

A STUDY ON EPIDEMIC MODELLING USING DETERMINISTIC AND
STOCHASTIC SIR COMPARTMENTAL MODELS



SUBMITTED
BY

LIONEL LIM RUI QI

DIVISION OF PHYSICS & APPLIED PHYSICS
SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A final year project report
presented to
Nanyang Technological University
in partial fulfilment of the
requirements for the
Bachelor of Science (Hons) in Physics / Applied Physics
Nanyang Technological University

April 2023

Acknowledgements

I would like to thank my supervisor, Assoc Prof Chew Lock Yue, for suggesting the topic on epidemic modelling and guiding me through the entire project patiently. I enjoyed making amendments to my models after our discussions and most importantly conversations with you that taught me how to appreciate problems from different perspectives. I hope that I have met your expectations and I will do you proud!

For my family and friends who encouraged me throughout the course of my work, I thank you for your unwavering support. Lastly, for my peers who are going through the final lap of FYP, I hope that you enjoyed the process and I wish you all the best!

Abstract

Epidemic modelling is imperative as it provides predictions that aid timely interventions during a disease outbreak. Most works in the field adopt the SIR compartmental model as a foundation, before employing various deterministic and stochastic strategies to simulate the dynamics of an epidemic. In our research, we have chosen to apply the deterministic ODE and stochastic complex networks (with HOIs) approaches. Through experiments using ODEs on different r_0 and initial conditions, we identified a threshold r_0 that could result in the infection of an entire population, thus serving as an indicator for epidemic monitoring. However, as ODEs assume homogeneity in interactions, they are generally considered unrealistic. To circumvent this, the stochastic complex networks (with HOIs) were developed to study virus reproducibility. Among all networks considered, the BA network in particular has shown to consistently demonstrate spikes in epidemic reproduction at the onset, suggesting higher vulnerability to epidemic spreading for real-world communal structures with properties akin to the BA network. In the model, an infectability parameter, r_t was also found to be closely related in estimating actual virus transmission, and larger values of r_t were shown to potentially outweigh effects by an epidemic infectious period. Lastly, HOIs were integrated on the networks and a group transmissibility (p_{sc}) > pairwise ($p_T = 0.1$) was shown to steer the course of an epidemic and change actual virus reproducibility for different networks, highlighting the need to reduce group interactions to curb epidemic spreading. Through the constructs of our project, we hope to further develop a more comprehensive epidemic modelling framework that would provide insightful estimates on virus reproducibility when an outbreak ensues.

Table of Contents

Chapter 1: Introduction	6
1.1 Compartmental modelling	6
1.2 Deterministic compartmental models	6
1.3 Stochastic compartmental models.....	7
1.4 Reproduction number.....	7
1.5 Focus of project.....	8
Chapter 2: Deterministic Model (ODE).....	9
2.1 Ordinary Differential Equations (ODE).....	9
2.2 Basic reproduction number (r_0).....	11
2.3 Phase portraits for different initial conditions	12
2.3.1 Plots of phase portraits	12
2.3.2 Results and observations	13
2.4 Mathematical derivations and interpretations of 2.3.....	13
2.5 Summary of ODE method	15
Chapter 3: Stochastic Model (Complex Networks)	17
3.1 Complex networks	17
3.1.1 Features of the constructed networks.....	18
3.2 Simulation methodology	19
3.3 Reproducibility of virus in complex networks.....	23
3.4 Canonical average of Monte-Carlo simulation	24
3.4.1 Observations of canonical average.....	26
3.5 Effects of different r_ℓ and $D_{recover}$	27
3.5.1 Epidemic with different r_ℓ	27
3.5.2 Epidemic with different $D_{recover}$	30
3.6 Summary of complex social network model	32
Chapter 4: Stochastic Model (Simplicial Complex).....	34
4.1 Simplicial Complex	34
4.2 Construction of 2-simplex.....	35
4.3 Modified simulation methodology.....	35
4.4 Effects of r_{sc} on networks.....	38
4.4.1 Epidemic with $r_\ell = 3$ and 9.....	38
4.4.2 Epidemic with $r_\ell = 1$	39
4.4.3 Epidemic with different $D_{recover}$	41

Chapter 5: Conclusion	44
References	45
Appendices.....	49
Annex A: Code availability	49
Annex B: Data availability.....	49
Annex C: Detailed plots for 2-simplex	50

Chapter 1: Introduction

The coronavirus disease (COVID-19) in 2019 caught the world by surprise and till today has caused a combine death toll of approximately 7 million individuals across the globe¹. With high infectibility and fatality, governing bodies like Singapore and China resorted to stringent control measures, such as national wide lockdowns², to curb the spread of the virus allowing the epidemic to be contained successfully^{3,4}. These interventions were guided by numerous predictions from epidemic modelling frameworks that gave useful insights to aid decision^{5,6} making among policy makers. However, not all models give accurate forecasts⁷, and their effectiveness is largely determined by the context of use. Therefore, the careful study of epidemic modelling is of great importance. We shall begin by exploring current methods that are used in the field.

1.1 Compartmental modelling

In most epidemic models, the concept of compartmental modelling^{8,9} is adopted, starting from the simplest Susceptible (S), Infected (I) and Recovered (R) framework^{10,11}, which can be extended to include other compartments such as Exposed (E) and Vaccinated (V) depending on the nature of the virus. The idea behind the SIR model is to segregate individuals in a population to either the S, I or R compartment depending on their current state of health and characterize the evolution of an epidemic with transitions between compartments. In the SIR framework, transitions are unidirectional, that is individuals who are infected will not be susceptible to the disease and those who have recovered cannot be re-infected again, and they are governed by transmission and recovery parameters as shown in Figure 1. This approach provides a straightforward and realistic representation of an epidemic, serving as a backbone to construct several predictive models that are deterministic and stochastic in nature.



Figure 1: Transitions between S, I and R compartments

1.2 Deterministic compartmental models

A deterministic compartmental approach is usually used in epidemic modelling if we want to project the evolution of an epidemic with full certainty, that is the outcome at every time step

is fixed. This is usually done through a set of Ordinary Differential Equations (ODE) which entail a mean-field representation of interactions between individuals in a population. Deterministic ODEs are fairly simple to construct, and they produce the same results consistently for the same set of inputs. Therefore, they are widely studied to predict the outcomes of several diseases^{10,11,12,13,14}. For instance, Cooper et al¹³ studied the spread of COVID-19 on countries like China and India using the SIR models, while Batista et al¹⁴ used other variants of deterministic compartmental model such as SEIR and SIS, to summarise and forecast epidemic spreading for different viruses. These models proved to give useful predictions of the rough evolution of an epidemic, allowing timely interventions to be meted to curb infection peaks.

1.3 Stochastic compartmental models

Other than deterministic approaches, there has been numerous research devoted to epidemic studies using stochastic techniques. Unlike the former which gives fixed outcomes for the same set of inputs, stochastic models produce an ensemble of outputs that simulates alternative representations of an epidemic. In previous studies, stochastic compartmental models were centred on extensions to ODEs^{15,16}, such as the continuous-time Markov chains (CTMCs) and stochastic differential equations (SDEs). However, there has been a shift in contemporary approaches to incorporate them into complex social network systems in order to understand the dynamics of an epidemic^{17,18} better. This is as network science¹⁹ has been proven to better describe social connectivity in the real world and hence, able to show realistic interactions among individuals of different compartments. Furthermore, with recent advances in complex networks to include Higher-Ordered Interactions (HOIs)^{20,21}, newer and richer phenomenon have emerged, allowing epidemic spreading to be modelled more comprehensively.

1.4 Reproduction number

In both deterministic and stochastic approaches, a key ingredient for accurate predictions of an epidemic is the model's associated reproduction numbers. Reproduction numbers are essential as they characterize the spread of a virus in a community.

In most literatures, a commonly used reproduction number is the basic reproduction number, r_0 , which is defined as the expected number of secondary infections produced by an index case in a completely susceptible population^{22,23,24,25}. Depending on the modelling technique employed, r_0 can be derived or estimated through the transmissibility of a disease during an

initial outbreak²⁶. When $r_0 > 1$ an epidemic will ensue whereas for $r_0 < 1$, it will die down. However, r_0 is non-universal across different epidemic frameworks as the manner of construction in each model is unique. Hence, its interpretation and performance is only useful with respect to the model explored^{27,28}.

1.5 Focus of project

With the foundations laid, we now investigate epidemic modelling using both deterministic and stochastic SIR compartmental models. In Chapter 2, we first experiment on the deterministic approach using classical ODEs with the purpose of observing the dynamics that arises when different initial conditions and r_0 are used. In Chapter 3, we improve on the limitations of the ODE method by constructing a stochastic agent-based modelling framework for observing epidemic evolutions in complex networks. We shall tune transmission and recovery parameters and explore their effects on the true virus reproducibility in hindsight. In Chapter 4, we modify the complex network models to include HOIs in the form of simplicial complexes with the goal of improving the realism in our model. By studying the reproducibility of the virus, we shall discover new and interesting phenomenon due to simplicial complexes. Finally we conclude in Chapter 5 by summarizing key insights of our models and discuss future works for consideration.

Chapter 2: Deterministic Model (ODE)

To model the spread of an epidemic in a community, we start with the deterministic approach in the form of ODEs. This is the simplest representation of an epidemic as it gives a mean field understanding of interactions between individual and has a fixed outcome for the same input parameters.

2.1 Ordinary Differential Equations (ODE)

The dynamics between individuals in the S, I and R population are modelled using the set of ODEs with units of time in terms of days. In our work, we make two important assumptions:

1. The total population at any point in time is closed: $N = S(t) + I(t) + R(t)$, that is we assume no birth or deaths (no vital dynamics²⁹) throughout the epidemic.
2. At any point in time, the S and I population mix homogeneously. This means that both S and I individuals are evenly distributed, and each S individual will experience the same amount of interaction with his I counterpart.

With that, we can write the set of ODEs describing an epidemic as follows:

$$\frac{dS}{dt} = \frac{-\beta SI}{N} - (1)$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I - (2)$$

$$\frac{dR}{dt} = \gamma I - (3)$$

There are two parameters that are used to characterise the transitions of individuals between different compartments. These parameters are assumed to be constant and are defined as follows^{30,31}:

1. β is known as the effective contact rate and is the product of the average rate of contact between susceptible-infected individuals and the probability of infection given a contact between susceptible-infected individuals. Intuitively, this can be thought of as the number of infectious contacts by an infected individual per unit time.
2. γ is known as the recovery rate and is the inverse of the average time of recovery (or average duration of infectiousness).

With the parameters defined, we can understand why the ODEs were constructed in the above manner.

For equation (1) the product of β and I gives the total number of infections producing contacts per unit time, and thus, the entirety $\frac{\beta SI}{N}$ gives us the number of susceptible individuals who will transit from the S to I population per time. Therefore, $\frac{dS}{dt}$ is the negative of $\frac{\beta SI}{N}$, which describes the decrease in susceptible individuals over time.

As for equation (2), apart from the added $\frac{\beta SI}{N}$ infected individuals explained earlier there are those who have recovered and will be removed from the infected pool. This is characterized by the product γI which gives the average number of infected individuals who recover per unit time. Hence, $\frac{dI}{dt}$ gives the equation $\frac{\beta SI}{N} - \gamma I$.

Lastly, for the γI individuals who have recovered, they will transit to the recovered pool, giving $\frac{dR}{dt} = \gamma I$ as in equation (3).

Though the systems of ODEs characterize an epidemic in a simplistic manner, they are in fact non-linear coupled differential equations with no analytical solution. However, we can employ numerical methods to determine its evolution. In our case, we choose the Runge-Kutta (4th Order) methods (see Annex A: Code availability). An example of the numerical results with the parameters: $N = 500$, $S_0 = 499$, $I_0 = 1$, $R_0 = 0$, $\beta = 4$ and $\gamma = \frac{1}{3}$, on a time period of $t_0 = 0$ to $t_f = 20$ days is shown in Figure 2.

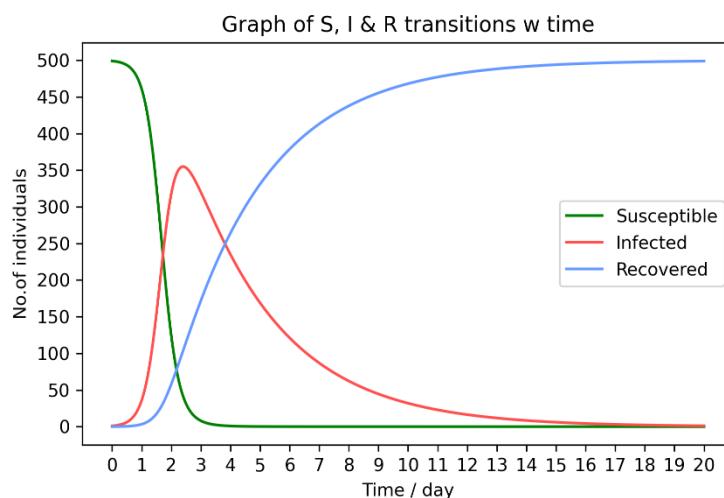


Figure 2. Example of SIR transitions with parameters: $N = 500$, $S_0 = 499$, $I_0 = 1$, $R_0 = 0$, $\beta = 4$ and $\gamma = \frac{1}{3}$, on a time period of $t_0 = 0$ to $t_f = 20$ days

2.2 Basic reproduction number (r_0)

The basic reproduction number, r_0 , is a commonly studied quantity in ODE models as it determines the transmission potential of a disease at the onset.

In the ODE model, r_0 can be derived in the following manner^{32,33}:

$$r_0 = \frac{\beta}{\gamma}$$

This is as the product of β (the number of infectious contact) and $1/\gamma$ (infectious period) gives the average number of new infections throughout the epidemic, and thus at the initial stage where the population is susceptible, will be the expected number of secondary infections caused by index cases. This fulfils the definition of r_0 described earlier.

For $r_0 > 1$, the epidemic will spread, and the number of infected individuals will increase at the next time step. On the other hand, for $r_0 < 1$, the epidemic slowly dies down and the number of infected individuals will decrease. Intuitively this can be understood as if $r_0 > 1$, the transmission rate β is larger than the recovery rate γ and thus, more individuals will get infected than recovered in the population. The inverse can also be observed for $r_0 < 1$.

We can verify this mathematically by analysing equation (2). For the epidemic to outbreak:

$$\frac{dI(0)}{dt} > 0$$

This can be re-expressed as:

$$\frac{\beta S_0 I_0}{N} - \gamma I_0 > 0$$

$$\gamma I_0 \left(\frac{r_0 S_0}{N} - 1 \right) > 0$$

As γ and I_0 are always greater than 0:

$$\frac{r_0 S_0}{N} - 1 > 0$$

$$r_0 > \frac{N}{S_0}$$

Since $N > S_0$, it is clear that $r_0 > 1$ at the onset of an epidemic, corroborating with the idea above.

An example is shown in Figure 3.

