Community structure and its effects on multi-strain disease diversity and structure

COMP90055 COMPUTING PROJECT REPORT
Supervisors: Nic Geard, Rebecca Chisholm
Conventional research project
Credit Points: 25
Spike Chun Yi Lee

Abstract:

Humans have throughout history, had to deal with the consequences of infectious diseases moving around our communities. Understanding how these diseases, their dynamics and how they interact with us and our communities can help us better understand and control the outbreaks. Since we are constantly connected to each other, it is also important to understand how the ways our communities are connected can influence the diseases that travel through them. Computational modelling allows us to simulate these behaviours over different situations that would be infeasible in real-life. Previous research has created models investigated how host connect networks could affect the diversity and dynamics of these infectious diseases. This paper attempts to replicate and extend functionality from those models. The implemented model was used to also investigate if community connectivity structure within networks or changing the number of connected communities could affect the diversity of multi-strain disease population. Results showed that community connectivity structure can affect diversity at both the total population level and individual community level. Changing the number of connected communities also affect the diversity at an individual community level. These results can help us better understand how infectious diseases could potentially move through our communities.

Table of Contents

Declaration:

I certify that

-this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person where due reference is not made in the text.

-the thesis is 5,058 words in length (excluding text in images, table, bibliographies and appendices).

Date: 4/11/19 *Signed:* 

Introduction:

Diseases plagued humanity throughout our history. Whether it be the common cold or something more serious such as malaria, these diseases have an impact on our lives and the communities we live in. It is therefore important we understand the pathogens (viruses, bacteria, etc.) that cause these diseases. Understanding these pathogens can allow us to produce vaccines that can provide immunity to a certain disease.

However, the problem of controlling these diseases is made harder by the fact that some of these pathogens can have multiple strains. This means that a pathogen can have highly diverse genotypes which can result in vaccines possibly being ineffective to certain strains. Studies have also shown that many important pathogens have diverse genotypes. As Bianco, Shaw and Schwartz (2009) states, one significant example of endemic multi-strain disease is dengue fever. Dengue has four dominant strains that through the years have spread geographically to "virtually" all tropical countries. As mentioned by Bianco et al. (2009), "Approximately 2.5x10^9 people are at risk of contracting dengue, and between 50x10^6 and 100x10^6 cases are reported each year." We can see that a multi-strain disease can be far-reaching and diverse. It is therefore essential that we understand how the multi-strain diseases move and oscillate through regions and communities.

Previous studies (Buckee, Koelle, Mustard, & Gupta, 2004) have investigated how host connect networks could affect the diversity and dynamics using their computational model. Buckee, Danon and Gupta (2007) were also able to investigate how connectivity between neighbouring communities could affect the diversity in pathogen populations. They found that a high diversity of disease population can be maintained within neighbouring communities when cross immunity is high and if the connectivity between them is very low. What has not been investigated in detail are the changes that would occur to the diversity and dynamics of a pathogen population when a population consists of more than two communities. Moreover, the question of whether the way these communities are connected could potentially have any resulting effect.

The aim of this project is to investigate the relationship between community connectivity structure and multi-strain disease population. How does the number of connected

communities and community connectivity structure affect the dynamics and diversity of strains within its disease population? This will be done by firstly replicate and extend a computational model from previous work by Buckee et al. (2004), Buckee et al. (2007) and Xiao Peilin (2019). This would allow for the modelling of multi-strain disease dynamics within a host network of more than 2 communities with multiple different connectivity structures.

This report will firstly discuss background information on the topic and software used to assist in creating the model. Background information will include discussion on multi-strain disease dynamics, relevant information found from the previous research done and how computational modelling is useful in context to this research problem. Next, the model implemented will be described in detail. This will include model parameters, logic and design decisions made. Experiments and results will then be discussed to investigate questions posed at the start of the report. Using the results, we will look at what we can learn and in turn contribute to the problem area. We will then reflect on the strengths and limitations of the study and describe future work that can be done.

Background

In this section, we will discuss relevant background knowledge on multi-strain disease dynamics and computational modelling. We will also investigate disease modelling and previous relevant research such as the models and results found by Buckee et al. (2004) and Buckee et al. (2007) that forms the foundation of this study. Lastly, the software environment used within this study is described.

Multi-strain Disease Dynamics and Modelling

In this study, we will be specifically looking into a set of behaviour and dynamics of infectious multi-strain diseases:

- **Infection:** People could possibly be infected by multiple strains of the same disease at the same time.
- **Recovery:** Once a person recovers from an infection from a specific strain, they could possibly be immune to infection from the same strain.

- **Mutation**: Strains could potentially mutate to different strains.
- **Recombination**: Two strains could potentially combine with each other to produce a different "child" strain as a result.
- **Cross-immunity**: If a person has immunity to a strain, they would be less vulnerable to infection from strains sharing similar traits to the immune strain.

