Using Agent-based Modelling to Research the Effects of Transportation Networks on Epidemic Outbreaks

Yue Li

This dissertation is submitted in part requirement of
the requirements for the degree of MSc in the Centre
for Advanced Spatial Analysis, Bartlett Faculty of
the Built Environment, UCL

Abstract

It deserves to include transportation network in epidemic researches. If human is the vector of an infectious disease, then it is possible to be spread out following the trajectory of the people hosting it. Transportation is the sum of trajectories of people's movement. But it is hard to simulate such dynamic environment that incorporates both macroscopic system and microscopic individuals. This work attempts to use NetLogo to construct an Agent-based model that simulates the epidemic outbreaks under the effects of transportation networks through SIR model and urban-rural migration theory. Experiments of different scenarios are employed, whereby different types of transportation networks, different types of diseases and different initial outbreak places. Outputs are compared among each other. The finding is the transportation with larger capacity would lead to a lower infection level and higher restriction level within the time-fixed testing. But the distributions of individuals are less even.

Declaration

I, Yue Li, hereby declare that this dissertation is all my own original work and all resources have been acknowledged. It is 10335 words in length.

Table of Contents

List of Figures

List of Tables

List of acronyms and abbreviations

ABM	Agent-based Model/Modelling
AF	Agent Focussed Concepts
G	General Concepts
ODD	Overview, Design Concepts and Details
pop	Population
SD	System Dynamics method
SIR or S/I/R	Susceptible/Infectious/Resistant

Acknowledgements

I would like to express my gratitude to my supervisor, Dr Sarah Wise, and acknowledge her invaluable expertise and guidance for helping me frame the project topic and providing advice throughout my entire research process. Her enthusiasm and understanding of the topic and kindness in helping me has been incredibly inspirational and motivating.

Chapter 1

1 Introduction

On 31 December 2019, a cluster of cases of pneumonia of unknown cause was detected (Hui et al. 2020). A terrible pandemic that still make bad influence on people was declared to have begun. Since then, people are becoming more and more concerned about the relevant knowledge of epidemic. The vectors of such infectious diseases are often human. Therefore, the spread of it to some extent comes from the movement of humans. In various ABM studies, people are normally set as agent. However, transport, which is essential if people need move, seems being omitted, as well as the transportation network, which contains the sum of trajectories of people's movement. This research attempts to catch some patterns from the effects of transportation networks on epidemic outbreaks in order to raise the awareness of the importance of transport networks in the study of epidemic transmission.

For this topic, the difficulty is to simulate a dynamic environment that incorporates both macroscopic system and microscopic individuals. Due to the complementary properties of SD and ABM, their combination is proposed for constructing the model. And it is going to set different scenarios as experiments for presenting results. Therefore, the research questions would be generated as:

Research Question 1: Under different types of transportation tools, how does the outbreaks of epidemic occur in the network change?

Research Question 2: Would the results be different if the type of disease changes?

Research Question 3: Would the results be different if the initial outbreak place changes?

In summary, these questions would be addressed throughout this report as follows: In **Chapter 2**, review academic literature relevant to simulate the epidemic outbreaks and human migration, identify the research gap and discuss how to combine them and apply to this model. In **Chapter 3**, outline the methodology by carefully describing the approaches and techniques used in the research. In **Chapter 4**, present results of experiments based on disease types. In **Chapter 5**, discuss the limitation of the algorithm and design in model and put forward possible improvements. In **Chapter 6,** make conclusion.

Chapter 2

2 Literature Review

2.1 Epidemic Simulation and SD

When the study is to discuss the spread out of infectious diseases, there is a vast amount of relevant research. Among these studies, there are several models which have been commonly used since their development to study different stages of infection. And they are still important even for recent infectious diseases. For example, in the ongoing pandemic of coronavirus disease, there is a study applying the Susceptible-Infected-Recovered (SIR) model for predicting the COVID-19 epidemic in a specific country, Algeria (Boudrioua, M.S. & Boudrioua, A. 2020). In addition, the Susceptible-Exposed-Infected-Recovered (SEIR) model has been applied to research on the contribution of community awareness to eradicate the COVID-19 successfully in Indonesia (Aldila, Ndii & Samiadji 2020). Except for these 2 models, another study has used the Susceptible-Infected-Recovered-Death (SIRD) model which is to estimate and simulate the impacts of COVID-19 on wider ranges of areas types (Fernández-Villaverde, J., & Jones 2020).

Figure 1. Sketch of SIR model
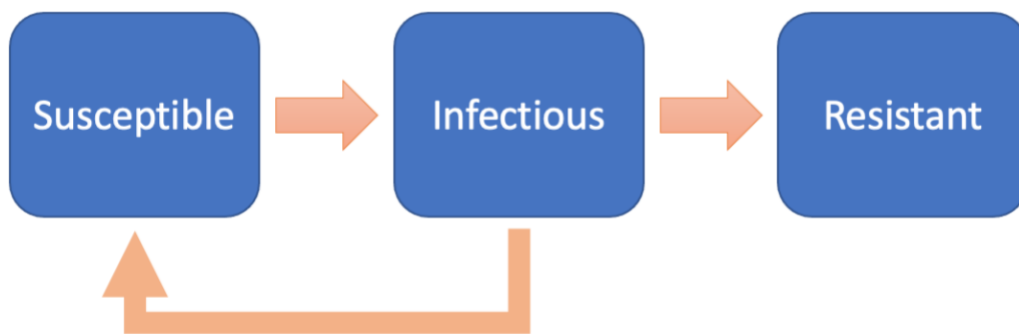
In this work, it would use the simplest model with only 3 components, i.e., Susceptible-Infected-Recovered (SIR). The essential mathematical theories needed for the model, such as the stochastic processes and differentiations, were perfectly discussed within 4 published studies (Kermack & McKendrick 1927, 1991a, b, c). In order to provide comprehensive understanding and progress the application from theories to real epidemic situations, there

was a paper reviewed various mathematical theories like this and carefully explained them and their rumours (Dietz 1967).

With the development of technology, a research suggested to build computer programs for making detailed predictions at different stages (Keeling & Rohani 2011). As the properties of the spread of epidemic could be nicely described by a system of differential equations, the dynamics of that process would then be clearly simulated. That is to say, the future state of the process could be derived from tis actual state (Gilbert, Nigel & Troitzsch 2005). This technique is called as System Dynamics method (SD). It is to deal with the problem in macro level. Aggregating the interactions of individuals, it breaks people down into interconnected groups and consider the dynamic patterns of how they shift people in different stages.

Because the assumptions of the model are part of its properties, the SD method should also follow that. In SIR model, the demographics, such as births, deaths and migration are excluded, and complete immunity is conferred by a single infection. Thus, applying SIR model by SD could work nicely to simulate the epidemic outbreaks within a fixed area but fail to model the process of migration using transportation networks.

2.2 Human Migration and ABM

To study the impacts of transportation on epidemics outbreaks, it is necessary to include the spreading of disease across the areas in the model. In reality, many infectious diseases are transferred to uninfected area by the movement of people. And transport is the indispensable tool if people want to travel, and transportation network is then formulated by the trajectory of people. However, the reasons to travel are varied and this phenomenon has been occurring from ancient times to the present. It could be natural reasons (e.g., resource scarcity, climate change) or caused by people (e.g., war, slave trade). There is an extensive literature on this topic. In this research, it would follow the classic rural-urban migration theory. It formulated an economy with 2 sectors and the migration decision is centred on balancing the expected income between the rural and urban sectors (Harris & Todaro 1970). Because they thought the main factors of migrants' decisions are economic incentives (e.g., earnings differentials, job opportunities at destinations). Based on this, another research developed further and suggested to also consider social relationships at destinations (Yap 1977). For example, people would tend to migrate if their relatives and friends stay at the destination places.

12

With the development of studies, the attractiveness of an area can be divided into many aspects. Thus, in this study, it would differ from the tradition but won't depart too far. The attractiveness is set to highly related to the population size of the area. To avoid being fully decided by population, the approach of normalisation would be used. Then it is a place with more population is more likely to attract more migrants. Similarly, as the tradition, places are divided into urban areas and rural areas. At the beginning, the urban area is set to have more people than rural area. To divide them into these 2 types, the principle to compare their number of links. Notice the network is made up by several nodes and links, with node representing area and link meaning road. Because road is created by people. For a road existed, it is either used or not yet. For used roads, they are past movement trajectories. For not yet used roads, they are made for future transporting. Therefore, the area with more links is defined as city, while the one with less links is countryside.

The difficulty and also the key point to simulate migration is because the result of this process are the emergences through microscopic actions and interactions of migrants, not simply the sum of them. And at this situation, the dependencies among the elements of a system become important, removing one such element destroys system behaviour to an extent that goes well beyond what is embodied by the particular element that is removed (Miller & Page 2009). The process is presented from bottom to up. The desirable tool to study this is called Agent-based Modelling (ABM). According to a research, this technique is used for the situation when individual entities actions would produce higher-level dynamics (MacAl & North 2010). By creating agents who are autonomous entities and interact with one another and the environment, ABM is able to capture the emergence phenomena that cannot be predicted by the combination of individual-level interactions. The reason is that the average outcomes are not sufficient, and it should include the randomness in individual level. In addition, people are divided into 3 groups using SIR model introduced in last section. The population is heterogeneous, which means people can be different from others. While modelling migration, this feature should also be included.

2.3 NetLogo and Sample Model

NetLogo is a widely used ABM simulation environment and was originally developed at the Centre for Connected Learning and Computer-Based Modelling at North-western University (Wilensky 1999). There are mountains of useful sample models provided by NetLogo.

A sample model named as "Virus on a network" (Stonedahl, F. and Wilensky 2008) using SIR model is an inspiration of this research. In that model, the author sets up a network which formulated by several nodes and links, which provides the baseline of how to design a virtual world with places and roads.