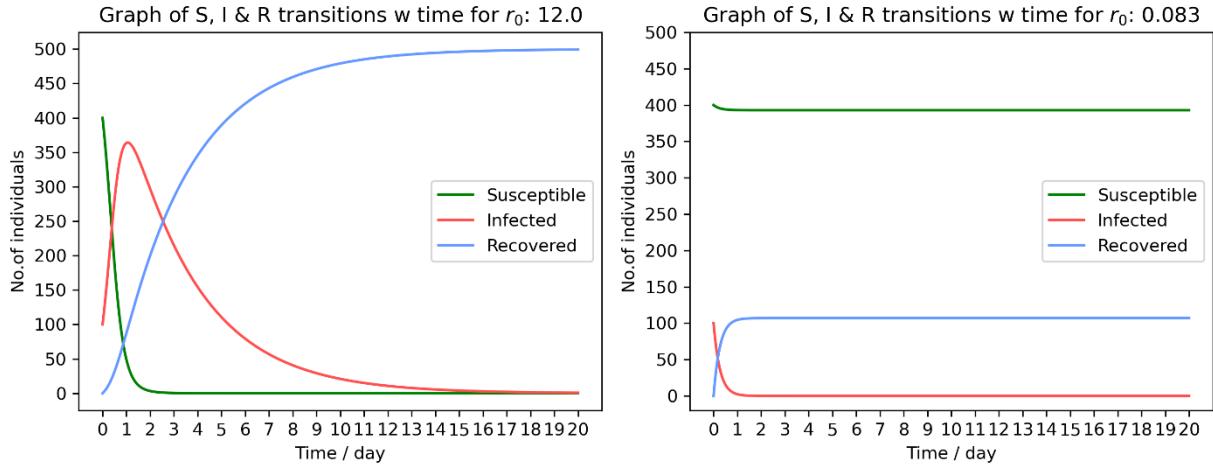


Figure 3: Comparison of SIR transitions for $r_0 > 1$ (left) and $r_0 < 1$ (right).

On the left, we have $(\beta, \gamma) = (4, 1/3)$ and on the right $(\beta, \gamma) = (1/3, 4)$. This gives r_0 (left) equals 12 and r_0 (right) equals 1/12. The dynamics of the SIR transitions are plotted with $S_0 = 400$ and $I_0 = 100$. From the plots, we can clearly see that the infected population increases then decreases for r_0 (left) > 1 whereas it decreases throughout for r_0 (right) < 1 .

2.3 Phase portraits for different initial conditions

Phase portraits give a geometric representation of the trajectories of a dynamical system when different initial conditions are used. In our context, we seek to observe changes in the dynamics of S and I individuals for different S_0 and I_0 . Apart from different initial conditions, we shall also use different r_0 values for comparisons. In its entirety, we hope to get insights on the dynamics at the end of the epidemic, namely as $(S, I) \rightarrow (S_\infty, I_\infty)$.

2.3.1 Plots of phase portraits

To plot the phase portraits, we first choose a set of S_0 and I_0 values, keeping in mind the relation $N = S_0 + I_0$ (assuming $R_0 = 0$). For r_0 , we choose the values 2, 4 and 6, derived by fixing $\gamma = 1/2$ while changing β to 1, 2 and 3 respectively for different plots.

We input these parameters into our Runge-Kutta4 method for a population of $N = 500$, and plot I against S. The results are shown in Figure 4.

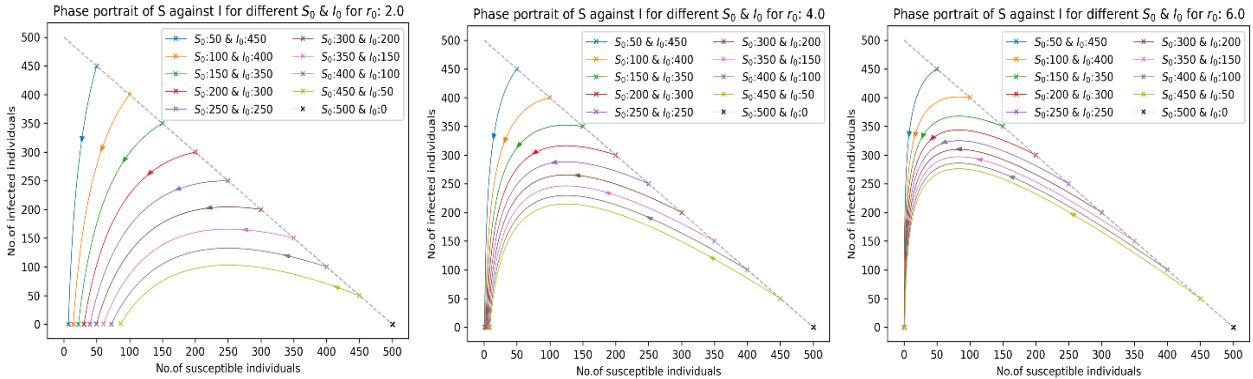


Figure 4: Phase portraits for different S_0 and I_0 for $r_0 = 2$ (left), $r_0 = 4$ (center) and $r_0 = 6$ (right)

2.3.2 Results and observations

There are three main observations we can make.

1. For the case of $S_0 = N = 500$, the plot remains as a point unlike other S_0 which are curves pointing towards 0. This is understandable since $I_0 = 0$, implies no infected individuals at the onset and hence, no susceptible individuals will become infected.
2. At $t \rightarrow \infty$, I_∞ tends to 0 for all initial conditions S_0 and I_0 (apart from $S_0 = 500 \& I_0 = 0$). This corroborates with the intuition that for an epidemic to end, there must be no infected individuals.
3. As r_0 increases, S_∞ tends to 0 regardless of the initial condition S_0 and I_0 . This can be explained by understanding that a higher r_0 suggests a larger outbreak and thus, more susceptible individuals will get infected. Hence, most of the susceptible individuals would contract the virus resulting in a value close to 0 at $t \rightarrow \infty$. In our observation, it seems that $r_0 > 6$ is an approximate range where all S_∞ converges 0.

2.4 Mathematical derivations and interpretations of 2.3

The results of the phase portraits in 2.3 could be derived mathematically if we analyse equations (1) and (2).

We first observe the consequence when initial conditions, $S_0 = N = 500$ and $I_0 = 0$ are chosen. By substituting these into (1) and (2), we find that it gives:

$$\frac{dS_0}{dt} = \frac{-\beta S_0 I_0}{N} = 0$$

$$\frac{dI_0}{dt} = \frac{\beta S_0 I_0}{N} - \gamma I_0 = 0$$

This suggests that there are no dynamics at the onset, which justifies why a point is observed for $(S_0, I_0) = (500, 0)$.

For other initial conditions, their evolutions can be observed if we take equations (2)/(1). This gives us:

$$\frac{dI}{dS} = -1 + \frac{N\gamma}{S\beta} - (4)$$

By substituting $\frac{\beta}{\gamma}$ with r_0 , and integrating $\frac{dI}{dS}$ over the initial conditions S_0 and I_0 , we obtain:

$$\int_{I_0}^I dI = \int_{S_0}^S -1 + \frac{N}{r_0 S} dS$$

$$I - I_0 = [-S + \frac{N}{r_0} \ln(S)]_{S_0}^S$$

$$I = I_0 - S + \frac{N}{r_0} \ln(S) + S_0 - \frac{N}{r_0} \ln(S_0)$$

$$I = N - S + \frac{N}{r_0} \ln\left(\frac{S}{S_0}\right) - (5)$$

With equation (5), we plot I against S using the same set of r_0 and S_0 parameters as in 2.3 (excluding $(S_0, I_0) = (500, 0)$). To ensure that the results are physical, we constrained all S and $I \geq 0$ and enumerate S up to the respective S_0 values ($S \leq S_0$). The results are shown in Figure 5.

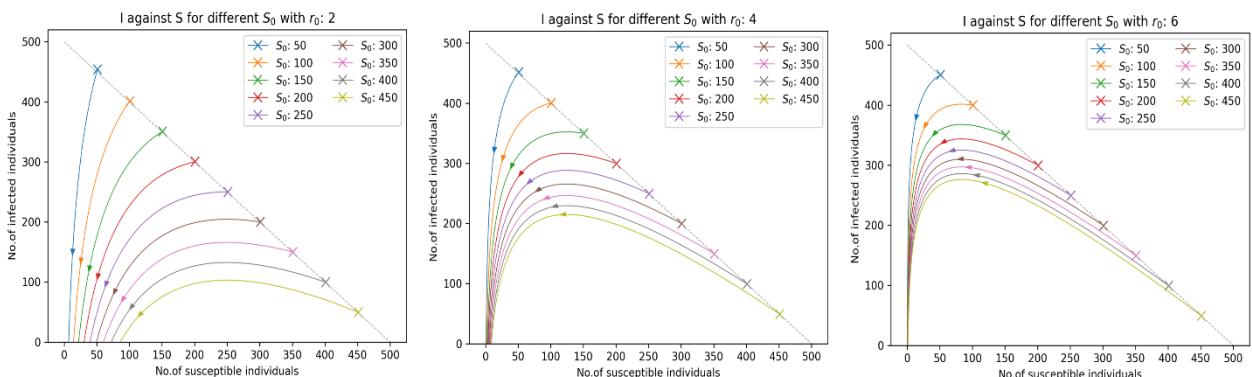


Figure 5: Mathematical plots of I against S for different S_0 and I_0 for $r_0 = 2$ (left), $r_0 = 4$ (center) and $r_0 = 6$ (right)

As expected, the plots in Figure 5 corroborate with the results in 2.3.

Apart from similarities in the plots, we can also explain the trend behind the phenomenon of S_∞ as r_0 increases. To do so, we rearrange equation (5) into the following:

$$N = S + I + \frac{N}{r_0} \ln\left(\frac{S_0}{S}\right)$$

As we have assumed a closed population in our analysis, $N = S(t) + I(t) + R(t)$, this implies that at the end of an epidemic, t_∞ , we have

$$R_\infty = \frac{N}{r_0} \ln\left(\frac{S_0}{S_\infty}\right)$$

Making S_∞ the subject, we get:

$$S_\infty = S_0 e^{\frac{-R_\infty r_0}{N}} - (6)$$

As $R_\infty \leq N$, we rewrite (6) as:

$$S_\infty = S_0 e^{-\sigma r_0} - (7)$$

Where $\sigma = R_\infty / N$, $\sigma \leq 1$.

As $S_0 < N$ and $\sigma \leq 1$, increasing r_0 indefinitely causes the exponent in (7) to converge to 0 regardless of S_0 values, consequently resulting in $S_\infty \rightarrow 0$. This corroborates with what was observed.

We can also approximate a threshold value of r_0 that will cause the entire population to be infected. This is found by taking $R_\infty \rightarrow N$, $\sigma \rightarrow 1$. Therefore,

$$S_\infty \approx S_0 e^{-r_0} - (8)$$

For $S_0 \leq N - 1$, we observe that the first instance of r_0 that leads to $S_\infty < 1$ occurs at:

$$r_0 > \ln(N - 1)$$

For $N = 500$, $r_0 > 6.22$ will cause all $S_\infty \rightarrow 0$, which supports the approximation made in 2.3.

2.5 Summary of ODE method

In our attempt, we have successfully studied the evolution of an epidemic using the ODE method. From our experiments, we have seen how the dynamics of an outbreak changes when different initial conditions S_0 , I_0 and basic reproduction numbers r_0 are used. Notably, for a

population of $N = 500$, a $r_0 > 6.22$ is sufficient to infect everyone regardless of the S_0 and I_0 . These insights give valuable predictions when we seek to ascertain the course of an epidemic over time.

However, the greatest limitation in the ODE model is its lack of consideration for the demography and structure of a population. As the model fundamentally assumes that interactions between S and I individuals are homogeneous, the dynamics of the epidemic is restricted by a fixed average virus reproducibility. This is not true in practice as individuals have their own social circles with varying degrees of connectivity, and therefore, not everyone will receive the same amount of infection inducing contacts over the course of an epidemic.

To circumvent this constraint, we relax the assumption of homogeneity by using complex networks with stochastic transmission processes to represent more realistic interactions between individuals in a community.

Chapter 3: Stochastic Model (Complex Networks)

Stochastic modelling is an alternative approach to study epidemic spreading as it gives a probabilistic representation of an epidemic's dynamics. In contemporary research that focuses on complex networks, stochastic processes are built into viral transmissions between each S-I pair, consequently predicting an epidemic with vastly different outcomes for the same set of inputs. In our study of epidemic spreading, we shall construct a Monte-Carlo agent-based simulation on discrete times and explore the reproducibility of an outbreak in different complex networks.

3.1 Complex networks

Complex networks are relevant in the study of epidemic spreading as they reflect realistic heterogenous interactions between individuals in a community. In the networks, nodes (vertex) represent individuals that can either be in the S, I or R compartment, while links (edges) illustrate pairwise interactions between each individual. There are many properties that characterize a network, such as the degree distributions, average shortest path lengths and clustering coefficients (elaborations can be found in Lee³⁴) . However, the focus is usually on degree distributions which is defined as the probability distribution of degrees in a network, where degree means the number of neighbours each node has. This quantity is important as it identifies possible hubs (individuals who are highly connected with others) which have been proven to have the potential to cause large outbreaks during an epidemic.

To model complex networks in the real world, we chose three types of networks in our study: Erdos-Renyi (Random), Watts-Strogatz (Small-World) and Barabasi-Albert (Scale-Free), each with a total population of $N = 500$ individuals. These networks are inherently random graphs as they are generated probabilistically. However, the manner of construction gives it specific properties that differentiates one from another. To create these networks, we use the NetworkX³⁵ package in the Python library (see Figure 6). A brief explanation of their formation is given below:

- (a) Erdos-Renyi (ER): The Erdos-Renyi network is an example of a purely random network. The network is constructed by iterating over every node and allowing each one of them to establish an edge with another with probability, p_{ER} . The programme ends when all nodes are connected to at least one other node. In our setup, $p_{ER} = 10/499$.

(b) Watts-Strogatz (WS): The Watts-Strogatz network is an example of a small-world network, that is most nodes are not neighbours of one another, but the neighbours of any given node are likely to be neighbours of each other^{36,38}. In the construction, each node originally has k_{ws} number of edges with its closest neighbours. The algorithm then iterates over each edge and rewrites them to other nodes with probability, p_{ws} . The programme ends when all nodes are connected to at least one other node. In our setup, $k_{ws} = 10$ and $p_{ws} = \frac{1}{2}$.

(c) Barabasi-Albert (BA): The Barabasi-Albert network is an example of a scale-free network, that is its degree distribution follows a power law^{37,39}. To construct the network, new nodes with m_{BA} edges are preferentially attached to existing nodes. In other words, the more connected a node is, the more likely the m_{BA} edges will connect to it. The programme terminates when all nodes are connected to at least one other node. In our setup, $m_{BA} = 5$.

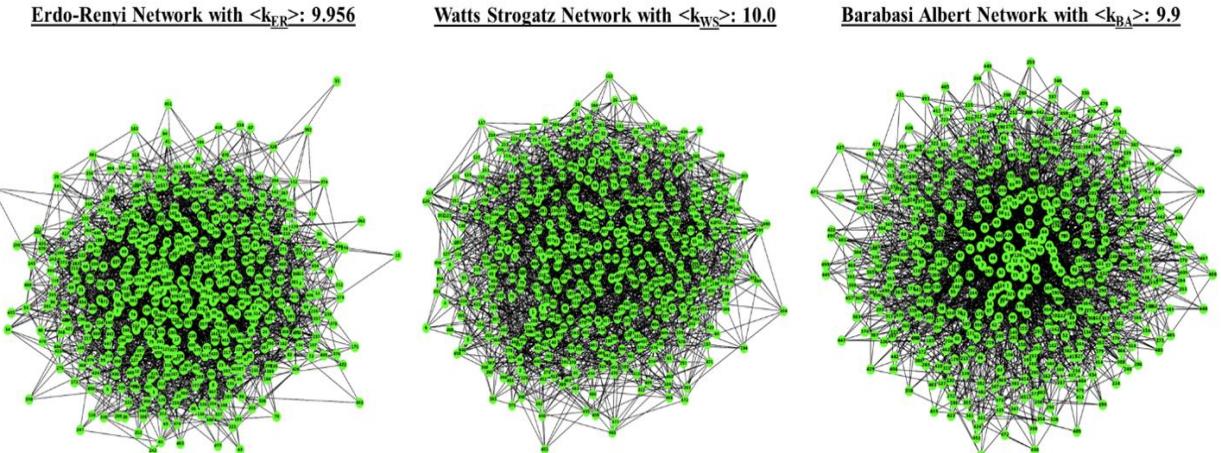


Figure 6: Plots of complex networks, ER (left), WS (center) and BA (right).