One way for us to better understand how a multi-strain disease can spread through communities is for us to computationally model the spread of multi-strain diseases. It would be immoral and impractical to run real-world experiments. Moreover, models can help us validate or challenge our understanding of previous epidemics or biological phenomenon. Populations, communities and geographical environments can also all be diverse and complex. For this study, we will define communities as a group of strongly connected individuals. A computational model can, therefore, help us attempt to simulate these complex systems and study how different factors can influence the spread of disease in different environments. However, modelling strain dynamics and diseases can have its challenges. For example, it can be challenging to translate from looking at a single host to population strain model (Wikramaratna et al., 2014). As Wikramaratna (2014) explains, many models make simple assumptions on how to design "partial-cross immunity". It can be difficult to choose how we abstract complex population and disease dynamics to models. It is also important that we understand how our abstraction choices and affect the results of our investigations.

Previous studies done by Buckee et al. (2004) has created a "stochastic, spatially heterogeneous analogue" of the deterministic model made by Gupta et al. (1996). Buckee et al. (2004) were able to demonstrate how different host contact networks can affect strain diversity and dynamics using their stochastic model. For randomly mixed populations, their model confirms that cross-immunity can play a significant role. It may also result in structuring pathogen populations into discrete, nonoverlapping strains. The model was also able to show, as Buckee et al. (2004) states, an increasingly diverse pathogen population appears as "contacts between hosts become more localized and the assumption of random mixing is relaxed". Buckee et al. (2004) were able to conclude that

host contact network structure plays a large role in influencing pathogen strain structure and dynamics.

Further studies were also done to investigate how contact between neighbouring communities affect multi-strain disease diversity (Buckee et al. 2007). Models incorporating separate host communities were implemented. This was done to be able to find out how these neighbouring communities and the different interaction between them could affect the diversity of multi-stain disease populations within them. Buckee et al. (2007) found that neighbouring communities can have community structure that can consistently maintain diverse subsets of pathogen strains when cross-immunity is high and connectivity between them is low.

By moving from just observing neighbouring communities, we can investigate the relationship between community connectivity structure and multi-strain disease population. We can also investigate how the number of connected communities and community connectivity structure could potentially affect the dynamics and diversity of strains within its disease population.

Software environment

The software/code environment used to create the model and to investigate the results was python 3.7. Python libraries/packages were also used to help implement features of the model. These libraries were also used to create experiments and graph the results of the model for observation and analysis.

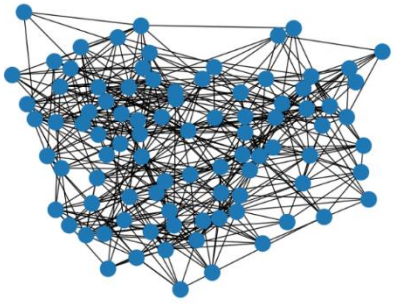

Figure 1 Example of the python package NetworkX's node network used to implement communities and population

Model Design

In this section, we will also look at all the different parts of the model. This includes how the networks, pathogen and their dynamics are designed among others. After that, all model parameters will be listed and a brief description of how all the different parts work together to make up the model.

Time: In this study, progression through time was implemented as iteration through timesteps. During a timestep, actions can be undertaken using values based on the start of that timestep. After all actions are taken, the results of those actions will be saved into the model and used for the next timestep. The flow of states can be seen below:

1. Start of a timestep
2. For each action:
    a. Use values from start of timestep
    b. Save result of action
3. After all actions:
    a. Update all results to relevant values
4. End of timestep

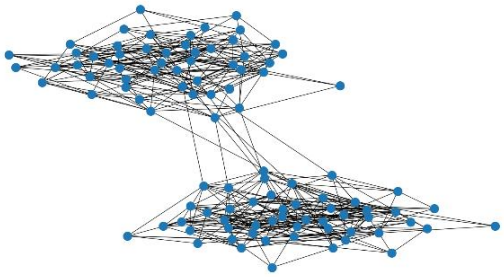

Figure 2 Two communities within a network population. Each community is a group of strongly connected nodes

Community Networks: Populations were represented as a network of connected nodes where each node would represent a person/host. Each node will have connections to other nodes on the graph. A community within the network would then be a group of strongly connected nodes. To simulate differently connected communities, two

community connectivity structures were implemented. A two-community structure can be seen from fig. 2 where we can see two strongly connected groups of nodes with sparse connections between them. To determine connections between nodes (either within or out of a community), a probabilistic method is used. On average, each node will have a set number of connections(*z_total*). This total will include average inside community (*z_in*) and outside community (*z_out*) connections per node. To achieve this, average number of inside and outside connections were the result of multiplying the number of nodes within a community (N_c) by the probability of that type of connection happening ($P_{in/out}$).

Two community connection structures were used within this study, a linear connection structure and a random connection structure.

