow to monitor the simulations for analysis. In addition, Epydemic offers some built-in networks, like ER networks, but can also create own networks needed with NetworkGenerator. In our project, we build SIR models using the Epydemic library. We simulate disease dynamics through stochastic processes over two self-built networks. The details of implementation will be introduced in later sections.

2.4.2 Networkx

Networkx is a python package that allows people to create and manipulate complex networks [34]. It supports generating networks when using the Epydemic or other applications and has powerful methods to analyse and draw networks. In our study, the modular network and the core-periphery network are generated using the Networkx library, whose implementation will be present in the following sections.

3 Design and Implementation

In this paper, we simulate the spread of different diseases based on the SIR model over modular networks and core-periphery networks. We change the size and the connectivity of networks to observe how network properties affect disease spread. In this chapter, we will present how we design and implement our simulation system and two networks.

3.1 Design

Two aspects should be considered to reach our aims. The first one is designing various disease models. The second one is designing networks over which diseases spread.

3.1.1 Disease Modelling

Our models of disease dynamics are based on the SIR model. To simulate the spreading process, three parameters, $P_{infected}$, P_{infect} and P_{remove}, mentioned in Section 2.2.1 should be carefully considered. Each disease has different characteristics and properties. For example, some can transmit among the population more easily, while some have higher virulence causing a longer time for individuals to recover. Since we wanted to simulate the spread of different diseases in the network, we use different reproduction numbers $\mathcal{R}$ to describe diseases with different properties. Here, we calculate $\mathcal{R}$ as the ratio of P_{infect} and P_{remove} [5]. P_{infect} is the probability that a susceptible gets infected after interacting with an infected person, which indicates how easily the disease can transmit to individuals. P_{remove} is the probability of recovering or dying after getting infected.

$$\mathcal{R} = \frac{P_{infect}}{P_{remove}}$$

This ratio can describe how fast a disease can spread over populations. If $P_{infect} > P_{remove}$, $\mathcal{R} > 1$. This shows that the probability of infection is greater than the probability of remove, which indicates that the disease will grow dramatically and infect more and more people. If $P_{infect} < P_{remove}$, $\mathcal{R} < 1$. This indicates that the

disease will finally die out. To create different $\mathcal{R}$ in this project, we fix P_{remove} and change P_{infect} alone.

3.1.2 Networks

In this paper, we simulate the spread of two different diseases based on the SIR model on modular networks and core-periphery networks. Both networks are developed based on the traditional Erdos-Rényi (ER) network introduced in Section 2.2.2.

Modular Networks

A modular network is formed by several independent ER networks of different sizes and densities. The modular network in this paper includes a major central ER network with the largest scale and highest density and three surrounding networks with smaller size and lower density, simulating the real-life population distribution, where most people live in the central town and other small communities live surrounding the town, like small villages. In Figure 4, we show an example of the modular network including a central ER network of 1000 nodes and three small surrounding ER networks of 200 nodes. The density of them is 0.03 and 0.025, respectively.

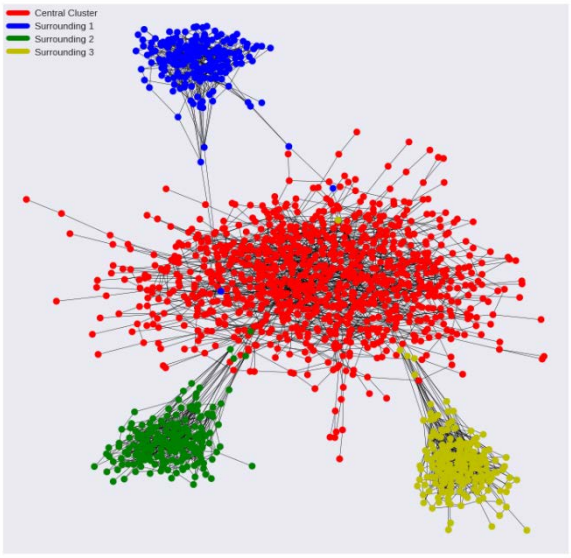

Figure 4 An example of a modular network

For a modular network, modularity is used to measure the strength of the division of a network into different modules. Networks with higher modularity mean fewer edges that connect different module exist, which means modules tend to be more independent by themselves. Newman gives a mathematical way to calculate modularity in a network in 2006 [35].

Different from the modular networks in the research of Laurent [31], where only one edge exists to connect different clusters, our project constructs various modular networks with a different number of edges connecting clusters and uses a self-defined number $P_{modularity}$ to represent the modularity of a modular network, whose reverse represents the connectivity between surrounding clusters to the central cluster. Higher $P_{modularity}$ means clusters in a modular network are more separated, so the connectivity of surrounding clusters to the central cluster is lower.

We define $P_{modularity}$ in the following way. Consider the number of edges within all surrounding networks is E_{sur} ($E_{sur} > 0$), And the number of edges that connect surrounding networks to the central network is E_{con} ($E_{con} \geq 0$). Here, we define the $P_{modularity}$ as 1 minus the fraction of edges that connect the surrounding networks to the central network. Therefore, $P_{modularity}$ is

$$P_{modularity} = 1 - \frac{E_{con}}{E_{sur}+E_{con}} = \frac{E_{sur}}{E_{sur}+E_{con}} \qquad (1)$$

The range of $P_{modularity}$ is

$$0 < P_{modularity} \leq 1$$

Here, we use an example to illustrate how to calculate $P_{modularity}$.We construct a modular network including a central cluster of 6 nodes and 15 edges, a surrounding cluster of 5 nodes and 10 edges, and the other surrounding cluster of 3 nodes and 3 edges, as shown in Figure 5.

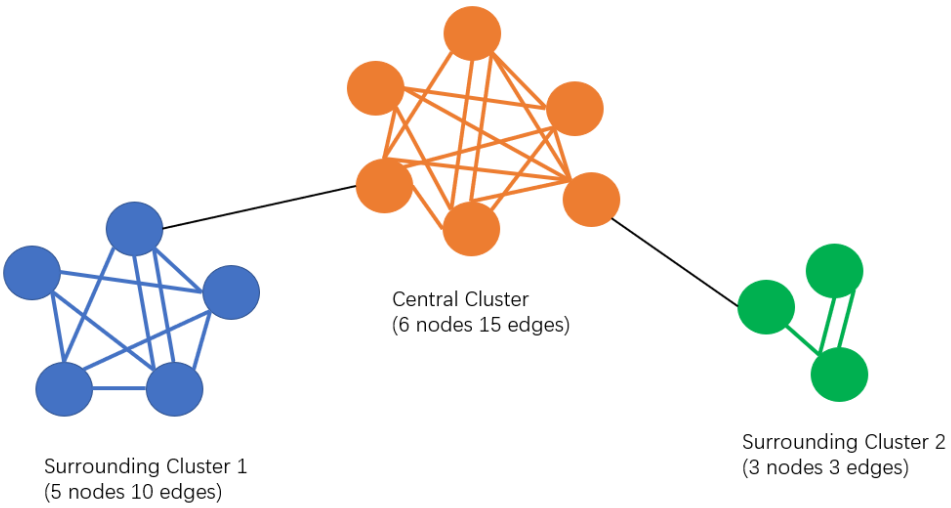

Figure 5 An example of modular network with $P_{modularity}$=0.867

In this example, E_{sur} equals to the number of edges in both surrounding clusters, which means $E_{sur} = 13$, the number of edges that connect surrounding networks to the central network is 2, which means $E_{con} = 2$. Therefore, $P_{modularity}$ in this example is $\frac{E_{sur}}{E_{sur}+E_{con}} = \frac{13}{13+2} \approx 0.867$.

If we increase the number of edges that connect surrounding networks to the central network to 6, which means $E_{con} = 6$, as shown in Figure 6 below, the $P_{modularity}$ will reduce to $\frac{E_{sur}}{E_{sur}+E_{con}} = \frac{13}{13+6} \approx 0.684$.

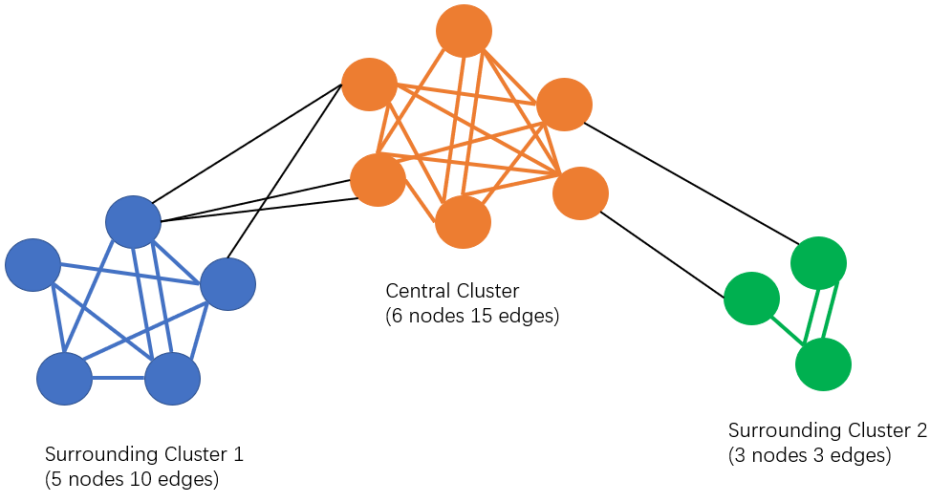

Figure 6 An example of modular network with $P_{modularity}$=0.684

Core-periphery Networks

A core-periphery network is formed by two nested ER networks with different numbers of nodes and densities, a core ER network and a periphery ER network. In this paper, the core-periphery network simulates the population of a town centre. In real life, there are some people who are more popular than others, which means they have more friends and connect with more people. Or there are some people interacting with more individuals at their positions, such as waiters and delivery staff. Besides, the number of people in this group usually is not quite large. Therefore, the core network aims at simulating this group of people who interact with others a lot, while the periphery network simulates others in the population. In network topology, the nodes of the core network usually have higher degrees. Hence, consider the numbers of nodes in the core network and the periphery network are N_{core} and N_{peri}, respectively. The density of the core network and the periphery network are ρ_{core} and ρ_{peri}, respectively. We have $N_{core} < N_{peri}$ and $\rho_{core} > \rho_{peri}$. In Figure 7, we show an example of the core-periphery network including a core ER network of 100 nodes and a periphery ER network of 1000 nodes. The density of them is 0.15 and 0.003, respectively.