Figure 2. Interface setting up of model "Virus on a Network"
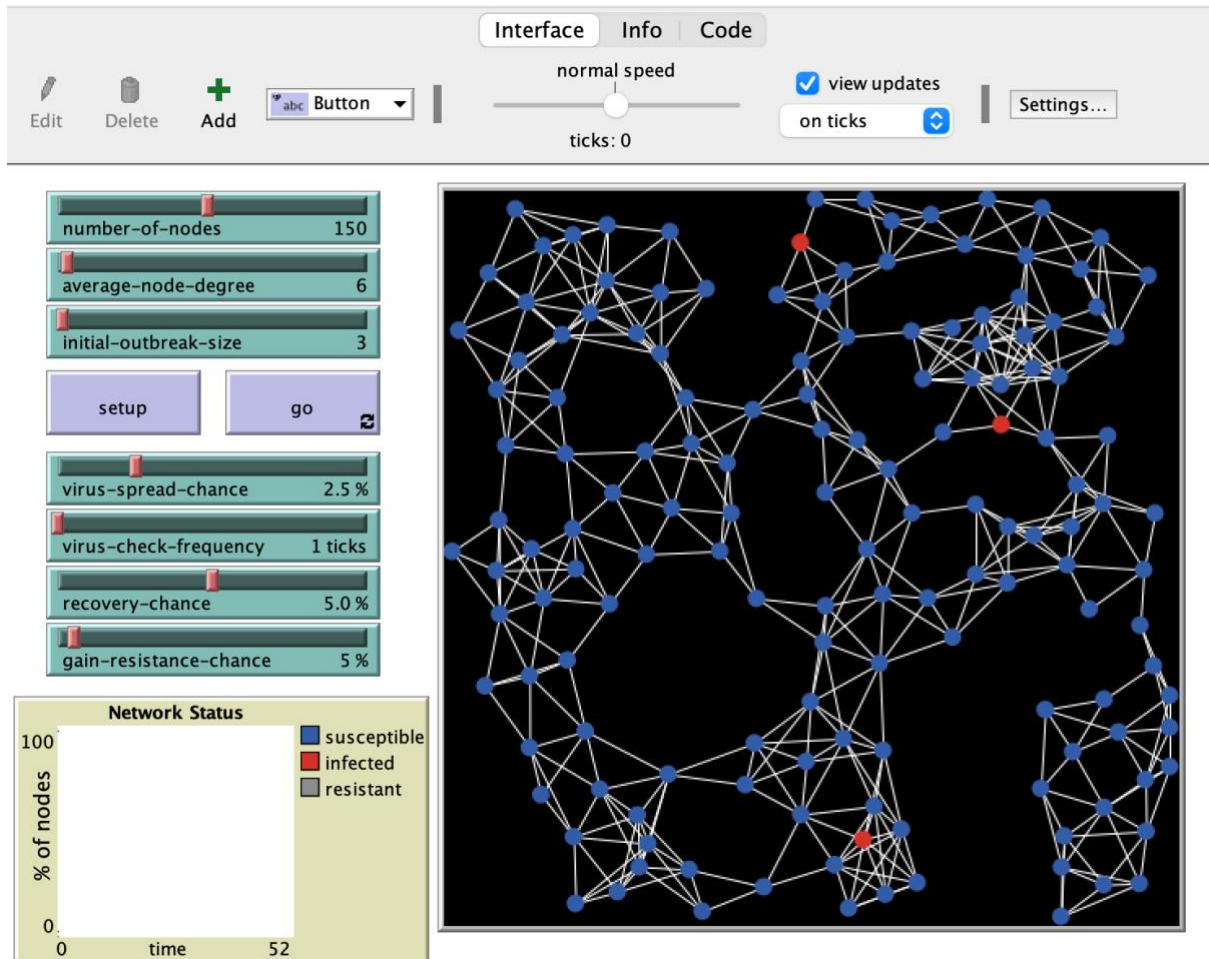
However, with a network already set, no further discussion about nodes. All nodes are treated fairly in this sample model. From the figure below, it is obvious that some nodes are clustered, while some are located lonely. In addition, while considering the processes of epidemic, many details of processes are missed. For example, since a node is infected and marked as red, no more information would be provided to describe the detailed situation. That is to say, there is no difference among these red nodes. And this also works for both susceptible and infected nodes. Besides, no migration is considered. The spread of disease is decided by a fixed figure of spread chance. What's more, when the nodes become grey, they are fully resistant, and it is impossible to be infected. But it conflicts with the setting of spread chance. Notice if there are people transferring among nodes, none of the node would be fully resistant. Because the infected tourists might move to this area. And if no tourist

moving, it is hard to figure out how to spread the virus. The more reasonable explanation is that the node is to represent people and this model is to study how the infection happens in a comparatively static environment.

To develop further while thinking about the limitation of this model, one hidden reason could be considered as the limitation of ABM method. Because the ABM model is focus on an individual level of a system, it uses inductive reasoning to present the highly disaggregated views. Therefore, it has strength in modelling endogenous interactions of individuals, but obviously it is lacking visions of macro system and its causal relationship. Nevertheless, the SD model concentrates on the dynamics of a system level, it uses deductive approach to present the highly aggregated views. This makes it suited for inferring the emergence of a certain behaviour to represent a large system, but it has to fix its structure of simulation and ignores the behaviours among users (Martinez-Moyano et al. 2007). It is generally agreed that they are both important to understand complex systems, but pursue different viewpoints (Phelan 1999). Thus, the different mechanisms are not the hindrance to research. Conversely, they can complement each other and enhance the understanding of complex systems. And there are many attempts to combine SD and ABM, such as decision making in automotive industry (Kieckhäfer et al. 2009), management in supply chain (Schieritz & Größler 2003) and aging in immune system (Figueredo & Aickelin 2010). These literatures make a solid background of this research.

In this research, the first part is modelled as a SD model to represent the epidemic outbreaks using SIR model and generate the information that would transfer to the agents. Then it is the second model, which is modelled using ABM model to illustrate the migration by applying rural-urban migration theory. The agents would absorb information and make decisions. With inspiration from the sample model "Virus on a Network", different transport would be represented by setting different parameters in initialization step. At the end, the results of simulation would be used to measure and compare the effects of transportation networks on epidemic outbreaks. Notice NetLogo is designed for ABM simulation. Due to the relationship between people and their living area, the SD part is embedded in the ABM part while coding.

Chapter 3

3 Methodology

To describe the simulation model, this section would follow a popularly used format in ABM studies, which is fully called as Overview, Design Concepts and Details (Grimm et al. 2006, 2010, 2020), and is normally abbreviated as ODD description. It has 3 subsections and those are already included in its name. Based on the logic of how they introduced, they would be used as title names and be carefully explained in its original order in the following. Notice that the authors developed their studies and there are some differences among the versions, this report would be more dependent on the latest edition.

3.1 ODD - Overview

The Overview part consists of 3 elements. For the experience of reading, they would be used as titles and follow the order as introduced in Grimm's research.

3.1.1 Purpose and Patterns

The *purpose* of the model is to understand how different scenarios of transportation affect the dynamic process of a virus spreading out within a network. In this simulation, it takes into account the iteration of epidemic within each area. The objective is to properly operate the changes of the status (S/I/R) of people while virus spreading out among them. After that, population migration among nodes is considered, while nodes are to symbolize cities or countryside. Because it represents how the infection transferring from an area to another to attack the network.

Pattern 1: Iteration of epidemic within node. This pattern reflects how the SIR status of people changes as the size of SIR population changes. In the simulation, the total number of populations is made up of the 3 SIR group of people, which follows the rules of SIR model. Recall that the model is discussed in last chapter. The main idea of it is that, at the beginning, all are susceptible. Then, the infection starts to expand, and more and more people are infected, while people who might recover from the disease and possibly gain resistance after recovery are defined as resistant group. If they cannot recover, they stay in the infection

group. And if they can recover but are not lucky to have the ability of resistance, they would send back from infection group to susceptible group.

Pattern 2: Migration of population among nodes. To illustrate how people migrate among areas, firstly, the number of tourists is selected under a rate. While deciding where to go, the size of population of each neighbour region is the major factor. The relevant theory has been introduced in last chapter. Generally, it suggests that cities would have higher attractiveness than countryside. Then, the SIR status of each tourist is influenced by the situation of their departure region. Larger infection group in the origin place would possibly cause more tourist who are able to spread out disease after arriving in new area.

3.1.2 Entities, state variables, study scales

The following *entities* are included in this study: agents representing: (a) region (i.e., city, countryside); (b) population group (i.e., SIR group, tourist group); (c) transportation links, grid cells (i.e., virtual geographical location of region); (d) the global environment representing the real state consisting of several cities. Regions represent departure and arrival places where tourist leaving and arriving. Population groups represent the situation of epidemic and migration. It is not fully satisfied the definition of entity as it is embedded within the entity of region. But for understanding the model, it should be specified as entity. Transportation links represent how people moving from one area to another. Grid cells represent the size and location of regions. The global environment is the only entity on the system level, defining how epidemic and migration happen.

Region has 1 *state variable* for types. It can be: (a) city, meaning it has higher level of links than average degree and it is marked as the shape of house in the model; (b) countryside, meaning it has lower level of links than average and it is designed as the shape of triangle in the model.

For population group, they should have state variable to describe the situation of SIR and Migration in normal case. But as embedded within region, they are not allowed to set state variables as usual. Therefore, the substitution way for illustrating its SIR situation is to set different colours for nodes. It is: (a) blue, meaning the area is fully susceptible; (b) red, meaning there exists infection and most cannot resist the disease; (c) green, meaning most are resistant. Another substitution way to describe the Migration is to set different colours for links. It is: (a) blue, meaning all tourists transporting through the line are susceptible; (b) red, meaning there exists at least one tourist is infected; (c) green, meaning all tourists migrating by the line are resistant. Notice all links are undirected. To explain a bit more is that for 2

nodes, say A and B, it is possible that A is the destination of tourists from B and at the meanwhile, B is also the destination of tourists from A. The rule is even at least one of the 2 tourist groups is infected, the link would be coloured by red. Besides, due to embedded reason, the agentsets of population group are stored in region. In details, they are the number of all types of population groups, the destination, and the status of tourists.