3.1.1 Features of the constructed networks

We highlight some properties of the constructed networks in 3.1: degree distributions (and average degree), average shortest path lengths and clustering coefficients. The former is plotted in Figure 7 while the rest are calculated using NetworkX functions in Table 1.

	Average degrees $\langle k \rangle$	Average shortest path length	Average clustering coefficient
ER Network	9.96	2.95	0.0231
WS Network	10.00	3.04	0.0940
BA Network	9.90	2.78	0.0585

Table 1: Properties of complex networks

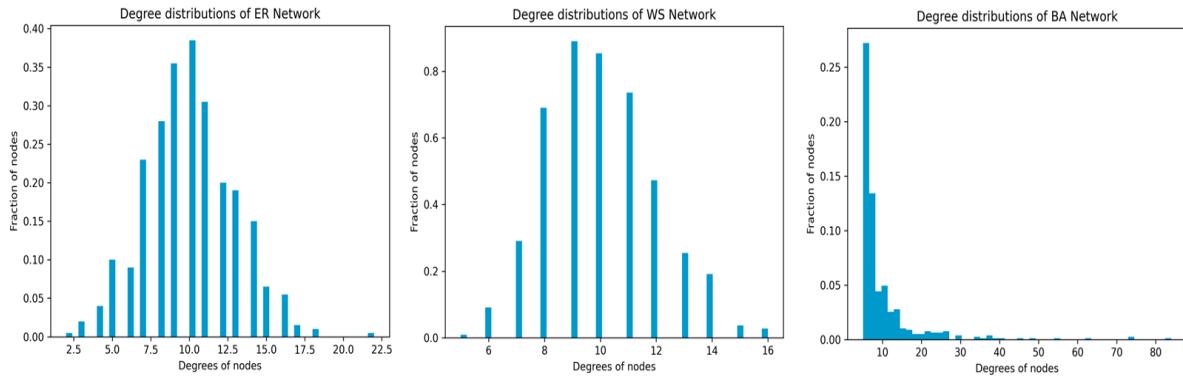


Figure 7: Degree distributions of complex networks, ER (left), WS (center) and BA (right).

The small world properties^{36,38} of the WS network can be observed in Table 1, where the average clustering coefficient is the highest compared to the ER and BA networks, and its average shortest path length is rather small. On the other hand, scale free features^{37,39} of the BA network is seen in Figure 7, as seen where the degree distribution follows approximately a power law compared to the ER and WS networks. These attributes affirm the validity of our constructions.

3.2 Simulation methodology

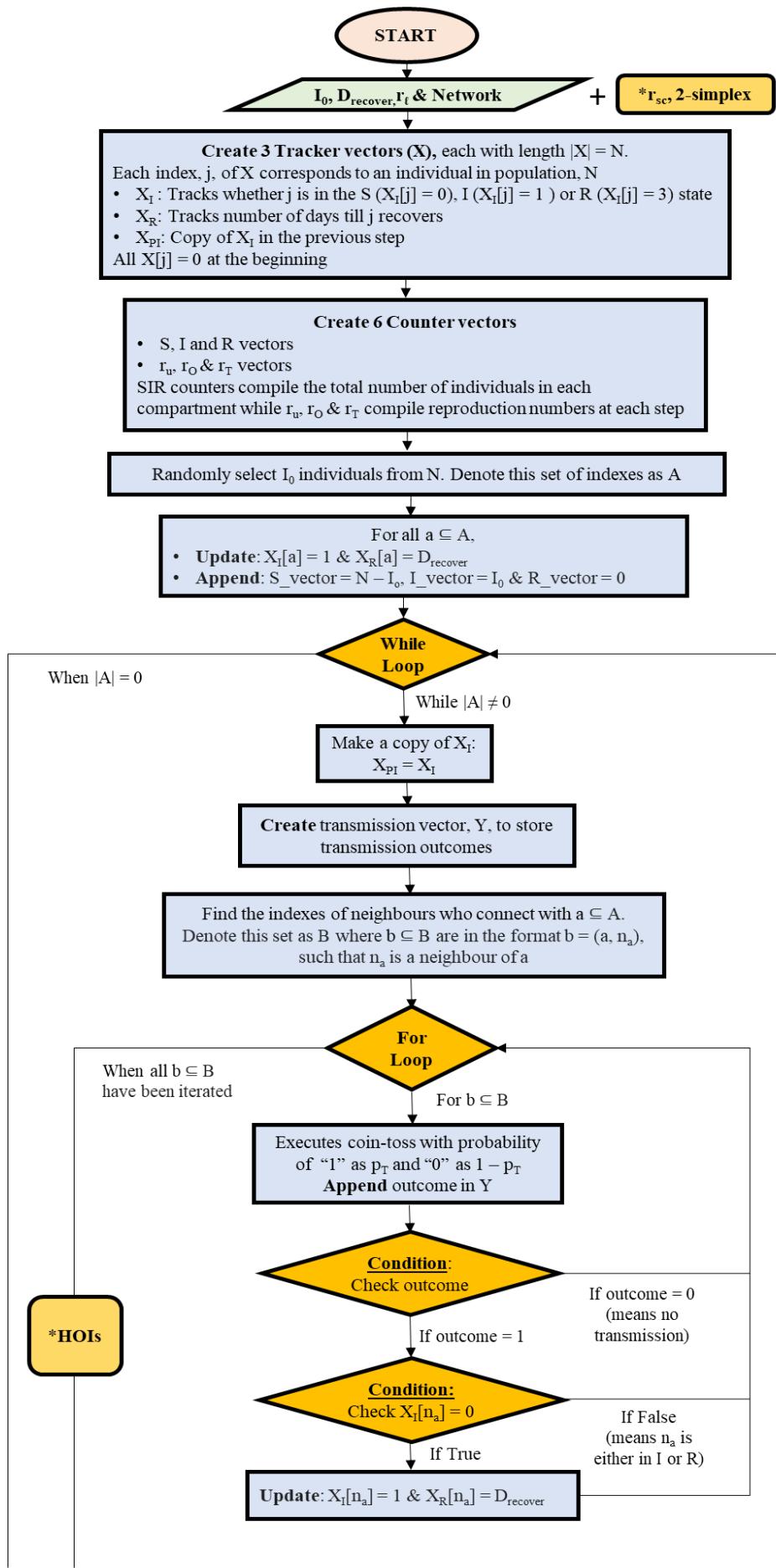
With the foundations laid, we now construct the agent-based Monte-Carlo simulation on the complex networks to study epidemic spreading (see Annex A: Code availability).

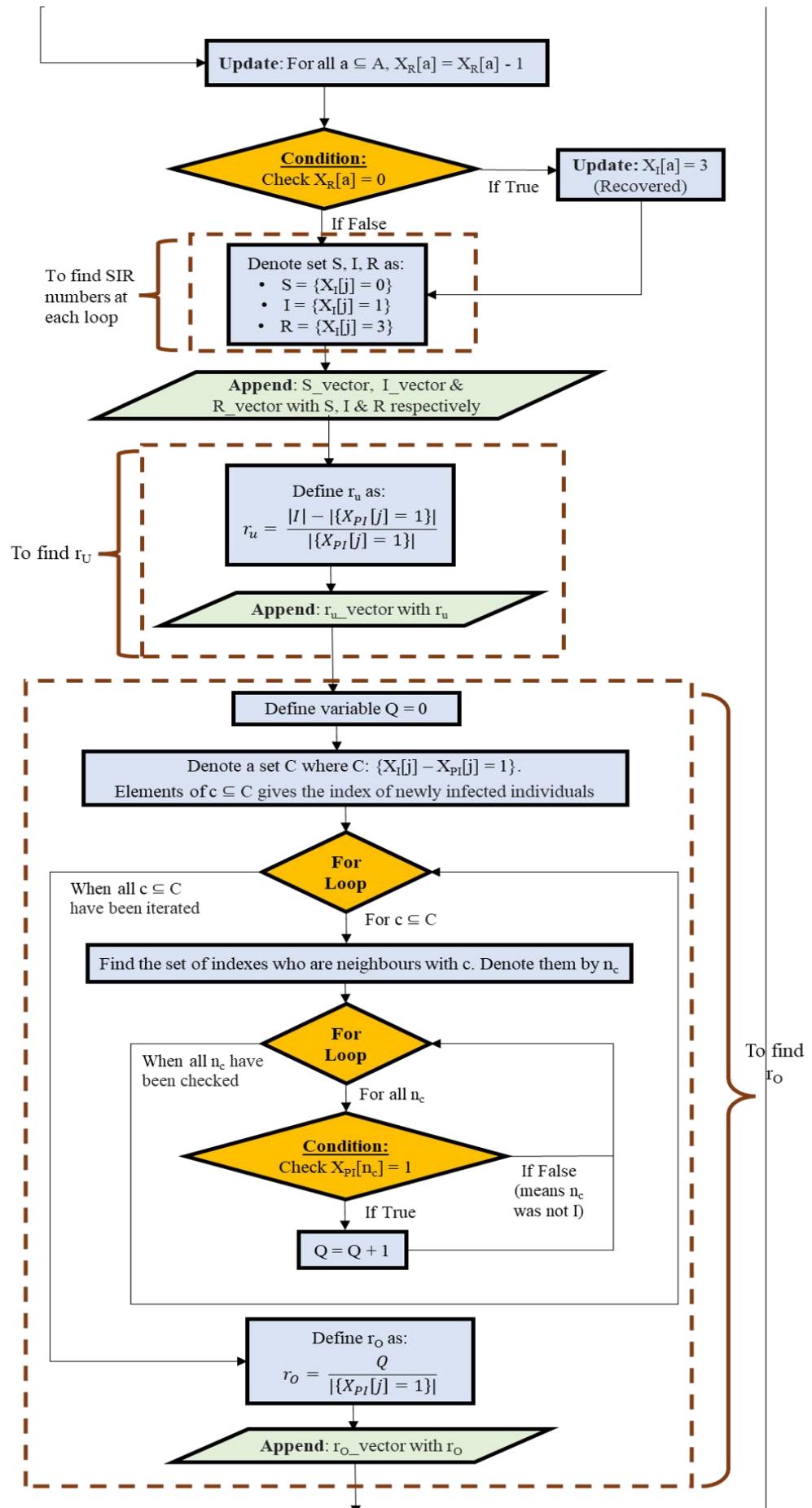
The simulation has four inputs and a flowchart describing it is shown in Figure 8:

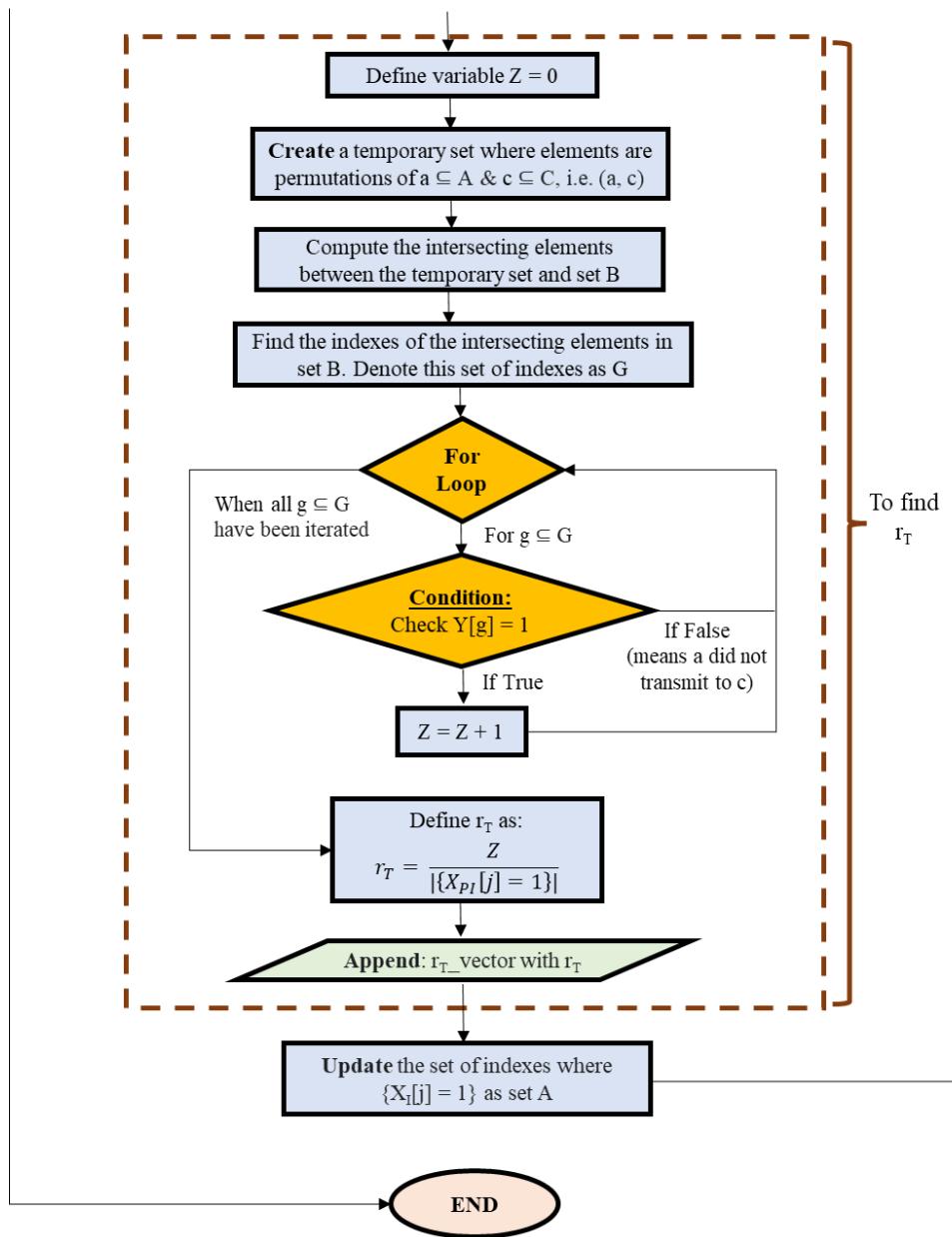
- Pairwise infectability (r_l): r_l is a transmission parameter to characterize the probability of virus transmission between S-I links (p_T). It is defined as the product of p_T , and $\langle k \rangle$ of a network, and acts as a standardizing parameter²⁰ to compensate for the difference in $\langle k \rangle$ in a network. p_T can be found as follows:

$$p_T = \frac{r_l}{\langle k \rangle}$$

- Days to recover ($D_{recover}$): Number of days it takes an infected individual to recover (or number of days an infected individual stays infected)
- Initial number of infected individual (I_0): 1 by default
- Network type: Erdo-Renyi, Watts Strogatz or Barabasi Albert







Legend

→ - Process flow	○ - Terminal	◇ - Conditions / loops
- Length of vector		
{ } - Set of elements	■ - Process	■ - To be added in Chapter 4
	[/] - Input/output	

Figure 8: Simulation flowchart

Apart from capturing S, I, R transitions, the simulation calculates the dynamics of the reproducibility of a virus (see section 3.3) after every time step.