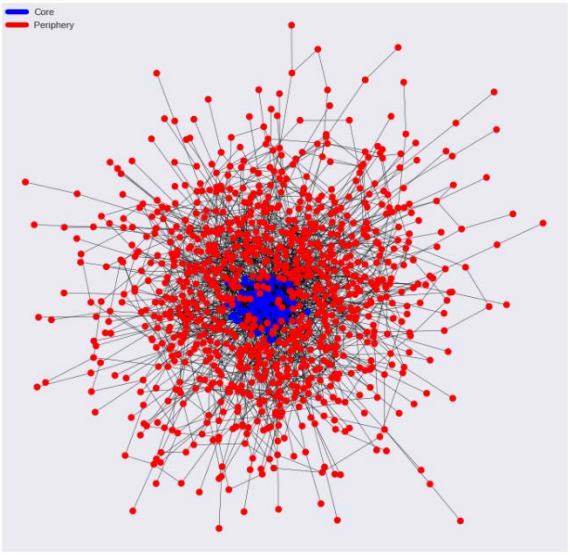

Figure 7 An example of a modular network

To describe the connectivity between the periphery cluster to the core cluster, we use $P_{connectivity}$ in our project here. For every edge between the core network and

periphery network, it has a probability of connecting both networks. $P_{connectivity}$ is used to describe this connecting probability between the two nested networks. Higher $P_{connectivity}$ indicates that more edges will exist to connect the periphery network to the core one, so the connectivity of the network is higher. In Figure 8 below, we draw an example to show how it stands.

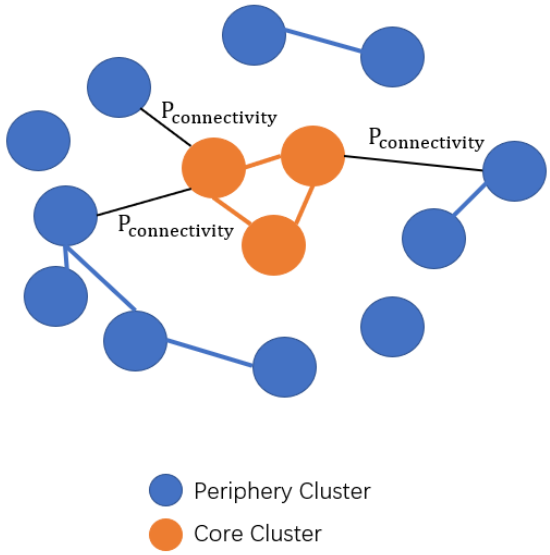

Figure 8 An example of core-periphery network to show what $P_{connectivity}$ represents

3.2 Implementation

Three classes are included in this project: ModularNetwork class, CorePeripheryNetwork class, and MonitoredSIR class.

ModularNetwork

This class is mainly responsible for generating modular networks. It inherits from the NetworkGenerator class in the Epydemic library. Three parameters are needed to build a modular network. The first one is the number of clusters in a modular network, named N_CLUSTER. The second one is the $P_{modularity}$ of a modular network, named P_MODULARITY. The last parameter is the size and density of each cluster. Here we use two lists to describe this information, named nodes_info and phi_info. Nodes_info includes the nodes number in each cluster, while the phi_info includes the density in each cluster. They are relating to each other according to the indices of the lists.

The _generate() function is used to generate a modular network. Firstly, we create all clusters using gnp_random_graph() from the Networkx library and compose them together as a single large modular network. Secondly, we find the cluster with the maximum number of nodes as the central large cluster and regard the others as the surrounding clusters. Thirdly, we calculate the E_{con} according to equation (1) and given P_MODULARITY. Ceil() function is used here to get an integer since E_{con} cannot be a float. Finally, we get the number of new edges required in each surrounding cluster according to its edge proportion and connect the randomly selected nodes of the surrounding cluster and the central cluster.

CorePeripheryNetwork

This class is mainly responsible for generating core-periphery networks. Similar to ModularNetwor, this class also inherits from the NetworkGenerator class in the Epydemic library. Five parameters are required for constructing a core-periphery network: the number of nodes of core and periphery networks named N_1 and N_2, the density of the core and periphery networks named PHI_1 and PHI_2, and the

$P_{connectivity}$ named P_CONNECTIVITY.

In this class, the _generate() function is used to generate a core-periphery network. Here, we use gnp_random_graph() to generate the core and the periphery networks and compose() to compose both. Then, we join the periphery network to the core one according to $P_{connectivity}$. We select a node in each network randomly. Then, we use the random() function to select a random value between 0 and 1. If the value is smaller than input $P_{connectivity}$, an edge will be added between these two nodes.

MonitoredSIR

To monitor the simulation process for our further analysis, ModitoredSIR class is addressed. It inherits from the SIR class in the Epydemic library. We track the number of nodes in three compartments: Susceptible, Infected, and Removed in different timesteps using trackNodesInCompartment() function provided by Epydemic. This class helps us to analyse the impact of disease on the population change of each compartment.

4 Experiments and Results

In this chapter, we will discuss how we conduct our experiments and show the results. We simulate diseases spread over modular networks and core-periphery networks. The targets of our experiments are from three perspectives. We did experiments on observing effects of (a) different reproductive number $\mathcal{R}$, (b) different network connectivity, and (c) different network sizes on the spread of diseases. We will introduce our parameters setting at first and show our experiment results from these three perspectives in this section. Due to the randomness of the simulation, we run each experiment three times and take the average value as the result.

4.1 Experiments Parameters

We will introduce the parameters we set for our experiments here. We create diseases of different reproduction numbers $\mathcal{R}$, and we construct networks with different connectivity and two different sizes, a large size and a small size.

First of all, we show the disease parameters in Table 1 below. Four diseases of distinct reproductive numbers $\mathcal{R}$ are created. As mentioned in Section 3.1.1, $\mathcal{R}$ is defined as the ratio of P_{infect} and P_{remove}. We fix $P_{remove} = 0.002$ in all of our experiments and change P_{infect} to create different $\mathcal{R}$. The reason why we create $\mathcal{R} < 1$ is because if $\mathcal{R} > 1$, diseases will spread among all population dramatically. It is difficult to observe their differences in this situation. We also limit $\mathcal{R} \leq 0.5$. This is because diseases spread dramatically when $\mathcal{R}$ stands between 0.5 and 1. When R $\leq$ 0.5, it helps us to observe the results more clearly.

Table 1 Parameters of diseases with various $\mathcal{R}$

$\mathcal{R}$	P_{infect}	P_{remove}
0.1	0.0002	0.002
0.25	0.0005	0.002
0.45	0.0009	0.002
0.5	0.001	0.002

Secondly, we will introduce the sizes of modular networks and core-periphery networks we use in our project. To observe whether network sizes affect disease spread, we create networks of a large size and a small size. The small size is ten times less than the large size. Parameters setting for network sizes are shown in Table 2 and Table 3 below. 20000 nodes and 900 edges are included in both networks in large size, while 2000 nodes and 90 edges are included in small-size networks. In our experiments, we control ten people who get infected initially. Hence $P_{infected}$ is 0.0005 in a large-size network, while $P_{infected}$ is 0.005 in a small-size network.

Table 2 Parameters of modular networks of various sizes

	Large size	**Small size**
Number of clusters	4	4
Nodes of each cluster	[16000, 2000, 1000, 1000]	[1600, 200, 100, 100]
Density of each cluster	[0.05, 0.03, 0.01, 0.03]	[0.05, 0.03, 0.01, 0.03]

Table 3 Parameters of core-periphery networks of various sizes.

	Large size	**Small size**
Nodes of core cluster	5000	500
Density of core cluster	0.06	0.06
Nodes of peri. cluster	15000	1500
Density of peri. cluster	0.04	0.04

Thirdly, we will introduce the connectivity parameters of modular networks and core-periphery networks we use in our project. To observe whether network connectivity affects disease spread, we create networks with various connectivity. As mentioned in Section 3.1.2, $P_{modularity}$ is used to describe the connectivity of a modular network from an opposite side, and $P_{connectivity}$ is used to describe the connectivity of a core-periphery network. We select five numbers from 0.1 to 0.99 and assign them to $P_{modularity}$ and $P_{connectivity}$. The parameters are shown in Table 4 below.

Table 4 Parameters of networks of various connectivity

$P_{modularity}$	$P_{connectivity}$
0.1	0.1
0.3225	0.3225
0.545	0.545
0.7675	0.7675
0.99	0.99

In summary, this section shows all parameters we set for our experiments. Firstly, we create four diseases of different reproduction numbers $\mathcal{R}$ as disease parameters. Secondly, we create both networks of a large size and a small size, where the small size is ten times smaller than the large one. Thirdly, we use five different values for setting the connectivity of both networks. Next, based on our three objectives, we will present our results according to them.