For the rest, none has *state variable*, but has lists (in NetLogo, agentsets). The observer controls the global variables and similarly has no state variable but agentsets. And its agentsets is to keep the record like some ratios of population which change over time.

Apart from above, while considering the *scale* in the model, it is fully specified by users. In details, the spatial scale in the simulation is the range of "world" in NetLogo and it in the reality could vary. Within this research, it is set as a no wrapping "world", which means neither node nor link could across the border. However, it is possible to set it as wrapped if the user wants to discover. And for time scale, this model is set to stop when time tick is 100 and then explore the results. The definition of time tick in the model is defined as the number of runs. Similarly, the figure and definition could also be changed through codes. If the user wants to, they can even change the stop condition.

3.1.3 Processes and scheduling

The model is developed to cover the whole cycle of a virus spreading out. It is structured in 3 *processes*: one related to the initial outbreak, one concerning SIR population group within area, the last one performed by tourist population group across places. Initial outbreak only occurs at the beginning of the simulation. The SIR and tourist population group update their amount every time step over the whole simulation.

The simulation is *scheduled* to start at initialization set up by the user through the parameters. The nodes of network are scheduled first because the links depend on the distribution of nodes. After creating the network, it is able to decide the type of the region, urban or rural area, because they are selected by their number of links. It then setting up the initial outbreak areas by selecting vertexes because infection cannot happen out of thin air. At the same moment, all nodes should be set as susceptible as in this simulation the previous infection and resistance would not be discussed. Spreading out is the first action because it could be assumed as the primary activity of a virus. Then it can pick up tourists. Migration would exist among all nodes. If there is a tourist coming from the infected area and is carrying the virus, then the SIR population group in the new area is affected. By tourist, virus could attack new places. When the resistance group is getting larger and larger, the range of

places that virus can stay is getting smaller and smaller. For observer, its actions like taking record and plotting take place at every time tick and action like writing output files would happen at the final stage.

3.2 ODD - Design Concepts

It is known to all that a model consists of many concepts. This section is going to explain them. In Grimm's opinion, they could be divided into 2 types, general and agent focussed. One annoying thing is some could be both general (G) and agent-focussed (AF). So, it would be messy to list and discuss them by types. It should be better to introduce them one by one following the logic that is from the start of the simulation to the end and use G and AF to mark their type.

3.2.1 Basic Principles (G)

This model is extended from a classic model of epidemiology which is known as SIR model. This problem is to estimate the risk of the population of people changes within a certain time of period under the threat of a virus spreading out, and how that risk could be reduced by means. Additionally, this model considers people migration. Therefore, the idea that it was designed to illustrate transportation type could be necessary to consider how epidemic would spread out within a network. Understanding this can be critical to deal with the threat of virus.

3.2.2 Emergence (G)

One expected outcome of the model is people would be more likely migrate to the areas with larger size of population group. These areas might be cities at the early stage of model. To illustrate this further, it is necessary to consider another possible result – the gap of population size between urban and rural areas might decrease to a low level. As areas are divided by the number of their links rather than their amount of people, it is possible to have some countryside with larger size of population than some cities. And those outcomes might have impact on the results of the SIR model.

3.2.3 Adaptation (AF)

No adaptive behaviour would happen through the whole process.

3.2.4 Objectives (AF)

In this research, people are set as no objective, as well as the places. Because infection and migration are controlled by global variables.

3.2.5 Learning (AF)

Individuals would never learn in this simulation.

3.2.6 Prediction (AF)

There is no prediction in this model.

3.2.7 Sensing (AF)

For people in the same place, they are assumed to perfectly sense the current situation of the epidemic. However, the sensing range is fixed overtime, that is where they are. So, they are not able to sense the situation of the epidemic in the other areas. The only exception is for tourists. They would be allowed to sense the population size of the linked neighbours. That is for them to choose destination. As same as the other people in the origin, they know nothing about the SIR information of other places. Note that for a tourist group transferring from area A to area B, before moving they belong to area A and after moving, they belong to area B. The migration is designed to happen immediately and once a time step.

3.2.8 Interaction (AF)

There are 2 kinds of interaction in this model: among urban and rural areas, and among population within a node. The places interact directly with tourists by deleting them from its population if it is the origin or adding them to its population if it is the destination. At each time tick, every node would send their tourist out. However, not every node would have tourist coming. The more attractive places might have several tourist groups coming. Within every node, the second interaction that occur among people would happen indirectly and the results of infection iterations are calculated using the SIR model. Note that all calculation about tourists should be generated after the SIR process. And for nodes, the process of SIR iterations is all calculated independently.

3.2.9 Stochasticity (G & AF)

Stochasticity would occur in 3 forms. First, the coordinates of each node and their links are initialized stochastically because there is no fixed pattern to locate nodes and formulate a network. It follows the same way of the sample NetLogo model "Virus on a network". The other 2 forms are the type of AF and not completely random. They could be influenced by some factors. One is to choose the destination of each tourist group. It is decided under the influence of the population size of the neighbour areas around its departure place. Notice that in the model it allows decimals in calculating population. But there are extra rules to make the final results as integer numbers. Note that when population is lower than 1, no tourist group would be selected. The last form of stochasticity exists in deciding the SIR status of

each tourist. The decisions of individual are made independently. Similarly, as choosing destination, the status would be influenced by the SIR population size of their origin area.

3.2.10 Collectives (G & AF)

This model includes 2 kinds of collectives, SIR groups that formulate the whole population of the place and tourist groups that are generated from the SIR group in each area. The collectives are represented as specific kinds of model entity with their own agentsets and behaviours, which are included in above section. These collectives are included in the model because turtles are already occupied by areas. If need to consider both, it is easier to model these as internal entities.

3.2.11 Observation (G)

Because the observations of this model cannot be derived directly, adjustments are needed. In specific, the direct result would be population but due to different population size, making comparisons directly on population is meaningless. Therefore, please refer to 3.3.5 for details of explanation and calculation.

3.3 ODD - Details

3.3.1 Initialization

Figure 3. NetLogo Interface of this model after completing initialization
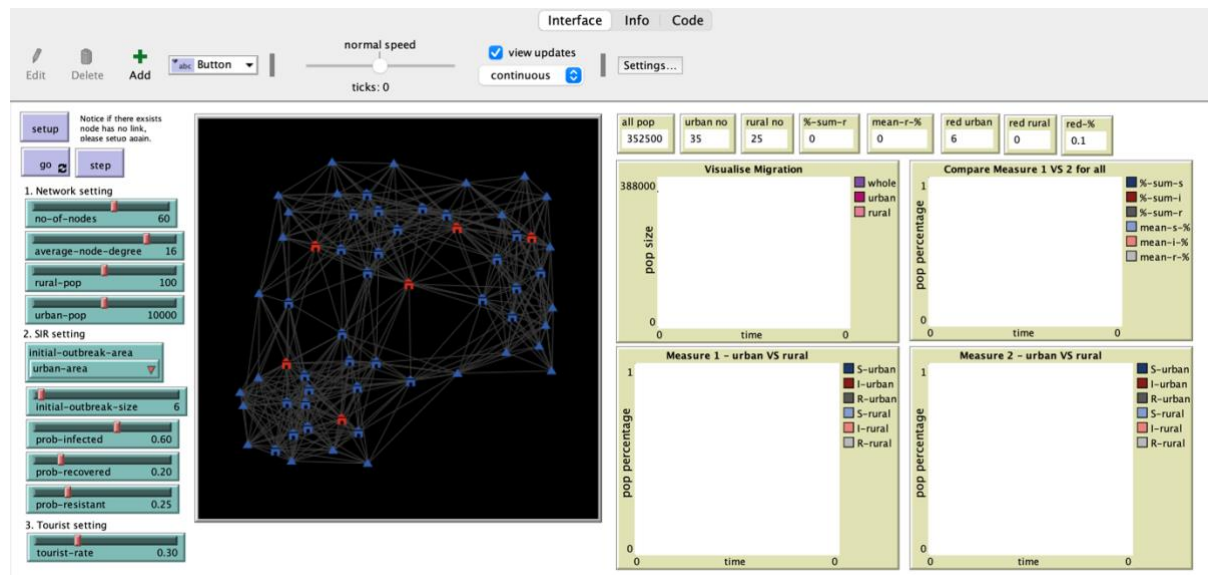
To operate the model, it is necessary to set up values for several variables beforehand. This part is to introduce this step, the start of the simulation. The variables to be set initially in this research are all called as global variables in ABM and a screenshot of the model's interface in NetLogo after finishing this step is shown above.

The 10 variables coloured in green and listed on the right-hand side are the parameters to set in initialization step. The 3 purple ones above them are the commands. The picture with black background would be the place to show the animation of this model, which is called "world" in NetLogo. The 4 large squares on the left-hand side are to plot the line graphs of this research. The 8 squares above them are the monitors to show some figures. Therefore, these 12 squares are the results of the simulation. However, the way to present results in this report would be different from these. As the plots and monitors are just from a single plot, they are inconclusive. Thus, for the stability of results, the model would be run for 30 times and use the average value of the outputs. The tool for this is called "BehaviorSpace" in NetLogo. It would run the model as different experiments. The variables needed to be set up in initialization for experiment are the same as those for a single model. Recall the aim of this research is to study the effects of transportation networks on epidemic outbreaks. According to the research questions discussed earlier, there are 3 things that should be set differently at beginning for making comparison later. One is different transportation methods, and one is different types of disease, the last one is different initial outbreak places.

To provide an overview, it firstly lists the parameters with definitions and range, then explain how to represent different types of transport method, virus, and initial outbreak place.