3.3 Reproducibility of virus in complex networks

The reproduction number is commonly sought after as an indicator of epidemic growth. As different networks fundamentally exhibit varying structures, it is imperative that we explore how virus transmit within each network. Unlike most research which aims to predict the reproduction number, we determine them through our simulations in hindsight (Figure 8) and observe their differences across each network.

We shall formally define three versions of reproducibility: underestimate, overestimate, and true reproduction number, and explain the motivations behind their creation.

- (a) Underestimate reproduction number (r_U): r_U is defined as the average number of new infections in a population due to infectious individuals at the time t . It is dynamic and can be found by dividing the new changes in infected individuals at t_{i+1} over the number of infected individuals at t_i , that is:

$$r_U = \frac{I_{i+1} - I_i}{I_i}$$

In the context of complex network modelling, r_U is useful if we have no information on the type of networks investigated. However, it is an underestimate as it only accounts for raw changes in the individuals of the I compartment (which may be restricted by the total number of S individuals as the epidemic progresses) and does not actually capture the true number of transmissions between each S-I pair. Nevertheless, it is expected to give a lower bound in our model.

- (b) Overestimate reproduction number (r_O): Unlike r_U which characterizes raw changes in infected numbers, r_O describes the mean number of possible infections by studying the number of I neighbours around a S individual. Formally it is defined as the number of possible infections at t_{i+1} over the number of infected individuals at t_i . To find this, we require knowledge of the network structure and the number of infected neighbours (denoted by Q_j) at time t_i , that surrounds a newly infected individual at t_{i+1} (denoted by j). r_O can then be calculated by taking:

$$r_O = \frac{\sum_j Q_j}{I_i}$$

r_O meaningful estimator if we want a representation of the worst-case scenario during an epidemic since it assumes that newly infected individuals contract the virus from all

their neighbours regardless of whether the transmission actually occurs. In reality, this can also be tracked if active contact-tracing and early reporting procedures were in place. However, values of r_0 may be out of proportion since it fundamentally does not capture actual virus transmissions as well. Nevertheless, it is postulated to give an upper bound in our model.

- (c) True reproduction number (r_T): r_T was constructed to improve on r_U and r_0 by accounting for actual infectious inducing contacts. It is defined as the number of infectious inducing contacts between each S-I pair at t_{i+1} over the number of infected individuals at t_i . It requires perfect information on all pairwise interactions between each newly infected individual at t_{i+1} (denoted by j) with his infected neighbour at t_i . Only infectious inducing interactions on j (denoted by Z_j) is counted and r_T can be found by:

$$r_T = \frac{\sum_j Z_j}{I_i}$$

In our model, r_T shall be the focus of study as we are interested in the most accurate description of the reproducibility of the virus. Though this quantity is extremely challenging to track, we shall see in later sections, how it relates to r_U and r_0 for different sets of input parameters.

3.4 Canonical average of Monte-Carlo simulation

To provide an accurate description of the epidemic across different networks for the input parameters in section 3.3, it is imperative that we do a sufficiently large number of Monte-Carlo simulations before taking the canonical average of all runs. However, there should be a gauge to characterize if the number of simulations (M) is adequate so that we have a good representation of the epidemic.

In our approach, we use the closeness of r_U , r_0 and r_T , to the pairwise infectability parameter r_t at the onset as a benchmark. This is as r_t can also be viewed as the theoretical number of initial infections assuming that each I_0 individual has a degree of connectivity, $\langle k \rangle$. Therefore, although each run may start from different I_0 , each with non-identical number of neighbours, their canonical average after M runs should approach $\langle k \rangle$. Thus, the reproduction numbers calculated at t_1 should converge to r_t for a sufficiently large M . For our experiment, we define the “closeness” relation to be 95% of the input r_t .

Once results of all M runs are completed, we standardize each run to have the same time period by finding the maximum time T_{\max} , where T_{\max} is defined as:

$$T_{\max} = \max(T_1, T_2 \dots T_j \dots T_M)$$

And T_j is the end time of the j^{th} run. For j^{th} runs that end before T_{\max} we extrapolate their last S, I, R, r_U , r_O and r_T data point by $T_{\max} - T_j$ number of times. With this, all time frames will be normalized, and we can find the canonical mean and standard deviations of each compartment and reproduction numbers.

As an example, we execute 5000 simulations and plot the dynamics of the SIR transitions and reproduction numbers, for input parameters: $r_U = 4$, $I_0 = 1$ & $D_{\text{recover}} = 3$ (see Figure 9).

Note that for the plots of reproduction number, the last time step will always be $T_{\max} - 1$ since the epidemic ends at T_{\max} . We shall denote $(T_{\max} - 1)$ as T_{end} for convenience later on. Data used for plotting are recorded in Annex B: Data availability.

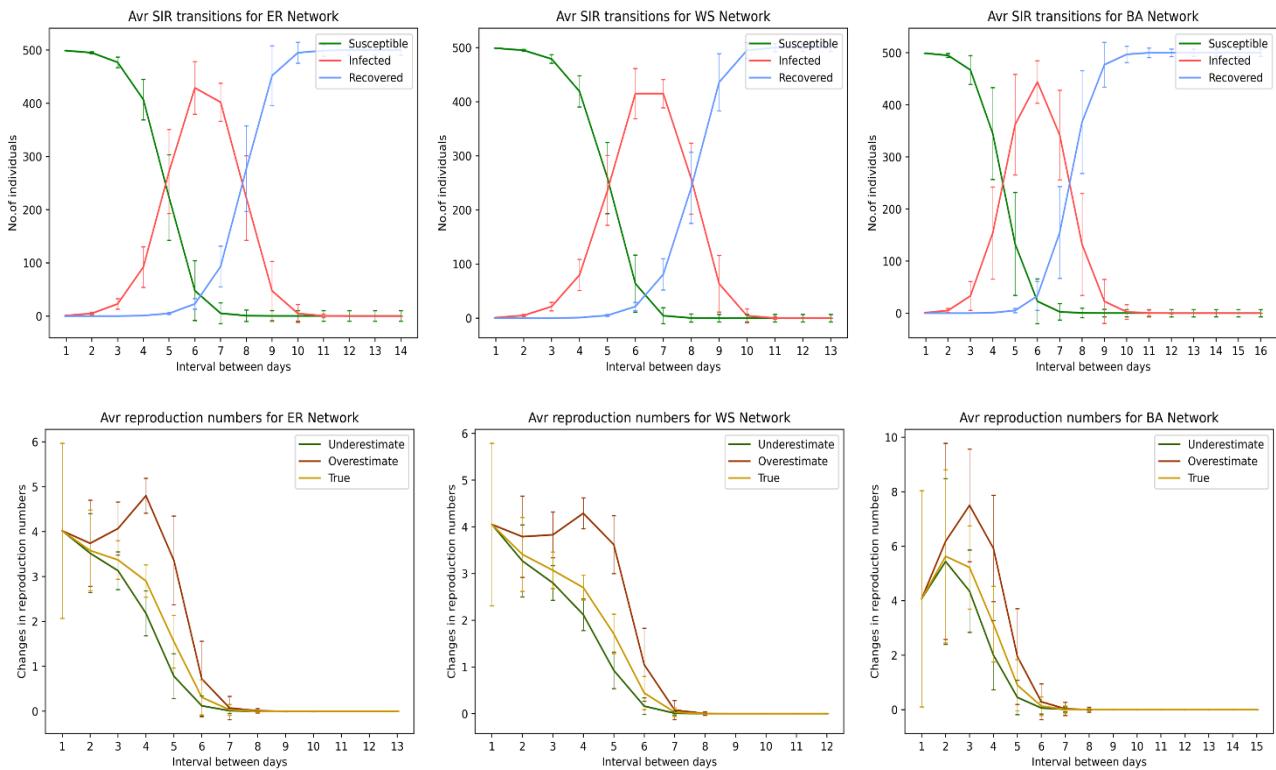


Figure 9: Canonical average of SIR transitions and reproduction number plots for ER (left column), WS (center column) and BA (right column) networks for 5000 simulations with parameters: $r_U = 4$, $I_0 = 1$ & $D_{\text{recover}} = 3$.

3.4.1 Observations of canonical average

Across all networks, we observe that the SIR transitions give similar results in terms of its trend. However, it seems that the BA network experiences larger standard deviations at each time step compared to the ER and WS networks. Albeit it is normal to have standard deviations since different starting points in each run have different connectivity, it seems that the power law degree distribution in the BA network gives it a more profound effect. This may suggest that a communal structure that is akin to the BA network in the real world may be more prone to higher uncertainties when modelling epidemic spreading.

In the SIR transitions, we also note that all networks experience a maximum infection peak > 400 at time t_6 . Leading up to the peak, we investigate the reproducibility at t_5 (see Table 2).

	ER Network	WS Network	BA Network
r_U	0.78	0.93	0.45
r_0	3.36	3.62	1.95
r_T	1.55	1.71	0.90

Table 2: Reproduction numbers at t_5

As the peak is at t_6 we would expect all reproduction numbers > 1 at t_5 so that the number of I individuals will grow. However, we see from Table 2 that r_U across all networks are < 1 . This may be due to the underlying depletion of S individuals as an epidemic progress and therefore the “reproducibility” of the virus, using r_U , is not captured. Nevertheless, we see that all $r_0 > 1$ while $r_T > 1$ for both ER and WS networks and close to 1 for the BA network, which is in line with expectations.

Turning our attention to the trends in virus reproducibility, we see that across all networks, r_T is encapsulated within the bounds of r_U and r_0 , that is, $r_U \leq r_T \leq r_0$. This is within our expectations since we define and the reproduction numbers in the manner in 3.3.

Next, we observe that all reproduction numbers at t_1 fulfil the closeness of 95% to r_T , (99.6%, 98.8% and 98.2% for the ER, WS & BA networks respectively) suggesting that these results are a good depiction of the epidemic. However, their standard deviations are large, though unsurprising since each I individuals have different degrees of connectivity. This is again the most apparent for the BA network which may be due to the power law degree distributions.

Finally we focus on r_T and find that it exhibits a declining trend throughout the entire epidemic for both ER and WS networks, whereas it spikes for the BA network at t_2 before decreasing

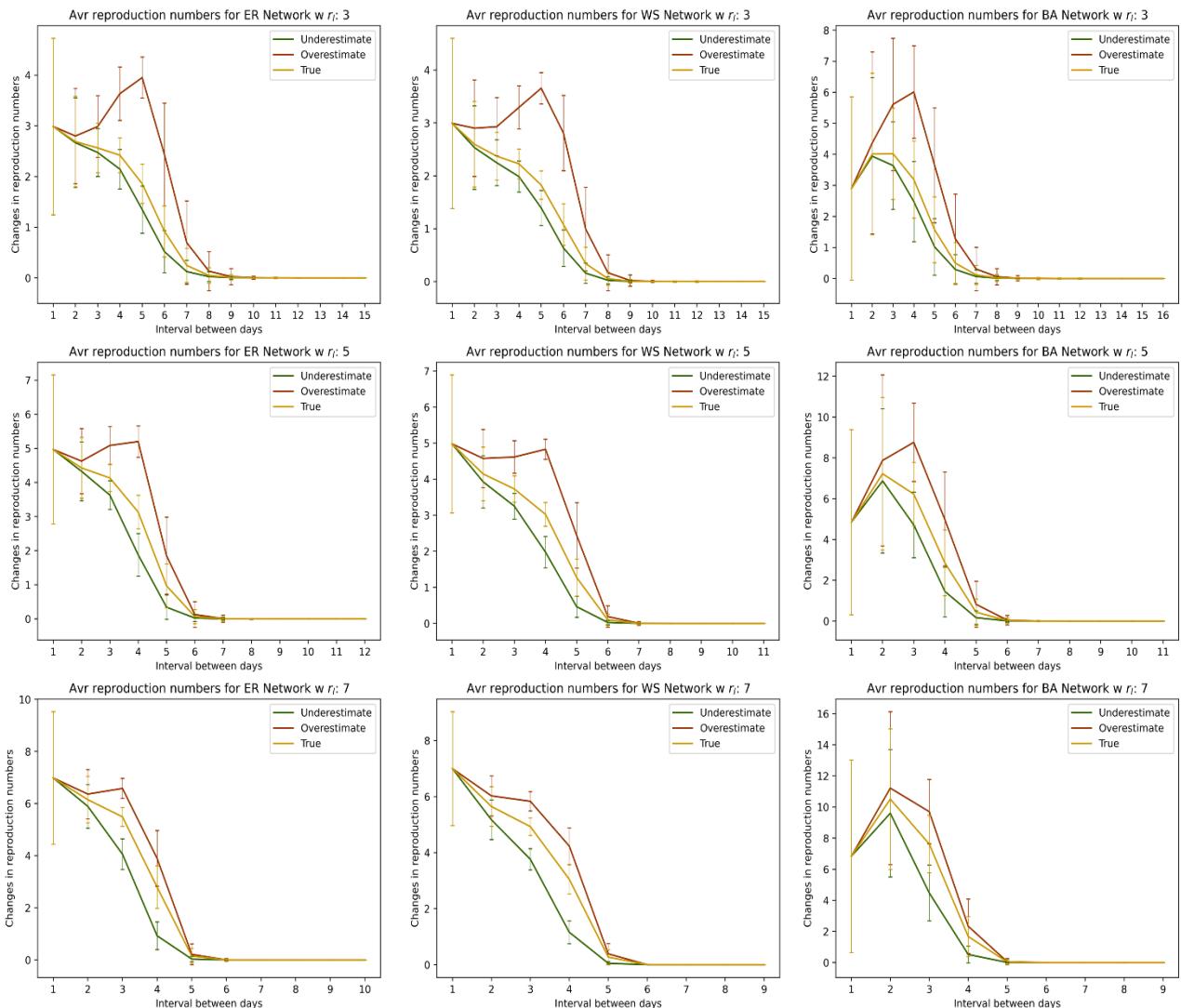
gradually. This suggests that at the onset of the epidemic, the BA network will show a steeper increase in I numbers, which can be verified if we observe the SIR trends in Figure 9. This may be caused by larger differences in the degrees of connectivity of each node in the BA network (especially at hubs), thereby creating the initial spike observed.

3.5 Effects of different r_t and $D_{recover}$

As observed in 3.4, an epidemic's reproducibility has impacts on the SIR transitions. Here we study the effects on virus reproducibility when different r_t and $D_{recover}$ are used to simulate different epidemic outbreaks. We begin our experiment using different r_t .

3.5.1 Epidemic with different r_t

We shall compare the effects on the set of $r_t = \{3, 5, 7, 9\}$ across all networks, with chosen $I_0 = 1$ and $D_{recover} = 3$. The results are shown in Figure 10, with closeness of reproduction numbers at the onset verified to be at least 95% of each r_t . Data points can be found in Annex B.