4.2 Results and Discussion

In this section, we will show the results of our experiments. We will present the general spread process of diseases at first in Section 4.2.1. Then we will show the results of changing different parameters in Sections 4.2.2, 4.2.3, and 4.2.4. Finally, we will show the relationship between Epidemic Size and the P_{infect} in Section 4.2.5.

4.2.1 Spread process of diseases

At first, we use an example to show how a disease spreads over networks generally. We simulate a disease of $\mathcal{R}$ equals to 0.45 spreads over a modular network of the small size described in Table 2 with $P_{modularity} = 0.1$. At the same time, we monitored the population changes in susceptible, infected, and removed compartments. The results are shown in Figure 9 below.

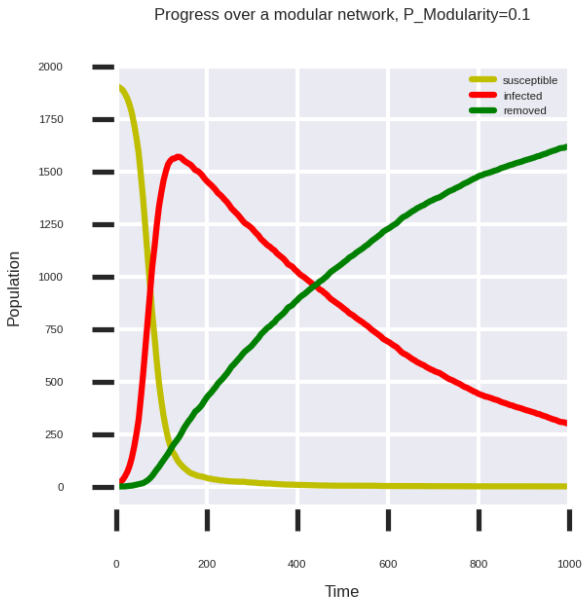

Figure 9 An example to show the spread process of diseases

At the beginning of the simulation, almost all people are susceptible. Then, the number of the susceptible populations decreases dramatically as the number of infected increases. This is because some susceptible individuals get infected, so they move from susceptible compartment into infected compartments. According to the figure, the number of removed stays empties at a very short time in the beginning. As the number of infected people increases rapidly, the number of removed people increases slightly.

This is because there are some infected people who have already recovered or died, so they move from the infected compartment to the removed compartment. In addition, we find that there is a time when the number of infected people can reach a peak. This indicates that about 1550 people out of 2000 overall can get infected in this example After the peak, no more people will get infected and the number of people in the infected compartment decreases.

It is obvious that the most interesting part of the simulation results is the first 200 time series. Therefore, to compare differences clearly, we will focus on the first 200 times of the following experiment results for analysis.

4.2.2 Effects of $\mathcal{R}$

To find out the effects of reproduction number $\mathcal{R}$ on the disease spread process, we simulate diseases with different $\mathcal{R}$ spreading over a modular network and keep network connectivity constant. We compare the infected population changes in various $\mathcal{R}$, and we show the results in Figure 10. The results tell that the disease with a larger reproduction number $\mathcal{R}$ spreads faster and infects more people, which means it has greater infectivity and transmissivity.

In the left graph in Figure 10, we show the population changes in the infected compartment for the entire simulation process. We use lines of different colours to represent diseases with different $\mathcal{R}$. It is obvious that the numbers of infected people increase rapidly for all values of $\mathcal{R}$. After reaching the peak, all numbers decrease similarly.

We find that the most interesting part locates in the first 200 time series. To observe their differences clearer, we plot the results from the first 200 times series in the right graph of Figure 10. According to this graph, we find that diseases with higher reproduction numbers spread faster among populations and infect more people. As $\mathcal{R}$ increases, the peak of each line reaches a higher number. In addition, the peak shifts to the right place when $\mathcal{R}$ rises, which means the peak arrives at an earlier time when $\mathcal{R}$ is larger.

It is easy to think of the reasons for these phenomena. A disease with larger $\mathcal{R}$ means it has higher infectivity. Hence, the probability that a susceptible person catches the pathogen from an infected person is higher, and people can become infected easily. This causes the number of infected people increases more dramatically when $\mathcal{R}$ is higher, so it reaches the peak at an earlier time.

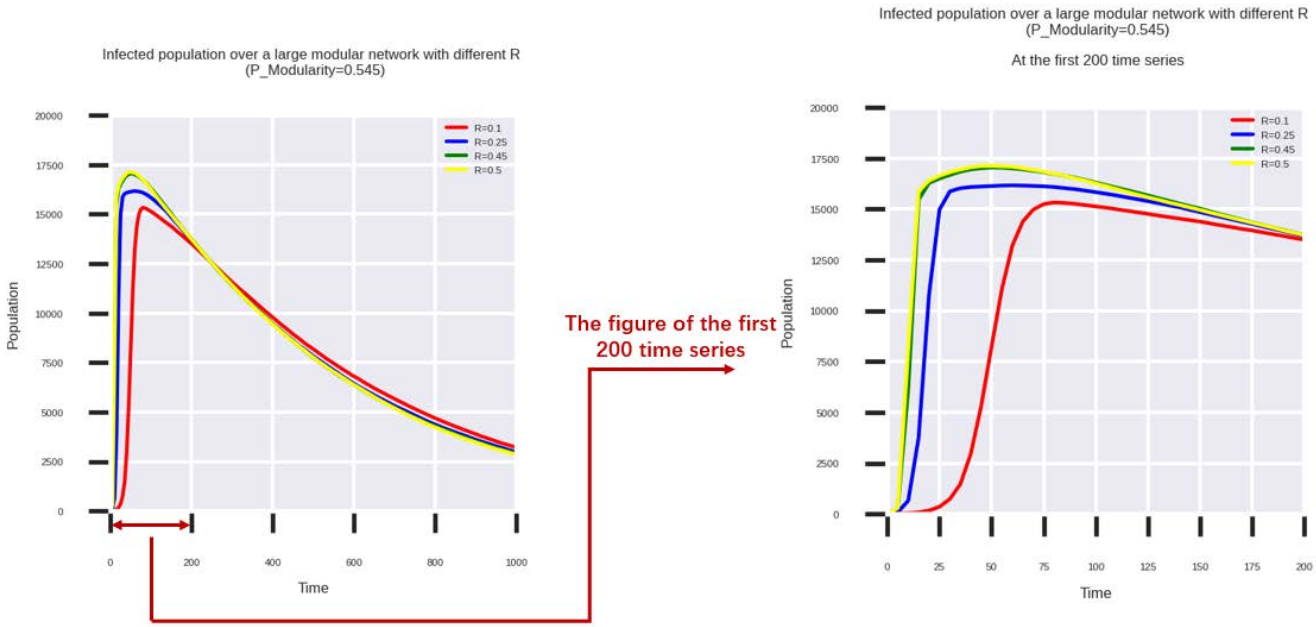

Figure 10 Result of diseases with different R over a large-size modular network

In our graph, we find that the yellow line and green line are really close to each other. This is because the $\mathcal{R}$ values of these two lines are really close, where one is 0.45 and the other is 0.5. Hence, the small differences in $\mathcal{R}$ cause insignificant differences in the graph. This implies that diseases with similar reproduction numbers obtain similar infectivity, so their spread processes are also similar.

In summary, the disease of larger reproduction number $\mathcal{R}$ not only can spread among populations faster but can also infect more people. Besides, diseases with similar values of $\mathcal{R}$ behave similarly. This indicates that the infectivity and transmissivity of disease guide its spread process among populations.

4.2.3 Effects of network connectivity

To find out the effects of network connectivity on the disease spread process, we simulate a disease spreading over a modular network and a core-periphery network with different connectivity. Since we use $P_{modularity}$ to represent the connectivity of a modular network from the opposite side and use $P_{connectivity}$ to represent the connectivity of a core-periphery network, we will show their results separately.

First, we show the results of various $P_{modularity}$ in modular networks in Figure 11. As $P_{modularity}$ represents the connectivity of a modular network from the opposite side, higher $P_{modularity}$ means lower connectivity. As usual, we show the results of the entire simulation in the left graph of Figure 11 and show the results of the first 200 time series in the right position for analysis clearly. We find that the infectivity of the disease grows if the network is more connected.

According to the right graph in Figure 11, when the same disease spreads over modular networks with various connectivity, the number of infected people in these networks grows at a similar rate at the beginning, which means that the transmissibility of the disease stays similarly. However, the maximum number of infected people in these networks differs. From the graph, when $P_{modularity} = 0.1$ where the network is highly connected, the disease can infect about 17500 people at most. But the number reduces to around 15000 when $P_{modularity} = 0.99$. This indicates that network connectivity can affect the infectivity of the disease. A disease will have higher infectivity in a more connected network, so causes more people may get infected.

The reason for these phenomena can be considered in the following way. In a more connected network, more connections exist between nodes. Diseases can easily spread from one node to another through connections. But in a sparse network, there are fewer edges in the network. It is difficult for diseases to transmit to another person through a limited number of connections.

