THE UNIVERSITY *of* EDINBURGH

MPhys Project

The impeded effects of descendent variants as a result of cross-immunity on a network.

Jabeth Musumba
May 2024

Supervisors: Professor Graeme Ackland and Professor Rowland Kao **Acknowledgements:**Aeron Sanchez

Contents

Abstract

In this paper, we introduce a multi-strain SIR model with strict cross-immunity and investigate how the dynamics differ between 3 types of network. We vary the rate of cross-immunity, the incubation period of the first disease on the network and the transmission rate of the second disease to study how for scale-free networks, cross-immunity shields portions of the population. We also look at ongoing and newly emerging disease variants infecting the population at the same time and how specific conditions can guarantee the sucesss of either.

1 Introduction

With the rise and fall of the COVID-19 pandemic, there has been increased focus on studying epidemiological systems with more complex and nuanced behaviour that can potentially exhibit real-world behaviour. One of the distinguishing factors in the prolonged nature of the pandemic is the existence and spread of disease variants which all had their own levels of transmission rates and would cause re-emerging epidemics. When we consider the short time-span between the emergence of the COVID-α and COVID-omicron we recognise that the variants of the original strain are spreading on a network that 'remembers' it via cross-immunity. This brings about interesting dynamics between initial and successive strains which have real-world implications [12]. We develop and study a multi-strain SIR model which aims to look at the infection spreads of successive disease variants given an initial epidemic that is still occurring. We then apply our model to 3 different network structures to discern any noticeable differences that network structure has on a secondary epidemic brought about through the introduction of a new strain. We define cross-immunity as a random event which occurs after Recovering from some variant α and has a fixed likelihood of $p_{\alpha,\beta}$ for every combination of variants. This cross-immunity grants full immunity to the other strain and is equivalent to being in the Recovered state for both diseases. This allows us to utilise the standard SIR model with the small modification that individuals can jump from Susceptible to Recovered without ever being infected. Our focus is limited to 3 types of static networks: Random, Small-World and Scale Free and we hope to determine if specific network layouts give rise to better or worse descendent-strain success. Considering this model as systematic site percolation the goal of this paper is to study the conditions that allow and prohibit variant percolation given the existence of an ongoing disease.

2 Method

SIR

Compartmental models in computational epidemiology are frequently utilized for examining disease dynamics within populations, the most famous of these is the Susceptible-Infected-Recovered (SIR) model [1]. This model categorises individuals into one of three states at any given time. As the disease progresses individuals transition from Susceptible to becoming Infected and once infected they will eventually Recover. Unlike some other compartmental models, the SIR model describes a one-way linear progression for the states of individuals, once Infected they can never be Susceptible again and once Recovered they can never be Infectious again. In the context of cross-immunity, we add an additional progression track that goes from Susceptible to Recovered and covers the case where a node gains cross-immunity to a disease it has never seen before. We stick to this simplified model, as opposed to SIRS and other modified models as it allows us to investigate how the spread of a descendent strain operates on a deteriorating network. Although this simplification may deny us directly observing certain types of

outbreaks, like those which exhibit periodic behaviour, we will hopefully observe analogous behaviour when looking at multiple diseases that 'take over'. Specifically, we anticipate observing instances where subsequent diseases persist even after the initial outbreak has subsided. It should also be noted that we look at a fixed system size N (representing the number of individuals) which is predetermined and the sum of the number of individuals in each compartment should add to N. Analytically the dynamics of the SIR model are determined by 3 sets of differential equations representing the rate of change of the number of individuals in each compartment[7]:

$$\frac{dS}{dt} = -\beta \cdot \frac{S \cdot I}{N}$$
$$\frac{dI}{dt} = \beta \cdot \frac{S \cdot I}{N} - \gamma \cdot I$$
$$\frac{dR}{dt} = \gamma \cdot I$$

where S(t), I(t), and R(t) are the number of individuals in these compartments, such that $S(t) + I(t) + R(t) = N$; β (Transmission rate) is the probability of transmission from an infected individual to a susceptible one per unit time and γ (Recovery rate) is the rate at which infected individuals become recovered per unit time. [1]
We implement this model stochastically and pay special attention to the way in which we define β and γ; in the equations above we have a generic β and γ which describe how the disease propagates on a holistic level but when we implement this with respect to networks some individuals have more neighbours than others. Here we define β to be the transmission rate between any pair of infected and susceptible individuals such that an individual that has many neighbours will most likely sooner pass on the disease over an individual that has few neighbours. These definitions of β are effectually the same but we take caution in accounting for how this will be implemented on a network with varying levels of connectivity.

The values β and γ are important because they define properties which play a crucial role in the propagation of the disease they are defined for. One such property is the basic reproduction number R_0 which measures the effectiveness of an infectious agent to spread in a totally susceptible population [6] and we define it to be $R_0 = \frac{\beta}{\gamma}$.

2.1 Networks

We look to implement our model on a few different network structures. For the purposes of generality, we do not predetermine these networks but rather generate them at the start of each run based on certain properties we want them to have. When it comes to the set of all networks there are many scales we can use to define them, in Fig.1 we show a 3-dimensional space accounting for Randomness, Heterogeneity and Modularity.

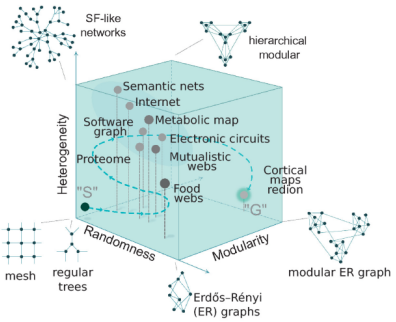


Figure 1: A portion of the Network-Space varied with Randomness, Modularity and Heterogeneity[6]

Randomness is the most trivial; when generating a network randomness is proportional to the probability that any two nodes on the network are connected. Meshes and regular trees are shown in Fig.1 to be the least random as nodes are only ever connected to their immediate neighbours, the

[1] Here we denote the susceptible, infected and recovered nodes as S, I, and R, respectively but in later contexts when speaking about multiple variants it will be useful to consider which diseases they are Susceptible, Infected or Recovered to/by/from. So when considering a disease α we denote these with a subscript; for instance we might say the Susceptible to α are S_α

probability that a node A is connected to another node B which is not a neighbour is zero. Heterogeneity refers to the degree distribution of a network[8]. The degree of a node is simply the number of other nodes it is connected to and so high heterogeneity describes a network with some nodes with small degrees and other nodes with large degrees, the degree distribution is wide. In contrast, a low heterogeneity describes a degree distribution that is narrow, most if not all nodes have the same degree. Finally, we take into account Modularity which is a measure of how well a network can be divided into distinct communities, groups or modules [10]. The network space is large and can be defined with respect to many more properties within this report we stick to three types of generated networks. Firstly we look at the extremely random **Erdos-Reyni** networks which connect a node to any other node given an input probability p[3]. With a large enough p these networks can be extremely robust in the face of percolation which is what will be happening when we run SIR on it. The average degree of these networks is then the size N multiplied with p. We then compare the results that we get from this with Watts-Strogatz generated graphs and Barabasi-Albert generated graphs. **Watts-Strogatz** is an algorithm for generating graphs with small-world properties; small world graphs are networks which tend to have a high clustering coefficient and a small average path distance between any two nodes. The way this is done is by generating a ring of nodes of size N [5]. The mean degree k which is input into the algorithm as a parameter is used to connect each node to its k nearest neighbours ($\frac{k}{2}$ to the left and $\frac{k}{2}$ to the right if k is even). There is also another parameter p which defines the level of randomness this graph will have. For each connection between node u and its neighbour v, β is the probability that the connection will be rewired to another node w (which is uniformly randomly chosen) [11]. In this project we stick to small world graphs built with a randomness parameter of $p = 0.5$.

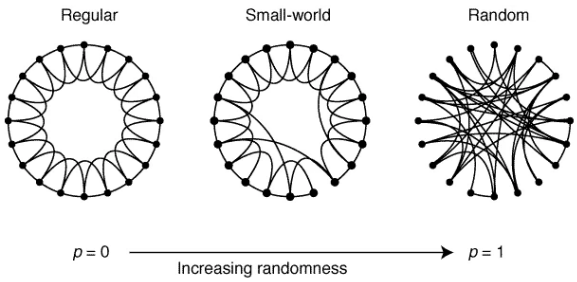


Figure 2: Similarities small-world networks have with random and regular graph structures when built using the Watts-Strogatz method[11]

The **Barabasi-Albert** algorithm is what we'll be using to generate scale-free networks. These are networks whose degree distribution follows a power law such that the proportion of nodes connected to k other nodes, $p(k)$, is inversely proportional to the number of connections: $p(k) \propto k^{-\gamma}$ where gamma is a unique positive number[2]. This results in few nodes that have many connections and many which have few and is modeled after networks that grow. The generation algorithm works by 'growing' the network from 1 to N; with an input parameter m and starting with 1 node, add nodes sequentially to the network by connecting m edges from the new node to existing nodes. These edges are connected 'preferentially' to nodes with higher degrees; we create a group G which contains every existing node and a copy of each node for the number of connections they have such that a node v with degree 10 will be repeated in G 10 times [5]. In other words $|\{n \in G; n = v\}| = 10$. When a new node w is entered into the system, we uniformly pick m nodes from G to connect it to.

2.2 Gillespie Algorithm

In this model, we implement SIR stochastically but there are multiple ways we could choose to implement iterations of the simulation; we go with a method coined as the Direct Approach of the Gillespie Algorithm[4], it works like so:

1. We search through the entire system and compile a list of all possible events. There are only two types of events: a) an infected node transmits the disease to its susceptible neighbour and b) an infected node recovers. This means we only ever need to look at the list of all infected nodes and their susceptible neighbours to know everything that *could* happen next.

2. We then weigh each event by its' likelihood. This likelihood is primarily determined by the transmission rate β for all events of type a and the recovery rate γ for all events of type b. This weighted list of all events is what we hereby refer to as the Gillespie Line.
   If we imagine that we have a disease α with a transmission rate β_α and recovery rate γ_α then the contribution disease α makes to the Gillespie Line (or rather all the weighted events involving nodes infected by α) is given by:

$$l_\alpha = \sum_{n_i \in I_\alpha} [m_{\alpha,n_i} \cdot \beta_\alpha + \gamma_\alpha] \quad (1)$$

   where l_α is the contribution to the Gillespie Line from variant α,
   I_α is the total number of nodes infected with α and
   m_{α,n_i} is the number of nodes neighbouring node n_i that are susceptible to α.
   It should be noted here that when we consider multiple disease variants acting on the network, an infected node n_i can also Recover from disease ω with fixed probability $r_{\alpha,\omega}$[2] but we have omitted to include it in the equation above for two reasons;

   - it should not affect the Gillespie Line proportions (otherwise the recovery rate for a disease would change based on the other disease simply existing in the system); so
   - it only occurs when a node $n_i \in I_\alpha$ recovers with probability γ_α [3]

3. Randomly pick an event from the Gillespie line using the weightings for the probability of each event. If the event, n_i transmits the strain α is chosen then uniformly choose 1 out of all m_{α,n_i} neighbours of n_i to infect. If the event, n_i Recovers from strain α then pick an outcome described in 2 based on their respective weightings. Reset the Gillespie line to empty once an event has been picked.

4. Taking every disease $\alpha \in \Omega$ we calculate the size of the contributions l_α and sum them such that $L = \sum_{\alpha\in\Omega} |l_\alpha|$. We also generate a random number, x, from the uniform distribution and then calculate the next time step τ with the equation:

$$\tau = \frac{1}{L} \cdot \ln\left(\frac{1}{x}\right) \quad (3)$$

5. Update the state of the node targeted by Step.3 and repeat from Step.1 until there are no more possible events

As we have briefly touched upon, our model follows the base principles of the Gillespie Algorithm but when we consider multiple variants acting on our network, there are more events that can occur than those outlined by the basic SIR.

[2] cross-immunity likelihood between α and ω

[3] If we were to list these within the Gillespie it would look like:

$$\text{Weighted list of recoveries given recovery from } \alpha = \sum_{S\subseteq\Omega} \left(\prod_{n\in S} r_{\alpha,n} \times \prod_{m\notin S} (1 - r_{\alpha,m}) \right) \quad (2)$$

which we would normalise to 1 and multiply by the γ_α in Eq.1. Ω is the set of all variants

2.3 Model modifications

The model we have so far consists of a disease that follows the SIR state transition scheme on a randomly generated network. The model will then iterate through events using the Gillespie Algorithm and will do so until there are no more events (no more infected nodes). To allow us to examine the effects of multiple diseases we make a few small modifications to this model.

The **first** is that we allow for multiple diseases to run on this network. This means that for n diseases, every node has n states which are either S, I or R for each disease. Contributions to the Gillespie Line sum linearly such that if there are n diseases, the Gillespie Line $= \sum_{\omega\in\Omega} l_\omega$ where $|\Omega| = n$. Currently this is equivalent to running n diseases on the same network structure in parallel (they do not interact). Our next modifications will add interactions between diseases.