Table 1. Parameters in Initialization

VARIABLES	DEFINITION	RANGE	Value for Test
no-of-nodes	The number of areas to be created in the network	[1, 100]	60
average-node-degree	The average degree of links for nodes in the system	[1, 20]	10
rural-pop	The amount of population in each rural area	[1, 200]	100
urban-pop	The amount of population in each urban area	[1, 20000]	10000
initial-outbreak-area	The type of area where the initial outbreak would happen	{urban-area, rural-area}	/
initial-outbreak-size	The number of areas occurring the outbreak at start	[1, 100]	6
prob-infected	The probability of being infected if susceptible	[0, 1]	/
prob-recovered	The probability of being recovered after infection	[0, 1]	/
prob-resistant	The probability of being resistant after recovery	[0, 1]	/
tourist-rate	The percentage of people that would migrate	[0, 1]	/

From the table above, the first 2 columns introduced the name of those variables and their definitions. The 3rd column, named as "RANGE", shows the full range of value that can be chosen. For a single which could be initialized in the Interface of Netlogo, they are marked as "slider" and "chooser". But the range would not be included in BehaviorSpace. Notice not all range of values need to be used. Therefore, the 4th column, named as "Value for Test", indicates the values used in experiments. The sign of "/" means the corresponding variables would be used for representing different scenarios. And the reason to set the rest variables as fixed figures is for the simplicity of further analysis. Recall this simulation is formulating a virtual world. It is only required to set figures relatively reasonable. Notice the codes would be provided in the appendix, if user want to amend them, it is completely editable.

For considering different transportation methods, one thing should be discussed is that the development level of transportation in different countries varies a lot. For developed states, they are generally and normally having more complexed transportation networks with larger capacity and higher effectiveness than less developed countries. Therefore, it is impossible to refer to real-world data while setting parameters. The values of parameters are considered sufficient as long as they can make sense in logics.

Table 2. Scenarios for transportation methods

VARIABLES	DEFINITION	Scenario I	Scenario II	Scenario III
tourist-rate	The percentage of people that would migrate	0.05	0.15	0.3

In this research, transportation methods are defined by "tourist-rate", with 3 scenarios from low, medium to high levels to represent 3 types. Without considering extreme cases, such as wars, people won't migrate on a large scale. Therefore, the high level is set as 0.3 and the medium is set as its half. For setting the low level, it shouldn't be too small otherwise it would be too hard to capture the feature of migration.

At early stage of this study, it attempts to set different values for the parameter "average-node-degree", because different modes of transport have different levels of density. However, it is finally abandoned due to its huge influences on network structure. For example, the pattern of urban and rural areas would then be different. Recall the way to divide urban and rural area is to compare their number of links with this parameter. That is to say, if still use it, it would be unacceptable to compare the results from different scenarios. The substitution way is to also consider this while selecting values of "tourist-rate". Specifically, the figures of different scenarios are set to not too close to each other.

For different types of epidemics, they could be expressed by different groups of SIR probabilities. The idea to design these 3 viruses is they all have to infect people easily. People would not care a less likely infected disease as it is hard to harm them. Followed by that, there should be 4 types of scenarios. However, the scenario which is easy to recover and gain resistance is ignored. Because it has much less harm to people than the others. Then for the 3 scenarios, Virus 1 represents the disease which people could easily recover from but hardly be resistant after recovery, Virus 2 indicates the illness which people would hardly recover from but easily gain resistance and Virus 3 means people would both hardly recover and have resistance. The detailed figures are shown in the table below.

Table 3. Types for types of disease

VARIABLES	DEFINITION	Virus 1	Virus 2	Virus 3
prob-infected	The probability of being infected if susceptible	0.7	0.7	0.7
prob-recovered	The probability of being recovered after infection	0.55	0.2	0.2
prob-resistant	The probability of being resistant after recovery	0.2	0.55	0.2

For considering the places where would occur infection at the beginning, the option is either all in urban areas or all in rural areas. The table is shown below.

Table 4. Scenarios for initial outbreak places

| VARIABLES	DEFINITION	Scenario I	Scenario II
initial-outbreak-area	The type of area where the initial outbreak would happen	urban-area	rural-area

3.3.2 Input data

This part is to introduce and explain the input datasets. Notice in last part, there are several parameters introduced. But in ODD description, the input for initialization is different from the input in this part. The meaning of input data in this section is about external source or figure. One thing to emphasize is that all data in this research are synthetic. After setting up the step of initialization, the model can run properly. So, no input data is requested in this simulation.

3.3.3 Sub Model 1- Internal SIR

This process calculated internally should always be finished at the beginning of running simulation as the following calculation need the updated value from this process. At each time step, the people within a place are experiencing the process described below.

Figure 4. SIR option tree for an area at each tick
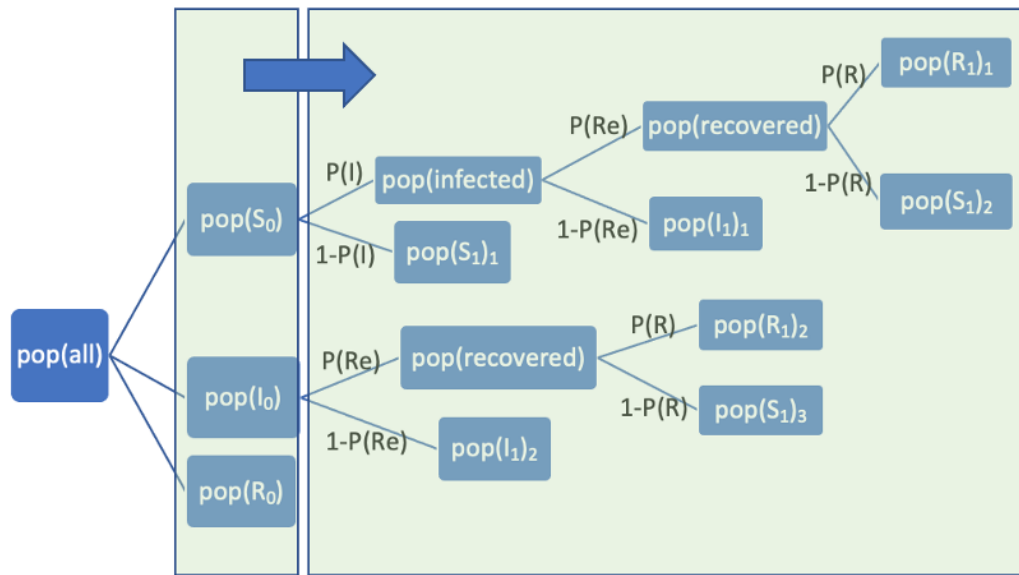
At first, "pop(x)" used above means population of x and "P(x)" means the probability of x. The rest expressions are defined below:

1. pop(all) the number of people in this area.
2. pop(S_0) is the number of people who is susceptible at the beginning of the time tick.
3. pop(I_0) is the number of people who is infectious at the beginning of the time tick.
4. pop(R_0) is the number of people who is resistant at the beginning of the time tick.
5. P(I) is the probability of getting infected by the disease. (Recall it is the global variable "prob-infected" in table 1):
6. P(Re) is the probability of being recovered from infection. (Recall it is the global variable "prob-recovered" in table 1).
7. P(R) is the probability of receiving resistance after recovery. (Recall it is the global variable "prob-resistant" in table 1).
8. pop(infected) is the number of people who were susceptible and now are infected.
9. pop(recovered) means the number of people who were infected and now are recovered. (Note it has been used twice for people from different initial group).
10. pop(S_1)$_1$, pop(S_1)$_2$, pop(S_1)$_3$ means the number of people who are susceptible after running the SIR calculation. (The figure inside the bracket means the time is at the end of the tick. And the one outside is to distinguish new susceptible groups).
11. pop(I_1)$_1$, pop(I_1)$_2$ means the number of people who are infected after running the SIR calculation.
12. pop(R_1)$_1$, pop(R_1)$_2$, pop(R_1)$_3$ means the number of people who are resistant after running the SIR calculation.

25

With those definitions above, it is able to present and explain the equations illustrated in the table 1 and the last process made at the end of this tick.

Equation 1 For people who come from group S at the start of the tick

$$pop(S_1)_1 = pop(S_0) * [1 - P(I)]$$
$$pop(I_1)_1 = pop(S_0) * P(I) * [1 - P(Re)]$$
$$pop(R_1)_1 = pop(S_0) * P(I) * P(Re) * P(R)$$
$$pop(S_1)_2 = pop(S_0) * P(I) * P(Re) * [1 - P(R)]$$

For people who were in group S: (a) If they are not infected, then the 1st new group of susceptible people is defined; (b) When they become infected but cannot recover, then the 1st new group of infected people is generated; (c)When they recover from infection and gain resistance, they are the 1st new group of resistance people; (d)If they recover from infection but cannot have resistance, they are the 2nd new group of susceptible people.

Equation 2 For people who come from group I at the start of the tick

$$pop(I_1)_2 = pop(I_0) * [1 - P(Re)]$$
$$pop(R_1)_2 = pop(I_0) * P(Re) * P(R)$$
$$pop(S_1)_3 = pop(I_0) * P(Re) * [1 - P(R)]$$

For people who were in group I: (a) If they cannot recover, they would stay in the status of infection and become the 2nd new group of infected people; (b) If they get both recovery and resistance, they would be the 2nd new group of resistant people; (c) If they only have recovery and miss resistance, they are defined as the 3rd new group of susceptible people.