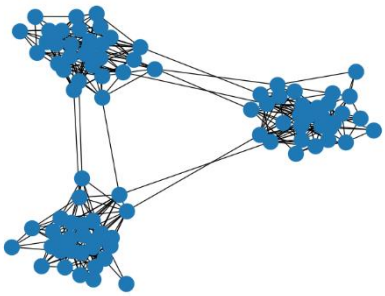

Figure 3.1 Random connection structure with 3 communities

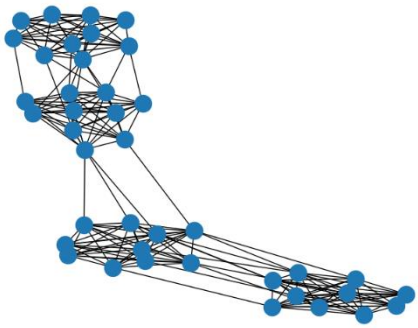

Figure 3.2 Linear connection structure with 4 communities

A randomly connected community structure was designed based off a random partition graph. Each community can be connected to any other community randomly. Fig 3.1 shows a visual representation of a randomly connected 3-community network. This was implemented in the model by using a pre-set python library function as seen in fig 4.

`random_partition_graph`(*sizes*, *p_in*, *p_out*)

Figure 4 random connection code

Fig 4 Parameters:

- **sizes** (*list of ints*) – Sizes of groups (each size is N_c)
- **p_in** (*float*) – probability of connection within communities
- **p_out** (*float*) – probability of connection between communities

In a linearly connected community structure, each community is only connected to the community to its left or right. This means that communities within this network could be considered connected in a "straight-line" as can be seen in fig 3.2. Fig3.2 also shows that a community at the left most end of the network will never be connected to the community at the right most end of the network. This was implemented through a custom-made function based of the logic of the function in fig.4. The Pseudo-code and parameters of the custom-made function can be seen below:

```
linear_community_network(n_communities, num_people,z_in,z_out):
   1. Calculate P_in using other attributes
      (z_in/out,n_communities,etc…)
   2. Calculate P_out using other attributes
      (z_in/out,n_communities,etc…)
   3. Create community of nodes  within network.
   4. For each community, set connections:
   5. If partition is left-most:
         a. Use P_in  to set connections within partition
         b. Use P_out to set connections to community to its right
   6. If partition is right-most:
         a. Use P_in  to set connections within partition
         b. Use P_out to set connections to community to its left
   7. Else:
         a. Use P_in  to set connections within partition
         b. Use P_out to set connections to community to its left and
            right
   8. Return created network
```

linear_community_network Parameters:

- **n_communities** (*int*) – Number of communities in network
- **num_people** (*int*) – Total number of nodes in network
- **z_in** (*float*) – Average number of inside community connections per node
- **z_out** (*float*) – Average number of outside community connections per node

Hosts: This agent-based model simulates a person/host as its own individual entity within the network (a node). Each host will then store information about its contacts, strains it is infected with and an immune memory (strains it has recovered from and now has immunity). A list of all information stored by the host is detailed below:

Host information:

- **contacts**: Connections with other hosts in the network
- **current infected**: A list of currently strains it is infected (based on start of time step)
- **current immune**: A list of strains it is immune to (based on start of time step)
- **newly infected**: Strains it has been infected with during the timestep and to be added to currently infected at the end of timestep
- **newly recovered**: Strains it has recovered from, to be removed from current infected and added to current immune at the end of timestep
- **next immune**: A list of strains the host has not lost immunity to by the end of the timestep (based on current immune).

Multi-strain Diseases (Pathogens): Pathogens are represented as bit-strings with each bit representing an immunodominant locus. To keep the model simple, the number of alleles per locus is limited to 2. These alleles are represented as either "0" or "1". Therefore, each pathogen can have 2^n different configurations (strains). On top of this, each strain is given a community identifying marker to represent the community the strain started it. For example, strain 01 from community 1 would then look like this, "`(1,01)`" where the community marker is in front of the strain. The community marker does not affect any pathogen dynamics (mutation, etc.) and is only used for identification. For this study, the number of locus was limited to just 2, meaning 2^2 configurations per community. A two-community network will then have 8 possible strains.

To help measure the dynamics and variability of strains within a network, a metric from Buckee et al. (2004) was used. Diversity (D) measures the "evenness with which a pathogen population is partitioned into all of its possible different strains" (Buckee et al.,

2004). As Buckee et al. (2004) states, this is calculated by "dividing the entropy of the disease population (Shannon-Weaver Diversity Index) by the maximum possible entropy of the population". Fig. 5 shows the formula used to derive diversity. For a disease population, diversity (D) can therefore only range between 0 and 1 with a diversity of 1 stating that all possible strains are equally present in a population.

$$D = \frac{\sum_{i=1}^{Ns} p_i \log(1/p_i)}{\log(N_s)},$$

Figure 5 Diversity Formula

Fig. 5 Parameters:

- p_i is the frequency of strain i in the population
- N_s is the number of strains