The **second** modification is disease progeniture. If the Gillespie Algorithm chooses a neighbour $m \in m_{\alpha,n_i}$ to infect, there is a fixed (and very small) probability p [4] that neighbour m is infected with a new disease ω instead of α. If this happens, m is considered as patient-zero for the new disease ω and its' neighbours *should* all be susceptible. This is not always the case, however, due to modification 3. Since a disease variant is directly spawned from a previous variant, we only need to start our simulation with one disease. After a certain number of iterations, a new disease will spawn off of this and eventually we will have a family tree of variants.

The **third** modification regards node Recoveries. When a node $n_i \in I_\alpha$ is chosen to recover, there is a fixed probability $r_{\alpha,\omega}$ that n_i will also become Recovered from disease ω, even if $n_i \notin I_\omega$. The fixed probabilities are set for every combination of 2 diseases $\alpha, \omega \in \Omega$ and they are symmetric such that $r_{\alpha,\omega} \equiv r_{\omega,\alpha}$. This represents cross immunity where diseases α and ω are 'related' by $r_{\alpha,\omega}$ which we call the Relationship Coefficient. This coefficient is reminiscent of the intersection coefficient between the set of Recovered nodes R_α and R_ω:

$$\lim_{|R_\alpha|,|R_\omega|\to\infty} \frac{|R_\alpha \cap R_\omega|}{|R_\alpha|} = r_{\alpha,\omega} = \lim_{|R_\omega|,|R_\alpha|\to\infty} \frac{|R_\alpha \cap R_\omega|}{|R_\omega|} \tag{4}$$

It should be noted that under this definition $r_{\alpha,\alpha} = 1$ which is exactly what we expect, a disease α is totally and completely related to itself.

2.4 Random.vs.Predetermined Variables

When we start the simulation, all we need to define is the type of network we want to generate and the β and γ parameters of the initial disease. The simulation will then choose a random node as patient zero and run the model as defined in Sec.2.3 and Sec.2.3. This is the Random regime of the model as information about the potential offspring are completely unknown to us until the simulation is running[5]. For an offspring variant ω we do not presume that the transmission and recovery rates β_ω and γ_ω share any relation with the parent diseases', β_α and γ_α In contrast, we see it safe to assume that $r_{\alpha,\omega}$ is likely to be larger than $r_{\alpha,\delta}$ or $r_{\omega,\delta}$ where $\delta \in \Omega$ but δ is neither descendent nor offspring of either α or ω. To generate an $r_{\alpha,\omega}$ with such properties we assign each disease a unique ID number at spawn, offspring ID numbers are chosen from a Gaussian distribution centred about their parent disease. The relationship coefficient between parent and offspring diseases is then determined by the output of another Gaussian function with $\Delta(\mathrm{ID}_{\mathrm{parent}}, \mathrm{ID}_{\mathrm{offspring}})$ as the input.

In the predetermined variables regime, we define the total number of diseases that will appear, the times at which subsequent diseases are spawned, the relationship coefficients between all diseases and the transmission and recovery rates for all diseases

[4] p is fixed for every variant in the system and for every run in this project we stick with a value

[5] That said we can define an $|\Omega|$ cap such that we don't have too many variants on the system. After the cap has been reached, no more variants will be spawned

as well as all the variables we define in the random regime. This gives us a greater level of control over the simulation as we will be able to examine the effects brought on specifically due to the cross immunity of variants and the properties of the networks they run on.

2.5 Analysis

SIR metrics

For this simulation there is a lot of data we could gather and analyse but with the focus being the initial diseases' impact on subsequent variants, we aim to examine metrics we could use to define the "success" of these variants and the conditions which maximise (or dampen) these metrics. The metrics that pertain to pandemics in general are easy to spot and track but are vital for characterising a disease. The peak number of infected individuals at any given time the time taken (from inception) for a disease to reach its peak are and the time taken for the disease to infect 50% of the population are but a few.

The **peak number of infected** individuals provides an indication for the maximum intensity of the disease. The mantra "dampen the curve" is in direct reference to this peak as it indicates whether the outbreak could overwhelm real-world healthcare systems (due to the large volume of infected people).

The **time taken until the peak** number of people have been infected tells us about the diseases' *effective* transmissibility. Whilst each disease will have its own transmission rates, which will be known to us, this rate defines interactions on a node-node level. To glean information about how virulent our variants are, we look at the time until peak infection and monitor how this changes.

The **time taken until 50%** tells us about the diseases' transmissibility but frames it through the lens of the entire network such that we can easily classify successful diseases opposed to unsuccessful ones

Percolation

Other metrics we look at are more specific to this model but serve well in illustrating the impact of preceding variants and the impact on proceeding variants. Just before we get to those we acknowledge that individuals who Recover to a disease, can no longer interact with that disease. By the nature of our SIR model; the size of R_α only increases, every $n \in R_\alpha$ will never again be a part of S_α or I_α and the only way a node $n \in R_\alpha$ can interact with the system again is through the lens of a new disease. This means that for all intents and purposes the set of nodes $\{n \in R_\alpha\}$ no longer exist in the eyes of α. Any connections these nodes have are impassible for α and without another route, nodes at the end of these connections are unreachable. This situation is called site percolation and brings about interesting consequences for the network structure after a pandemic has passed.

At the start of each simulation run, a new network is generated and these networks are static: they do not inherently change. When an initial disease α is run we know that, so long as α infects the entire population, every node will eventually be within the set R_α. By Eq.4 we also know that for large enough populations, $|R_\alpha| \cdot r_{\alpha,\omega} \approx |R_\omega|$ for a new variant ω that no one has recovered from yet ($R_\omega \subseteq R_\alpha$). This means that the network structure the initial disease α saw is completely different to the one ω now sees as certain connections have been completely destroyed and whether a disease ω can percolate through the network is dependent on the strength of the relationship coefficient. For a disease δ where $r_{\alpha,\delta} = 0$, the network has not changed at all.

Network Metrics

By treating the ancestor disease run as a form of systematic site removal we can focus our attention on the changing properties of the network and use these properties as our aforementioned metrics. The network structure at the start of a secondary out-

break (as seen by the secondary disease) contains information about the influence the primary disease will have. Similarly the network structure as seen by the secondary disease at the end of the secondary outbreak contains information about the influence of the secondary on any tertiary disease. Three variants is a more complex case, however, as $R_\delta \subset R_\alpha \cup R_\omega$ meaning that the network structure as seen by δ takes influence from both diseases. We take the time to reinforce the idea that although R_ω started out as a subset of R_α [6], if R_α does not infect the entire system, there could exist a nonempty set of nodes $\{n \in R_\omega; n \notin R_\alpha\}$. The situation becomes even more unpredictable when diseases spawn before their parents' infect most of the network as now descendants will affect ancestors just as ancestors affect descendants.

To examine "network structure" we will look at the degree distribution of the network at key times, the component size distribution at key times, the degree assortativity against time and the Largest, Mean and Least number of connections against time as seen by both a primary and secondary disease.

2.5.1 Conditions

To truly understand how descendent diseases are impacted by their ancestors we want to track our previously defined metrics as we vary certain conditions/parameters. The parameters that immediately draw our attention are the R0 values of the diseases and the Relationship Coefficients between the ancestors and descendants.

When it comes to the **R0 value**, the infectiousness of the initial disease holds little significance, as a lone disease on the network leads us back to the basic SIR model. Instead our focus lies in varying the R0 of the second disease while keeping the first fixed, allowing us to discern the extent to which disparities in raw infectivity contribute to the success or mitigation of following variants. It is easy to imagine a scenario where an incredibly infectious secondary disease surpases the initial, in terms of the number of infected individuals, before the initial has peaked. However we can also imagine a scenario where the secondary disease is totally out-competed.

Although we forego studying the R0 value of the initial variant alone, *when* the second disease spawns in relation to the spread of the first is vital as the network will look different to secondary diseases at different times as described in the previous segment. With that in mind we will vary the time at which a secondary disease is spawned.

The Relationship Coefficient is probably the most intriguing to look at here as it directly dictates the level of immunity a population should have to disease ω after the spread of disease α but depending on the initial network type, the infectiousness of the secondary variants w.r.t the initial and the time at which the secondary variants spawn, it is unclear how substantial $r_{\alpha,\omega}$ will be in curbing secondary, or even tertiary diseases.

3 Results

In Sec.2 we detailed how we have constructed our model to investigate the behaviours, interactions and dynamics of related diseases acting concurrently and sequentially. Within the following section, we will highlight some of our findings and interesting observations.

3.1 SIR Curves:

To get a sense of what is happening here we run the random regime [7] for 10,000 nodes with a cutoff of 5 variants to see if we can observe any distinct patterns not common to standard SIR infection curve plots. In Fig.3 we see two simulation runs with the exact same starting parameters. Both sim-

[6] assuming a nonzero $r_{\alpha,\omega}$

[7] as described in Sec.2.4

ulation runs have 5 random variants running on a Barabasi Albert generated network (scale-free) and we can clearly observe that very different events happen on these two runs despite the initial conditions being the same. This is to be expected as the initial conditions in the random regime fix only the characteristics of the network generated and the disease spawned at $t = 0$. This means that the specific connections for any given node on the network and the properties of subsequent variants that are not the initial will vary from run to run. In Fig.3(a) we see both variants A and B rapidly spread through the population, infecting well over 60% before the other variants get off the ground. Their twin success is very telling as it suggests that these variants barely interact/compete with each other. To dually infect close to 70% of the population at the same time, must mean that nodes that are immune to A are seldom immune to B and vice versa.

Their peak infection points are also very telling as variant B manages to surpass A. We might expect that even with very little relation to each other, if $r_{A,B}$ is non-zero then A, as the predecessor, will dampen B to some degree but this seemingly doesn't happen here. Looking at the basic rates of reproduction (calculated using the rate of infectivity β divided by the rate of recovery γ) for both diseases we see that $R0_B > R0_A$with $R0_A \approx 1.22$ and $R0_B \approx 1.49$. We can take this to mean that although disease A spawned first the diminishing factor of $r_{A,B}$ wasn't large enough to prevent the secondary variant B from outperforming A in the network.

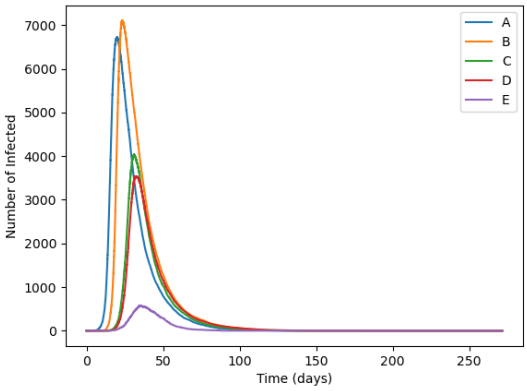
We can confirm the competitiveness of the diseases by looking at the Relationship Coefficient matrix which describes the probability that a victim from one disease will become immune to another when it recovers from the first for all combinations of diseases. In Fig.4 we see the Relationship Coefficient matrix generated for Fig.3.(a).

$$\begin{bmatrix} 1 & 0.01033 & 0.34553 & 0.28081 & 0.01291 \\ 0.01033 & 1 & 0.00004 & 0.00002 & 0.99721 \\ 0.34665 & 0.00004 & 1 & 0.99080 & 0.00006 \\ 0.28081 & 0.00002 & 0.99080 & 1 & 0.00003 \\ 0.01291 & 0.99721 & 0.00006 & 0.00003 & 1 \end{bmatrix}$$

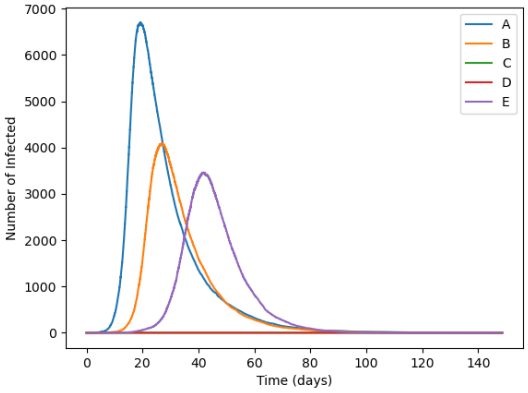
Figure 4: Relationship Coefficient Matrix for the 5 variants in the Fig.3.(a) simulation. The symmetry of the matrix is described by eq.4.

In the first row, which describes relations between A and all other diseases, $R_{A,B} \approx 0.01$. This is very low as it suggests that if A were to infect the entire population unperturbed approximately 100 of the total 10,000 nodes would be immune to B which explains the non-competitive nature of the A and B infection curves seen in Fig.**??**.(a). The other rows in Fig.4 can help us understand the dynamics pertaining to the other variants as well. Variants C and D take off near the same time as each other and share similar levels of relation with variants A and B which causes their curves to look very similar, though we must note that C and D themselves share a strong relationship coefficient: $r_{C,D} \approx 0.99$. Their strong levels of relation results in neither $|I_C| > 0.5N$ nor $|I_D| > 0.5N$ being true; the variants are actively competing and most nodes that recover from one become immune to the other. Despite $r_{A,C} > r_{A,D}$, Fig.3.(a) shows C to be more infectious, this must be a result of the fact that $R0_C \approx 1.29$ whilst $R0_D \approx 1.22$. Variant B has very negligible relationship coefficients with both C and D so we exclude diminished curves as a result of B as a possibility when discussing C and D. With that said, however, we observe in the infection curve plot that variant E does not infect any sizeable portion of the population. Fig.4 shows us that this is the result of being completely dampened by the effects of variant B: $r_{B,E} = 0.99721$.