Equation 3 Groups at the end of the tick

$$pop(S_1) = pop(S_1)_1 + pop(S_1)_2 + pop(S_1)_3$$
$$pop(I_1) = pop(I_1)_1 + pop(I_1)_2$$
$$pop(R_1) = pop(R_1)_1 + pop(R_1)_2 + pop(R_0)$$

The first two is easy to generate from above. For the 3rd line, the last element on the right-hand side is the number of people who are already in group R at the beginning of the time step. Due to the definition of resistance, they would keep status in resistant. At this stage, no need to calculate and update the overall population as it won't change. However, if without further changes, the problem is the number of people generated from above processes would not be integer numbers. Therefore, to sort it out, the equation to update population for group S/I/R should be like:

Equation 4 Updated SIR population groups in integers

$$pop(I_1) = floor[pop(I_1)_1 + pop(I_1)_2]$$

$$pop(R_1) = [pop(R_1)_1 + pop(R_1)_2 + pop(R_0)]$$

$$pop(S_1) = pop(all) - pop(I_1) - pop(R_1)$$

The updated population of group I and R would be calculated using floor function to get the integer numbers. In order to keep the overall population in this area as the same before running the SIR process, population of group S would be calculated at last step by minus new population size of I and R from the overall population. Thus, all figures would be integer numbers at the end. However, this way would cause some bias and more discussion about it would be in chapter 5.

### 3.3.4 Sub Model 2 - External Migration

For migration, first step is to generate the total number of tourists for each node. Its equation is shown below:

Equation 5 Count the number of tourists

$$no(tourists) = floor(TouristRate * population)$$

The size of tourist group is the product of the rate multiplying the overall population of the node.

Figure 5. Flow chart for deciding tourists
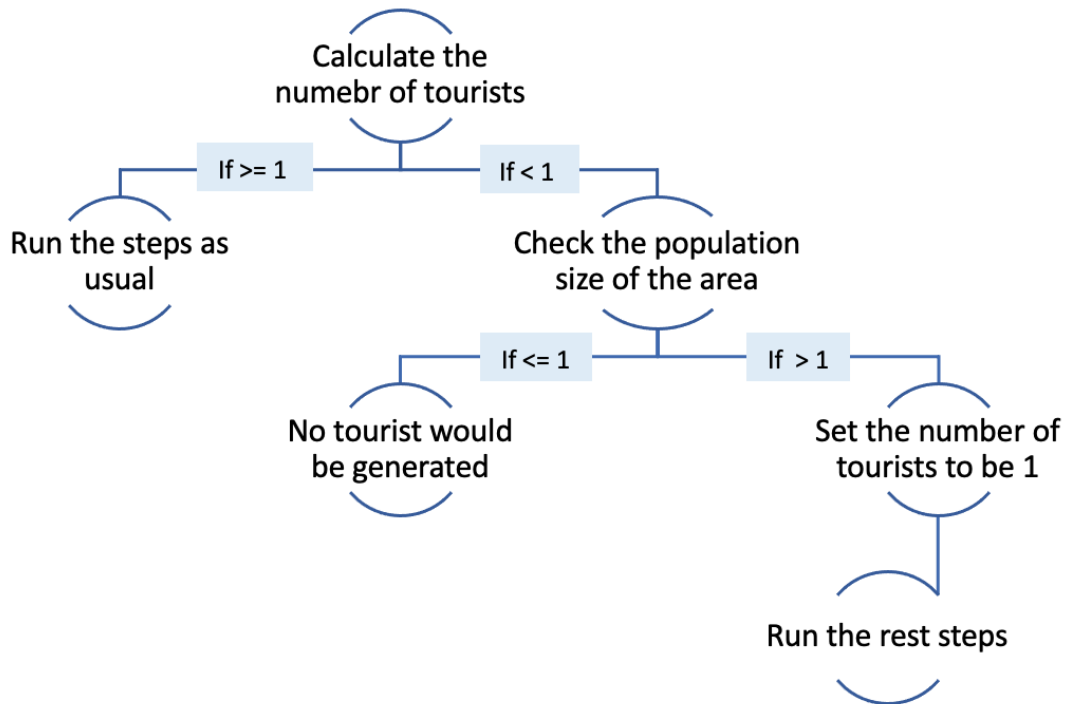
However, to ensure the integer number for population, it should use floor functions for the product. However, there is a hard part in this step. The case would happen when the results of the product inside the floor function is less than 1. Then the number of tourists would be 0. The places would then stop to send tourists out. For example, if there are 8

27

people in an area and the tourist rate is 0.1, if no change, there is no tourist. But it is still able to have tourists. In order to solve this problem, it would firstly check whether the result of the floor function is bigger than or equal to 1, if yes, no further step needed. If not, check whether the current population at this area is bigger than 1, if yes, directly set the size of tourists as 1. If not, no tourist would be generated. A flow chart of how to check is shown above.

Then it is important to figure out 2 things: (a) the destination of each tourist group; (b) the status of individual in each group.

As mentioned before, destination is decided after considering the population size of all linked neighbours. For easy explanation, for every node, assume they have 2 lists. One stores all population of the linked neighbours. The other contains all index number of the linked neighbours. And the number of populations in the first list could correspond to the index of node in the second one. Define them like: N_{POP} = {pop_1, pop_2, pop_3, …} and N_{index} = {i_1, i_2, i_3,…}. The reason to set the index list is because there is no way to complete such process using the current features in NetLogo.

The next step is to normalise the list N_{POP}. Firstly, every element would be divided by the sum of the list. Then, from the 2nd element, they would add all elements in front of them. The normalised list would be like: (Notice the last element must be 1)

Equation 6 Normalise the population of linked neighbours

$$normalisedN_{pop} = \frac{pop_1}{sum(pop_1 + pop_2 + \cdots)}, \frac{pop_1 + pop_2}{sum(pop_1 + pop_2 + \cdots)}, \ldots, 1$$

Then, set normalised-N_{POP} = {$npop_1$, $npop_2$, $npop_3$, …}. Notice N_{POP} is not in sorted order, but normalised-N_{POP} is in ascending order. Pick a random value p between 0 and 1, for the first element in the normalised-N_{POP} list that is bigger than the value, the corresponding neighbour is the destination. However, there is also no ready-made function to finish this process in NetLogo. The substitution way is to firstly insert the random value to the normalised-N_{POP} list, sort the new list ascendingly and get the position of p in the new list. Because the normalized list is sorted, this process would let p replace the position of the first element that is larger than p. Then it is easy to get the position index of p in the new list and find the corresponding index in the N_{index} list.

To choose the status for tourist, it is important to make the decision independently. In other words, the status of individual in tourists' group is only related with the current SIR population size in the origin place, not about the status of the other tourists. For easily describing the process, define 2 lists for each vertex: $Rate_{sir}$ = {S%, I%, R%}; $Indicator_{sir}$ =

{S, I, R}, where S%/I%/R% is the ratio of susceptible/infected/resistant people divided by the overall population. Then the normalised list would be calculated as:

$$normalisedPOP_{sir} = S\%, (S\% + I\%), (S\% + I\% + R\%)$$

$$(S\% + I\% + R\%) = 1$$

Similarly, like discussed in last paragraph, normalize the Rate$_{sir}$ list and select a random value q between 0 and 1. And find the first element that just larger than q. The alphabet that has the same position in the Indicator$_{sir}$ list is the status for this tourist. The way to decide independently is to run the process above for several times. The number of tourists is the number of repetitions. After that, store the status of all tourists from the same origin to a new list and count the occurrence of the 3 alphabets. The occurrence of each is the number of tourists in each status.

The calculation to update the population of both origin and destination places is easy to make. For the node who is origin, delete the number of SIR tourists from the SIR group, and update the total number of people in that node. For the node who is destination, add the number of SIR tourists to its SIR group and sum to have the current number of overall populations.

3.3.5 Sub Model 3 - Monitoring

This simulation would stop when the time tick is 100. Then it is generally to compare the level of infection and resistance. And there are 2 possible measurements, and both need extra calculations.

One is from the general level. For different types of area: (a) all area; (b) all urban areas; (c) all rural areas, get the number of people who is I/R and divide it by the overall population at the end tick. The detailed calculations are listed below.

$$\%SumPopI = \frac{overall\ infected\ population}{overall\ population}$$

$$\%SumPopR = \frac{overall\ resistnat\ population}{overall\ population}$$

$$\%SumPopIurban = \frac{overall\ infected\ population\ of\ whole\ urban\ areas}{overall\ population\ of\ whole\ urban\ areas}$$

$$\%SumPopRurban = \frac{overall\ resistant\ population\ of\ whole\ urban\ areas}{overall\ population\ of\ whole\ urban\ areas}$$

$$\%SumPopIrural = \frac{overall\ infected\ population\ of\ whole\ rural\ areas}{overall\ population\ of\ whole\ rural\ areas}$$

$$\%SumPopRrural = \frac{overall\ resistnat\ population\ of\ whole\ rural\ areas}{overall\ population\ of\ whole\ rural\ areas}$$

By using these, it is easy to know the situation of the whole (urban/rural) places. But it is not enough. For example, 90% of people are resistant in the system. Is it able to say that's the end of the outbreak? Of course not. If the rest 10% of infection population is clustered in one node, that would ruin that place and cause the second outbreak of the disease. Compared to the overall population of the whole network, the population of each node is a small number, especially for node who is rural area. Therefore, the observations caring more about the level of individuals are necessary.

That measurement is calculated based on the level of each agent. Before summing up, each node calculates its own rate of I/R and then get the mean value of them based on different types of areas. The detailed calculations are listed.

Equation 9 Measurement 2

$$AvgPopI\% = mean(I_1\%\quad I_2\%\quad …)$$
$$AvgPopR\% = mean(R_1\%\quad R_2\%\quad …)$$
$$AvgPopI\%urban = mean(I_{urban1}\%\quad I_{urban2}\%\quad …)$$
$$AvgPopR\%urban = mean(R_{urban1}\%\quad R_{urban2}\%\quad …)$$
$$AvgPopI\%rural = mean(I_{rural1}\%\quad I_{rural2}\%\quad …)$$
$$AvgPopR\%rural = mean(R_{rural1}\%\quad R_{rural2}\%\quad …)$$

Notice $(I_1\%\quad I_2\%\quad …)$ means the infected rate of place 1, place 2 and the rest and $(I_{urban1}\%\quad I_{urban2}\%\quad …)$ means the infected of place 1 in urban are, place 2 in urban area and the rest place in urban area. Similarly for the rest.