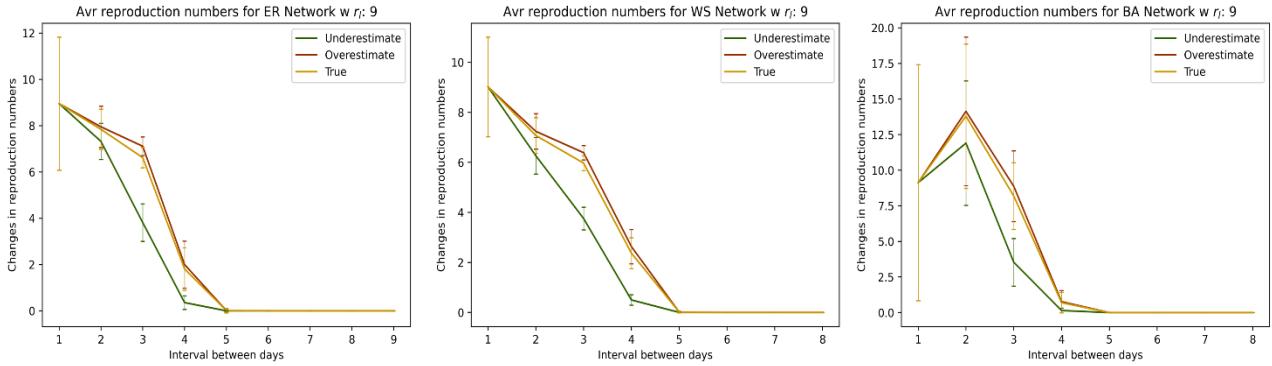


Figure 10: Reproduction numbers of ER (left column), WS (center column) and BA (right column) networks for $r_t: \{3, 5, 7, 9\}$ with parameters $I_0 = 1$ & $D_{recover} = 3$.

From the plots, we notice a similarity for r_T across different networks as we increase r_t . For $r_t = 3$ all networks have r_T closer to r_U . However, as r_t increases, all r_T deviates away from r_U and tends closer towards r_O . This is the most apparent for $r_t = 9$. This is interesting though unsurprising as r_t is directly proportional to p_T , and thus, a larger r_t will signify a larger probability of virus transmission. This suggests that most susceptible individuals in the networks will contract the virus from his infected counterparts, therefore registering a true reproduction number closer to that of the overestimate. The inverse can be easily understood as well.

Next, we replot the results in Figure 10, and compare only changes to r_T as they represent true viral transmissions across each S-I edge. This is shown in Figure 11.

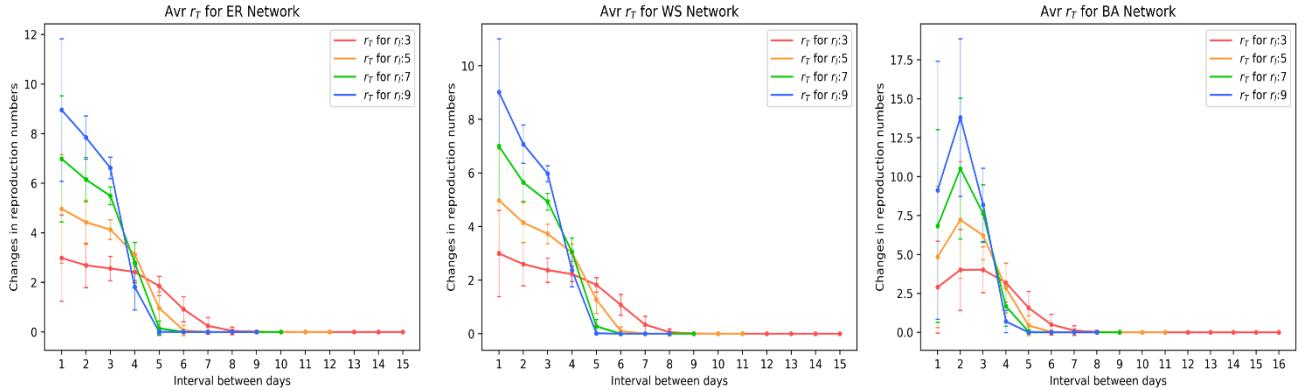


Figure 11: Comparison of r_T for different r_t across the ER (left), WS (center) and BA (right) networks.

By comparing r_T across different networks for the same r_t (see Figure 11), we can confirm the consistency in the spike of the true reproduction number of an epidemic outbreak at t_2 for the BA network. This jump at t_2 is even more dramatic for a larger r_t . On the other hand, the r_T for both ER and WS network always subsides over time.

Lastly, we make comparisons of r_T for different r_t , within the same network. This is shown in Figure 12.

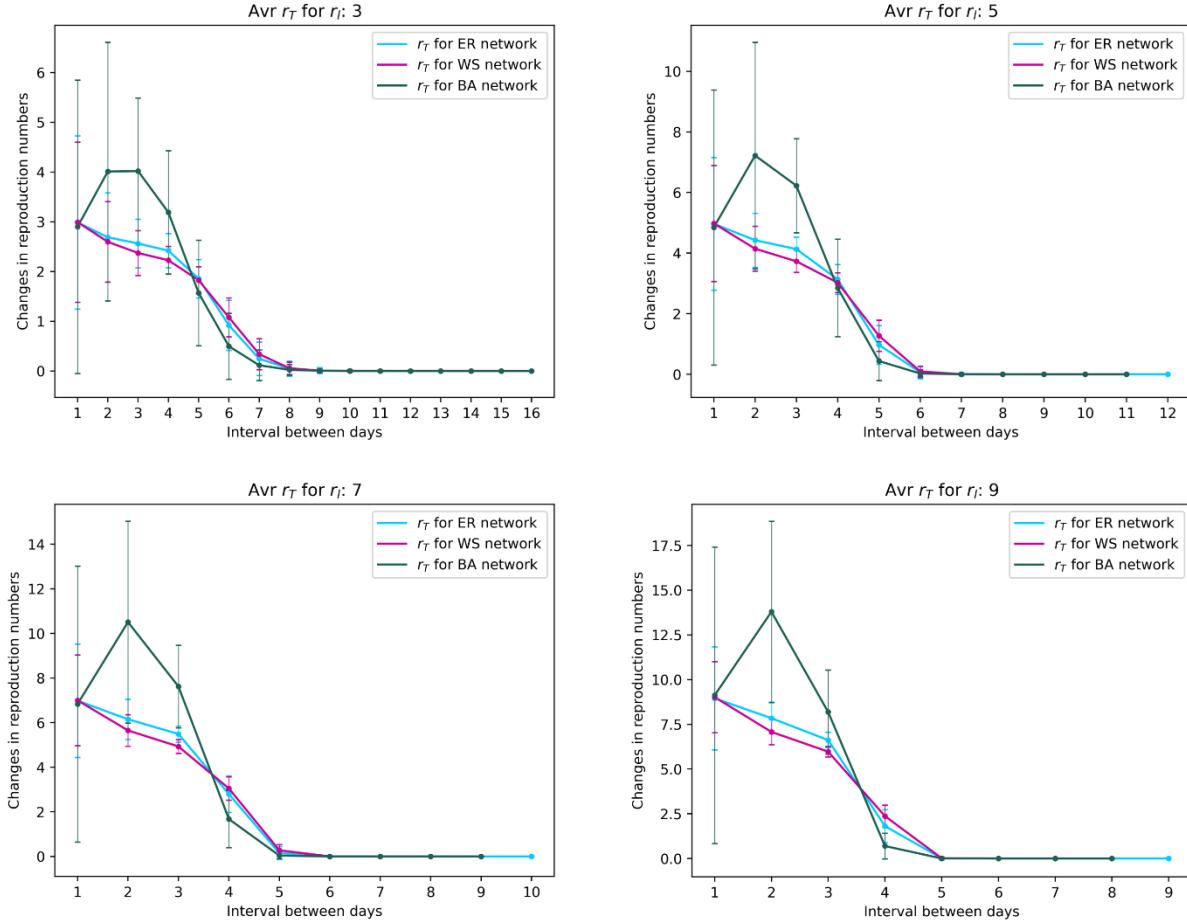


Figure 12: Comparison of r_T across each network for $r_t = 3$ (top left), 5 (top right), 7 (bottom left) and 9 (bottom right)

From the plots, we observe the shortening of the epidemic duration with increasing r_t . This observation is consistent for all networks. The idea behind this could be that a higher r_t , implying a larger p_t , has allowed more susceptible individuals to contract the virus at a quicker rate. This can be verified by observing that r_T is consistent larger for higher r_t at the initial few time steps ($\sim t_1$ to t_4) and dies down much quicker afterwards, consequently ending the epidemic sooner. From this, we may also deduce that the duration of an epidemic is more closely related to the transmissibility of a virus at the onset rather than the social construct of a community (type of network), since all network displays the same trend for a larger r_t .

3.5.2 Epidemic with different $D_{recover}$

We now determine the effects of $D_{recover}$ on the set: $\{1, 3, 5\}$ across all networks. As we have shown earlier that the dynamics of the reproduction numbers is vastly different for the infectability parameters $r_t = 3$ and $r_t = 9$, we shall test the set of $D_{recover}$ with these r_t as well.

The results are shown in Figures 13 and 14 with closeness of reproduction numbers at the onset verified to be at least 95% of each r_t . Data points can also be obtained in Annex B.

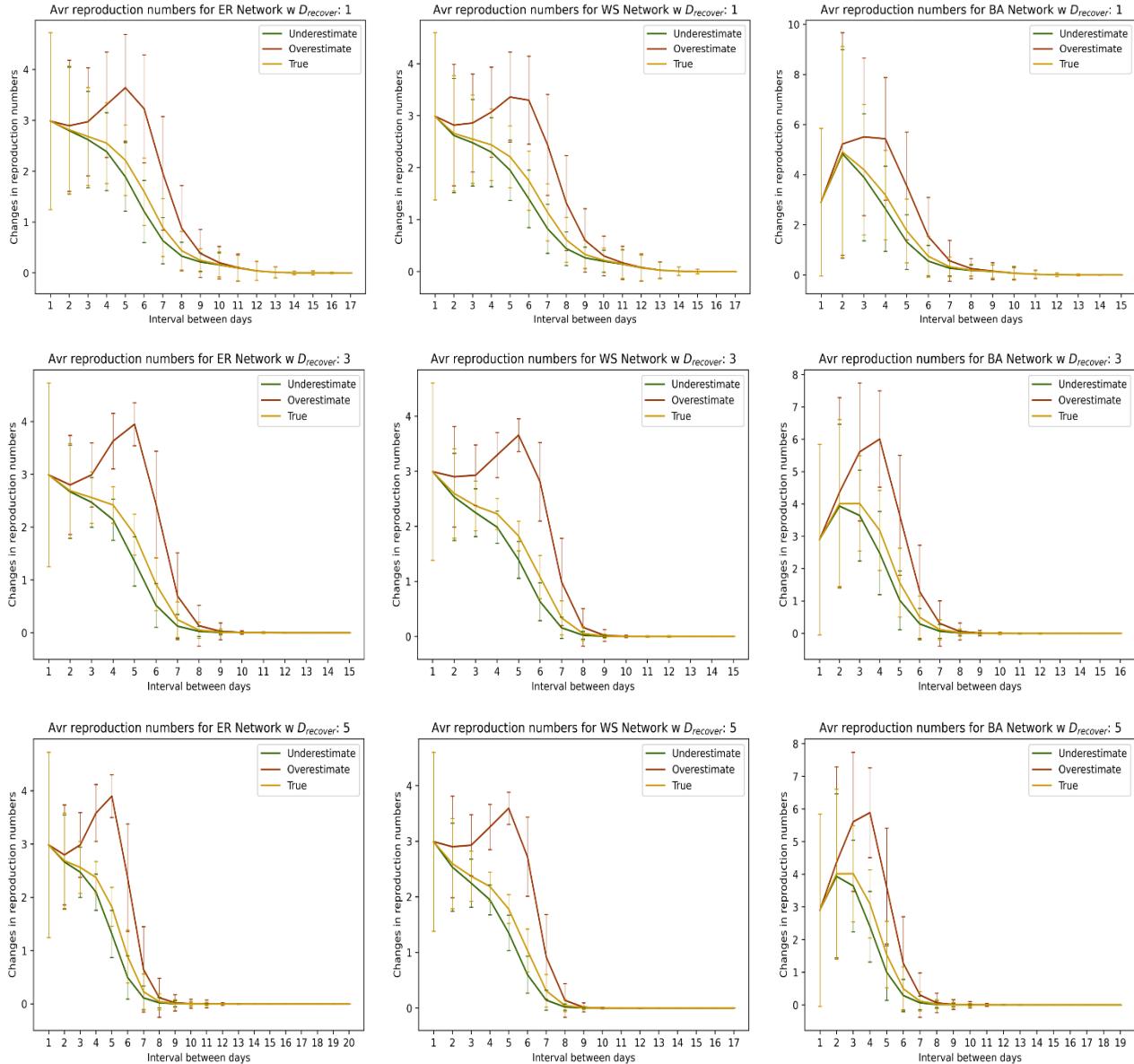


Figure 13: Reproduction number plots for ER (left column), WS (center column) and BA (right column) for $D_{recover}: \{1, 3, 5\}$, with parameters $r_t = 3$ and $I_0 = 1$.

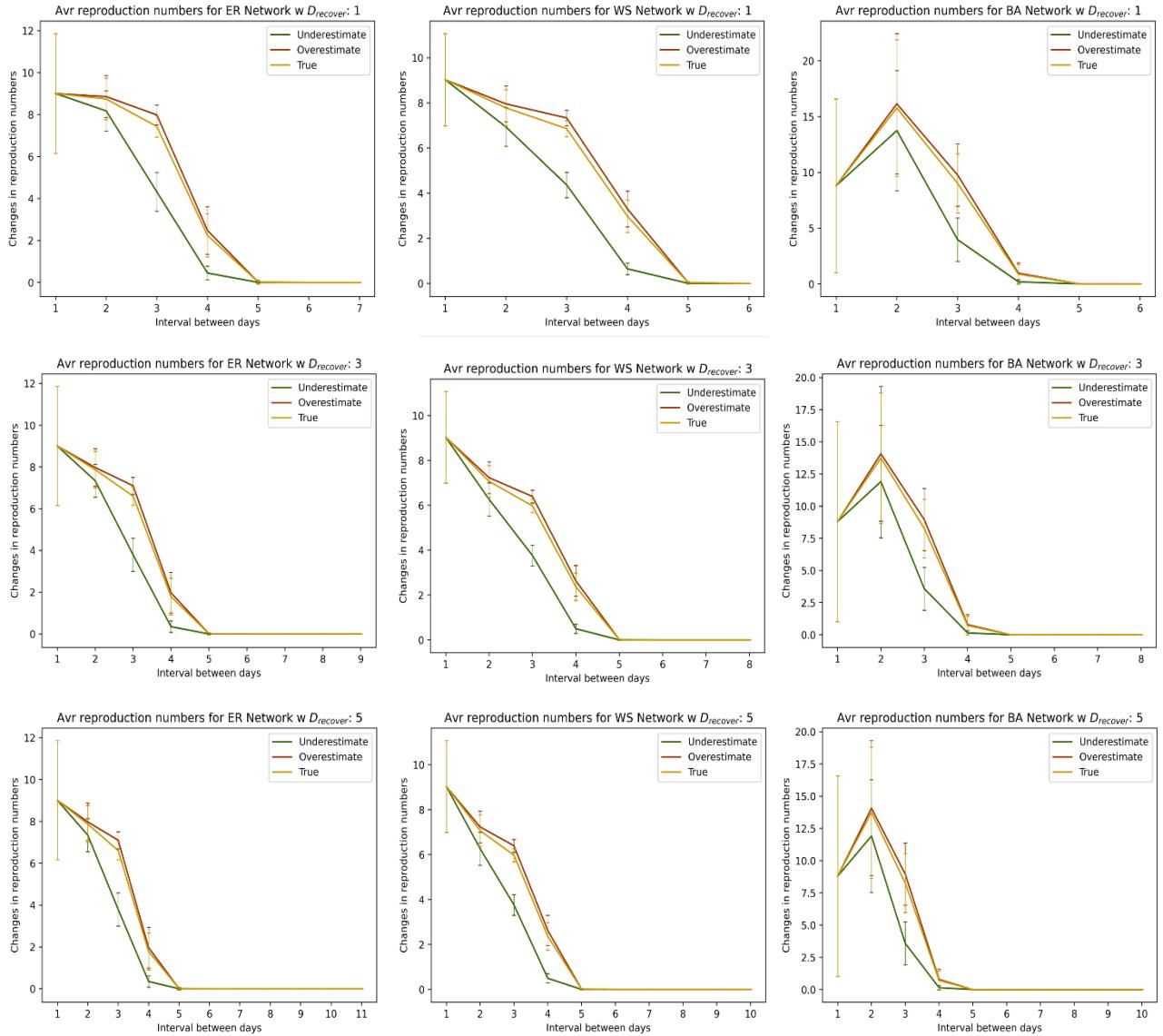


Figure 14: Reproduction number plots for ER (left column), WS (center column) and BA (right column) for $D_{recover} \in \{1, 3, 5\}$, with parameters $r_t = 9$ and $I_0 = 1$.