Moving our attention toward Fig.3.(b) we see that only 3 variants contribute to the dynamics of this simulation. The most obvious initial guess, especially after having looked at Fig.3.(a) is that severe dampen-



(a) Variants C & D are closely related; Variants B & E are closely related; A is relatively independent.



(b) Variants A & B are fairly related ($r_{A,B} \approx 0.432$) whilst variant E is barely related to either.

Figure 3: Infection.vs.Time plots for 5 variants on a Barabasi-Albert network with 10000 nodes. Both runs have identical initial parameters with $R0_A = 1.22$ and new node connections for generating the network, $m = 5$

ing of descendent variants is occurring as a result of high relationship strength coefficients but when examining the relationship matrix for this system we find this not to be the case. For Fig.3 we find that:

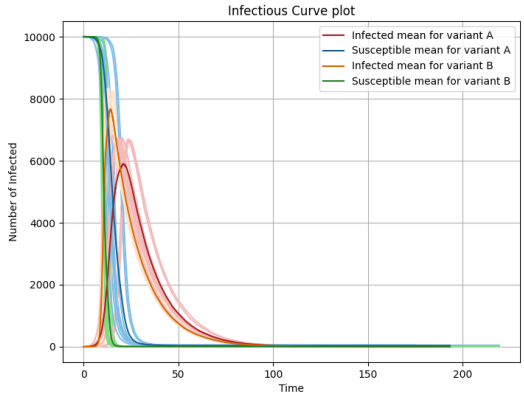
$$r_{C,\omega} < 0.45 \quad \forall\, \omega \in \Omega \quad \text{where } \omega \neq C \quad (5)$$

$$r_{D,\omega} < 0.45 \quad \forall\, \omega \in \Omega \quad \text{where } \omega \neq D \quad (6)$$

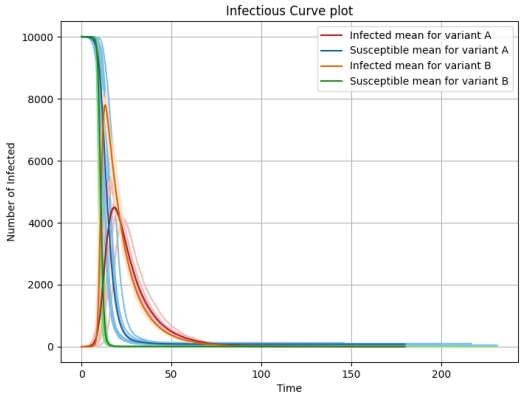
This leads us to investigate other causes and we realise that reason variants C and D do not get off the ground is because both of their respective R0 values are less than one. With a sub-unitary R0 value, it is more than plausible that the patient zero for both diseases recovered before infecting another node.

Unlike Fig.3.(a), the diminished infection curves of B and E in Fig.3.(b) are not the result of direct competition with each other ($r_{B,E} = 0.04607$). Variant B shares a relationship coefficient with A of $r_{A,B} = 0.43724$ which explains why B is so reduced but variant E has an $r_{E,\omega} < 0.05\ \forall \omega \in \Omega$; $\omega \neq E$. This suggests that $R0_E < 1$ which we find to be true but we also find that $R0_E < R0_D < R0_C < 1$ and yet remarkably E is the variant that 'takes off'. While initially surprising, it's imperative to acknowledge that the R0 value does not universally dictate every interaction within the system. This phenomenon can be attributed, in large part, to the heterogeneous structure inherent in the network which is preserved due to the low $r_{E,\omega}$ value. In this scenario, we can easily imagine a hub node carrying the infection to many parts of the system but due to the sub-unitary R0 value, it will eventually die out[9]. Whilst we cannot say that this behaviour is exactly happened or that it cannot occur on other network structures, it is a helpful way to understand the existence of 'localised' epidemics in the face of sub-unitary R0 values.

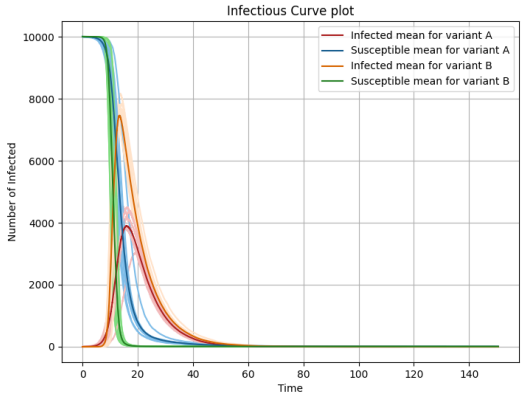
For the rest of this report we shift our focus towards controlled variables within this simulation framework as outlined in Sec.2.4 and Sec.2.5. We hope that by simplifying our system to 2 variants and imposing controlled conditions that we vary, we can gain a deeper understanding of the causal relationships and mechanisms that determine disease spread. With a first look at Erdos-Reyni networks, we examine the effects of fluctuating the relationship strength, the R0 value of the secondary variant and the time at which the secondary variant is spawned.



(a). $r_{A,B} = 0$, $R0_B = 3, t_D = 5$



(b). $r_{A,B} = 0.5$, $R0_B = 3, t_D = 5$



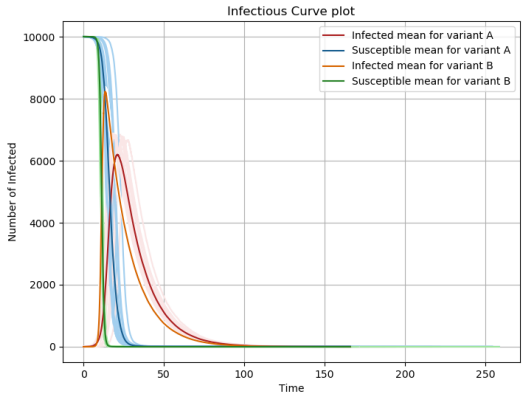
(c). $r_{A,B} = 1$, $R0_B = 3, t_D = 5$

Figure 5: Infection.vs.Time plots for 2 variants running on the randomised Erdos-Reyni network (with an average degree of 10). All three plots are run with the same initial conditions bar the relationship coefficient. Each of these plots are run over 25 simulations (which the faded lines represent) and averaged to give mean infection and susceptibility curves (bold lines).

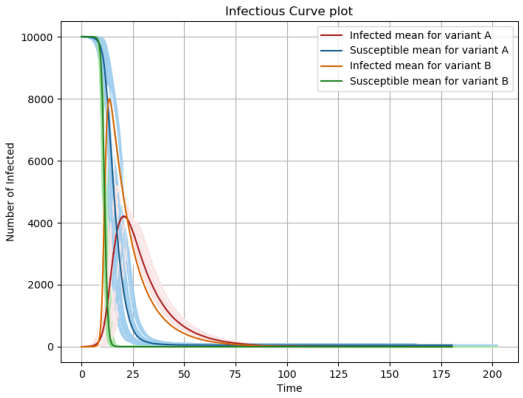
Varying Relationship Coefficient

The three plots in Fig.5 demonstrate the extent of the influence the relationship coefficient has on the spread of a secondary variant that is much more infectious than its predecessor. Considering only the basic rates of reproduction for both variants, $R0_B = 3.0$ & $R0_A = 1.22$ we expect that the secondary variant B will be far more infectious in a shorter time span than variant A which is exactly what we observe to be true in all 3 figures. Focussing on Fig.5(a) we see that the peak infection for variant B hangs around 8000 nodes whilst the peak infection of variant A sits at around 6000 nodes. Despite the fact that the initial variant A had a small headstart, the rate of infection for variant B is so much larger that it almost suddenly surpasses variant A in infections. As the relationship strength increases from $r_{A,B} = 0$ in Fig.5.(a) to $r_{A,B} = 0.5$ in 5.(b), we observe the overall infectious curve for variant A diminishing. The spread of A is being actively inhibited which is a phenomenon we've seen already in Fig.3.(a) where the dominant variant in the network impedes the spread of related variants. We also note that the 'spread' of the subdued variant does not fall with a rate proportional to $r_{A,B}$.

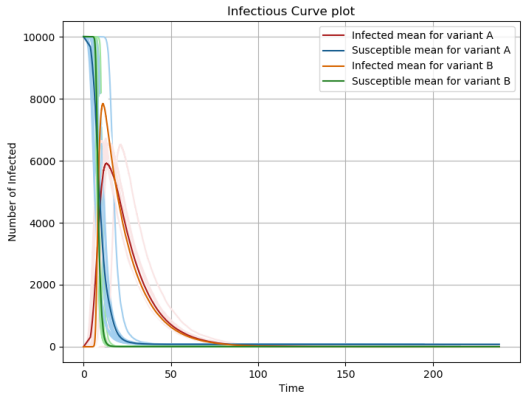
In Fig.6 we see how the same conditions for Fig.**??**.(a) and Fig.**??**.(b) applied



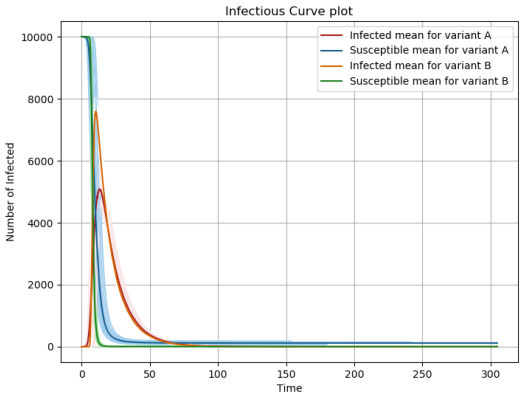
(a). Watts Strogatz network with initial conditions: $r_{A,B} = 0$, $R0_B = 3, t_D = 5$



(b). Watts Strogatz network with initial conditions: $r_{A,B} = 0.5$, $R0_B = 3, t_D = 5$



(c). Barabasi Albert network with initial conditions: $r_{A,B} = 0$, $R0_B = 3, t_D = 5$



(d). Barabasi Albert network with initial conditions $r_{A,B} = 0.5$, $R0_B = 3, t_D = 5$

Figure 6: Infection.vs.Time plots with the same conditions as Fig.5.(a) and Fig.5.(b) but the network structures are different. Plots (a) and (b) represent runs on a Watts-Strogatz network with $r_{A,B} = 0$ and $r_{A,B} = 0.5$ respectively. Similarly plots (c) and (d) represent runs on a Barabasi-Albert network with $r_{A,B} = 0$ and $r_{A,B} = 0.5$.

on different network structures manifests slightly different curves whilst conforming to the same overall behaviour. Looking at both Fig.6.(a) and Fig.6.(c) we observe that the secondary variant (as a result of its extremely high R0 value) completely overtakes the primary variant and has a peak infection around 8000 nodes. This is the classic SIR result given that $r_{A,B} = 0$ and is equivalent to running the simulation for each variant separately. Interestingly enough, in Fig.6.(b) we see that the peak of variant B is diminished as a result of the new relationship coefficient being $r_{A,B} = 0.5$. This contrasts the Erdos-Reyni simulation in Fig.5.(b) where the much more infectious variant B inhibits the spread of variant A. The same behaviour is observed for the Barabasi-Albert simulation in Fig.6.(d) which implies that the significance of the ancestor variant on the peak-infections for the descendent variant is greater for scale-free and small-world networks than completely random Erdos-Reyni networks. We also note the jaggered infection curver of variant B in Fig.6.(d) which seems to be a result of a few individual runs immediately finding hub noddes and quickly spreading across the system in comparison to the average time taken for this variant to find a

densley connected hub node. Ultimately we see the same type of behaviour across all three networks but with Erdos-Renyi networks we observe that the most dominant variant on the network impedes the spread of the other whereas with Watts-Strogatz, the earliest spawning variant appears to hold more weight in impeding the spread of the later variants and with Barabasi-Albert this effect seems to only be more significant.

Increasing time delay

As we've seen aluded to when discussing some of the previous plots, the foothold the first variant has on the network can greatly determine the outcome of the second variants' spread and peak infection but without directly changing any of the properties regarding the initial variant we can use the time delay between variant spawns to give the initial variant more or less of an advantage. Changing the time delay when $r_{A,B} = 0$ is meaningless as the two variants do not interact but when $r_{A,B} > 0$, a larger time delay results in a higher percentage of nodes being impassible for the second variant. This is demonstrated in Fig.7. Not only can we see that the initial variant A has greater rise and peak infection number in Fig.7 in comparison to Fig.5.(b) but we can also see that just by increasing the time delay, the curve of variant B in Fig.7.(b) never surpasses A. This shows us that the relationship coefficient must be considered with respect to the time delay to determine how much of the original network variant B can still percolate through.

Reducing the R0 of variant B has a rather trivial effect of reducing the peak infection and spread of variant B but when we consider varying $R0_B$ in tandem with one of the two previously explored variables we observe some non-trivial dynamics. Fig.8 provides a demonstration of the infection curves when we select the relationship coefficient, time delay and $R0_B$ parameters to drastically diminish the spread of variant B.