However, it is still possible to have outliers but hard to know, while just comparing the mean values. The variance is suggested. At later stage, it is most likely that most of the population are resistant. Then only calculate variance of the resistance group.

Equation 10 Measurement 2 extra

$$VarPopR\% = var(R_1\%\quad R_2\%\quad …)$$
$$VarPopR\%urban = var(R_{urban1}\%\quad R_{urban2}\%\quad …)$$
$$VarPopR\%rural = var(R_{rural1}\%\quad R_{rural2}\%\quad …)$$

With results from both general and individual, it could take the most factors into consideration and then draw reasonable conclusions.

Chapter 4

4 Results

In this chapter, it will present the simulation results for different scenarios using the tool "BehaviorSpace" from NetLogo. Recall there are 3 types of transport tools, 3 types of viruses and 2 types of initial breakout places. Therefore, it is to firstly list the results and make the comparisons by virus type. After that, there would be a part to summarise the intersections and differences generated before and discuss the findings.

For figures, it is going to use histograms to show the results. Notice the rectangles are coloured by green or blue to distinguish the outbreak whether initially happened at urban or rural areas. And there are 6 elements in legend. They are to represent the 3 types of transportation networks in 2 different scenarios of initial outbreak places, which is 6 in total. Recall the 3 types are: low; medium; high. They are the different level of capacity of transportation tools with their figures accompanying.

4.1 Virus Type 1

Virus 1 is defined as the epidemic that people could easily recover from but hardly be resistant after recovery.

Figure 6. Virus 1 - Visualise I level at end tick
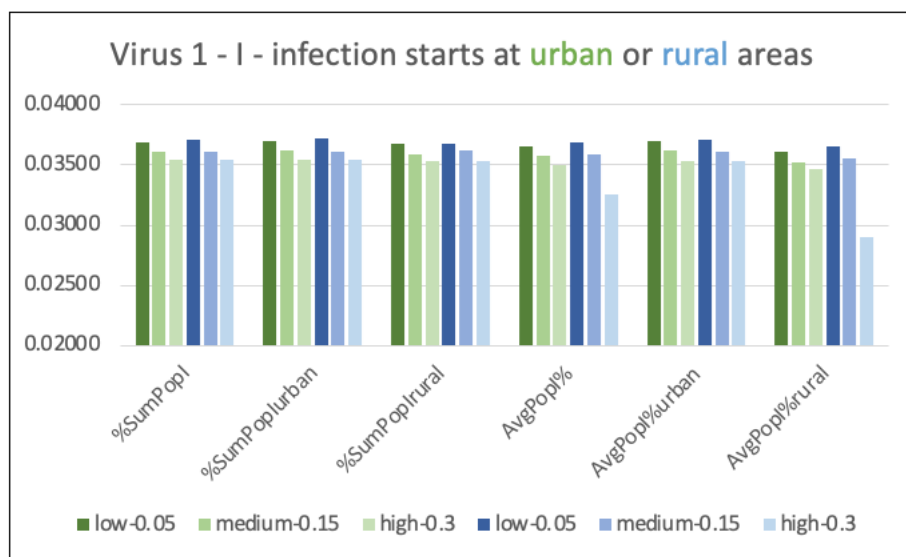
For better view of visualisation, the y axis is set from 0.02 to 0.04. Generally, the infection level of all scenarios using both measurement 1 and 2 would be around 0.03. To all extent, that could be considered as low level of infection. There are few people infected in the system by checking the left half of the graph. And they are not clustered but distributed averagely through comparing the right half of the plot. For general level of infection in either different initial outbreak places or different transportation types, the difference is very small.

However, while considering the infection from individual level, the difference is rapidly increasing, especially for the initial outbreak happening at rural areas and the high level of capacity of transportation. It seems like with the capacity level of transportation increasing, the level of infection tends to decrease. And if the epidemic starts at rural areas, this effect seems stronger than it happens at urban areas. To support this, it needs to also consider the level restriction in the system.

Figure 7. Virus 1 - Visualise R group at end tick
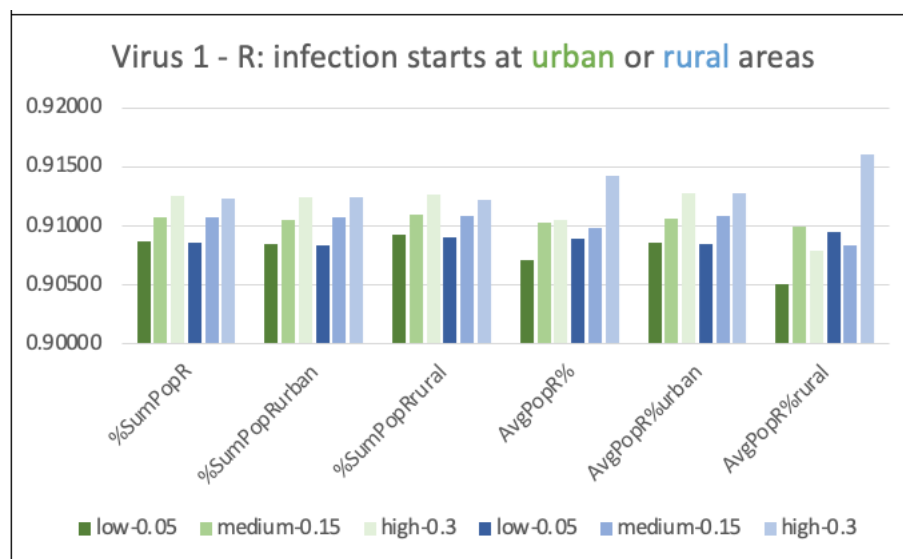
Similarly for the reason of improving the experience of viewing, the y axis is set not from 0 but 0.9 to 0.92. If the disease firstly occurs at urban areas, the restriction level would be around 0.91. And generally, the higher capacity the transport has, the higher level of restriction the population would be. And if it happens at rural areas, the restriction level is even higher. The only exception is to consider the restriction of rural areas from individual level. But generally, it could generate the similar conclusion as the one generated from considering infection level.

Table 5. Virus 1 - Variations of R groups at end tick

Outbreak Area	Transport type	VarPopR%	VarPopR%urban	VarPopR%rural
urban	low-0.05	0.001867293	0.000011729	0.004307262
	medium-0.15	0.000959049	0.000014499	0.002145676

|				
	high-0.3	0.002350632	0.000024885	0.005126764
	low-0.05	0.000025685	0.000010949	0.000044663
rural	medium-0.15	0.001430242	0.000017953	0.003292190
	high-0.3	0.006313227	0.000024945	0.014099042

Due to the exception occurs, it is necessary to consider the variation of R groups. From the whole table, all variations of measuring R level in urban areas are stable and much smaller than the others. And the larger values are often the high capacity of transport.

That is to say, for an epidemic, which people could easily recover from but hardly be resistant, within the same modelling time, the results of transportation that could migrate more people would generally better than the one with lower capacity, especially when the epidemic occurs in rural areas at the beginning. However, it always has a larger variation, which indicates its elements are not distributed evenly.

4.2 Virus Type 2

For Virus 2, people would hardly recover if they were infected, but could easily gain resistance if they can recover.

Figure 8. Virus 2 - Visualise I level at end tick
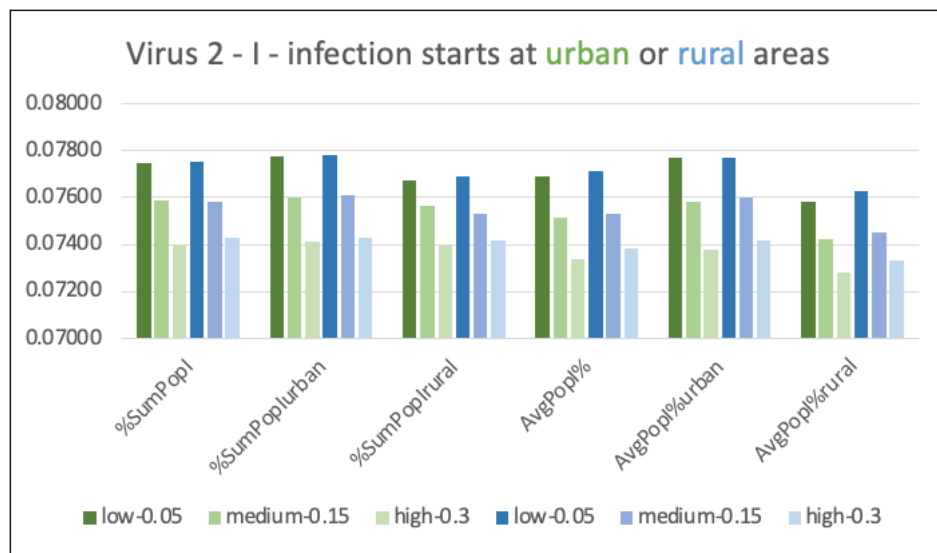
For better view of visualisation, the y axis is set from 0.07 to 0.08. And the infection level of all cases is around 0.075. Compared to Virus 1, it has increased a lot. But it is still a low level of infection in general. And different from Virus 1, the rectangles are not that close in values. The infection level in urban area is obviously higher than other areas. The difference between initial place at urban and rural areas are not evident.

However, while considering the scenario of transportation types, the pattern is similar. When the capacity level goes up, the infection level goes down. Differently, when the epidemic starts at urban areas, this effect is stronger.