Comparing Figures 13 and 14, we observe that r_T tends closer to r_U for a smaller r_t , and closer to r_O for a larger one, regardless of the infectious period chosen. This suggests that the values of r_T are strongly related to the infectibility between each S-I edge, rather than the period of infectiousness of the virus.

Next, we find that for both $r_t = 3$ and 9 , the first instance where the reproducibility of the virus equals 0 occurs at each network's $(T_{end} - D_{recover})$ day. This commensurate with our expectations since the last infected individual(s) require " $D_{recover}$ " number of days to recover and therefore, there should be no new infections by day $(T_{end} - D_{recover})$.

Lastly, we compare the duration of the epidemic for $r_t = 3$ and 9 separately. In the case for $r_t = 3$, we encountered an interesting scenario where we found that $(T_{\text{end}} - D_{\text{recover}})$ is the largest for $D_{\text{recover}} = 1$ across all networks. In other words, the epidemic persists the longest for the shorter infectious period. This is counter intuitive as we expect a smaller D_{recover} to signify a smaller infectious period, consequently ending the epidemic faster. In this case, the relevance of network paths may be applied to understand the situation. A possibility for this phenomenon could have been that the virus took a longer path to reach a susceptible neighbour rather than a direct and shorter path from an infected individual. This can happen if the period of infectiousness is short (one day) and each unsuccessful S-I transmission cannot be repeated.

On the other hand, we do not meet such anomalies for the case of $r_t = 9$, as seen where $(T_{\text{end}} - D_{\text{recover}})$ is a constant across all networks, for all D_{recover} . This suggests that a larger r_t (higher S-I transmission probability) has the ability to outweigh possible effects caused by the short period of infectiousness, D_{recover} , therefore may be a more important factor when characterizing the evolution of an epidemic.

3.6 Summary of complex social network model

We have seen how an epidemic behaves when we use a more realistic stochastic approach to model interactions through complex social networks, for a population of $N = 500$.

In our results we found that the BA network consistently gives r_T that increases sharply before decreasing at the initial stages, while the ER and WS networks shows r_T with a decreasing trend throughout. This suggests that real-world communal interactions may be more prone to the effects of epidemic spreading if they exhibit properties similar to the BA network (scale-free). We also discovered that the value of infectibility parameter r_t is crucial in understanding whether we may have under or overestimated the actual reproducibility of a virus in a community, thus a quantity that should be explored during an epidemic. Lastly, we concluded that larger magnitudes of r_t do shorten epidemic durations, regardless of the type of network, and has the ability to outweigh effects caused by short infectious periods, thus being a more important factor when modelling the evolution of an epidemic.

Through the study of virus reproducibility in complex networks, we have inarguably understood epidemic spreading from another perspective. However, networks constructed in our model are static and assumes that individuals always have the same pairwise interactions.

This is untrue in practice as people mingle around and thus, network structures with the same properties should be dynamic. Nevertheless, this could be a future work of research.

With our focus mainly on static complex networks, we now ask if we could make modifications in interactions to further improve on its realism. As a start, we note that individuals usually congregate in cliques rather than pairs, and therefore, our model should account for such group interactions. Henceforth, we shall build HOIs, in the form of simplicial complexes in the next chapter.

Chapter 4: Stochastic Model (Simplicial Complex)

HOIs are group interactions between three or more individuals. They are a focus of study in complex networks research as they have unravelled new phenomenon into the structures of real-world systems which cannot be explained with lower ordered dyads (pairwise interactions). An example of modelling HOIs is through the use of simplicial complexes. As its applications in epidemic spreading is rather new, we hope our research could provide new insights on its effects on virus reproducibility in different complex networks.

4.1 Simplicial Complex

A simplicial complex, denoted by Φ , is a topological space constructed by combining lower dimensional simplices, such as vertex (0-simplice), edge (1-simplice), triangles (2-simplice), tetrahedrons (3-simplice) and more. Formally, Φ on a given set of vertices V , is a collection of simplices, with the extra requirement that if simplex σ is an element of Φ , then all the sub-simplices $v \subset \sigma$ built from subsets of σ are also contained in $\Phi^{20, 40, 41}$.

An example with Φ as a 2-simplex is shown in Figure 15.

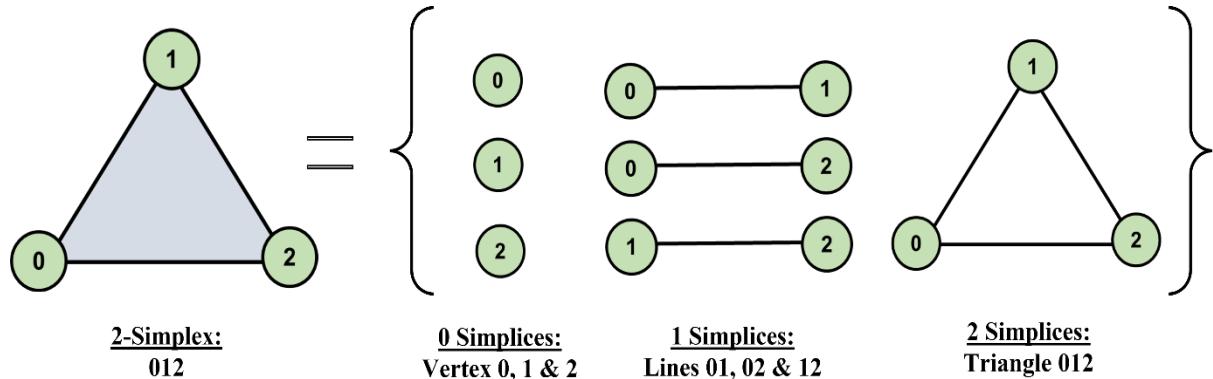


Figure 15: Elements of the 2-simplex, Φ

In Figure 15, Φ is made up of the set of three 0-simplices (vertex 0, 1 & 2), three 1-simplices (line 01, 02 & 12), and one 2-simplices (triangle 012). Each 1-simplex can be considered as a 1-simplex $\subset \Phi$, formed by its constituent elements: two vertex and a line. Evidently, they are also elements of the 2-simplex, Φ .

Simplicial complexes are built on complex social networks by identifying possible 1-simplices (edges), 2-simplices (triangles) and more, in an ascending order, thus considering interactions from bottom up. This represents an epidemic in a more realistic manner. In our model, we shall

adopt the idea of simplicial complexes, up to 2-simplex, and construct them on our ER, WS and BA networks defined earlier. Through this, we shall observe its effects on virus reproducibility in our networks.

4.2 Construction of 2-simplex

In each network, there are often cases where three vertices connect to form a triangle. By iterating all possible edges of the nodes, we can identify the total number of triangles in the network. In our construction, each triangle found is automatically a 2-simplex. The number of 2-simplex in each network is shown in Table 3.

	Number of 2-simplex	$\langle k_{sc} \rangle$
ER Network	192	1.152
WS Network	710	4.260
BA Network	655	3.930

Table 3: Number of 2-simplex in each network and its corresponding average degree.

Similar to finding the average degree of connectivity $\langle k \rangle$ for the pairwise case, we should also characterize the average degree of connectivity, denoted as $\langle k_{sc} \rangle$, for the 2-simplex case. This can be found by taking:

$$\langle k_{sc} \rangle = \frac{3 \times \# \text{ 2 simplex}}{N}$$

The numerator has a multiplication of 3 to account for 2-simplex interactions experienced by each vertex (individual) of a triangle, while the denominator is N so that we can find the average. Calculated $\langle k_{sc} \rangle$ of each network are shown in Table 3.

4.3 Modified simulation methodology

To include 2-simplex transmissions, we define another input transmission parameter, known as group infectability, r_{sc} .

Similar to the pairwise infectability r_t , r_{sc} is defined as the product of the probability of virus transmissions in a 2-simplex, p_{sc} and the average degree of connectivity of the 2-simplex, $\langle k_{sc} \rangle$ in each network. With this, p_{sc} can be found by taking:

$$p_{sc} = \frac{r_{sc}}{\langle k_{sc} \rangle}$$

r_{sc} is used to standardize the reproducibility due to 2-simplex transmissions, across all networks and in our context, $r_{sc} \leq 1.125$ (a limit due to $\langle k_{sc} \rangle$ of the ER network)

In the modified simulation, transmissions between 2-simplex can only occur under the following conditions:

1. A susceptible individual is in a 2-simplex with two other infected individuals.

2. The pairwise transmission between each S-I edge within the 2-simplex has failed.

Both conditions are necessary in order to fulfil the idea of bottom-up interactions. An example of how a 2-simplex transmission works is shown in Figure 2.

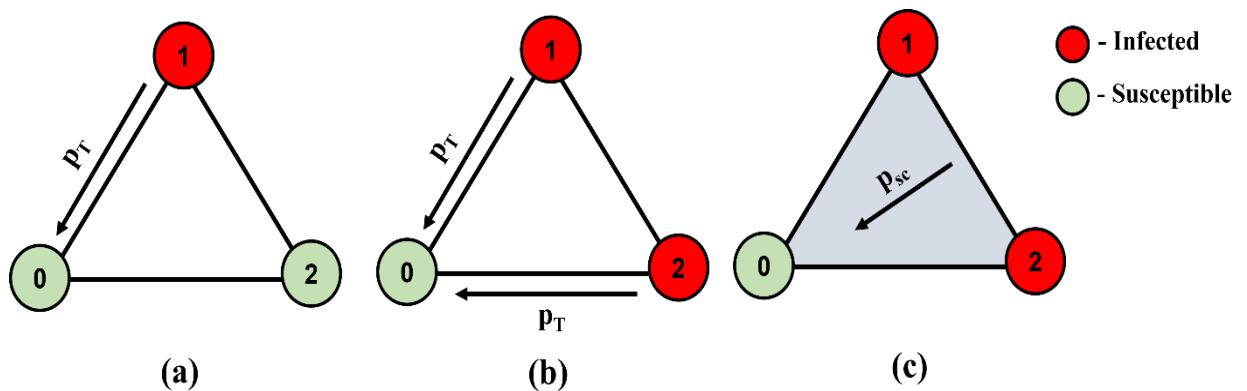
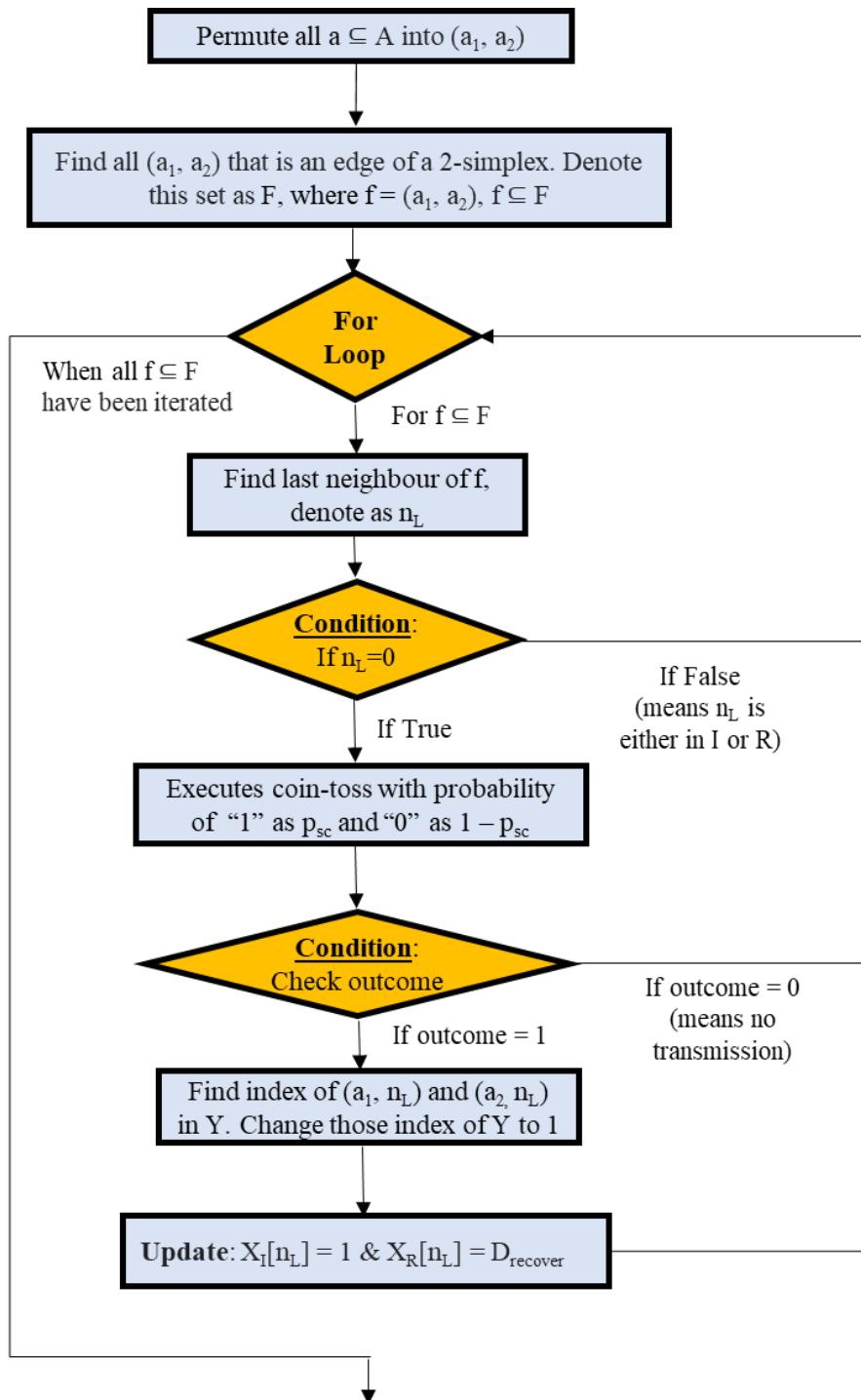


Figure 16: Example of 2-simplex transmission

In Figure 16(a), the S node 0 can only be infected through a pairwise interaction, with probability p_I , by the I node 1. On the other hand, in Figure 16(b), the S node 0 can be infected either by node 1 or 2 through a pairwise infection, again with probability p_I . However, both pairwise transmission fails, the 2-simplex interaction in Figure 16(c) kicks in and node 0 will be infected with probability p_{sc} .