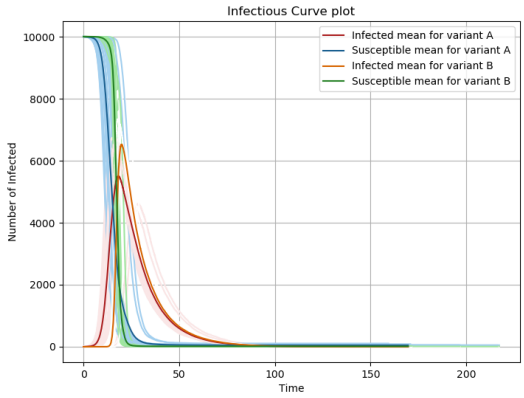
3.2 Contour Plots

To observe the differences of our infection life-cycles on a larger scale we draw a variety of contour plots from 2 of 3 axes: Relationship Coefficient - $r_{A,B}$, basic reproduction number of the secondary variant B - $R0_B$ and the time delay between the two variant spawn times - t_D. We draw these contour plots to evaluate several metrics of our simulation to see if we can observe any distinctive characteristics between and across network types. To narrow the focus of our scope we mostly discuss the effects and properties of variant B under changing input parameters.

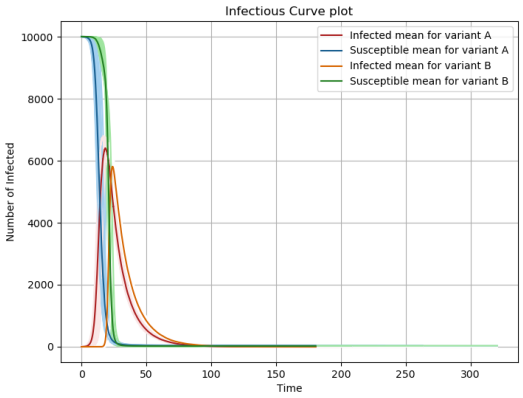
Peak Infections

Fig.9 illustrates how the number of nodes infected by the secondary variant at the height of its infection varies as a function of $R0_B = \frac{\beta_B}{\gamma_B}$ and $r_{A,B}$. As mentioned in Sec.2, the recovery rates, γ, of both variants will remain constant and equal. This means that any changes to $R0_B$ are only changes to β_B but with stick use the basic reporduction number for its clarity. Note also that $R0_A$ will remain unchanged for the duration of our analysis on the dual variant simulations. Aligning with, perhaps intuitive, expectations gathered from the prior SIR curves, Fig.9 shows a noticeable transition where variant B goes from peaking at 10% of the population to infecting over 80% at its peak. The isopleths which delineate regions of peak infection of equal size across the dual parameter space have a mostly positive gradient within these plots. This confirms the contrasting relationship that we conjectured $R0_B$ and $r_{A,B}$ to have on the spread of the secondary variant. Following a linear path from the highest point on these graphs to the lowest we can identify a pronounced gradient shift [8] prevalent in all 3 graphs that is steeper near the lower peak-infection values and shallower near the higher peak infection values. This implies that regardless of network type, the sensitivity to parameter changes differs at different regions in

[8] in the z-axis / along the peak-infection surface



(a). $r_{A,B} = 0.5$, $R0_B = 3$, $t_D = 11.\dot{6}$



(a). $r_{A,B} = 0.5$, $R0_B = 3, t_D = 15$

Figure 7: Infection.vs.Time plot with the same initial conditions as Fig.5.(b) but with an increased time delay between the first and second variants.

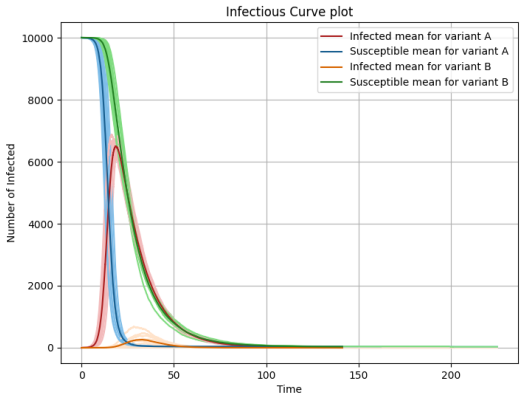
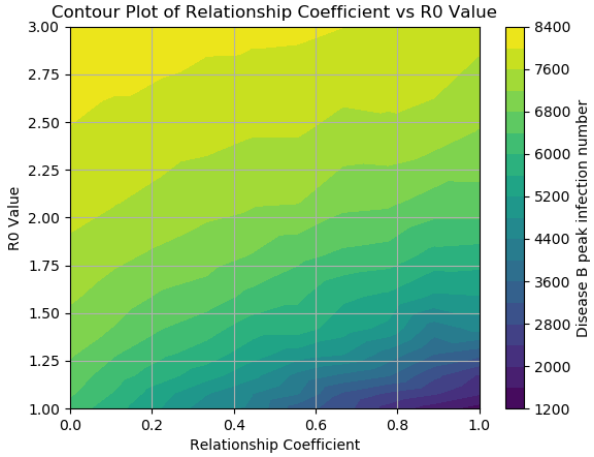
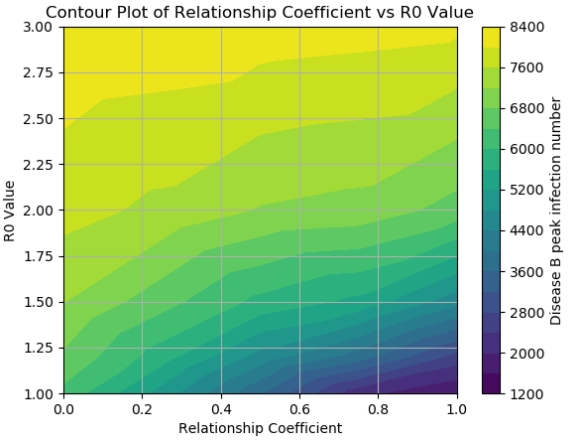


Figure 8: Infection.vs.Time plot for 2 variants on an Erdos-Reyni network with 10000 nodes. $r_{A,B} = 1$, $t_D = 10$, $R0_B = 1$

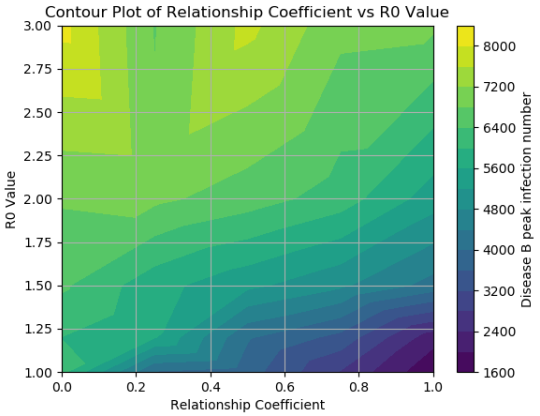
the parameter space of R0 .vs. Relationship Strength. Taking Fig.9.(a) as an example; the range of peak-infection values across $r_{A,B} = 0$ is $\Delta Peak \approx 487$ whereas the range of peak-infection values across $r_{A,B} = 1$ is $\Delta Peak \approx 4819$. Across all three plots in Fig.9 we notice slight differences in the gradients of the isopleths but note that there are no discernible distinguishing features between the contour plots. This suggests that irrespective of internal network structure the dual variant simulation contains characteristic dynamics which are consistent throughout. In Fig.9.(a) we spot that the scale of peak-infection values has a lower bound of approximately 1200 nodes. By increasing the time delay we might expect that only the scale of our plots would increase but instead, Fig.10 shows that our peak-infection surface is much more fluid than anticipated. In this plot we have taken Fig.9.(a) and increased the time delay between the spawning of both diseases from $\Delta t_D = 5$ to $\Delta t_D = 10$



(a). Erdos-Reyni generated network with a probability of edge generation at 0.001 for an average degree of 10



(b). Watts-Strogatz generated network with each node connected to 10 nearest neighbours and a rewiring probability of 0.5



(c). Barabasi-Albert generated network with the new node parameter $m = 5$

Figure 9: Contour plots of $R0_B$ vs $r_{A,B}$ for 3 different network types showing the peak infection number of disease B. The time delay between these two diseases spawning is low at $\Delta t = 5$ and each network should have a mean degree of 10. $R0_A = 1.22$

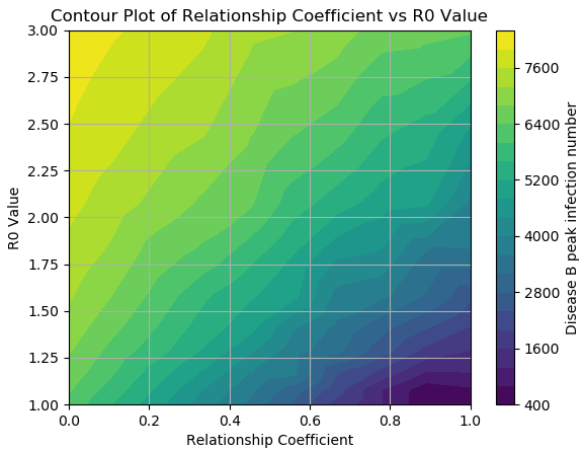


Figure 10: Infection.vs.Time plot for 2 variants on an Erdos-Reyni Network where B is spawned 5 units of time after A. Variant B is inherently 'more infectious' than A with $R0_A = 1.22$ and $R0_B = 3.0$. The variants are related by $r_{A,B} = 0.5$

Looking at Fig.10 we can identify two stark differences between this plot and Fig.9.(a). The first being the greater impact that $r_{A,B}$ now has on the peak-infection values of B. The increased gradients of the isopleth lines (of equal peak-infection value) imply that with a longer incubation time on the network for variant A, the success and spread of variant B becomes more de-

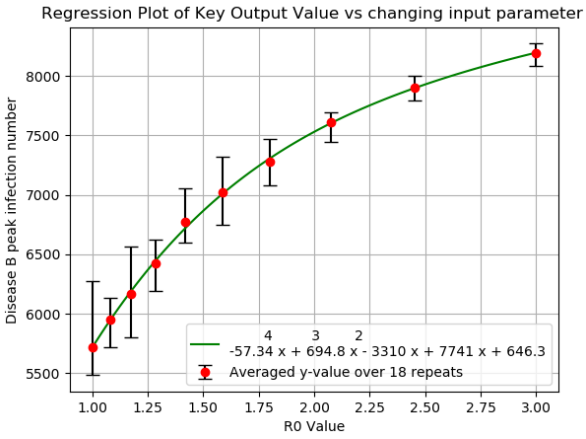
pendent on the relationship coefficient $r_{A,B}$. The other main difference is one we anticipated: the peak-infection value scale has lowered. For both Fig.9.(a) and Fig.10, the maximum peak infection number for variant B is roughly 8000 nodes but when the delay between spawn times is low, the peak infection number for variant B bottoms out at 1400 nodes. In contrast, increasing the time delay results in the spread of variant B being dampened to a greater degree. It should be noted here that much like Fig.9, the differences between Fig.10 and the same plot for a Watts Strogatz generated network are minimal. Ultimately, these plots give us a glimpse at how the behaviour we saw in Sec.3.1 applies for the plane of $R0_B \times rel_{A,B}$ and is in line with what we observed:

- For parameters where variant A is the dominant variant in the system, increasing $r_{A,B}$ will attenuate the intensity of the outbreak for variant B,
- The decrease in the peak-infection number of variant B has a diminished impact as $R0_B$ increases
- Increasing the delay between spawn times of the two variants has some causal impact on the sensitivity to $r_{A,B}$ that the spread of B has

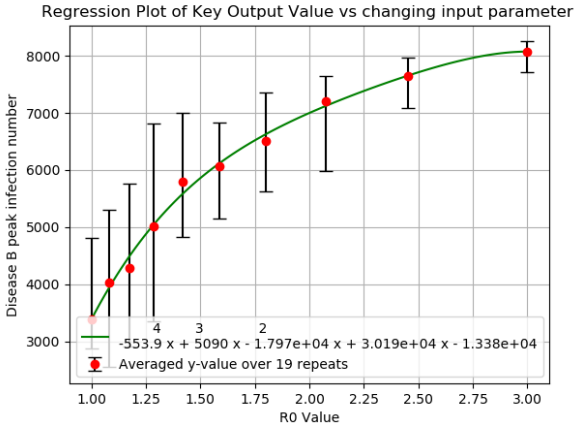
Fig.11 shows explicit cross sections of the data of the peak-infection surface in Fig.9.(a) for 3 main vertical intersections; $r_{A,B} = 0.1$, $r_{A,B} = 0.5$ and $r_{A,B} = 0.9$. In this plot, we perform polynomial regression with degree 4 to fit the mean of the data (seen in the Fig.9 contour plots) at 10 points along these intersections where $r_{A,B} = 0.1$, 0.5 and 0.9.