Multi-strain Diseases Dynamics: Pathogens cannot exist outside of hosts. A host can contain more than one strain but can only infected another host with one strain at a time. The infecting strain is chosen at random. If the host being infected does not have any immunity to the infecting host, it is infected with probability BETA. Once infected, a host may lose the pathogen with probability MU, this results in an average infection duration of 1/MU per strain. Once recovered from an infection, a host will remain immune to that strain. A host can lose its immunity to a specific strain based on probability SIGMA, resulting in an average immune duration of 1/SIGMA. The strains in a host's immune memory can possibly affect the probability of infection depending on the similarity between the infecting strain and strains in the immune memory. For example, if a host is immune to strain 00, it will be less vulnerable to infection from strain 01 as they both share the bit 0. This model's version of cross-immunity (GAMMA) is based on the version implemented by Buckee et al. (2004). Vulnerability to infection is then calculated by the equation in fig.6

$$v = (1 - f^{1/\gamma})^{\gamma},$$

Figure 6 vulnerability to infection

Fig 6 Parameters:

- f: the fraction of identical bits
- γ: is the level of cross-immunity

During an individual infection event, either mutation or recombination happens. The infecting strain could mutate into another strain with probability TAU. Mutation can introduce strains that were previously extinct back into the population. An example of this behaviour can be seen in fig 7 where strain 00 could potentially mutate into strain 10. The pseudo code below explains in more detail the logic of the mutation process:

```
mutation:
   1. For each bit in strain:
   2. If pass probability TAU check:
         a. Bit changes to different bit
   3. Else bit does not change
   4. Return mutated strain
```

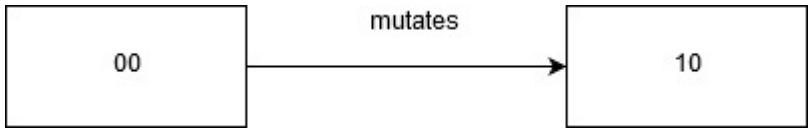

Figure 7 Mutation example

During recombination, the infecting strain could potentially recombine based on probability R with a random strain in the infected host producing a "child" infecting strain instead. The random strain chosen also does not get affected by this process. Fig 8 shows an example of this behaviour where strains 00 and 11 can recombine to produce child strain 10 that the host will be infected with. In this model, if two strains from different communities recombine, the child strain will keep the community marker of the infecting strain. The pseudo code below explains in more detail the logic of the recombination process:

```
recombination:
      1. For each bit in each strain:
      2. If pass probability R check:
            a. Choose bit from either strain to add to child
               strain
      3. Else keep bit from infecting strain
      4. Return child strain
```

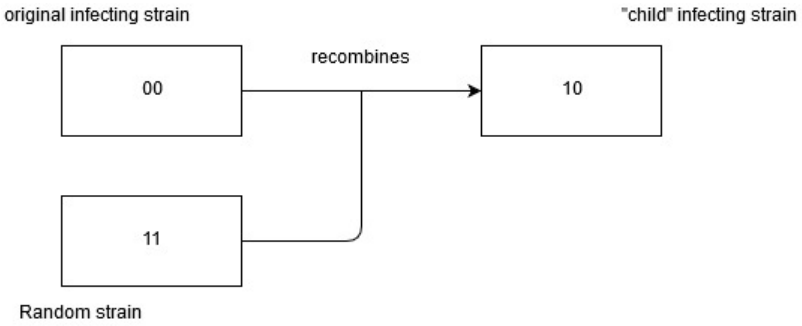

Figure 8 Recombination Example

Full Model Parameters & Logic: The table below shows all the important parameters used to make up the model and some that were discussed previously. These parameters will also be used and changed for model experimentation in the next section of the report.

Table 1 Model Parameters

Parameter	Description
z_total (z_in/z_out)	Mean number of contacts per host (inside/outside community)
N_communities (N_c)	Number of communities
N_Nodes	Number of people
COM_TYPE	Community connection type(linear/random)
MU	Recovery probability
SIGMA	Immunity lost probability
BETA	Infection probability
R	Recombination probability
TAU	Mutation Probability
GAMMA	Cross-immunity

When we add all the pieces of the model together, we can observe the full logic of the model. Below is the pseudo-code interpretation of everything that will happen when the model is run.

Model Pseudo code:

1. Assign network attributes (probability for infection, number of neighbours,etc)
2. Generate network
3. Generate strain space
4. Seed network with initial strains
5. For each time step:
6. For each host:
    a. If host has immunity memory:
    b. For each strain in memory:
        i. Check immunity loss based on probability SIGMA
    c. Save new immunity memory to next_immune attribute
    d. If host has any infection:
    e. For each neighbour:
        i. Attempt to infect neighbour with random strain from infection memory
        ii. If infected:
        iii. Attempt mutation based on probability TAU
        iv. Attempt recombination with strain from neighbour's memory based on probability R
        v. Add resulting strain to neighbour's next_infected attribute
    f. Attempt recovery from infections
        i. Update status of newly recovered to newly_recovered
    g. Update all memory attributes with the next attributes (updates from current timestep) :
        i. Set current immune memory to the next_immune | newly_recovered
        ii. Set current infected to current infected - newly recovered
        iii. Then set current infected to current infected | newly_infected
    h. Clear all next/newly attributes