With the foundations laid, we modify the flowchart in Figure 8 (yellow boxes) to include the 2-simplex transmission described above (see Annex A: Code availability). The appended process is elaborated in a flowchart as shown in Figure 17.

Methods to identify the SIR transitions and virus reproducibility following the modified process in Figure 17 remains unchanged and we use canonical averages to give a more accurate representation of the epidemic.



Legend

→ - Process flow [Light Blue Box] - Process ◊ - Conditions / loops

Figure 17: Appended flow for the 2-simplex transmission

4.4 Effects of r_{sc} on networks

With the new model, we can compare differences in the virus reproducibility when we include 2-simplexes in each network. For all experiments, we explore r_{sc} : {0, 0.5, 0.75, 1}, with $r_{sc} = 0$ implying only pairwise interactions. We will do our investigations on epidemic with different r_t and $D_{recover}$.

4.4.1 Epidemic with $r_t = 3$ and 9

In this section, we compare the effects of r_{sc} on two different epidemics with $r_t = 3$ and 9. For both experiments, $D_{recover}$ and I_0 have been set to 3 and 1 respectively. The reason for comparing these two r_t in particular is due to the contrasting observations we identified in Chapter 3, where r_t tends closer to r_U for the formal and r_O for the latter.

Results for the true reproduction number r_T are shown in Figure 18, while detailed plots with the under (r_U) and overestimate (r_O) reproduction numbers can be found in Annex C: Detailed plots for 2-simplex. Datapoints of each graph are also available in Annex B.

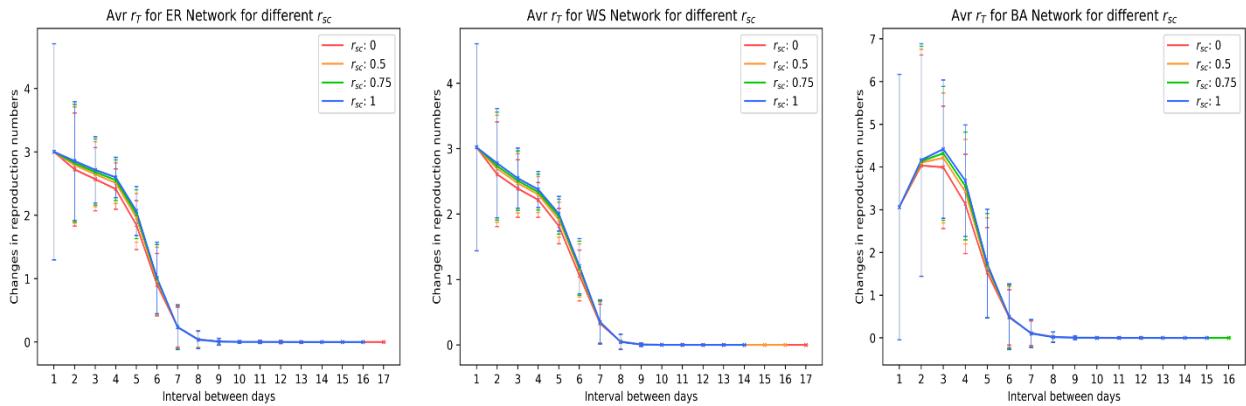


Figure 18a: Plots of r_T for different r_{sc} , with $r_t = 3$, $D_{recover} = 3$ and $I_0 = 1$ for ER (left), WS (center) and BA (right) networks.

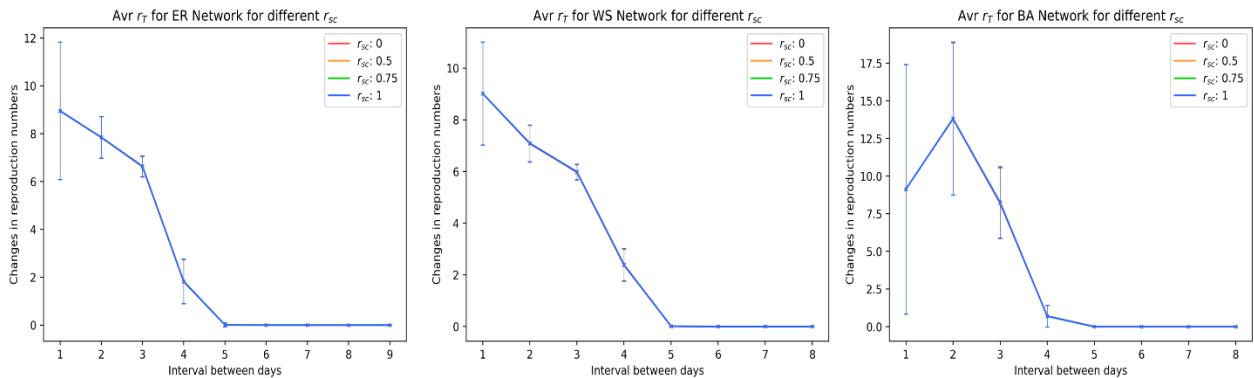


Figure 18b: Plots of r_T for different r_{sc} , with $r_t = 9$, $D_{recover} = 3$ and $I_0 = 1$ for ER (left), WS (center) and BA (right) networks.

From the detailed plots (see Annex C), we observe that $r_\ell = 3$ gives r_T that still tends closer to r_U even with 2-simplex effects, though there are very slight inclinations towards r_0 as r_{sc} increases. On the other hand, for $r_\ell = 9$, r_T remains close to r_0 , which is within expectations.

Shifting our focus to studying r_T for the case of $r_\ell = 3$ (Figure 18a), we notice that increasing the value of r_{sc} has very minimal effects on increasing r_T as a whole. This suggests that pairwise interactions may still be the primary form of viral transmission throughout the epidemic in this context. Next, we observe that the effects of r_{sc} sets in rather early, as seen where the increment in r_T begins at t_2 . This is unsurprising as each node has an average degree of connectivity for the 2-simplex interaction that is > 1 , thereby allowing potential group effects to start early. Lastly, we find that introducing 2-simplex gives an epidemic end time that is generally smaller compared to the one without ($r_{sc} = 0$). This corroborates with our intuition since $r_{sc} > 0$ increase r_T and hence should cause an epidemic to end faster.

On the other hand, in the case of $r_\ell = 9$ (Figure 18b), introducing 2-simplex has no effect on r_T at all. This is most likely due to high pairwise transmission probabilities that has almost certainly infected a susceptible neighbour in each S-I edge. Therefore, 2-simplex interactions do not take place which consequently gives the trend as seen in Figure 18b.

4.4.2 Epidemic with $r_\ell = 1$

In the previous section, we observed that the effects of 2-simplex is minimal ($r_\ell = 3$) to none ($r_\ell = 9$) and pairwise interactions seem to play a bigger part in virus transmissions. As there is an inherent limit to r_{sc} , we cannot increase the effects of group interactions any further. However, the pairwise effects can be lowered to allow 2-simplex transmissions to be more apparent. This can be done if we choose $r_\ell = 1$. The value is selected in particular as it implies a pairwise transmission probability of $p_T \approx 0.1$. Comparing p_{sc} for the set of r_{sc} across all networks (see Table 4), we observe that $p_{sc} > p_T$ for all cases.

	$r_{sc} = 0$	$r_{sc} = 0.50$	$r_{sc} = 0.75$	$r_{sc} = 1.00$
ER Network	0	0.434	0.651	0.868
WS Network	0	0.117	0.176	0.235
BA Network	0	0.127	0.191	0.254

Table 4: Probability of 2-simplex transmissions (p_{sc}) corresponding to the respective r_{sc} for each network.

Physically, this means that virus transmissions are more likely to take place when a S individual is exposed to a group of I neighbours simultaneously instead of each one-to-one interaction sequentially. This fits what we expect in reality.

We now study this in detail with $D_{\text{recover}} = 3$ and $I_0 = 1$ respectively. Results for r_T are shown in Figure 19, while comprehensive plots with changes to the under (r_U) and overestimate (r_O) reproduction numbers can be found in Annex C. Datapoints are also available in Annex B.

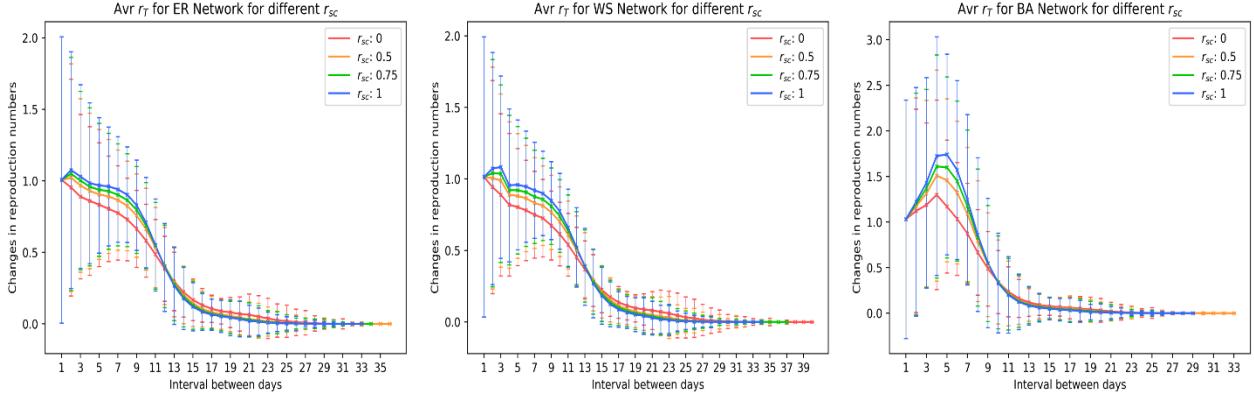


Figure 19: Plots of r_T for different r_{sc} , with $r_t = 3$, $D_{\text{recover}} = 3$ and $I_0 = 1$ for ER (left), WS (center) and BA (right) networks.

From Figure 19, we observe that the effects of 2-simplexes are much more obvious compared to those of section 4.4.1, having larger increments in r_T . We also note two similarities with the case of $r_t = 3$, that is the 2-simplex is prevalent at the initial stages, and in general, the inclusion of r_{sc} shortens the epidemic duration (except a small anomaly for BA network with $r_{sc} = 0.5$).

Apart from trends previously observed, we also notice that r_T with 2-simplex effects gives lower values than r_T without 2-simplex effects, for times after $t = 12, 13$ and 10 for the ER, WS and BA networks respectively. This is consistent for all non-zero r_{sc} , suggesting that most 2-simplex came into play by then, which consequently contributes to the shortening of the epidemic duration.

Lastly, we identified a very interesting finding on the initial dynamics of the ER and WS networks when 2-simplex effects are included. From all findings thus far, we observe that the ER and WS networks show a declining r_T trend throughout the entire epidemic. With 2-simplexes however, we see an initial spike for r_T at the initial outbreak (t_1 to t_2) similar to that of the BA network. This demonstrates that HOIs are capable of changing the onset dynamic of networks, consequently affecting the overall evolution as an outbreak progresses, for an epidemic with low pairwise transmissions.

4.4.3 Epidemic with different $D_{recover}$

In the previous section, we saw major differences in epidemic evolutions when we use a pairwise transmission probability that is lower than that of group. This is achieved with $r_t = 1$. Using this value, we test if epidemics with different infectious period (recovery days), $D_{recover} \in \{1, 3, 5\}$, would alter the dynamics of r_T .

We begin by comparing the results for r_T when $D_{recover} = 1$ for the set of r_{sc} chosen. The plots are shown in Figure 20 (data points available in Annex B).

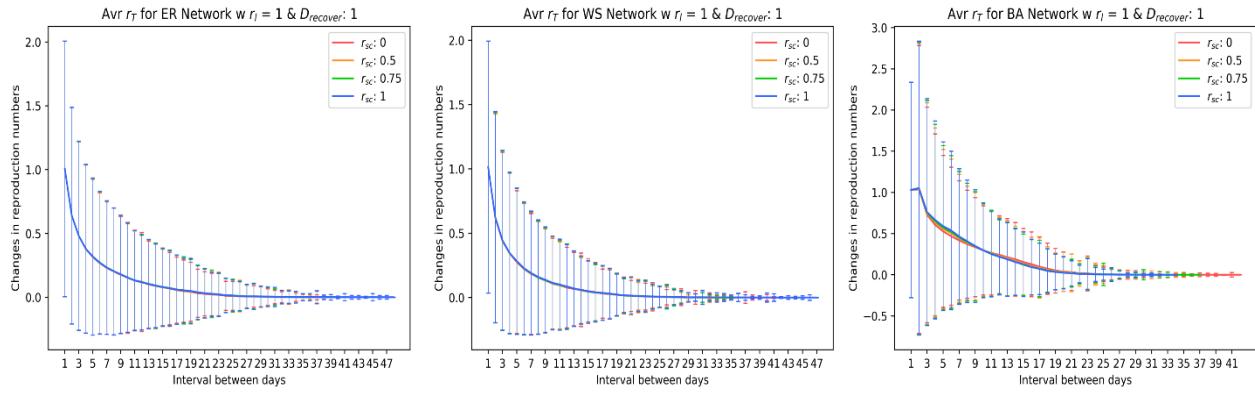


Figure 20: Plots of r_T for different r_{sc} , with $r_t = 1$, $D_{recover} = 1$ and $I_0 = 1$ for ER (left), WS (center) and BA (right) networks

From Figure 20, we observe that 2-simplex interactions have very minimal effects on r_T across all networks. This could be due to the fast recovery time of individuals that consequently results in groups having little to no opportunities of formation. Hence, group interactions play very little part in increasing the actual transmissibility of the virus.

We also saw that the epidemic, with the inclusion of 2-simplex interactions, ends at a later time for both ER and WS networks when $r_{sc} = 1$. This is surprising as group interactions should set in early and decrease end times in general. This could mean that the effect of r_{sc} in shortening epidemic durations is not applicable for short infectious period of $D_{recover} = 1$.

Next, we compare r_T for $D_{recover} = 3$ and 5 across each r_{sc} value. This is plotted in Figure 21 (data points available in Annex B).