The concavity seen in all four plots immediately confirms one of our first observations about the plots in Fig.9, changes in $R0_B$ for lower $R0_B$ result in greater changes to the peak infection number of B - the peak-infection surface gradient decreases. Fig.11.(d) takes the three surface regression fits and then plots them together. This figure is arguably the most interesting plot in the set as it confirms that, much like $R0_B$ itself, the sensitivity the peak-infection number of B has towards $r_{A,B}$ is greater when the $R0_B$ value is lower - the gradient of the peak-infection surface steepens as $r_{A,B}$ for lower $R0_B$. The spread between the lines $r_{AB} = 0.1$ and $r_{A,B} = 0.9$ at the points of $R0_B = 1$ and $R0_B = 3$ show that in contrast to the relationship coefficient having a reduced effect on B at higher infectivity rates, $r_{A,B}$ becomes a dominating factor at lower values of $R0_B$ which makes intuitive sense; if the two variants are relative in terms of infectivity (or the second variant is less infectious), variant B's ultimate success is going to be dependent on how much of the original network structure the primary variant destroys for the second. With a higher $r_{A,B}$ value much more of the network will be out of reach for variant B in contrast to a lower $r_{A,B}$ value.

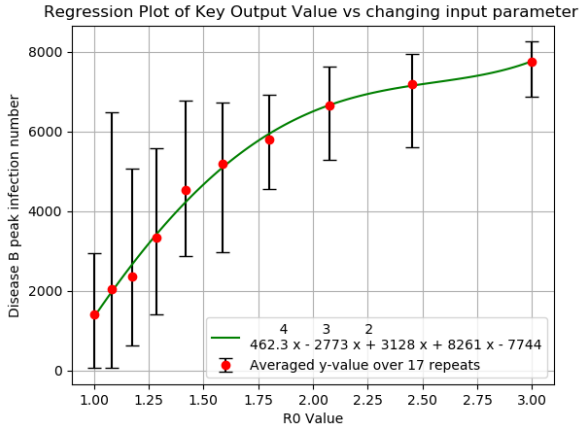
To assess the alignment of our regression fit with the contour data we use the Root Mean Squared Error (RMSE) which will tell us how far our fit lies from the data. For Fig.11.(b) the RMSE value we obtain for our fit is $\approx$ 109.25 which tells us that on average, our fit is within the range of 110 nodes when predicting the peak infection value of B for the parameters $r_{A,B} = 0.5$, time-delay = 5. Fig.12 plots this same regression fit and overlays it atop the data for the peak-infection surfaces of the Watts-Strogatz network and the Barabasi-Albert network produced in Fig.9 such that we can examine the extent of similarity these networks' contour plots have with the Erdos-Reyni contour plot. For both of these plots, we can see that our ER fit maps the data quite well with our curve going between every point range. We also observe that while the ER fit does map to our points, there is a tighter correlation with the mean values for the Watts-Strogatz plot than there is for the Barabasi-Albert plot. This observation is supported by the RMSE values of the ER fit for the WS data (RMSE $\approx$ 152) and the BA data (RMSE $\approx$ 403). While the RMSE for the BA plot is not obscene, given that the order of nodes at the low-



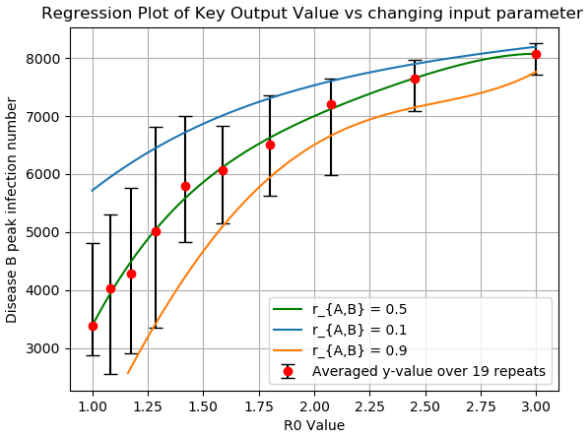
(a). 4 degree polynomial regression on Fig.9.(a) at $r_{A,B} = 0.1$



(b). 4 degree polynomial regression on Fig.9.(a) at $r_{A,B} = 0.5$

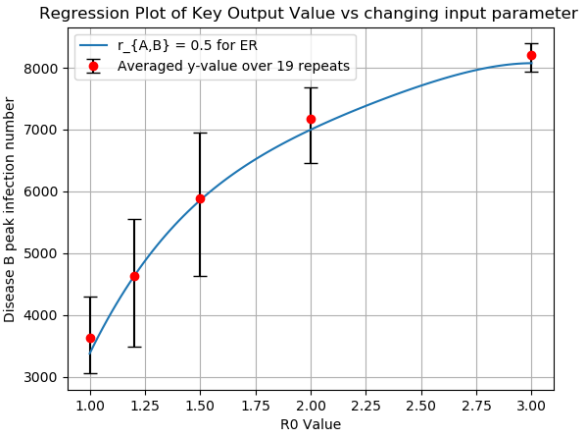


(c). 4 degree polynomial regression on Fig.9.(a) at $r_{A,B} = 0.9$

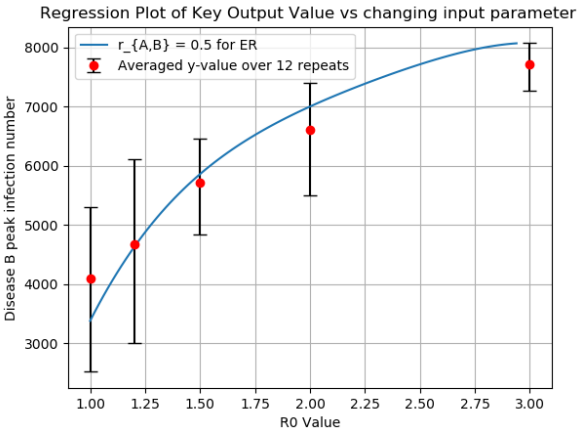


(b). Combined polynomial regression plots of (a), (b), (c) overlaid on (b)

Figure 11: Polynomial regression on Fig.9.(a) at different vertical $r_{A,B}$ intersections



(a). Erdos-Reyni generated network with a probability of edge generation at 0.001 for an average degree of 10



(b). Watts-Strogatz generated network with each node connected to 10 nearest neighbours and a rewiring probability of 0.5

Figure 12: 4 degree polynomial fitting Erdos-Reyni peak-infection surface of variant B overlaid on Watts-Strogatz peak-infection surface of B and Barabasi-Albert peak-infection surface of B

est end $\tilde{3}000$, this numerically tells us that the peak-infection surface of B exhibits the most dissimilarity on a scale-free network in comparison to the randomised Erdos Reyni network and pseudo-random, small world Watts Strogatz network.

As we saw with Fig.10, the time delay has a yet-unknown impact on the landscape of the $R0_B \times r_{A,B}$ plane but given its influence on the sensitivity of $r_{A,B}$ we recognise it as a variable worth investigating. This will involve further contour plots for the $R0 \times$ Time plane and the Time $\times r_{A,B}$ plane. In Fig.13 we show the peak infection numbers for variant B in both of the aforementioned planes. Whilst still in consideration, we have opted to omit the plots for the WS and BA networks due to their close likeness to the Erdos-Reyni plots (as seen with Fig.9).

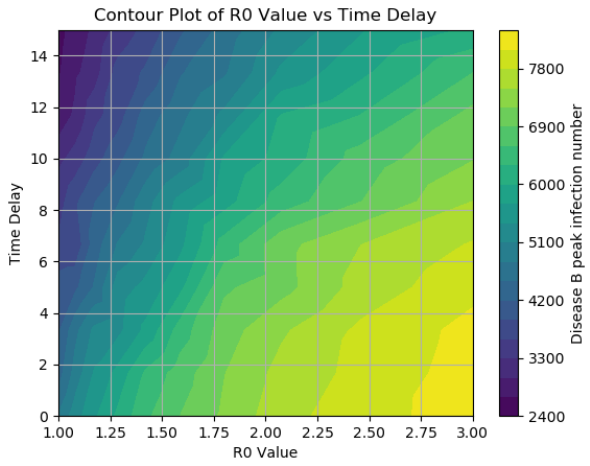
It is immediately apparent when looking at these plots that there is a clear non-linear relationship between the time delay and both $r_{A,B}$ and $R0$ in effecting the peak infections of B. For Fig.13.(a) the relationship coefficient is set at $r_{A,B} = 0.5$. We recall that in Fig11.(b) which plots the $r_{A,B} = 0.5$ intersection of the peak-infection surface for B on an Erdos-Reyni network, the peak-infection curve is much steeper for lower values of $R0$ and plateaus for larger values. We can see this same pattern in Fig.13.(a) for low values of Time Delay, $t < 8$. Choosing a low Time Delay and looking along the horizontal axis we can see that, much like in Fig.9, the gradient of the surface representing the peak infections of variant B is larger for lower values of R0. Due to the non-linear relationship between Time Delay and the R0 value, or in other words, because the surface that represents the peak infections of B is not flat but contorts the gradient as seen horizontally in Fig.13.(a) becomes more consistent across the range of R0 values. This could potentially explain the differences we see between Fig.9 and Fig.10.

In Fig.13.(b) the basic reproduction number of variant B is equal to that of A, $R0_B = R0_A = 1.22$. Much like we did for Fig.13.(a) we can see here that due to the curve of the peak infection surface, for a fixed Time Delay the gradient of said surface is lower for smaller Time Delays and higher for larger Time Delays. The plot Fig.13.(b) is particularly interesting as it is apparent to see here that when the Relationship Coefficient, $r_{A,B}$ is near zero, our simulation tends to the classic, individual SIR disease on a network, case. When $r_{A,B}$ is low, the time at which it is released becomes irrelevant as it no longer interacts with A; whether time-delay = 0 or 50, the moment variant B begins, it will follow the standard SIR curve based on the network size, type and $R0_B$. In contrast, the near-vertical isopleths at high $r_{A,B}$ show an extremely dependent relationship between the Time Delay and the peak number of infected nodes of B. This is in line with our previous assertions about the system dynamics at 100% relation; for every unit of time that passes disease A evolves undisturbed, which results in potentially another node becoming immune to B. At high levels of relation, (and low enough R0 values) the spread of B is contingent on how much of the original network remains intact.

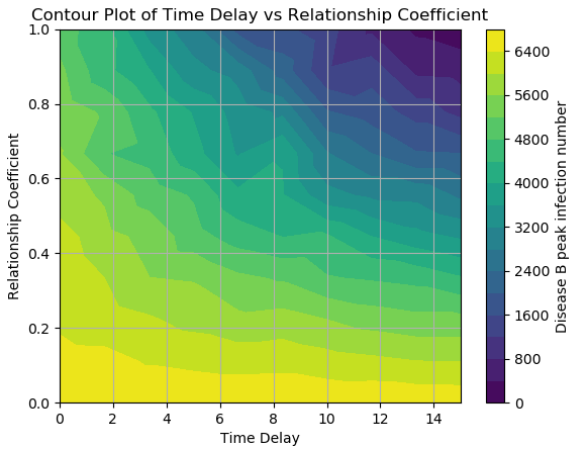
In Fig.14 we use polynomial regression again but this time instead of fitting the gradient of peak-infection surfaces we fit the isopleths in Fig.9, Fig.10 and Fig.13 to obtain functions which describe the relationships between Time-delay,$R0_B$ and $r_{A,B}$.

Without any additional analysis, these plots appear to be quite accurate in plotting the level surface of the Peak Infections of variant B, which is very useful in a more holistic context [9]. Understanding the relationship strength needed for a secondary virus with a specific Time Delay and R0 to be below some critical level could have practical significance regarding the IDS situation outlined in Sec.1. Unfortunately, since

[9] discerning the accuracy of these plots to produce an error value without turning to another metric will involve knowing the true function of the peak-infection surface (or modelling that) which we don't pursue in this report



(a).Contour plot of t_D vs. $R0_B$ where $r_{A,B} = 0.5$



(b).Contour plot of $r_{A,B}$ .vs. t_D where $R0_B = 1.22$

Figure 13: Contour plots of time delay t_D vs. $R0_B$ and t_D .vs.$r_{A,B}$ respectively for the Erdos Reyni network. $R0_A = 1.22$

we do not plot the entire 3-dimensional space of $R0 \times r_{A,B} \times$ Time-Delay [10], these regression plots are limited in their predictive capabilities. Additionally, since the gradient of the peak-infection surface is not uniform across the entire space, additional regression analysis would have to be done for varying isopleth levels but at the very least the regression plots in Fig.14 give us a good basis for what the variable relationships look like.