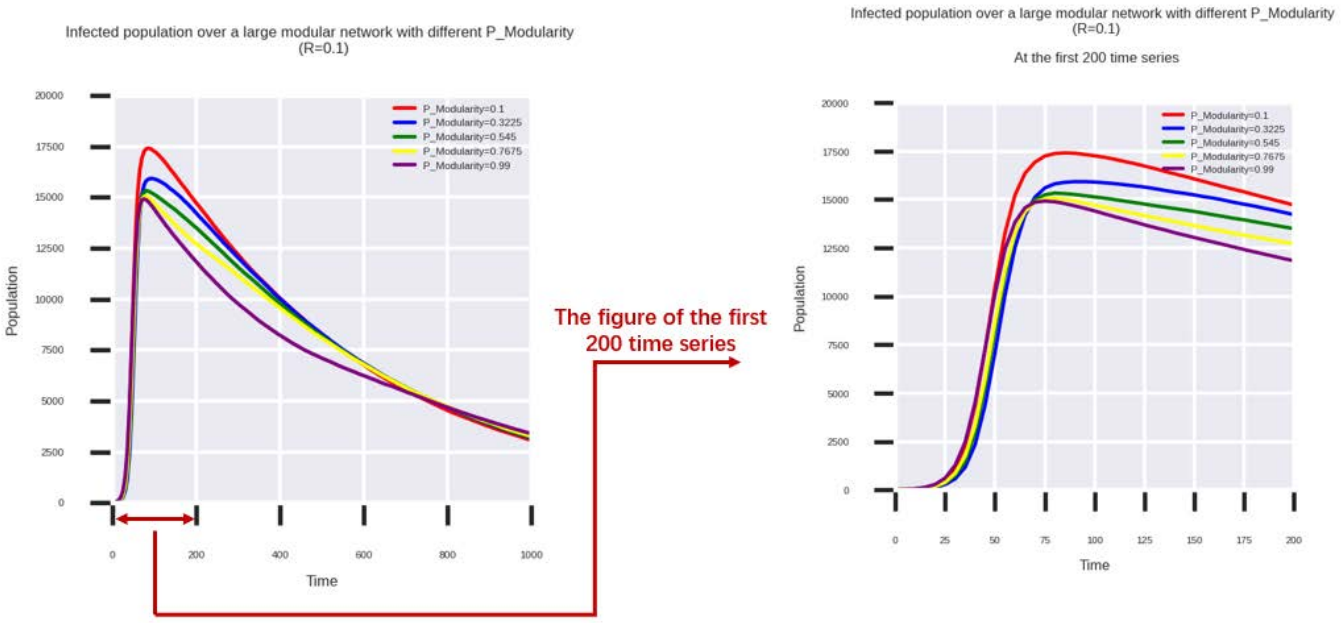

Figure 11 Result of a disease spreading over modular networks with different $P_{modularity}$

We did the same experiment on core-periphery networks to see how they behave compared to results in modular networks. As $P_{connectivity}$ representing the connectivity of a core-periphery network, higher $P_{connectivity}$ means greater connectivity. We show the results in Figure 12. We find similar results in core-periphery networks, showing that the disease infectivity can be affected by the network connectivity. But one thing that is different from modular networks is that the transmissibility of the disease also is affected by the core-periphery network connectivity. Both the infectivity and the transmissibility of the disease increase in the more connected core-periphery network.

Through the right graph in Figure 12, the purple line, which has the highest $P_{connectivity}$, reaches about 19900 population maximum, while the red line which has the lowest $P_{connectivity}$ reaches around 18500 only. Different from modular networks, the number of infected people in five different $P_{connectivity}$ does not grow at a similar rate. The purple line of the highest $P_{connectivity}$ grows fastest among all five lines. We assume this difference may be due to the different ways we use to define $P_{modularity}$ and $P_{connectivity}$. Our definition for $P_{connectivity}$ in the core-periphery networks tends to create more edges than that in the modular networks. Hence, a disease transmits

from one person to another easily through connections between them. The diseases spend less time for their spread and infect more people.

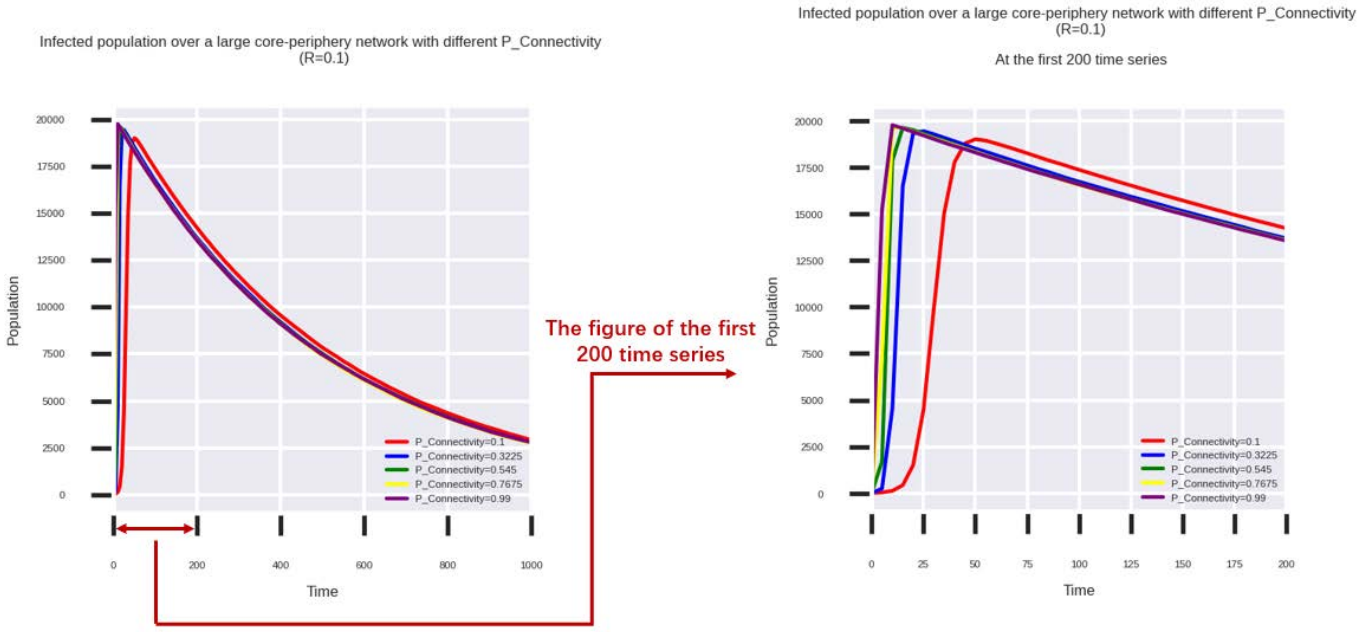

Figure 12 Result of a disease spreading over core-periphery networks with different $P_{connectivity}$

In summary, the disease properties and the network connectivity are not independent of each other. The results of our experiments show that, when using our way to define the connectivity metrics for networks, the disease infects more people when transmitting on a more connected modular network, and the disease not only grows faster but also infects more people when transmitting on a more connected core-periphery network.

4.2.4 Effects of network size

To observe the effects of network size on the process of the disease spread, we simulate the same disease spreading over networks with the same connectivity, but only change their sizes for comparing results. We did experiments on both modular networks and core-periphery networks, and we show their results in Figure 13. In the experiments, we simulate the same disease ($\mathcal{R} = 0.1$) over both networks, and we choose both networks in their highest connectivity ($P_{modularity} = 0.1$ and $P_{connectivity} = 0.99$ respectively). The results indicate that the network size influences the disease properties. For the same disease, a higher percentage of people will be infected in a network of larger size.

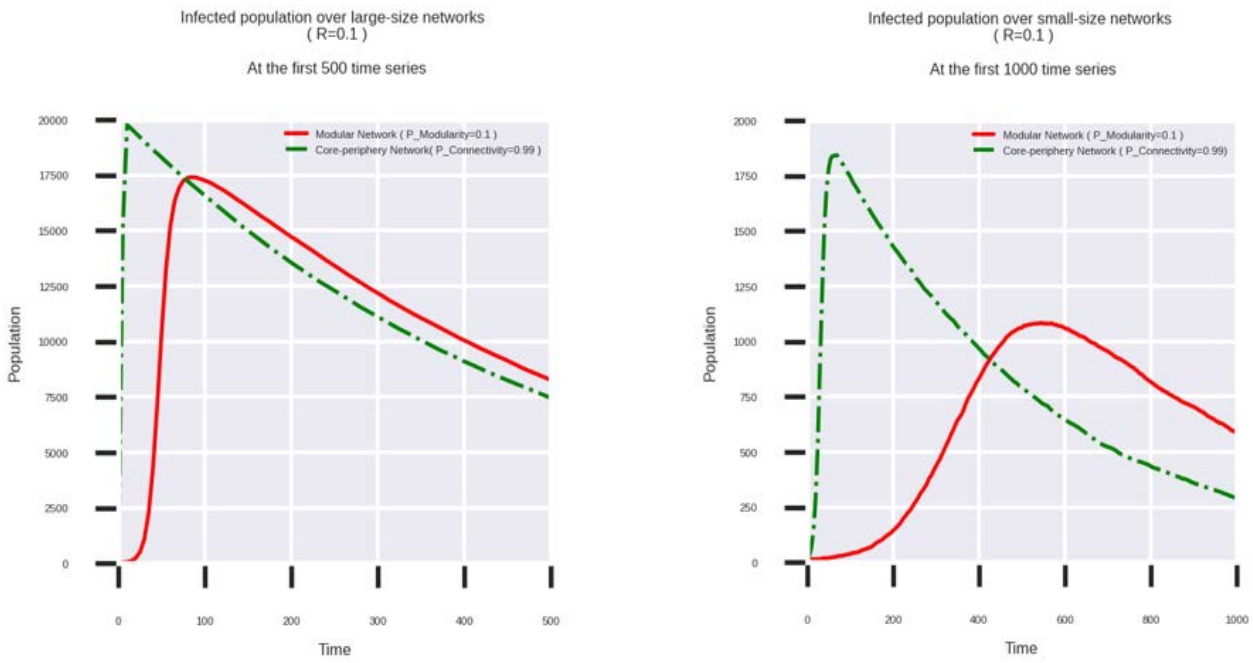

Figure 13 Results of a disease spreading over networks with different sizes

The left graph in Figure 13 shows results for both networks with large sizes, while the right graph shows results for networks with small sizes. The red lines represent the results in modular networks, while the green dashed lines represent the results in core-periphery networks. Looking at the red lines at first, the left graph shows that there are about 17500 people get infected at most among 20000 populations in total, so the

percentage of people who get infected is about 87.5%. On the other hand, the right graph shows that there is a maximum of around 1100 people get infected out of 2000 populations, so the percentage is only about 55%. Similar phenomena happen in the core-periphery networks (green dashed lines). In the large-size core-periphery network, almost all people may get infected by the disease, but the percentage decreases to 90% in the small-size core-periphery network. Hence, in both networks, the percentage of people getting infected by the same virus is higher in networks of larger size.