Figure 9. Virus 2 - Visualise R level at end tick
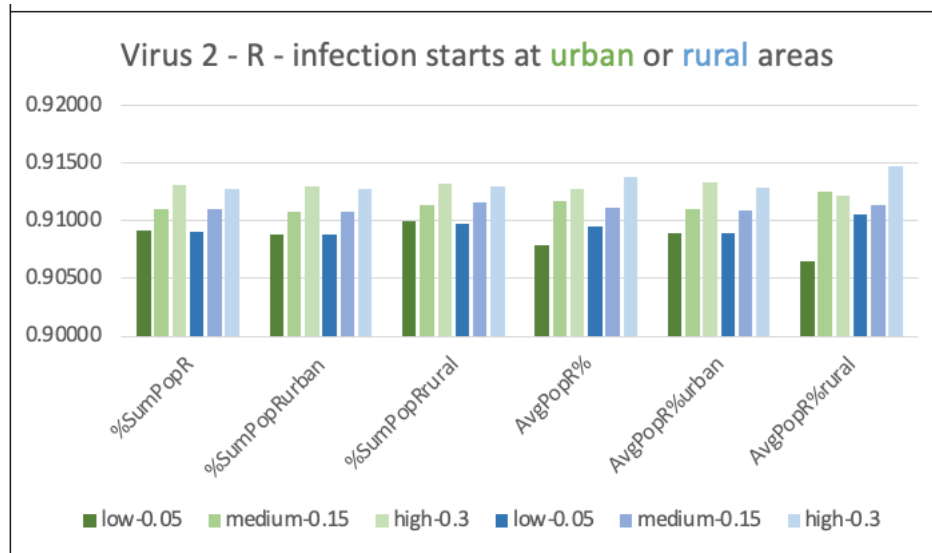
Similarly, the y axis is from 0.9 to 0.92. Generally, most results of R level are distributed evenly around 0.91. Only the one that measures the R level of rural areas from individual level varies larger than the others. However, all of them follow that: the R level would follow the increase of transport capacity to increase. Because high level of restriction in population means the relatively low level of infection in the system. The plots of R level and I level illustrate the same thing. Before making summary of the patter, it is necessary to review the variations of restriction.

Table 6. Virus 2 - Variations of R groups at end tick

Outbreak Area	Transport type	VarPopR%	VarPopR%urban	VarPopR%rural
	low-0.05	0.000029599	0.000012419	0.000053946
urban	medium-0.15	0.000499345	0.000015487	0.001176597
	high-0.3	0.000804626	0.000024417	0.001862758
	low-0.05	0.001412869	0.000011809	0.003423466
rural	medium-0.15	0.000069013	0.000016389	0.000135083
	high-0.3	0.000986566	0.000023550	0.002102158

From the table above, the first half of outbreak area in urban places follows the same pattern of Virus 1. The variation level would be higher if the capacity of transport is going up. However, if it happens in rural area, the results vary, and it is hard to illustrate the pattern. Because none of them is too big to reject the deduction made from last 2 plots of Virus 2. It is reasonable to conclude that for a disease with low possibility to recover but high possibility to resistant after recovery, the higher capacity transportation would contribute to higher

restriction level and lower infection level within the fixed simulation time. But it needs extra attention on uneven distribution of the infection and restriction level of individuals.

4.3 Virus Type 3

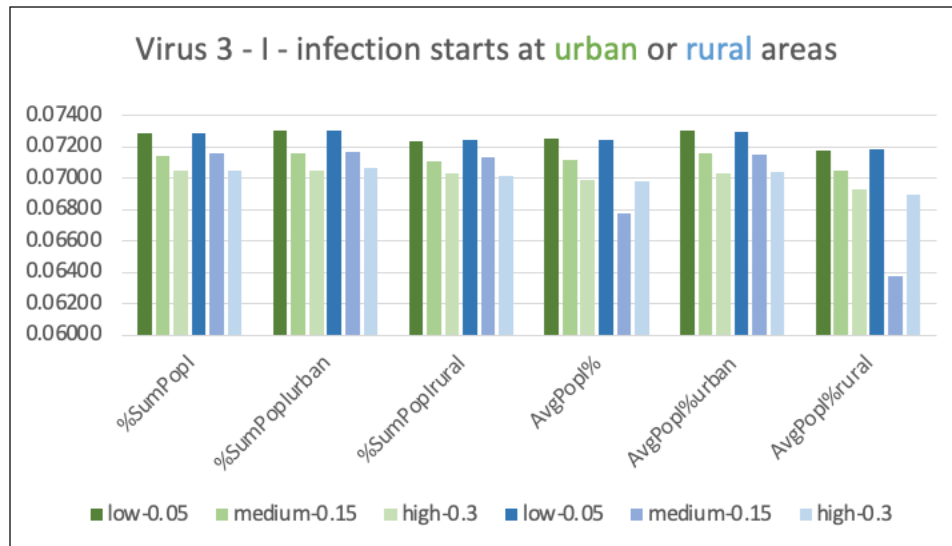For Virus 3, people would hardly recover if infected and be resistant after recovery.

Figure 10. Virus 3 - Visualise I level at end tick

For better view of visualisation, the y axis is set from 0.06 to 0.075. And the infection level of all cases is around 0.07. The results of measuring individual level that initial outbreak areas are rural places are different from the ones from overall level. For the rest, it is still that higher capacity of transport indicates lower infection level. To consider the exception, the rule would make sense if ignore the medium level of capacity.

Figure 11. Virus 3 - Visualise R level at end tick
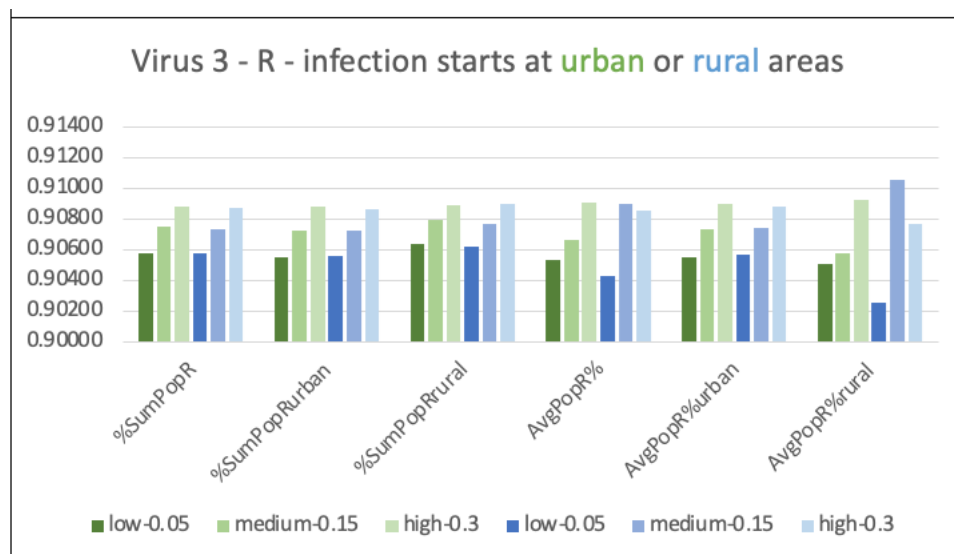
Similarly, the y axis is from 0.9 to 0.915. And for all cases, it seems around 0.905. It indicates the same pattern of the one illustrated from viewing the level of I. Without considering the medium level, the rest would follow that when the capacity of transport goes up, the restriction level would then follow. The reason of it might be found through checking the corresponding variations.

Table 7. Virus 3 - Variations of R groups at end tick

Outbreak Area	Transport type	VarPopR%	VarPopR%urban	VarPopR%rural
urban	low-0.05	0.001450194	0.000009256	0.003223543
	medium-0.15	0.010316549	0.000014377	0.020853534
	high-0.3	0.000982331	0.000021378	0.002625036
rural	low-0.05	0.000503797	0.000010283	0.001195378
	medium-0.15	0.000946616	0.000013893	0.002235126
	high-0.3	0.000593306	0.000022342	0.001280003

From above table, it is found that the variations of medium level that initially occurring places are rural areas is higher than the rest 2 types of transport. Besides, this still happens to the case that the outbreak areas are urban places, but this cannot be noticed from the plots of I and R level. Because these variations are small in values. The pattern introduced above would still make sense.

In *summary*, recall that for Virus 1, its I level is around 0.03 while R is about 0.91. And the high capacity always led to high variations. For Virus 2, its I level is nearly 0.075 and R is 0.91. The high capacity would lead to high variations only when the epidemic happens initially at urban areas. For Virus 3, the I level is around 0.07 and R is close to 0.905. Different from the previous 2, the medium capacity tends to lead larges variations. For them all, higher capacity is more likely to have lower infection level and larger restriction level, and none has their individuals distributed evenly in these 2 levels.

From these, one interesting thing is that the R levels for all 3 types of viruses are very close while the I level of Virus 1 differs. The reason might be Virus 1 has the highest probability of recovering from infection. However, Virus 2 with the highest probability of resistant don't have much larger restriction level than the others.

In addition, all y axis of histograms is adjusted. Therefore, the actual difference is less than it performed by plotting. The idea of design is to make all efforts to capture the difference no matter how small it is. Because even though the differences among rates are small, it still could be a large group of people if the population size is big.

Chapter 5

5 Discussion

In this chapter, it is to discuss the limitations of the model from 2 aspects. One is the bias might be caused by algorithm, which is included in the embedded SD model to study the spread of epidemic. The other is the drawbacks in the network design. It is about the ABM model to research on the migration. Within each section, it would present the discussion in detail and put forward the possible direction of improvements.

5.1 Epidemic Simulation

5.1.1 One Virus

In reality, it is common for people to have several different diseases at the same time. And these diseases are not always infectious. So, they cannot be simulated through epidemic models. Besides, they might interact with each other and make the condition worse. For example, people who has illness A would be more likely to be infected by epidemic B. Then the probability of being infected by epidemic B should be different for people with or without illness A. In addition, the original ability of resistance is different for different person. Generally, young people are likely to have stronger resistance than elder people. Therefore, the probability of resistance should vary. In this research, it uses the simplest epidemic model, the SIR model. If possible, the attempts of more complexed model could be considered, such as the Susceptible-Exposed-Infected-Recovered (SEIR) and the Susceptible-Infected-Recovered-Death (SIRD) model mentioned in literature review.