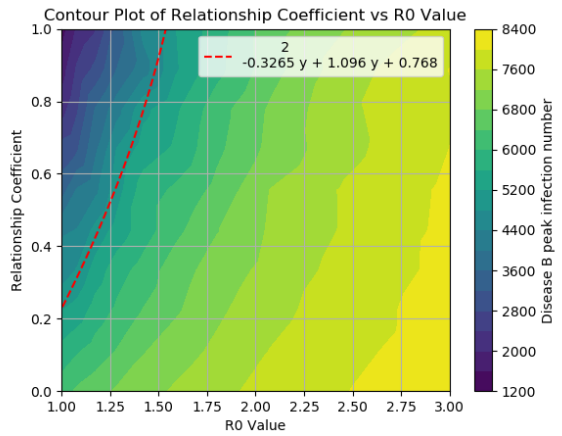
Critical time-point contour plots

We now briefly move away from looking at the peak infection of disease B to analysing another metric that gives us an insight into how the secondary variant B spreads; the time at which 50% of the population has been infected. Whilst the peak infection number of a disease is useful for understanding the extent of prevalence said disease has within the network, it doesn't tell us about the rapidity of its spread. In real-world settings, slower-spreading viruses are most commonly easier to track, manage and overcome therefore we dedicate this section to studying the conditions which give birth to slower-moving epidemics. We first note that some infectious curves for the secondary variant may never reach 50% of the population as a result of dampening caused by A. This means there exists a bounded region in the parameter space where the secondary variant cannot infect over 50% of the population at any given time. This boundary is equivalent to the isopleth which represents the peak infection of variant B being equal to 50% of the total population (5000) which Fig.15 substantiates.

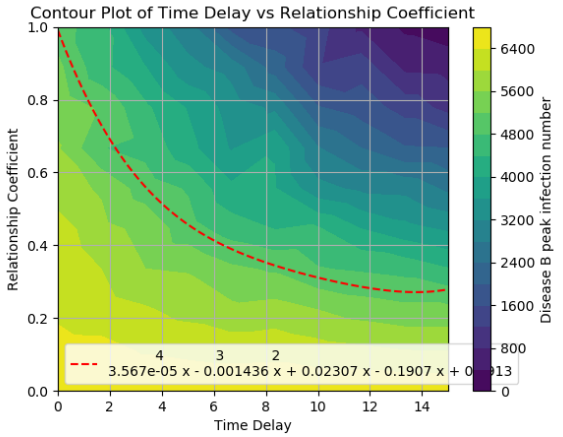
In Fig.15 we set simulation runs where a variant B will never reach 50% of a population (due to the peak being lower) as having a value for 50% infection time = 500, where the order of magnitude for most runs that do hit 50% of the population is between 0 and 30. As previously mentioned, these curves show a distinct and sharp boundary as a result of the set of nodes that do not infect over 50% of the population. We can see that for Fig.15.(a) this plot has axes $t_D.vs.R0_B$ and when we compare it with Fig.13.(a) the curve of the isopleths in Fig.13.(a) seem to follow the curve of the boundary in Fig.15. We note the same behaviour with Fig.15.(b) and Fig.13.(b).

In fact, we overlay the isopleth polynomial regression plots made in Fig.14 for the peak infection observed in Erdos-Reyni networks atop the time until 50% infec-

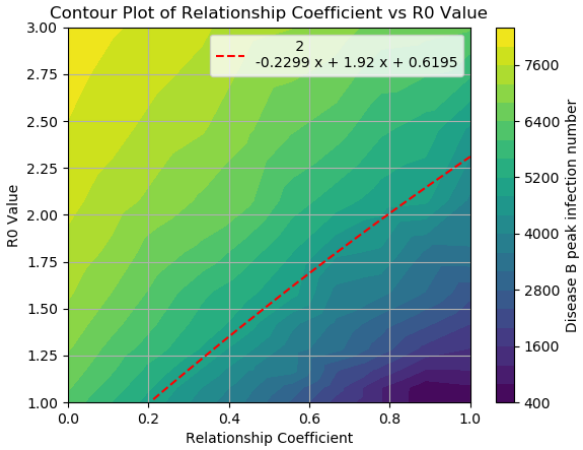
[10] for computational complexity reasons



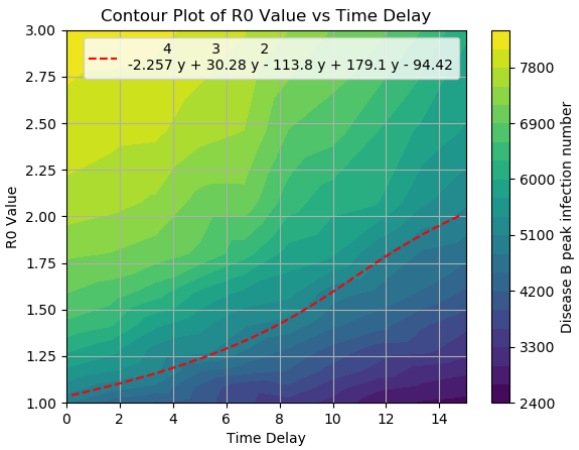
(a). Inverted Plot of Fig.9.(a). Using polynomial regression with degree 2



(b). Plot of Fig.13.(b). Using polynomial regression with degree 4



(c). Plot of Fig.10. Using a polynomial regression with degree 2.

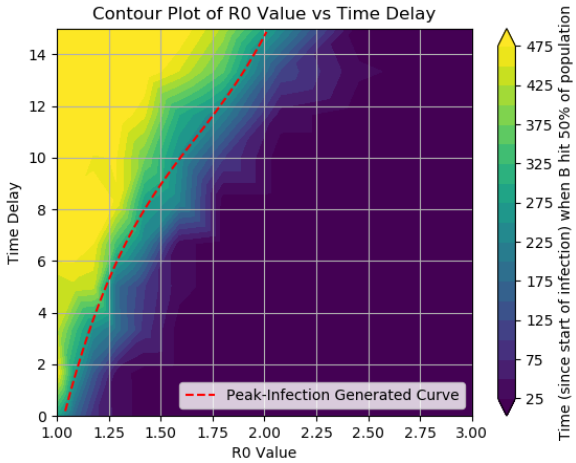


Inverted plot of Fig.13.(a). Using a polynomial regression with degree 4.

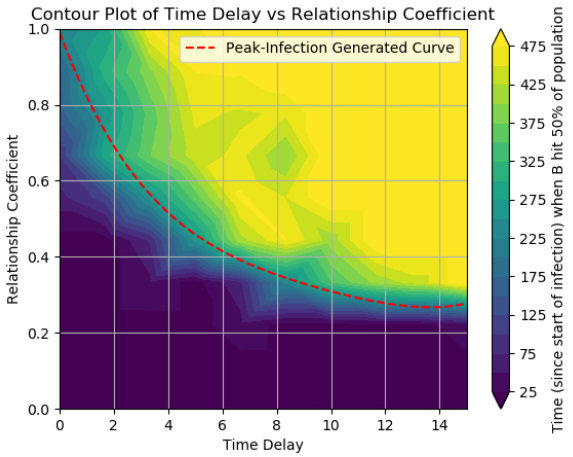
Figure 14: Fitting the isopleth level = 5000, we have 4 previous plots from Fig.13, Fig.9 and Fig.10 overlaid with a polynomial regression fit. Axes for plots (a) and (d) inverted for easier plot comparison. Plots (a) and (c) use degree 2 instead of 4 to avoid overfitting.

tion contour maps but for Erdos-Reyni, Watts-Strogatz and Barabasi-Albert networks. From these plots, and keeping in mind the RMSE values for our ER regression plot on WS and BA, we can see that the ER isopleth very accurately bounds the Erdos-Reyni and (to a lesser degree) the Watts-Strogatz simulations but begins to falter with the Barabasi-Albert network, particularly at high t_D values. This suggests that at the extremes of high time delays this model differs the most between networks (given that we begin to see differences arise between ER and WS for high t_D as well). This is a topic we discuss more when looking at the differences in network structure vs time.

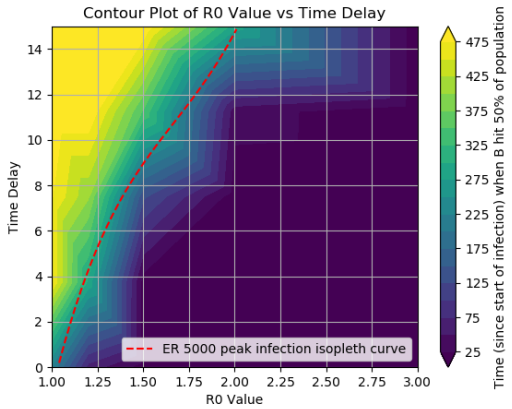
To get a more balanced contour map that is not overpowered by the variants that fail to reach 50% of the network we plot the time at which a peak infection occurs in Fig.16. In these contour maps, we can see that they share some resemblance with Fig.15 which makes intuitive sense. For most infection curves, the time at which 50% of the population has been infected (a major milestone for the disease) is usually very near to the time at which peak infection of that curve occurs (looking at Sec.3.1 we see variants with very sharp initial gra-



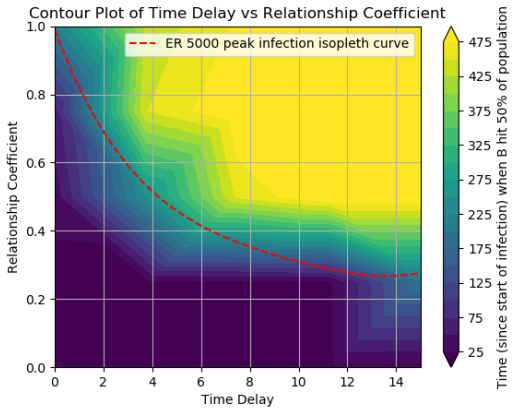
(a). ER - $t_D.vs.R0_BB$; $r_{A,B} = 0.5$



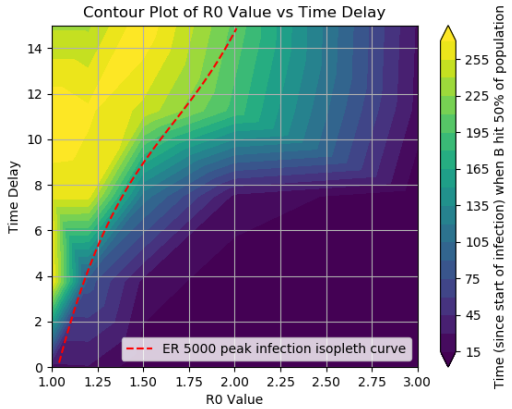
(b). ER - $r_{A,B}.vs.t_DB$; $R0_B = R0_A = 1.22$



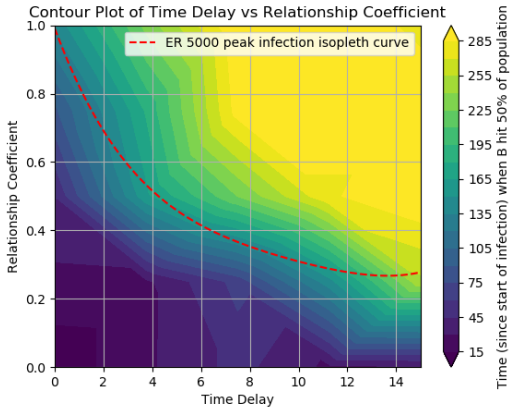
(c). WS - $t_D.vs.R0_B$; $r_{A,B} = 0.5$



(d). WS - $r_{A,B}.vs.t_D$; $R0_B = R0_A = 1.22$



(e). BA - $t_D.vs.R0_BB$; $r_{A,B} = 0.5$



(f). BA - $r_{A,B}.vs.t_DB$; $R0_B = R0_A = 1.22$

Figure 15: Time until 50% of the population has been infected. The left column of figures plots the time until 50% infection for the simulation described by Fig.13.(a), whilst the right column of figures plots the time until 50% infection for the simulation described by Fig.13.(b). Each row corresponds to a different network structure with (a) and (b) being Erdos Reyni, (c) and (d) being Watts Strogatz and (e) and (f) being Barabasi Albert. Each figure on the left column is overlaid with the fit created for Fig.14.(d) and each figure on the right column is overlaid with the fit created for Fig.14.(b). The time has been normalised w.r.t the time delay.

dients). Comparing these maps with Fig.13 like we had done with Fig.15 we see that Fig.16 much more closely resembles the inverse of the peak-infection numbers plotted by Fig.13. Intuitively this also makes sense because variants with a higher infection peak typically have an extremely high transmissibility rate as opposed to variants with smaller infection peaks and 'flattened curves' [11]. Interestingly, however, we notice that in Fig.16.(b) if both $r_{A,B}$ and t_D are high enough then the peak infection time begins to plot. This is a result of variants that have been almost completely impeded by the spread of the initial variant resulting in only a handful of nodes infected.

3.3 Network Connectivity

By now we are more than familiar with the concept that the connectivity of our networks changes as both diseases spread through the population. In this section we will examine the network properties against time and at key times to discern how the networks change and what a secondary variant B might see after a given incubation period for variant A. Before we examine some of the differences seen in the network structure we return our attention to some of the initial SIR curves to look at some cases where the differences across network type are particularly pronounced. As we've observed with our contour plots, the variant dynamics, for the most part, act very similarly across all 3 network structures but we can see in Fig.15, these similarities begin to break down when the time delay between the variants has increased. As a result of this, we will take the examples used in Fig.5.(b) and Fig.6 but increase the time delay[12].