Looking at every single graph in Figure 13, it is easy to find that the disease always spread faster and infects more people in core-periphery networks compared to that in modular networks. We assume that this is because we use two different ways to define network connectivity metrics for each network, which we present in Section 3.1.2. Our methods tend to add more edges to the core-periphery network implicitly.

In summary, the size of the network has an impact on the spread of disease. Our experiments show that a higher percentage of people can get infected by a disease in a larger population. In addition, due to the different ways of defining the metrics of network connectivity in our project, a disease spread faster and infects more people in a core-periphery network than in a modular network.

4.2.5 Epidemic Size and P_{infect}

In previous sections, we show how network properties affect disease spread. Here we try to find out how different values of P_{infect} affect the epidemic size.

We select 50 values for P_{infect} from 0.00001 to 0.00009 and simulate the disease-transmitting over a small-size modular network. We calculate the number of people in the removed compartment at the end of the simulation, which indicates how many people actually have been infected during the simulation. Due to the randomness that exists in the simulation process, we run each experiment 10 times and show the results in the left graph of Figure 14. Each red point represents the result of a single simulation. For every P_{infect} value, there are ten points, which refer to the 10-time running for each experiment. The height of the "line" formed by these points represents the experiment variance.

According to the left graph, when the disease has low infectivity (at the left part), almost none of the population gets infected, and the disease dies out eventually. When P_{infect} starts to become larger and within a certain range, the epidemic size will increase rapidly, until P_{infect} is greater than this range where the epidemic size becomes more stable. In our experiments shown in the left graph, the range locates between 0.00009 to 0.0002. We also find that more variances appear in this range, as the height of the "line" formed by nodes is larger. In addition, we did similar experiments over the network with lower $P_{modularity}$, as shown in the right graph of Figure 14. We find that the range stays the same, but it leads to a larger epidemic size.

When 10 people get infected initially

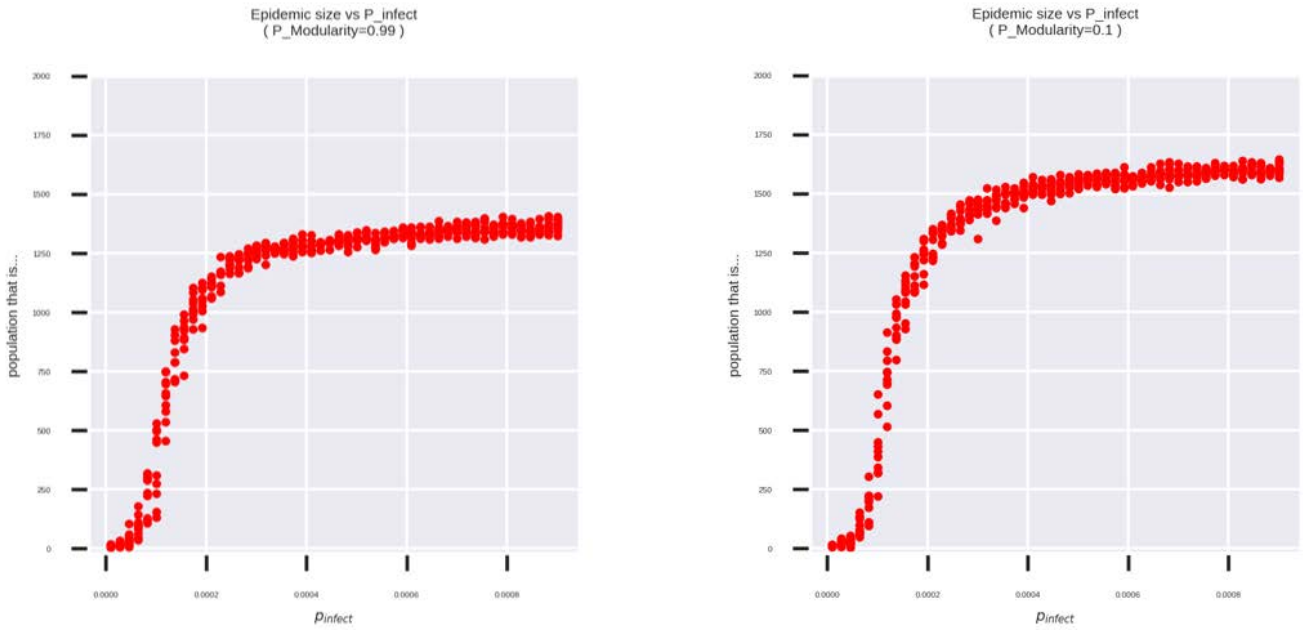

Figure 14 Results of epidemic size with different P_{infect} over modular networks (10 people infected initially)

Figure 14 shows the experiments of 10 people infected initially. We reduce the number to only 1 person becomes infected initially and show the results in Figure 15.

According to Figure 15, there are red points at the bottom of the graphs for every P_{infect}, which cannot be seen in Figure 14. This means that when only 1 individual gets infected initially, sometimes the disease may not infect more people and finally die out, even with the highest infectivity. Therefore, a big variance exists in the disease with the highest infectivity. There can be two extreme situations for the disease with the highest infectivity. It may infect more than half of the population, or it may not infect anyone and die out in the end.

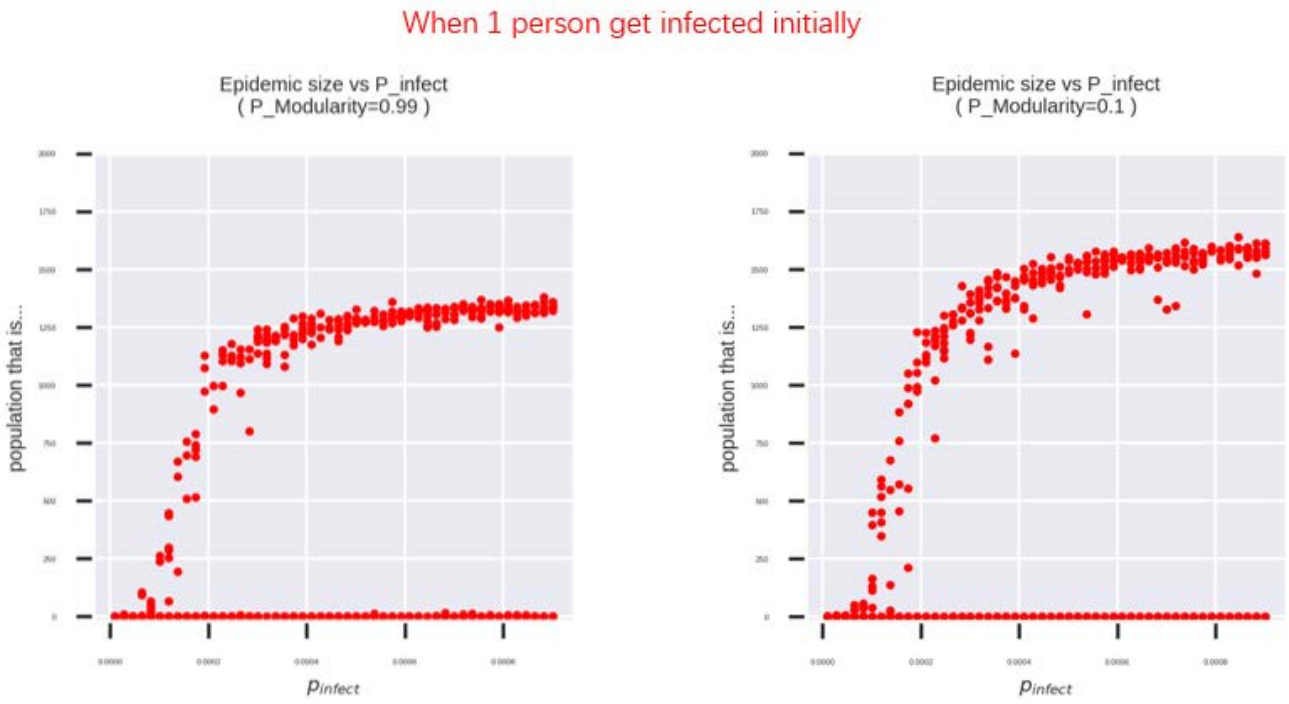

Figure 15 Results of epidemic size with different P_{infect} over modular networks (1 person infected initially)

The experiments above are based on modular networks. We also did similar experiments over the core-periphery networks, which have slightly sharper looks but show similar results. We show their figures in the Appendix A.1 at the end of the report.

We have shown that the disease has a higher probability of dying out when fewer people get infected initially. Next, we try to find out whether the modularity of a modular network can affect this probability of diseases.

We select five values for $P_{modularity}$ as usual and set only 1 person who becomes infected initially. Then calculate the number of points whose epidemic size is smaller than 15 for each $P_{modularity}$. Since there are 500 nodes in each graph (50 values of P_{infect}, 10 experiments for each value), we use the number divided by 500 to get the probability. The results are shown in Figure 16. It is obvious to find that as $P_{modularity}$ rises, the probability of the disease dying out is becoming larger. This indicates that disease tends to infect fewer individuals and die out in the less connected networks.