5.1.2 Population is integer

It is a common sense that population should be integer number. However, to achieve this, floor functions are used twice to get the approximate results of calculations. Because the second time of using it is for migration. That should refer to next section.

For using floor function in epidemic algorithm, it is applied to the last stage of updating the infection iteration among populations. Recall it is using SIR model in the research. There are 3 groups of people (S/I/R). While summing up the new groups of infected and resistant

37

people, the results are the approximation using floor function. And for susceptible people, it is the value of the overall population minus the group size of I and R. Therefore, the bias range for the group size of I and R is [0,1), while the range for group size of S is [0,2). If the result inside the floor function is integer, then the floor function would not change it. If not, it is going to set an example to explain it clearly. In extreme case, for I and R group, the original results of calculation are $100.\dot{9}$ and $20.\dot{9}$. It means they infinitely converges to 101 and 21. However, under floor function, they are 100 and 20. Notice, if they are slightly bigger, the bias would not be the biggest. Recall the overall population would not change within the epidemic model. Therefore, both group I and R miss 1 people and these 2 people would be moved to susceptible group.

One possible way to increase the accuracy of results might be setting random value between 0 and 1 to consider whether keep the extra people in original group or move to another group. However, it brings another issue. The logic behind it to ensure the increase of accuracy is the normal distribution of picking a random value between 0 and 1. Therefore, the sample size should be considered. The number of times of operating model might need to be large enough to ensure the stability.

5.2 Migration and Network Design

5.2.1 Population is integer

The floor function is used while calculating the number of tourists. Recall the tourist group size is the product of tourist rate and population size of the node. At this stage, the range of bias is only [0,1). In the following step, it designed to consider some options.

The bias range would enlarge when the result of calculation is less than 1. Because the tourist number would then be set as 1, which differs from the initial tourist rate. One possible method is to set an extra random value to consider whether send a tourist or not.

Except for that, when the population size of the node is 1, option would be hard to make. In this research, it is set no tourist would be sent at this stage. However, with proper adjustment of calculation, this people could also leave the area. And this would lead to the concern of setting tourist rate.

5.2.2 Same tourist rate and same destination

Followed by the discussion above, if there are only few people stay in an area, it is unable to know whether they would leave that area one by one or all together.

Besides, the migrate rate of big cities should be larger than the rate of an isolated village. Furthermore, there could be a group of people migrating between 2 areas frequently. For example, people who cannot afford the house price in area A and live-in area B, they would then have to migrate between A and B every day. This is not included in this research.

Additionally, the destination of migrants from same origin is set the same for all. Because people don't always travel together. To improve this, people could decide their destination independently. Then, another concern would occur, people also don't always travel singly. It might be possible to set a random value for every tourist to consider whether move in groups or individually, if individually, decide the destination independently and if in groups, decide the destination together.

Chapter 6

6 Conclusion

In this research, it formulates a model with System Dynamic embedded in ABM to simulate the epidemic outbreaks under the influences of transportation networks using NetLogo. The iteration of epidemic is modelled by applying the SIR model to SD method, while its spread is represented by simulating human migration based on the urban-rural theory through ABM. Notice human migration is also used for illustrating transportation networks. Therefore, SD and ABM are integrated to support this research.

To investigate the effects, the experiments of setting several scenarios, which consisting of different types of disease, different initial outbreak places and different types of transportation tools, are made in initialization step. With the BehaviorSpace, model has been run several times to ensure the stability of the outputs. Using the mean values of the results, they are visualized through histograms and tables for further analysis.

Learnt from the results, it is found that, within the time-fixed tests, a transportation network with larger capacity would contribute to lower infection level and higher restriction level of the system. But for individuals, the distribution is less even.

Followed by that, the limitations of algorithm and model design are discussed, and possible directions and methods of improvements are put forward.

It is hoped that this work will serve as a foundation for future investigations of including transportation networks in epidemic research. The further analysis might consider developing this research in more complexed and comprehensive cases to test whether the results could generalize to other scenarios.

Bibliography

Aldila, D., Ndii, M.Z. & Samiadji, B.M., 2020. Optimal control on COVID-19 eradication program in Indonesia under the effect of community awareness, *Mathematical Biosciences and Engineering*, årg. 17, nr. 6, s. 6355–6389.

Boudrioua, M.S. & Boudrioua, A., 2020. Predicting the COVID-19 epidemic in Algeria using the SIR model, *medRxiv*. Cold Spring Harbor Laboratory Press, s. 2020.04.25.20079467.

Dietz, K., 1967. Epidemics and Rumours: A Survey, *Journal of the Royal Statistical Society. Series A (General)*, årg. 130, nr. 4, s. 505.

Fernández-Villaverde, J., & Jones, C.I., 2020. Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities, *National Bureau of Economic Research*. Cambridge, MA, s. 1–58.

Figueredo, G.P. & Aickelin, U., 2010. *Investigating immune system aging: System dynamics and agent-based modeling*.

Gilbert, Nigel & Troitzsch, 2005. *Simulation for the social scientist. UK: McGraw-Hill.* 2nd ed. [Book]. Maidenhead: Open University Press.

Grimm, V. et al., 2006. A standard protocol for describing individual-based and agent-based models, *Ecological Modelling*. Elsevier, årg. 198, nr. 1–2, s. 115–126.

Grimm, V. et al., 2010. The ODD protocol: A review and first update, *Ecological Modelling*. Elsevier, årg. 221, nr. 23, s. 2760–2768.

Grimm, V. et al., 2020. The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism, *Jasss*. University of Surrey, årg. 23, nr. 2.

Harris, J.R. & Todaro, M.P., 1970. Migration, Unemployment and Development: A Two-Sector Analysis [Article], *The American economic review*. Menasha, Wis: The American Economic Association, årg. 60, nr. 1, s. 126–142.

Hui, D.S. et al., 2020. *The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China*. Elsevier.

Keeling, M.J. & Rohani, P., 2011. *Modeling infectious diseases in humans and animals*. Princeton University Press.

Kermack, W.O. & McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics [Article], *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. London: The Royal Society, årg. 115, nr. 772, s. 700–721.

Kermack, W.O. & McKendrick, A.G., 1991a. Contributions to the mathematical theory of epidemics-I, *Bulletin of Mathematical Biology*. Springer, årg. 53, nr. 1–2, s. 33–55.

Kermack, W.O. & McKendrick, A.G., 1991b. Contributions to the mathematical theory of epidemics-II. The problem of endemicity, *Bulletin of Mathematical Biology*. Springer, årg. 53, nr. 1–2, s. 57–87.

Kermack, W.O. & McKendrick, A.G., 1991c. Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity, *Bulletin of Mathematical Biology*. Springer, årg. 53, nr. 1–2, s. 89–118.

Kieckhäfer, K. et al., 2009. *Integrating agent-based simulation and system dynamics to support product strategy decisions in the automotive industry*.

MacAl, C.M. & North, M.J., 2010. Tutorial on agent-based modelling and simulation, *Journal of Simulation*. Palgrave, årg. 4, nr. 3, s. 151–162.

Martinez-Moyano, I. et al., 2007. *Design for a Multilayer Model of Financial Stability :*

Exploring the Integration of System Dynamics and Agent-based Models.

Miller, J.H. & Page, S.E., 2009. *Complex adaptive systems: An introduction to computational models of social life*. Princeton University Press.

Phelan, S.E., 1999. A Note on the Correspondence between Complexity and Systems Theory, *Systemic Practice and Action Research*. Springer, årg. 12, nr. 3, s. 237–246.

Schieritz, N. & Größler, A., 2003. *Emergent structures in supply chains - A study integrating agent-based and system dynamics modeling*. Institute of Electrical and Electronics Engineers Inc.

Stonedahl, F. and Wilensky, U., 2008. *NetLogo Virus on a Network model*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U., 1999. *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Yap, L.Y.L., 1977. The attraction of cities. A review of the migration literature, *Journal of Development Economics*. North-Holland, årg. 4, nr. 3, s. 239–264.

Appendix A - Netlogo model

For codes and outputs of BehaviorSpace, please refer to the GitHub link attached: https://github.com/amberyli/CASA-thesis. Also, the figures included in this dissertation are also uploaded. Notice, to ensure the reproducibility of codes, please download the file stored in folder "code" and end up with ".nlogo".

Appendix B - Research Log

DATE	TASK
04 May 2021	Sort out clash with another student; Narrow down research exploration; Begin literature review
21 May 2021	Discuss the family of epidemic models (e.g., SIR, SEIR); Start to explore the EMA workbench for possible applications; Keep reviewing relevant studies
14 Jul 2021	Seek useful packages in Python that could work for ABM; Attempt to create network and ABM in Python with basic packages
20 Jul 2021	Settle on final topic; Try RGB and color brewer to represent SIR status; Consider adding population in analysis; Discuss a NetLogo sample model "Virus on a Network"
23 Jul 2021	Discuss the relevant study of human migration; Attempt to adapt migration with epidemic; Final attempt to Python;
02 Aug 2021	Move back to NetLogo; Consider normalization in migration and epidemic; Consider System Dynamic for simulating epidemic
09 Aug 2021	Adjust code to both include System and ABM; Begin the design of dissertation structure; Start literature review section
13 Aug 2021	Update the simulation of epidemic using System Dynamic; Start to model migration using ABM; Begin the methodology section;
16 Aug 2021	Decide make System Dynamic embedded within ABM; Fix bugs of code and make updates; Decide the direction of writing methodology
20 Aug 2021	Continued research for methodology and literature review; Keep updating codes in NetLogo
25 Aug 2021	Edit outputs of model; Discuss the possible results and the way to present them; Keep writing methodology and literature review; Start to locate the topics of discussion and discuss them
01 Sep 2021	Final draft of code; Decide the kinds of results to present
05 Sep 2021	Final updates of code and writing