Unlike in some of the previous examples Fig.17 we see 3 completely different effects on our 3 networks that have the same initial conditions. In this situation variant A has a large incubation period of $t_D = 10$ which allows it to effectively destroy a significant portion of the network. What we see is that for random networks like Erdos-Reyni and Watts-Strogatz graphs is that a highly infectious variant can still spread fast enough to overcome and outcompete its predecessor. In Fig.17.(c) we see a completely different story, it appears that despite the high transmissibility of the secondary variant, enough of the network has become impassible such that variant B has a diminished spread. Without looking at any of the internal statistics we can reason that differences like these could arise from the primary variant infecting a significant portion of the highly connected class of nodes and the strong relationship coefficient means that a significant portion of that significant portion is immune to variant B before it even arises. We look at this in slightly more detail in Fig.18.

Connectivity.vs.time

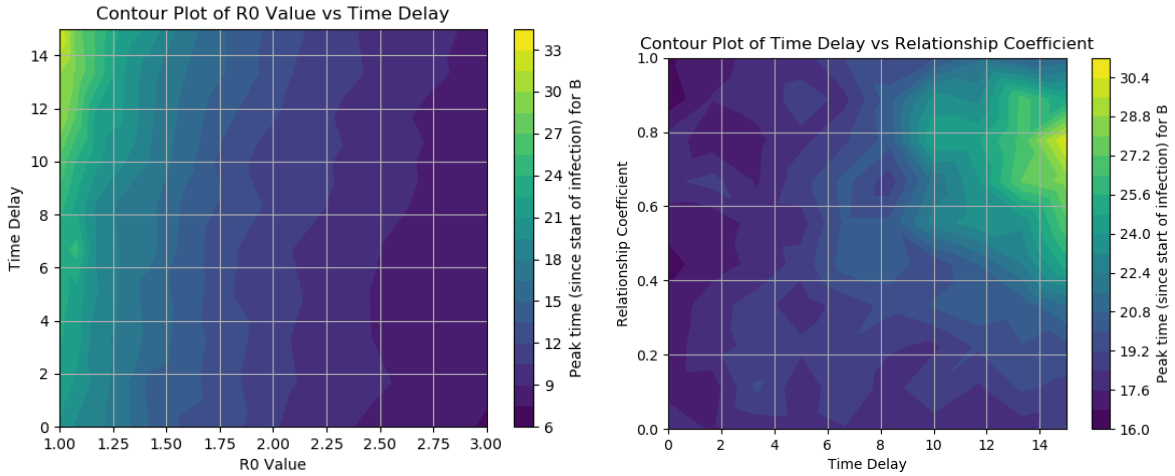
In Fig.18 we have plotted the average across multiple runs of the highest, mean and lowest degree [13] of the network. For the Barabasi-Albert plot recall that scale-free networks follow a power law for the distribution of connections [14] so we use the log scale; this results in the 'averaged least connected nodes' curve dropping to negative infinity when nodes become completely isolated. Looking at Fig.18 we observe a similar pattern that is present for both networks: the positive separation from the connectivity curves representing A to the connectivity curves representing B increases when the relationship strength increases. Furthermore, this is observed for both the mean node connections as well as the highest. In Fig.18.(a) we see that the mean number of connections within the network available to variant B starts higher than that for variant A but upon introduction,

[11] in reference to the COVID-19 rhetoric: 'flatten the curve'

[12] We also look at the points of $r_{A,B} = 0.25$ and $r_{A,B} = 0.75$ but for the sake of brevity we omit the former's SIR plot.

[13] number of connections/edges to other nodes

[14] the proportion of networks with degree $k, p(k) \propto k^{-\gamma}$



(a). ER - $t_D.vs.R0_BB$; $r_{A,B} = 0.5$ *(b). ER - $r_{A,B}.vs.t_DB$; $R0_B = R0_A = 1.22$*

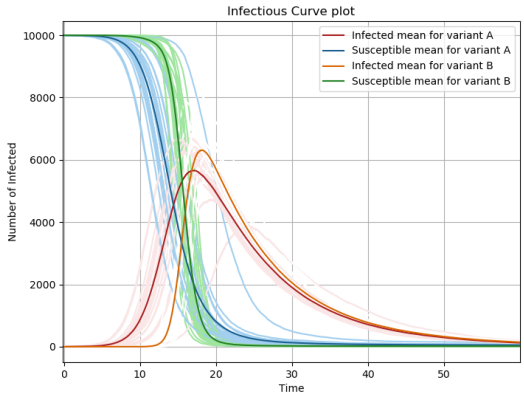
Figure 16: Time until peak infection for simulations on an Erdos-Reyni network with the same parameters as Fig.15 and Fig.14. The time has been normalised w.r.t the time delay.

quickly intercepts. This results in the mean number of connections available to B being lower than A [15] by the end of the simulation; a result which signifies that variant B propagated to parts of the network variant A never reached. We say this because there are nodes now Recovered from B that aren't in the state of Recovered from A. When we turn our attention to the same simulation but with a higher $r_{A,B}$ value (Fig.18.(b)) we see that the mean number of node connections for variant B is lower than that for variant A only briefly. This effect is much more pronounced when looking at the nodes with the highest number of connections. In the Barabasi-Albert network, we can see the exact same phenomenon where the only difference is that with a higher $r_{A,B}$ value, the mean node connection curves never touch, nor does the curve for variant B intercept A. The most connected curves for both network types look very similar to each other (when considering the differences across $r_{A,B}$ values) but given the contrasting structures of both networks, when variant B is prevented from infecting a highly connected node, the ramifications are that a larger portion of the network is unreachable.

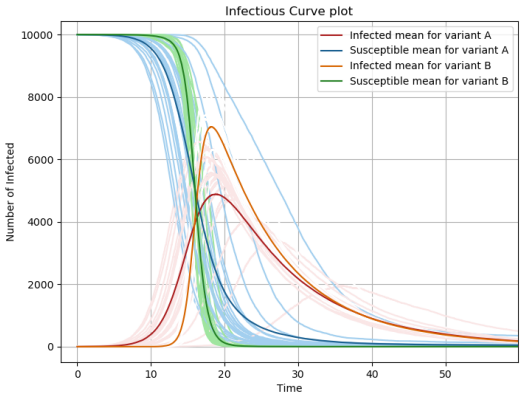
Degree Distribution

In Fig.17 we saw that for certain simulation input parameters, we observe differing results. When we then took Fig.18 into account we supposed that this must be the result of the different structures of the network which, in the case of Barabasi-Albert, allow for variants to travel fast via highly connected nodes or be impeded due to them. This is a phenomenon we initially couldn't see in Sec.3.2 but with a large enough incubation period, we observe that the network structure can aid in preventing pronounced secondary outbreaks. Fig.19 and Fig.20 support this supposition by confirming that our primary variant, when given enough of an incubation period, does not destroy the format of our initial network structure. We know that the degrees of nodes within Erdos-Reyni networks will follow a Gaussian due to the way in which they are generated and we can see as much in Fig.19 at time $t = 10$ (before the second disease has started as we can see from Fig.17). As variant B begins to take form within the network and both variants pass the peak-infection point of their life cycles we see at $t = 15$ and $t = 20$ that while the size of the distribution has reduced and

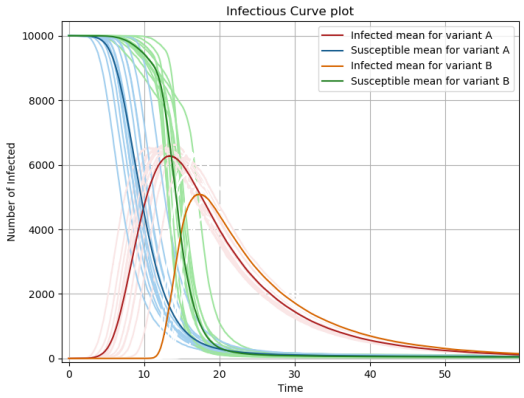
[15] negative separation from A to B



(a). Erdos-Reyni - $r_{A,B} = 0.75$, $R0_B = 3, t_D = 10$



(a). Watts-Strogatz - $r_{A,B} = 0.75$, $R0_B = 3, t_D = 10$



(a). Barabasi-Albert - $r_{A,B} = 0.75$, $R0_B = 3, t_D = 10$

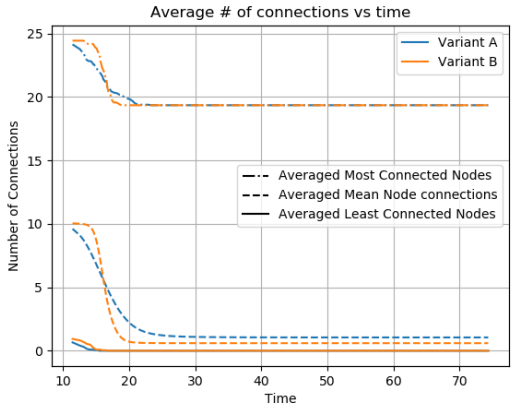
Figure 17: SIR curves showing the infection of 2 variants .vs. Time. All three plots have the exact same initial conditions barring the network structure. (a) has Erdos-Reyni, (b) has Watts-Strogatz, (c) has Barabasi-Albert

shifted, the shape remains constant for both variants. This essentially tells us that in spite of our initial thought that the network starts from some singular point, removes nodes in its path and destroys any underlying structure in its path, it tends to remove node connections in some uniform manner across the network.

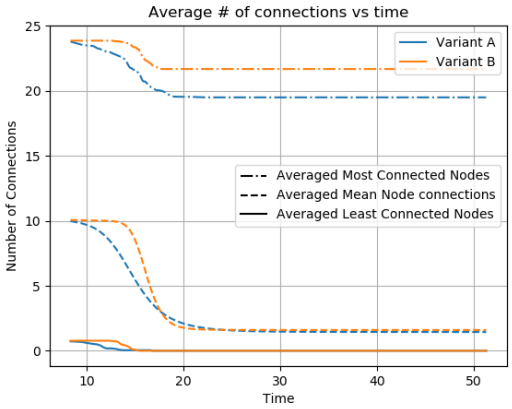
Much like with Fig.18, we see a similar pattern from the Erdos-Reyni simulation with the Barabasi Albert network in Fig.20. We reiterate that the degree distribution within Barabasi-Albert networks follows an inverse power law where the proportion of nodes with degree k, $P(k)$ is proportional to $k^{-\gamma}$ where γ is some positive constant [2]. Fig.20 shows us that as both variants progress in time, the only property of the network degree distribution that changes is the constant γ due to fewer nodes and hence fewer connections. This confirms that for both networks the underlying structure remains the same even in the face of long primary variant incubation times. By comparing Fig.21 and Fig.22 we confirm that the infection curves we see in Fig.17.(c) are a direct result of node-shielding caused by the prevalence of the primary variant A.

Component Sizes

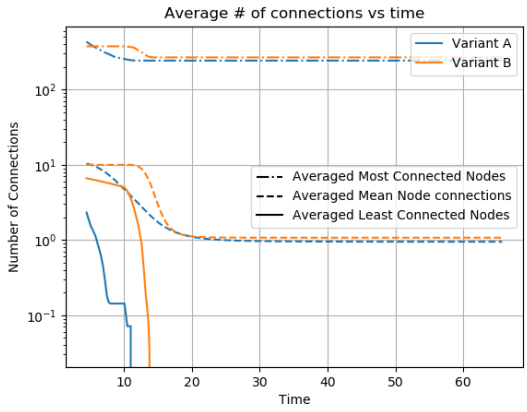
Within network theory, a component is commonly described as a group of con-



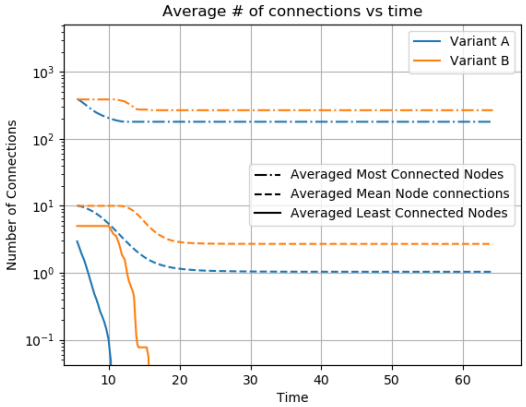
(a). ER - Same parameters as Fig.17.(a) but the relationship coefficient is smaller (which results in a more infectious B). $r_{A,B} = 0.25$



(b). ER - Same parameters as Fig.17.(a). $r_{A,B} = 0.75$



(c). BA - Same parameters as Fig.17.(b) but the relationship coefficient is smaller (which also results in a more infectious B). $r_{A,B} = 0.25$



(d). BA - Same parameters as Fig.17.(b). $r_{A,B} = 0.75$

Figure 18: The Most, Mean and Least connections in a network .vs.time. The Most, Mean and Least are averaged over multiple runs and we specifically plot the connections for Fig.17.(a) in the top row and Fig.17.(c) in the second row.

nected nodes. The size of a component, therefore, is just the number of nodes within said component. If we consider two completely separate networks of 5000 nodes, we could call them two components in a 10000-node graph. In Fig.21 we plot size of the components at two key times, $t = 10$ and $t = 20$. We use the log-scale and note that because we average our plots over multiple simulations, sub-unitary values are the result of specific component sizes showing up once and being normalised (imagine a component of size 9999 showing up for half the runs and component of size 9998 showing up in the other half). What we see in Fig.21 is that for Erdos-Reyni networks there initially exists a large connected component at $t = 10$ for both variants (and a significant sum of nodes completely isolated). We specifically note that the highly connected components for variant A at $t = 10$ are in the order of 10^-2 which indicates that the component is a random size ≥ 9000 nodes. In contrast With variant B, we see one component size of order 10^-1 which indicates that a component of that size appeared more than once but not very consistently. When the time has increased to $t = 20$ we