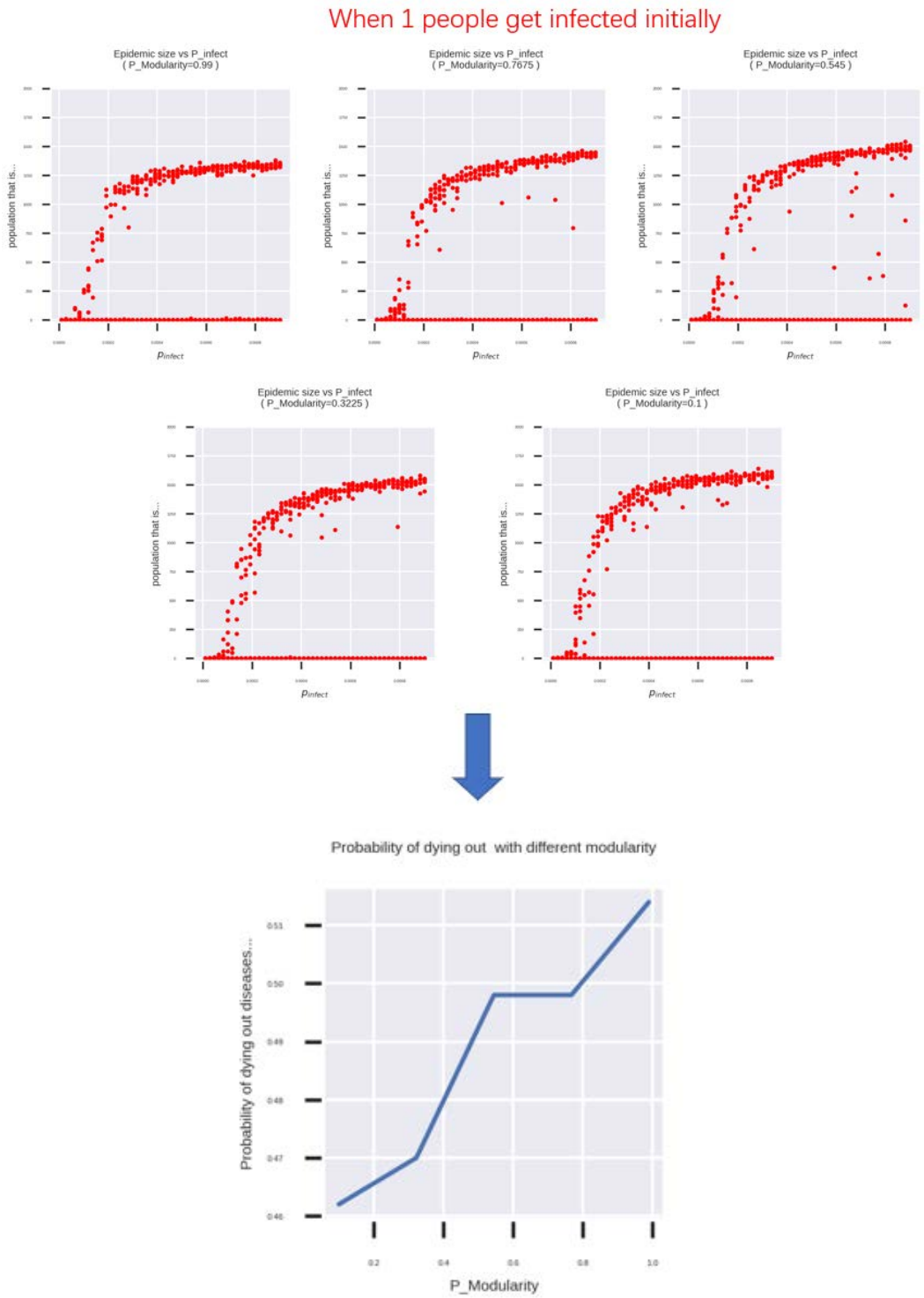

Figure 16 Probability that a disease dies out with different $P_{modularity}$

However, an unexpected shape exists in a core-periphery network. We did the same experiment over core-periphery networks and show the results in Figure 17. As $P_{connectivity}$ increases, the probability decreases significantly but rises a little bit, then decreases again. We assume this is caused by stochastic dynamics in simulation. Perhaps this phenomenon will disappear when we increase the number of times of running this experiment and take the average value.

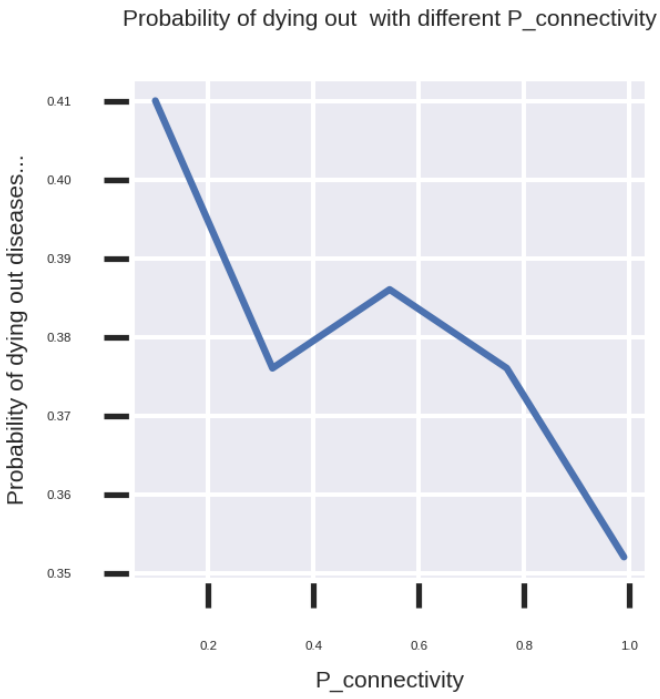

Figure 17 Probability that a disease dies out with different $P_{connectivity}$

In summary, P_{infect} has effects on the epidemic size. Diseases with higher infectivity can make more people infected. There is a certain range of P_{infect}, where the epidemic size can increase significantly as P_{infect} increases within the range. After this, the epidemic size increases stably. Secondly, we find that when fewer people get infected, diseases have a higher probability to die out eventually. This causes two extreme phenomena for diseases with high infectivity. The disease either dies out at the end or causes an epidemic among populations. Thirdly, we find that the modularity of modular networks and connectivity of core-periphery networks also affect the probability that the disease dies out. There is a higher probability of dying out of the

disease spreading over a more separated network. Even though a weird shape appears in the result of core-periphery networks, we suspect it is due to the stochastic dynamics in simulation.

5 Conclusion

In this chapter, we summarize our achievements and results and will discuss our future possible works.

5.1 Key achievements and Conclusions

Our project aims at finding out how network properties affect the transmission process of diseases, mainly focusing on the size and connectivity of the network. Our main achievements include:

(1) We build disease models based on the popular compartmented model - the SIR model, and the reproduction number $\mathcal{R}$ of disease is used to create diseases with different infectivity.

(2) We build two types of networks - the modular networks and the core-periphery networks, and different self-defined metrics are used to describe the connectivity of each kind of network.

In the experiments, we change different parameters and monitor the number of individuals in susceptible, infected, and removed compartments. Our experiment works include:

(1) We first observe how the reproduction number $\mathcal{R}$ of a disease affects its transmission process.

(2) Then, we did experiments on observing whether network properties have an impact on the spread process of diseases. We change the connectivity and sizes of modular networks and core-periphery networks to observe results.

(3) Finally, our experiments demonstrate how epidemic size changes with increasing P_{infect}.

(4) We also show how the probability that a disease will eventually die out is affected by the initial amount of infection and the connectivity of the network.

There are interesting findings we have discovered through our experiments. They include:

(1) The infectious ability of the disease dominates its transmission. A disease with a larger reproduction number $\mathcal{R}$ obtains higher infectivity. It can not only spread among populations faster but can also cause more people to become infected. Diseases with similar $\mathcal{R}$ show similar infectivity, and their spread process is also similar to each other.

(2) Networks with higher connectivity tend to select diseases with higher infectivity. In the more connected modular networks, results show that diseases infect more people. On the other hand, in the more connected core-periphery networks, the diseases also infect more people and can spread faster among the population.

(3) The size of a network has effects on the proportion of the infectious population. A higher percentage of people can become infected by the disease when spreading among a larger population.

(4) Epidemic size increases with increasing disease infectivity. There is a certain range where the size grows rapidly, but its growth becomes stable when exceeding the range.

(5) Diseases have a higher probability to die out when fewer people are infected initially and spread over a less connected network.

All those results show that the network properties and the spread process of diseases are not independent of each other and provide inspiration for human beings about how to control the spread of disease effectively when an epidemic appears. They explain the control measures taken by governments under the current COVID-19 situation. For example, people are required to maintain social distance and to self-quarantine at home when they are tested positive. Some countries have reduced the number of people entering and leaving the country, and some encourage citizens to stay locally and reduce the number of travelling during the epidemic. All those strategies are

aimed at controlling the spread of the disease by changing the connectivity of the network.

5.2 Future Work

Our projects discover the connection between network properties and the spread of diseases based on the SIR model. The SIR model assumes that people will not become infected the second time. But in real life, there are many diseases can do can cause secondary infection. So, the SIS and the SIRS models are proposed to address this situation. In modern life, if an epidemic occurs, vaccines will be applied to control its spread. Hence, we should also consider the effects of vaccines on the spread of the disease. There are some possible models related to this, such as the SIVR model [9].