Experiments and Results

In this section, the experiments run to explore the model will be detailed and the results found will be discussed. These experiments were designed to be able to observe how changes in both model parameters and population dynamics (changing community connection structure, number of connected communities, etc.) could lead to different resulting multi-strain disease populations. For each experiment, each network was randomly seeded with a set number of initial strains per community unless explicated stated differently. By default, each community starts with 12 hosts with initial strains. Each host will also had on average 10 connections. The logic of the seeding process can be seen from the pseudo code below:

```
seeding:
   1. For each community in network:
   2. Pick random set of initial host from community
   3. For each host in set:
         a. Seed with random strain from strain space with
            community marker of its community
```

Experiment 1: Replicating and testing behaviour from previous research

This experiment was done to observe if the implemented model was able to replicate behaviour found in the Buckee et al. (2007) paper. For this experiment, the model tested with relevant parameters (table 2) on a 2-community network with a random community structure. However, it is to be noted that in a 2-community network, changing the community connection structure would not matter as there are not enough connected communities to observe a structural difference.

Table 2 Parameters for experiment 1

Number of communities	Z_in	Z_out	GAMMA	MU	SIGMA	BETA
2	**0.5**	**9.5**	**4**	**1/7**	**1/23**	**0.2472**

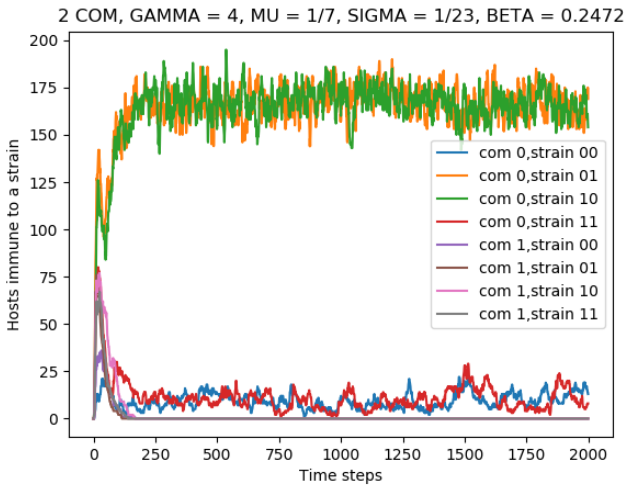

Figure 9 2 community randomly connected

Table 3 Diversity Values from fig. 9

Total diversity	Community 1 diversity	Community 2 diversity
0.36	0.24	0.25

We can see from fig 9 that when cross immunity is high (GAMMA = 4), two strains dominate the total population within the network with most other strains going extinct quickly. These two strains (10 and 01) can be considered discordant strains as they have no overlapping bits. The results from fig 9 (table 3) was also reinforced by the total diversity value being low at D = 0.36 and each individual community having similarly low diversity. This seems to confirm the behaviour found in the Buckee et al. (07) paper where it was stated that after a number of steps, one discordant set becomes dominant in the whole population and that when cross immunity was high, the within community diversities were expected to be relatively low. This experiment allows us to confirm that our model is behaving in a consistent manner backed by previous research done and that the behaviour can produce understandable results.

Experiment 2: Effects of community connection structure on pathogen diversity

For this experiment, the aim was to investigate if the two different community connection structures could lead to different resulting pathogen diversities. To achieve this, networks with three communities were used. Testing was done on two networks each with different community connection structure(linear/random) and a fixed total number of nodes. Model parameters (table 4) were identical for both networks.