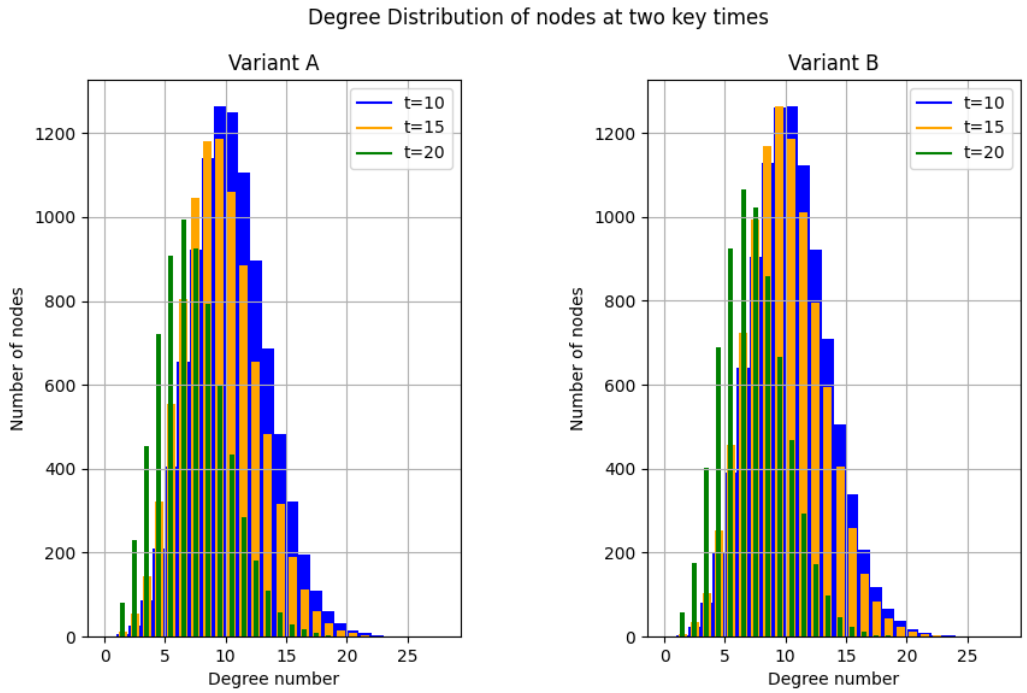


Figure 19: Degree distributions plotted for the Erdos-Reyni simulation shown in Fig.17.(a)

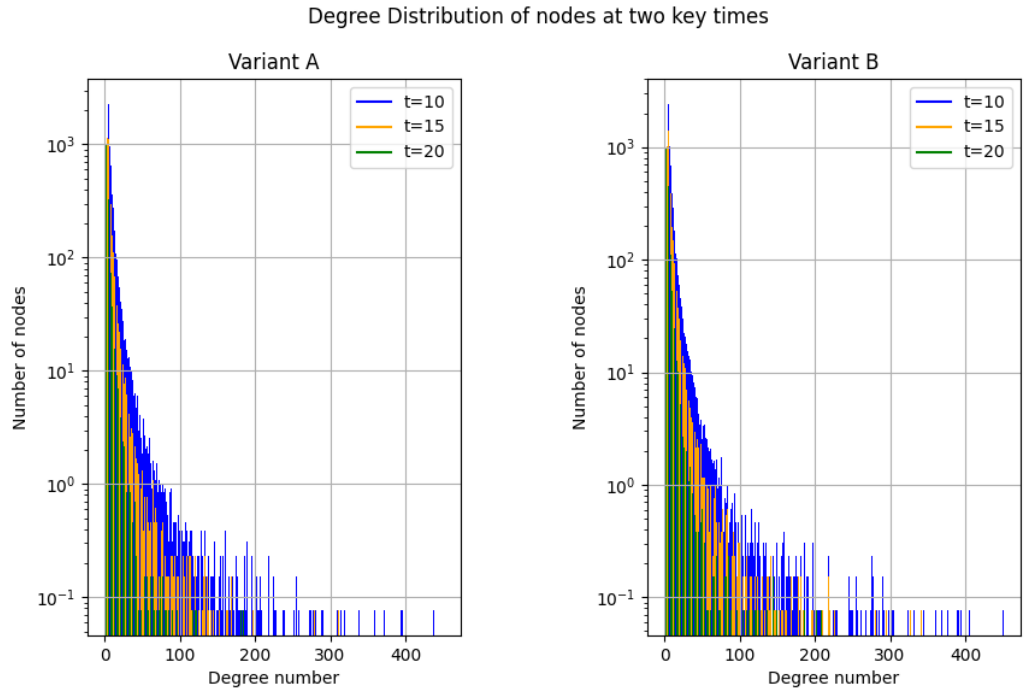


Figure 20: Degree distributions plotted for the Barabasi-Albert simulation shown in Fig.17.(c)

spot that there still remains only one giant component but its' magnitude has been reduced significantly whilst the number of isolated nodes has increased. What we learn from this plot is that on an Erdos-Reyni network, the nodes are not split apart into groups but instead picked from the highly connected set one by one.

When we compare this to Fig.22 we see a slightly different case. Even though both Fig.21 and Fig.22 have the same input parameters (barring the network itself), at $t =$

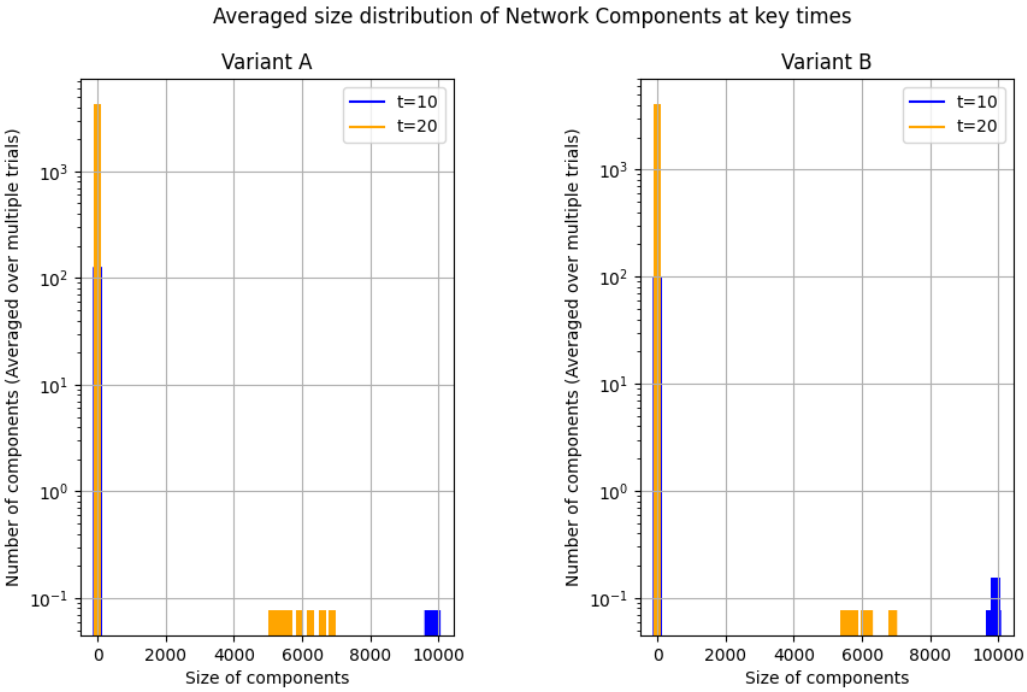


Figure 21: Sizes of all the components in the Erdos-Reyni simulation outlined in Fig.17.(a) for the times $t = 10$ and $t = 20$

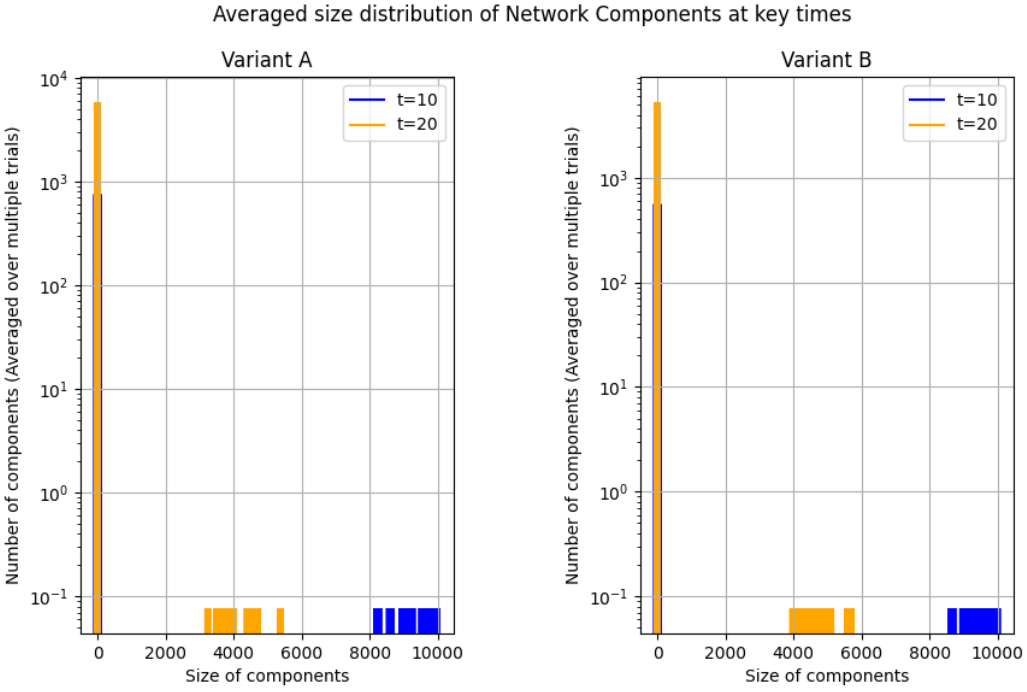


Figure 22: Sizes of all the components in the Barabasi-Albert simulation outlined in Fig.17.(c) for the times $t = 10$ and $t = 20$

10 there exists a larger spread of component sizes for both variants. The inherent transmissibility of variant A is not different between these two Figures and yet the larger spread of component sizes tells us that in a lot of our simulation runs a sizeable portion of the network is not connected to the giant component at t=10. The only explanation for this behaviour is that a highly connected component that connects with many nodes (and potentially their neighbours) that would be otherwise isolated has been removed from the network. If this node were to be removed by recovering from A and acquiring immunity to B, the isolated components might still spread A for a short while (assuming variant A was passed on before the highly connected node recovered). However, the node, now immune to B, and its isolated neighbours are unreachable to variant B.

4 Conclusion

Our study brings forth insight into the interactions and dynamics of disease variants spreading through a network with a strict cross-immunity enabled. By changing the basic reproduction number of the successive variants, varying the time delay between disease spawns and differing their 'closeness' we have explored regions of parameter space where the interplay between an ongoing epidemic and a newly emerging one exhibits variability across different network structures. Additionally, our results showcase the significance of cross-immunity in dampening the spread of secondary pandemics. This model opens up the opportunity for further study in various avenues. Noting that our model focuses on the ability of a secondary variant to percolate through a network that is continuously deteriorating additional study can be placed in examining when the percolation threshold has been crossed and how simulations about this boundary may or may not differ between networks. We also highlight that, although inspired by the COVID-19 pandemic, the applications of this model can extend to other network-based systems that are affected by systematic loss/deterioration; such as intrusion detection systems that manage a set of servers under attack from malicious parties.

References

[1] Anas Abou-Ismail. Compartmental models of the covid-19 pandemic for physicians and physician-scientists. *SN Comprehensive Clinical Medicine*, 2, 07 2020.

[2] Albert-Laaszloo Barabaasi and Reeka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[3] P Erdös and A Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[4] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[5] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[6] Davidson H. Hamer. Vaccinology: An Essential Guide. *Clinical Infectious Diseases*, 62(2):272–272, 12 2015.

[7] Herbert Hethcote. The mathematics of infectious diseases. 42:599–653., 01 2000.

[8] Rinku Jacob, K. Harikrishnan, R. Misra, and G. Ambika. Measure for degree heterogeneity in complex networks and its application to recurrence network analysis. *Royal Society Open Science*, 4:160757, 01 2017.

[9] Chun-Hsien Li, Chiung-Chiou Tsai, and Suh-Yuh Yang. Analysis of epidemic spreading of an sirs model in complex heterogeneous networks. *Communications in Nonlinear Science and Numerical Simulation*, 19(4):1042–1054, 2014.

[10] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[11] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[12] Nick Wilson, Jane Oliver, Geoff Rice, Jennifer A. Summers, Michael G. Baker, Michael Waller, and G. Dennis Shanks. Age-Specific Mortality During the 1918–19 Influenza Pandemic and Possible Relationship to the 1889–92 Influenza Pandemic. *The Journal of Infectious Diseases*, 210(6):993–995, 03 2014.