The simulation in this project is processed from a stochastic dynamic simulation. We run each experiment several times and take the average as the final results for analysis. Our experiments are run no more than ten times. To enhance accuracy, this can be improved by increasing the number of experiments running, and random seeds can be utilized at the same time for better analysis.

In our project, we use $P_{modularity}$ to describe the connectivity from the surrounding clusters to the centre cluster in modular networks, and we use $P_{connectivity}$ to describe the connectivity from the periphery cluster to the core cluster in core-periphery networks. These two metrics for describing connectivity are defined in different ways. Hence, this makes it impossible for us to accurately show whether the network topology will have an impact on the disease transmission process in this report. An improvement could be using a unified way to describe the connectivity of networks with distinct topologies, which we will focus on in our future work.

Appendix

A.1 Graphs of Core-periphery Networks

Here, we supplement some experiment figures of Core-periphery networks. In Section 4.2.5 Epidemic Size and P_{infect}, some figures about Core-periphery networks are not shown, so we show them here in the Appendix to support our conclusion.

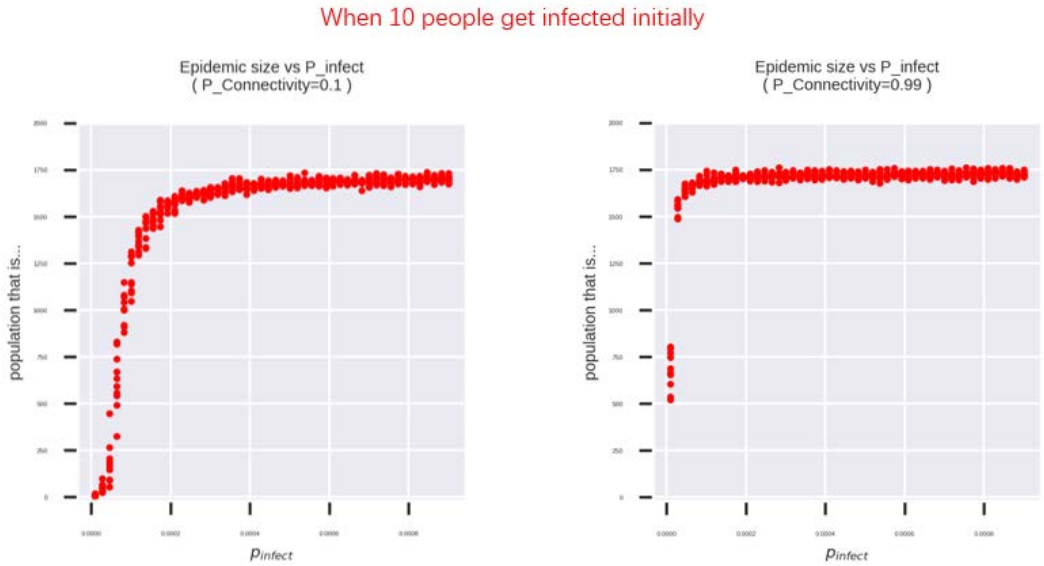

Figure 18 Results of epidemic size with different P_{infect} over core-periphery networks (10 people infected initially)

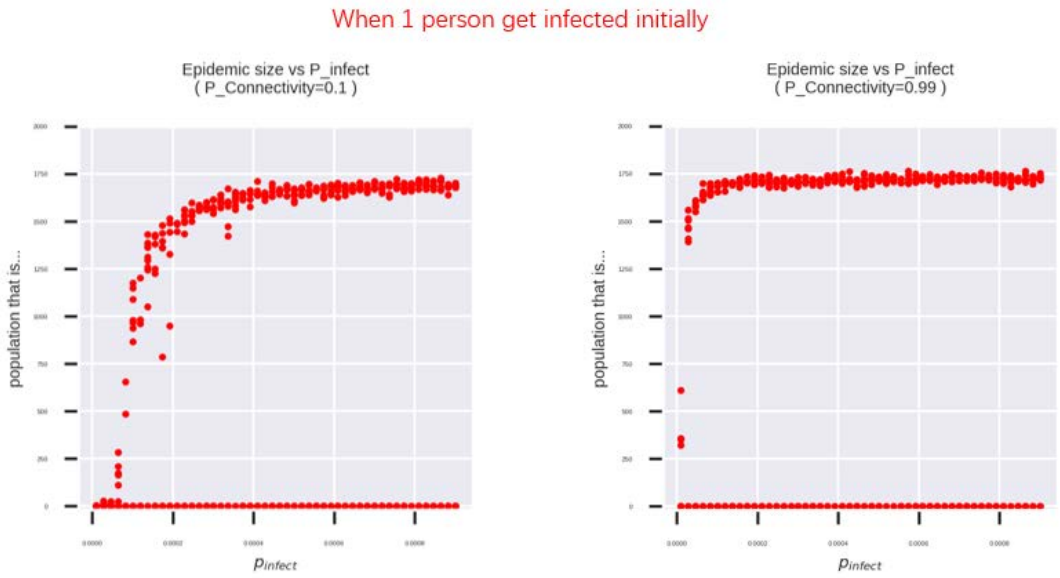

Figure 19 Results of epidemic size with different P_{infect} over modular networks (1 person infected initially)

A.2 User Manual

The development of our research is based on python 3.6. Jupyter Notebook is used to conduct our experiments. All results of our experiments are saved in a JSON file for further analysis. We use three Notebook files to support our research.

- ModularNetwork.ipynb

  This notebook is mainly responsible for constructing the modular network and performing experiments over it.

- CorePeripheryNetwork.ipynb

  This notebook is mainly responsible for constructing the core-periphery network and performing experiments over it.

- AnalyzeResults.ipynb

  This notebook is mainly responsible for our analysis and drawing graphs, which helps us to get our results.

The code is available in GitHub:
https://github.com/DaisyDDD/epycDissertation.git

A.2 Ethics

UNIVERSITY OF ST ANDREWS
TEACHING AND RESEARCH ETHICS COMMITTEE (UTREC)
SCHOOL OF COMPUTER SCIENCE
PRELIMINARY ETHICS SELF-ASSESSMENT FORM

This Preliminary Ethics Self-Assessment Form is to be conducted by the researcher, and completed in conjunction with the Guidelines for Ethical Research Practice. All staff and students of the School of Computer Science must complete it prior to commencing research.

This Form will act as a formal record of your ethical considerations.
Tick one box

- ☐ **Staff Project**
- ☒ **Postgraduate Project**
- ☐ **Undergraduate Project**

Title of project

Epidemic models realistic in time and space

Name of researcher(s)

Yujie Dai

Name of supervisor (for student research)

Simon Dobson

OVERALL ASSESSMENT (to be signed after questions, overleaf, have been completed)

Self audit has been conducted **YES** ☒ **NO** ☐

There are no ethical issues raised by this project

Signature Student or Researcher

Print Name

Yujie Dai

Date

26th May 2022

Signature Lead Researcher or Supervisor

Print Name

Simon Dobson

Date

9Jun2022

This form must be date stamped and held in the files of the Lead Researcher or Supervisor. If fieldwork is required, a copy must also be lodged with appropriate Risk Assessment forms. The School Ethics Committee will be responsible for monitoring assessments.

Computer Science Preliminary Ethics Self-Assessment Form

Research with human subjects

Does your research involve human subjects or have potential adverse consequences for human welfare and wellbeing?

YES ☐ **NO** ☒

If YES, full ethics review required
For example:
Will you be surveying, observing or interviewing human subjects?
Will you be analysing secondary data that could significantly affect human subjects?
Does your research have the potential to have a significant negative effect on people in the study area?

Potential physical or psychological harm, discomfort or stress

Are there any foreseeable risks to the researcher, or to any participants in this research?

YES ☐ **NO** ☒

If YES, full ethics review required
For example:
Is there any potential that there could be physical harm for anyone involved in the research?
Is there any potential for psychological harm, discomfort or stress for anyone involved in the research?

Conflicts of interest

Do any conflicts of interest arise?

YES ☐ **NO** ☒

If YES, full ethics review required
For example:
Might research objectivity be compromised by sponsorship?
Might any issues of intellectual property or roles in research be raised?

Funding

Is your research funded externally?

YES ☐ **NO** ☒

If YES, does the funder appear on the 'currently automatically approved' list on the UTREC website?

YES ☐ **NO** ☐

If NO, you will need to submit a Funding Approval Application as per instructions on the UTREC website.

Research with animals

Does your research involve the use of living animals?

YES ☐ **NO** ☒

If YES, your proposal must be referred to the University's Animal Welfare and Ethics Committee (AWEC)

University Teaching and Research Ethics Committee (UTREC) pages
http://www.st-andrews.ac.uk/utrec/

A.3 Testing Summary

The programming development of our research is conducted incrementally. We output the number of nodes and edges and we draw the graph of networks to show how they look to make sure they are correct.

A.4 Models UML

We show the UML class diagram of the two network classes. Both ModularNetwork and CorePeripheryNetwork inherit from the NetworkGenerator class. We override the _generator() method to create different networks for each class. Two attributes are included in the ModularNetwork. N_CLUSTER refers to the number of clusters in the modular network, and P_MODULARITY refers to the $P_{modularity}$ of the modular network. Four attributes are included in the CorePeripheryNetwork. N_1 denotes the number of nodes in the core network, while N_2 refers to the number of nodes in the core-periphery network. PHI_1 denotes the density of the periphery network, while PHI_2 refers to the density of the periphery network. P_CONNECTIVITY represents to the $P_{connectivity}$ of the core-periphery network.