Table 4 Parameters for experiment 2

Z_in	Z_out	GAMMA	MU	SIGMA	BETA	R	TAO
0.5	9.5	4	1/7	1/23	0.2472	0.0953	0.0042

Randomly connected communities

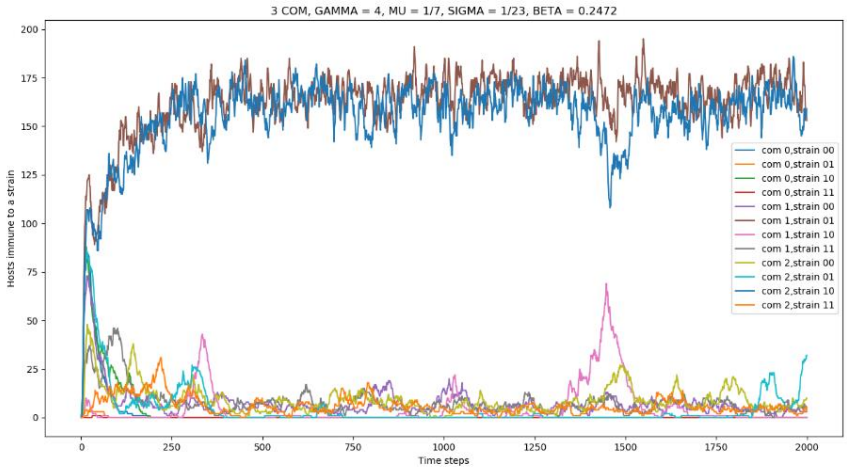

Figure 10 Randomly connected communities Total

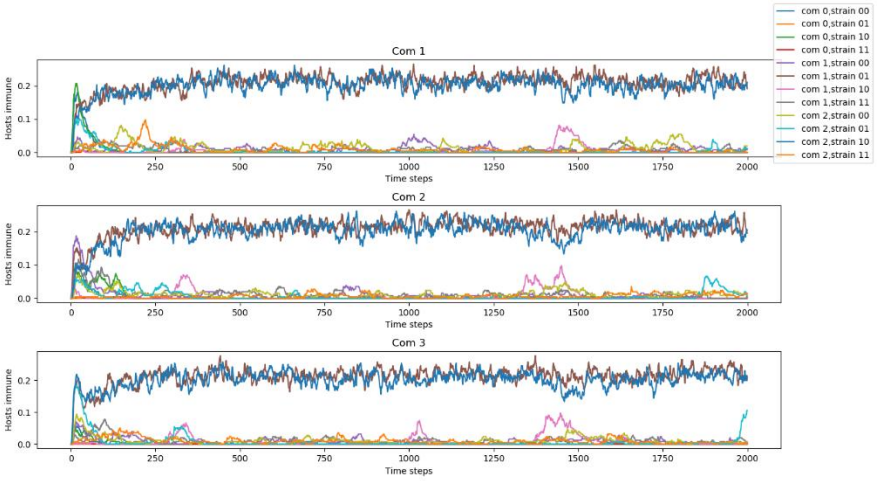

Figure 11 Randomly connected communities Each COM

Table 5 Diversity Values for randomly connected communities

Total Diversity	COM 1 Diversity	COM 2 Diversity	COM 3 Diversity
0.31	0.160	0.163	0.161

Fig 10 & 11 shows a very similar picture when compared to experiment 1. Increasing the number of communities in the network still results in similar behaviour and dynamics. In a randomly connected community structure network with a high cross-immunity value (GAMMA = 4), there will be two dominant strains with a low total diversity. We can also see that the two dominant strains are the same in every community and each community in the network maintains a very similar level of strain diversity.

Linearly connected Communities

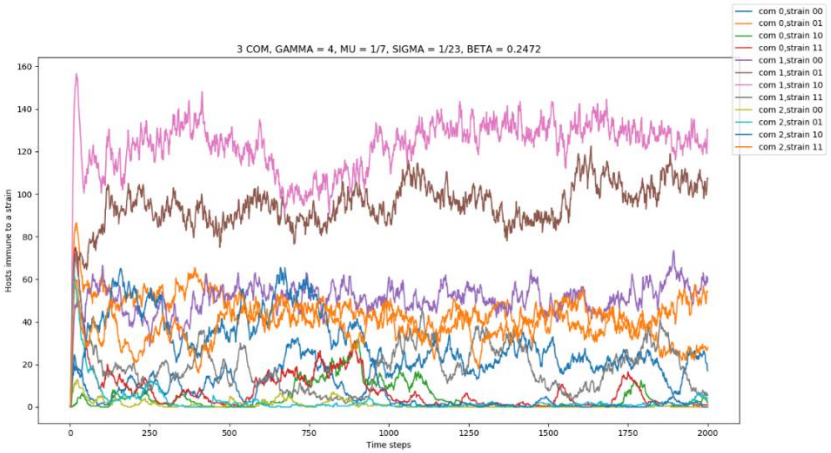

Figure 12 Linearly connected Communities Total

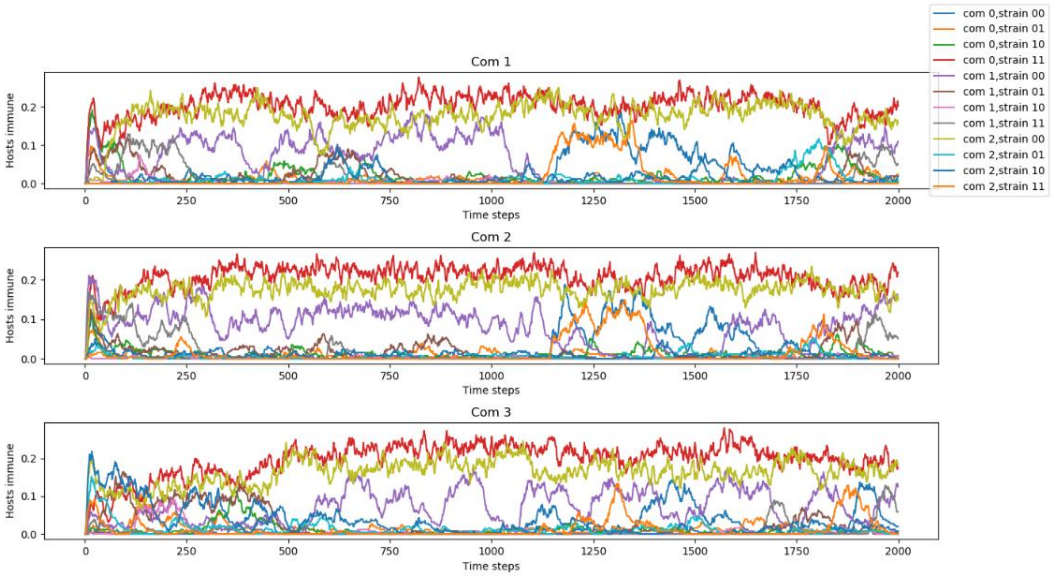

Figure 13 Linearly connected Communities Each COM

Table 6 Diversity Values for linearly connected communities

Total Diversity	COM 1 Diversity	COM 2 Diversity	COM 3 Diversity
0.45	0.220	0.228	0.217

Fig 12 & 13 shows us the results of a linearly connected community structure. From fig 12, we can see that two strains stay dominant through all the timesteps. However, we can see multiple other strains oscillate below the two dominant strains leading to a total diversity of D = 0.45. Fig 13 paints a similar picture, the two strains dominant in the total population are dominant in each community. We can also see other strains oscillate below the dominant strains; however, each community does not have the same strains oscillating at the same time even it results in similar diversity.

Result Comparison

If we compare the results of testing the two different community connection structures, we can see that how communities are connected can affect the diversity of the pathogen population of a network. By comparing the results of the two three community networks, we can see that linearly connected communities have a higher total diversity of 0.45 compared to 0.31 for randomly connected communities. Comparing fig 10 and 12 reinforces this observation. The relationship is similar if we compare individual community results. We can conclude that how communities are connected does influence both total

pathogen population and individual community pathogen population diversity with linearly connected communities producing higher resulting pathogen diversity compared to randomly connected communities.

Experiment 3: Effects of increasing number of communities of a given size within a network

The aim of this experiment was to observe if increasing the number of fixed size communities within a network would have any measurable difference in resulting pathogen diversity. Firstly, the size of each community was fixed to 100. For this experiment, the number of communities used was fixed to values between 3 and 10. This would mean that the total population will increase if the number of communities increases. Model parameters besides the number of communities were fixed and identical (table 6). A sampling method was used to make sure the model was run 20 times per each value (3 to 10) and the total and individual community diversity were recorded per run. Each run was ran over 1000-time steps. The results were plotted on an error plot where the dots(triangles) were the average diversity while the line going through the dots were the standard deviation.

Table 7 Parameters used for experiment 3

Z_in	Z_out	GAMMA	MU	SIGMA	BETA	R	TAO
0.5	9.5	4	1/7	1/23	0.2472	0.0953	0.0042

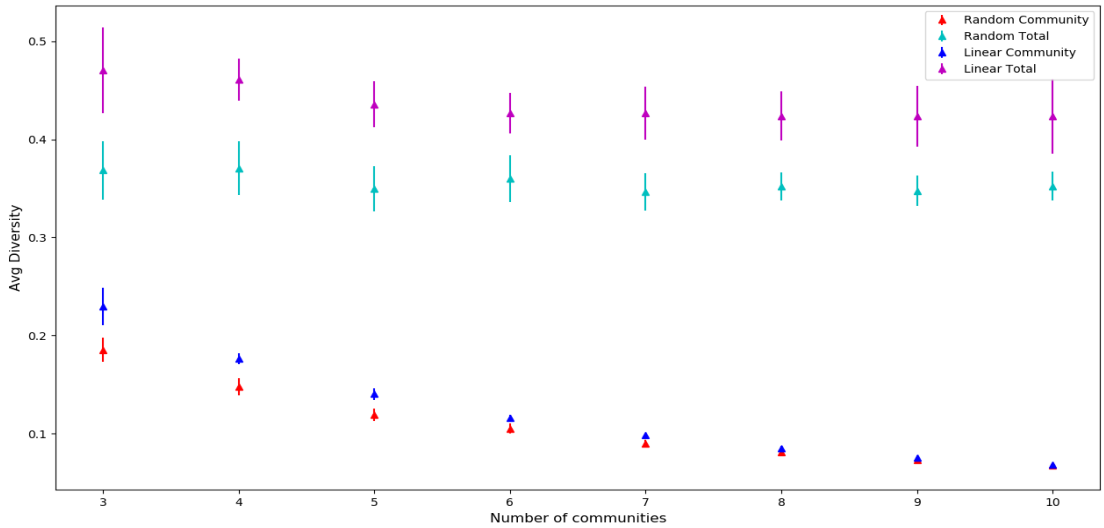

Figure 14 Different number of communities

We can see from fig. 14 that even with increasing number of communities, linearly connected communities result in higher average total pathogen population compared to randomly connected communities. We can also see from the standard deviation lines that even with the variance from multiple runs, the total diversity of linearly connected community networks were never below randomly connected community networks. However, increasing the number of connected communities does not drastically change the total diversity for either random or linear connection structures. Fig 14 also shows us that as we increase the number of the connected communities, the individual community diversity decreases for both connection structures. The difference between them also shrinks to the point where the average diversity is almost identical at 10 connected communities. We can also see that as the number of communities increase, the variance in results for individual community diversity decrease as the standard deviation line shrinks. We can conclude that changing the number of connected communities of a given size does not affect the total pathogen population diversity but does affect the individual community diversity.