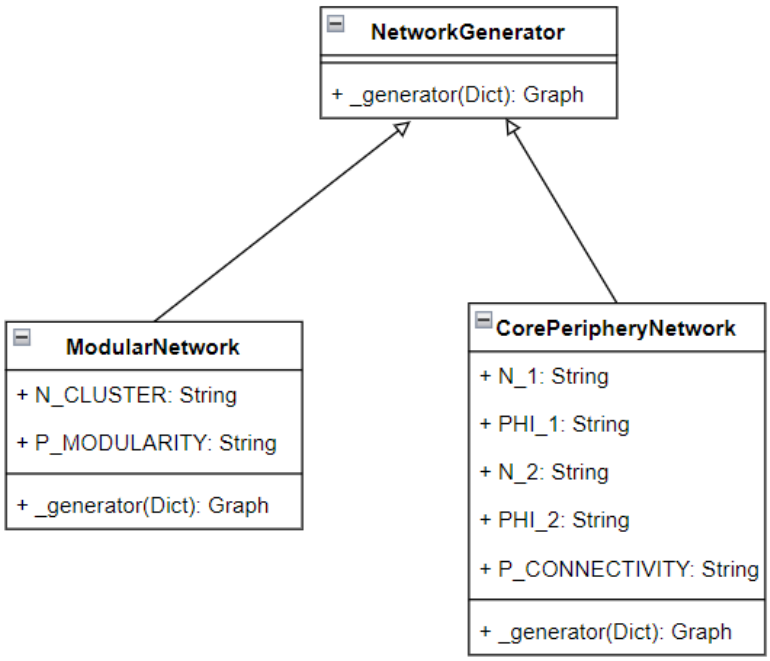

Figure 16 UML for network classes

References

[1]. Matrajt L, Eaton J, Leung T, et al. Vaccine optimization for COVID-19: Who to vaccinate first?[J]. Science Advances, 2021, 7(6): eabf1374. URL: https://www.science.org/doi/abs/10.1126/sciadv.abf1374

[2]. Courtemanche C, Garuccio J, Le A, et al. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States[J]. Health affairs, 2020, 39(7): 1237-1246. URL: https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2020.00608

[3]. Andersen M. Early evidence on social distancing in response to COVID-19 in the United States[J]. Available at SSRN 3569368, 2020. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3569368

[4]. Anderson R M, May R M. Infectious diseases of humans: dynamics and control[M]. Oxford university press, 1992. URL: https://books.google.com/books?hl=zh-CN&lr=&id=HT0--xXBguQC&oi=fnd&pg=PA25&dq=Infectious+diseases+of+humans:+dynamics+and+control&ots=Iem9PQVUkh&sig=oSPmTe5Jdx0w0V1Nd1qiLuxPwaQ

[5]. Hethcote H W. The mathematics of infectious diseases[J]. SIAM review, 2000, 42(4): 599-653. URL: https://epubs.siam.org/doi/pdf/10.1137/S0036144500371907

[6]. Keeling M J, Eames K T D. Networks and epidemic models[J]. Journal of the royal society interface, 2005, 2(4): 295-307. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsif.2005.0051?ssource=mfc&rss=1

[7]. Newman M E J. Spread of epidemic disease on networks[J]. Physical review E, 2002, 66(1): 016128. URL: https://journals.aps.org/pre/pdf/10.1103/PhysRevE.66.016128

[8]. Karrer B, Newman M E J. Competing epidemics on complex networks[J]. Physical Review E, 2011, 84(3): 036106. URL: https://journals.aps.org/pre/pdf/10.1103/PhysRevE.84.036106

[9]. Dobson, S. (2020). Epidemic modelling -- some notes, maths, and code. Independent Publishing Network. URL: https://simondobson.org/introduction-to-epidemics/index.html#

[10]. Shai S, Dobson S. Coupled adaptive complex networks[J]. Physical Review E, 2013, 87(4): 042812. URL: https://journals.aps.org/pre/pdf/10.1103/PhysRevE.87.042812

[11]. Mann P, Smith V A, Mitchell J B O, et al. Two-pathogen model with competition on clustered networks[J]. Physical Review E, 2021, 103(6): 062308. URL: https://journals.aps.org/pre/pdf/10.1103/PhysRevE.103.062308

[12]. Dodds P S, Watts D J. Universal behavior in a generalized model of contagion[J]. Physical review letters, 2004, 92(21): 218701. URL: https://link.aps.org/pdf/10.1103/PhysRevLett.92.218701

[13]. Li M Y, Graef J R, Wang L, et al. Global dynamics of a SEIR model with varying total population size[J]. Mathematical biosciences, 1999, 160(2): 191-213. URL: https://www.sciencedirect.com/science/article/pii/S0025556499000309

[14]. Dottori M, Fabricius G. SIR model on a dynamical network and the endemic state of an infectious disease[J]. Physica A: Statistical Mechanics and its Applications, 2015, 434: 25-35. URL: https://www.sciencedirect.com/science/article/pii/S0378437115003660

[15]. Biswas K, Khaleque A, Sen P. Covid-19 spread: Reproduction of data and prediction using a SIR model on Euclidean network[J]. arXiv preprint arXiv:2003.07063, 2020. URL: https://arxiv.org/pdf/2003.07063

[16]. Xia C, Wang L, Sun S, et al. An SIR model with infection delay and propagation vector in complex networks[J]. Nonlinear Dynamics, 2012, 69(3): 927-934. URL: https://link.springer.com/article/10.1007/s11071-011-0313-y

[17]. Arquam M, Singh A, Sharma R. Modelling and analysis of delayed sir model on complex network[C]//International Conference on Complex Networks and their Applications. Springer, Cham, 2018: 418-430. URL: https://link.springer.com/chapter/10.1007/978-3-030-05411-3_34

[18]. Boots M, Mealor M. Local interactions select for lower pathogen infectivity[J]. Science, 2007, 315(5816): 1284-1286. URL: https://www.science.org/doi/pdf/10.1126/science.1137126

[19]. Boots M, Sasaki A. ‘Small worlds’ and the evolution of virulence: infection occurs locally and at a distance[J]. Proceedings of the Royal Society of London. Series B: Biological Sciences, 1999, 266(1432): 1933-1938. Url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1690306/pdf/10584335.pdfMoore C, Newman M E J. Epidemics and percolation in small-world networks[J]. Physical Review E, 2000, 61(5): 5678. URL: https://journals.aps.org/pre/pdf/10.1103/PhysRevE.61.5678

[20]. Van den Driessche P. Reproduction numbers of infectious disease models[J]. Infectious Disease Modelling, 2017, 2(3): 288-303. URL: https://www.sciencedirect.com/science/article/pii/S2468042717300209

[21]. Heffernan J M, Smith R J, Wahl L M. Perspectives on the basic reproductive ratio[J]. Journal of the Royal Society Interface, 2005, 2(4): 281-293. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2005.0042

[22]. Newman M. Networks[M]. Oxford university press, 2018. URL:

https://books.google.co.uk/books?hl=zh-CN&lr=&id=YdZjDwAAQBAJ&oi=fnd&pg=PP1&dq=networks+an+introduction+newman&ots=V-LW-Ne4rt&sig=yso5YBvcTOiAauIIAjmvAMsFlXc#v=onepage&q=networks%20an%20introduction%20newman&f=false

[23]. O'Dea E B, Park A W, Drake J M. Estimating the distance to an epidemic threshold[J]. Journal of the Royal Society interface, 2018, 15(143): 20180034. URL: https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2018.0034

[24]. Jacquez J A, O'Neill P. Reproduction numbers and thresholds in stochastic epidemic models I. Homogeneous populations[J]. Mathematical Biosciences, 1991, 107(2): 161-186. URL: https://www.sciencedirect.com/science/article/abs/pii/0025556491900032

[25]. Newman M E J. The structure and function of complex networks[J]. SIAM review, 2003, 45(2): 167-256. URL: https://epubs.siam.org/doi/pdf/10.1137/S003614450342480?xid=PS_smithsonian

[26]. J Hopcrof. Large Graphs - Erdos-Renyi Model. Jan. 2006. URL: http://www.cs.cornell.edu/courses/cs485/2006sp/lecture%20notes/lecture1.pdf

[27]. Moore C, Newman M E J. Epidemics and percolation in small-world networks[J]. Physical Review E, 2000, 61(5): 5678. URL: https://journals.aps.org/pre/pdf/10.1103/PhysRevE.61.5678

[28]. Keeling M. The implications of network structure for epidemic dynamics[J]. Theoretical population biology, 2005, 67(1): 1-8. URL: https://www.sciencedirect.com/science/article/pii/S0040580904001121

[29]. Keeling M J, Eames K T D. Networks and epidemic models[J]. Journal of the royal society interface, 2005, 2(4): 295-307. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2005.0051

[30]. Lopes P C, Block P, König B. Infection-induced behavioural changes reduce connectivity and the potential for disease spread in wild mice contact networks[J]. Scientific reports, 2016, 6(1): 1-10. URL: https://www.nature.com/articles/srep31790

[31]. Hebert-Dufresne L, Allard A. Smeared phase transitions in percolation on real complex networks[J]. Physical Review Research, 2019, 1(1): 013009. URL: https://journals.aps.org/prresearch/pdf/10.1103/PhysRevResearch.1.013009

[32]. Paul Taylor. Susceptible, infectious, recovered. London Review of Books, May 2020. URL: https://www.lrb.co.uk/the-paper/v42/n09/paul-taylor/susceptible-infectious-recovered

[33]. Dobson S. Epydemic. URL: https://pyepydemic.readthedocs.io/en/latest/

[34]. Networkx. URL: https://networkx.org/documentation/stable/index.html

[35]. Newman M E J. Modularity and community structure in networks[J]. Proceedings of the national academy of sciences, 2006, 103(23): 8577-8582. URL: https://www.pnas.org/doi/full/10.1073/pnas.0601602103