Discussion

In this section, we will summarise and reflect on the results found through the experiments done, discuss how the results can help contribute to our understanding of multi-strain disease modelling and possible future extensions to the model. We will also reflect on the process of designing and implementing this study, what went well and what could have been improved.

Summary and Reflection of Results

From the experiments run on the model, we can see that in a two-community setting, two strains became dominant in the whole population when cross immunity was high and the within community diversities were relatively low. We can confirm that this behaviour stays the same when looking at a three community networks with high cross immunity. However, changing the connectivity structure within a three community networks shows us different diversity results. A linearly connected three community network has a higher total diversity (D = 0.45) than a randomly connected three community network (D = 0.31). The diversity in individual communities paint a similar picture, with linearly connected communities having a higher diversity rather than randomly connected communities.

If we increase the number of connected communities of a given size within a network, we can reinforce the observation that linearly connected community network result in higher total strain diversity. From three to ten community networks, linearly connected networks consistently have a higher mean total final diversity than randomly connected networks. However, as we increase the number of connected communities, individual mean final community diversity for both types of network decrease to the point where they are almost indistinguishable from each other.

The results from the two-community setting allows us to confirm behaviour described in previous research (Buckee et al. 2007) and show that the model is able produce reproducible results that fit into what we currently understand about multi-strain disease

dynamics. The other findings also show that how communities are connected in a multi-community network can affect strain diversity when there is high cross-immunity.

Significance of Study

This study and its findings contribute to our knowledge of multi-strain disease dynamics in multiple ways. We can see that how communities are connected can influence the diversity of disease strains within a population. This could help us better understand how multi-strain disease move within regions with differently connected communities such as more geographically isolated communities compared to cities or towns. The model implemented also extends from previous research and separates itself by investigating more than just two neighbouring communities. This model could be used to continue to explore how multi-community structure affects multi-strain disease dynamics.

Strengths and Limitations of Study

The model implemented in this study includes any changeable parameters and testing environments. This hopefully helps futureproof the design so that future research could be done easily. This study also details the design and methodology of experiments to make them easily reproducible.

Due to the ambiguity and lack of source code from the previous research referenced, many assumptions had to be made for the model implementation such as the recombination and initial strain seeding methodology. External libraries/packages were used help form parts of both the model and testing environment and how these libraries/packages could potentially affect the results were not investigated in detail. Lastly, more investigation and thought could be made for the reasoning beside some of the results found. As this study mainly focused on implementing a multi-community model and investigating if community connection structure can affect multi-strain disease diversity, there is a lack of investigation into the details of why exactly we observe such behaviour.

Possible Future Work & Extensions

A more detailed investigation into why differently connected communities produce different diversities at both total and community level. One could look at how diseases move been different communities in more detail. There could also be more research done to investigate how changing different parameters affect multi-strain disease dynamics and diversity. Additional types of community connection structures or more complex disease implementations could also be investigated. Lastly, different abstractions of multi-strain disease dynamics(mutation/recombination) or tools(libraries/packages) could be implemented to observe if there are any significant differences compared to the results found in this study.

Reference List

Bianco, S., Shaw, L. B., & Schwartz, I. B. (2009). Epidemics with multistrain interactions: The interplay between cross immunity and antibody-dependent enhancement. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *19*(4), 043123.

Buckee, C. O. F., Koelle, K., Mustard, M. J., & Gupta, S. (2004). The effects of host contact network structure on pathogen diversity and strain structure. *Proceedings of the National Academy of Sciences*, *101*(29), 10839-10844.

Gupta, S., Maiden, M. C., Feavers, I. M., Nee, S., May, R. M., & Anderson, R. M. (1996). The maintenance of strain structure in populations of recombining infectious agents. *Nature medicine*, *2*(4), 437.

Wikramaratna, P. S., Kucharski, A., Gupta, S., Andreasen, V., McLean, A. R., & Gog, J. R. (2015). Five challenges in modelling interacting strain dynamics. *Epidemics*, *10*, 31-34.

Buckee, C., Danon, L., & Gupta, S. (2007). Host community structure and the maintenance of pathogen diversity. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1619), 1715-1721.

Xiao Peilin. (2019). Effects of Host Contact Network Structure on Pathogen Population Structure. *Submitted to University of Melbourne COMP90055 COMPUTING PROJECT*