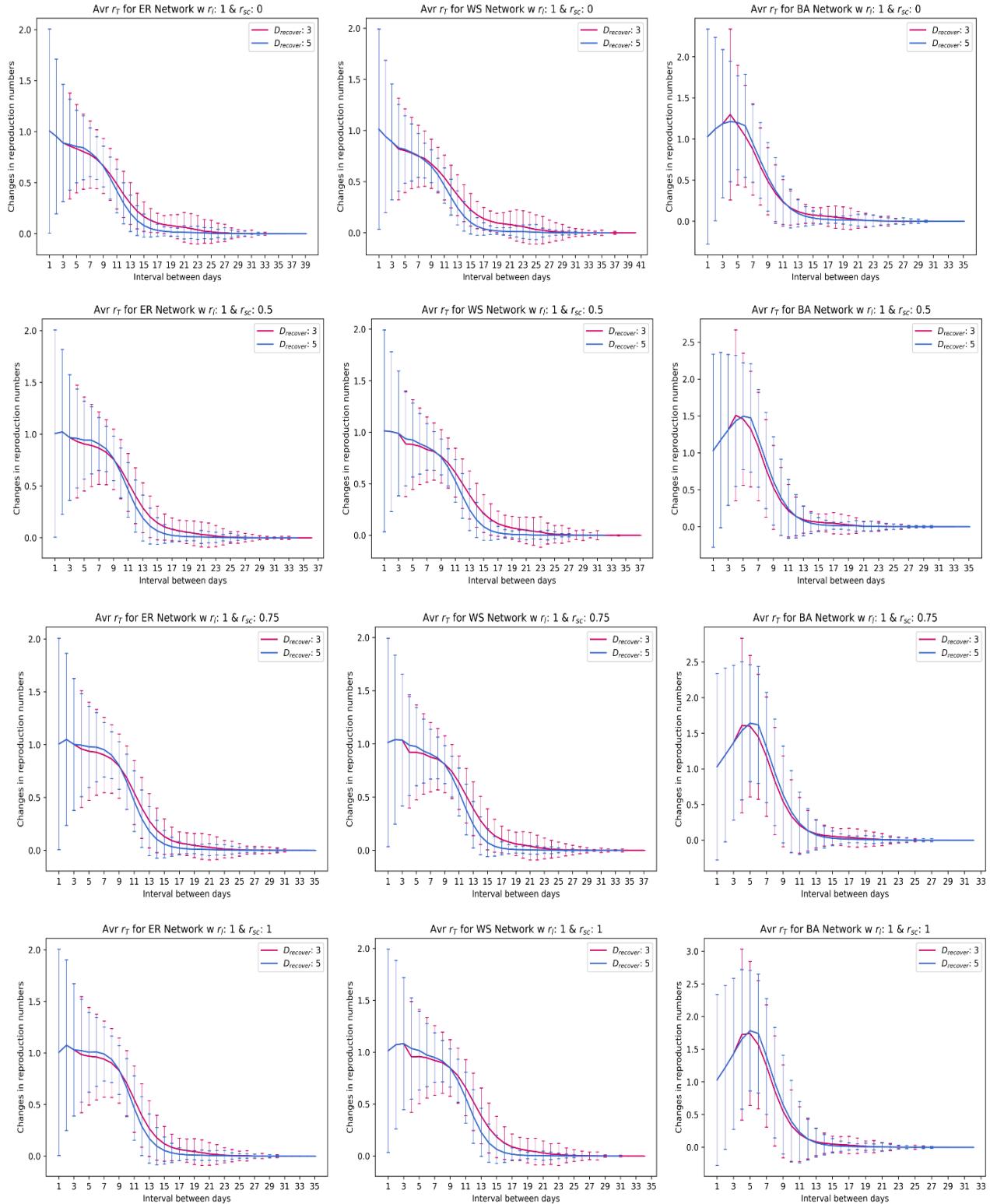


Figure 21: Comparison of r_T for $D_{recover} = 3$ and 5, with increasing r_{sc} values, for ER (left column), WS (center column) and BA (right column) networks.

In general, the inclusion of r_{sc} shortens the outbreak duration for epidemics with infectious period, $D_{recover} = 3$ and 5 . Moreover, all networks show the same trend in r_T till t_3 , with or without 2-simplex effects.

Diving into the details, we observe that without 2-simplex effects, the ER and WS networks show a r_T that is larger for a greater $D_{recover}$, up till t_9 and t_7 respectively. With its inclusion, the r_T values for both $D_{recover} = 3$ and 5 increases marginally and demonstrates trends that are similar to the case without 2-simplex. On the other hand, the BA network with a recovery time of $D_{recover} = 3$ gives a r_T that is larger at t_4 but smaller from t_5 onwards. Akin to the ER and WS networks, the 2-simplex only increases r_T values slightly and produces the same trend.

4.5 Summary of 2-simplex interactions

In this chapter, we studied the effects of 2-simplex HOIs on complex networks and discovered changes in the virus reproducibility when they are taken into consideration. From our findings, it seems that group interactions that have higher probabilities of transmission (p_{sc}) compared to pairwise ($p_T = 0.1$) will have more significant effects on the actual virus reproducibility as an epidemic evolves. Moreover, we identified how virus reproducibility can be altered at the onset if 2-simplex interactions play a more defining role. Lastly, we learnt that the inclusion of r_{sc} generally reduces an epidemic duration, provided that the infectious period $D_{recover} \neq 1$. These insights demonstrate the manner HOIs affect virus reproducibility in a community and allow us to gain a deeper understanding of epidemic spreading.

Though our objectives were met, the construct could be improved if we were able to generate an Erdos-Renyi network with higher average 2-simplex degrees of connectivity. This would allow us to consider larger r_{sc} as a whole and consequently, a higher probability of group transmissions, enabling us to model other realistic epidemics. Nevertheless, our model grants us sufficient insights that could be used as stepping stones for future works.

Chapter 5: Conclusion

In summation, our work has successfully modelled the spread of an epidemic for a population with $N = 500$, through the use of the deterministic ODE model and stochastic complex networks with HOIs. It has provided valuable insights into the dynamics of virus reproducibility, impact of network structures as well as HOIs on epidemic outcomes.

In the deterministic ODE model, we identified a threshold value of $r_0 > 6.22$ that can fully infect the entire population regardless of S_0 and I_0 numbers, thus serving as a good gauge for monitoring purposes. On the other hand, the stochastic complex network approach highlighted important dynamical behaviours of virus reproducibility in different network structures, notably on the BA network, and also showed the importance of r_t in determining if we should under or overestimate reproduction numbers during an epidemic. For epidemics with large r_t , we also confirmed their ability to outweigh effects caused by short infectious periods, thus imperative to isolate individuals who contracted such viruses during an epidemic. Lastly, by improving the networks with HOIs, we observed how 2-simplex transmissions ($p_{sc} >$ pairwise ($p_T = 0.1$) played a defining role in altering the onset of an epidemic as well as changing actual virus reproducibility for different networks. This affirms the need to reduce group interactions during an epidemic in order to successfully curb virus spreading.

In the future, it would be interesting to extend our work to study other compartmental models, such as SEIR, and for the complex network in particular, to include higher orders beyond 2-simplex. These will allow us to appreciate epidemics with varying nature and broaden our understanding of dynamical interactions in the real world.

References

1. World Health Organization (WHO). (2023). Weekly epidemiological update on COVID-19-30 March.
2. Koh, D. (2020). COVID-19 lockdowns throughout the world. *Occupational Medicine*, 70(5), 322-322.
3. Yan, Y., Pan, H., Shao, N., Xuan, Y., Wang, S., Li, W., ... & Chen, W. (2020). COVID-19 in Singapore: another story of success. *International Journal of Mathematics for Industry*, 12(01), 2050001.
4. Burki, T. (2020). China's successful control of COVID-19. *The Lancet Infectious Diseases*, 20(11), 1240-1241.
5. Heesterbeek, H., Anderson, R. M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., ... & Isaac Newton Institute IDD Collaboration. (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227), aaa4339.
6. Xiang, Y., Jia, Y., Chen, L., Guo, L., Shu, B., & Long, E. (2021). COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models. *Infectious Disease Modelling*, 6, 324-342.
7. Sofonea, M. T., Cauchemez, S., & Boëlle, P. Y. (2022). Epidemic models: why and how to use them. *Anaesthesia, critical care & pain medicine*, 41(2), 101048.
8. Brauer, F. (2008). Compartmental models in epidemiology. *Mathematical epidemiology*, 19-79.
9. Mehdaoui, M. (2021). A review of commonly used compartmental models in epidemiology. *arXiv preprint arXiv:2110.09642*.
10. Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772), 700-721.

11. Kermack, W. O., & McKendrick, A. G. (1932). Contributions to the mathematical theory of epidemics. II.—The problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 138(834), 55-83.
12. Giraldo, J. O., & Palacio, D. H. (2008). Deterministic SIR (Susceptible–Infected–Removed) models applied to varicella outbreaks. *Epidemiology & Infection*, 136(5), 679-687.
13. Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139, 110057.
14. Batista, A. M., de Souza, S. L., Iarosz, K. C., Almeida, A. C., Szezech Jr, J. D., Gabrick, E. C., ... & Caldas, I. L. (2021). Simulation of deterministic compartmental models for infectious diseases dynamics. *Revista Brasileira de Ensino de Física*, 43.
15. Allen, L. J. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2), 128-142.
16. Sani, A., & Abapihi, B. (2021, May). Connection of CTMC process, deterministic and stochastic differential equations in modeling of epidemics. In *Journal of Physics: Conference Series* (Vol. 1899, No. 1, p. 012111). IOP Publishing.
17. Aldous, D. (2003). A stochastic complex network model. *Electronic Research Announcements of the American Mathematical Society*, 9(19), 152-161.
18. Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of modern physics*, 87(3), 925.
19. Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.
20. Iacopini, I., Petri, G., Barrat, A., & Latora, V. (2019). Simplicial models of social contagion. *Nature communications*, 10(1), 2485.
21. Palafox-Castillo, G., & Berrones-Santos, A. (2022). Stochastic epidemic model on a simplicial complex. *Physica A: Statistical Mechanics and its Applications*, 606, 128053.

22. Jones, J. H. (2007). Notes on R0. California: Department of Anthropological Sciences, 323, 1-19.
23. Van den Driessche, P., & Watmough, J. (2008). Further notes on the basic reproduction number. Mathematical epidemiology, 159-178.
24. Miller, E. (2003). Potential and existing impact of vaccines on disease epidemiology. In The vaccine book (pp. 37-50). Elsevier Inc.
25. Foppa, I. M. (2016). A historical introduction to mathematical modeling of infectious diseases: Seminal Papers in Epidemiology. Academic Press.
26. Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., & Jacobsen, K. H. (2019). Complexity of the basic reproduction number (R_0). Emerging infectious diseases, 25(1), 1.
27. Heffernan, J. M., Smith, R. J., & Wahl, L. M. (2005). Perspectives on the basic reproductive ratio. Journal of the Royal Society Interface, 2(4), 281-293.
28. Li, J., & Blakeley, D. (2011). The failure of R_0 . Computational and mathematical methods in medicine, 2011.
29. Hethcote, H. W. (2000). The mathematics of infectious diseases. SIAM review, 42(4), 599-653.
30. Weiss, H. H. (2013). The SIR model and the foundations of public health. Materials matemáticas, 0001-17
31. Smith, D., & Moore, L. (2004). The SIR model for spread of disease-the differential equation model. Convergence.
32. Bai, F. (2020). Evaluating different epidemiological models with the identical basic reproduction number \mathcal{R}_0 . Journal of Biological Dynamics, 14(1), 849-870.
33. Kretzschmar, M. E. E. (2016). Measurement and modeling: Infectious disease modeling. Reference Module in Biomedical Sciences.
34. Bernick, L. (2018). Modeling Human Networks Using Random Graphs.

35. Hagberg, A., Swart, P., & S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
36. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *nature*, 393(6684), 440-442.
37. Barabási, A. L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4), 69-77.
38. Karyotis, V., & Khouzani, M. H. R. (2016). Malware-propagative Markov random fields. *Malware Diffusion Models for Wireless Complex Networks*. Morgan Kaufmann, Boston.
39. Barabási, A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific american*, 288(5), 60-69.
40. Frank, A. U. (2009). Simplicial Complex. In *Encyclopedia of Database Systems*. Springer Verlag.
41. KELLEHER, C., & PANTANO, A. Introduction to simplicial complexes. url: <https://www.math.uci.edu/~mathcircle/materials/MCsimplex.pdf> (cit. on p. 40).

Appendices

Annex A: Code availability

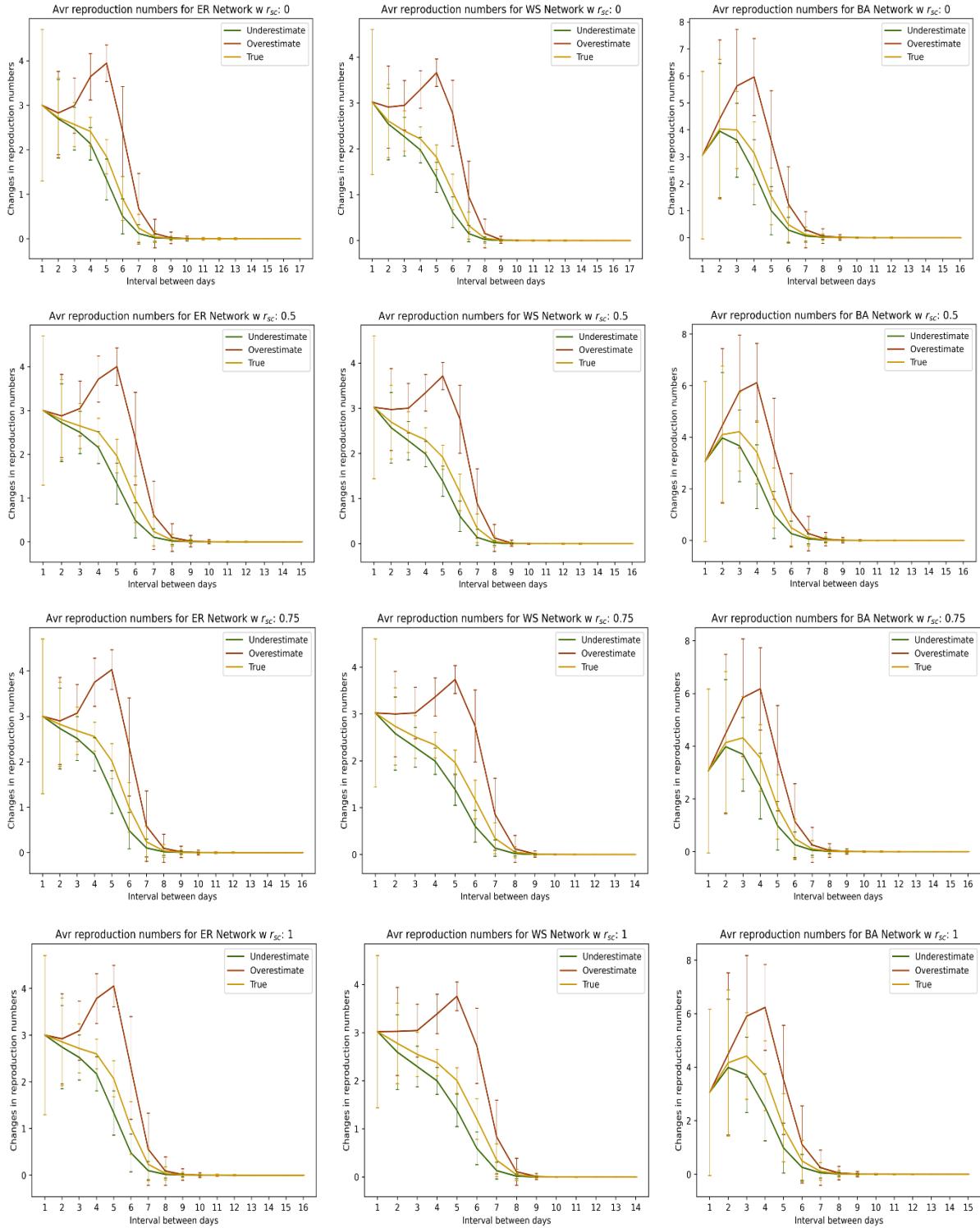
Codes used for plots and complex network simulations are available at:
<https://github.com/Lelion9865/Annex-A-Code-availability.git>

Annex B: Data availability

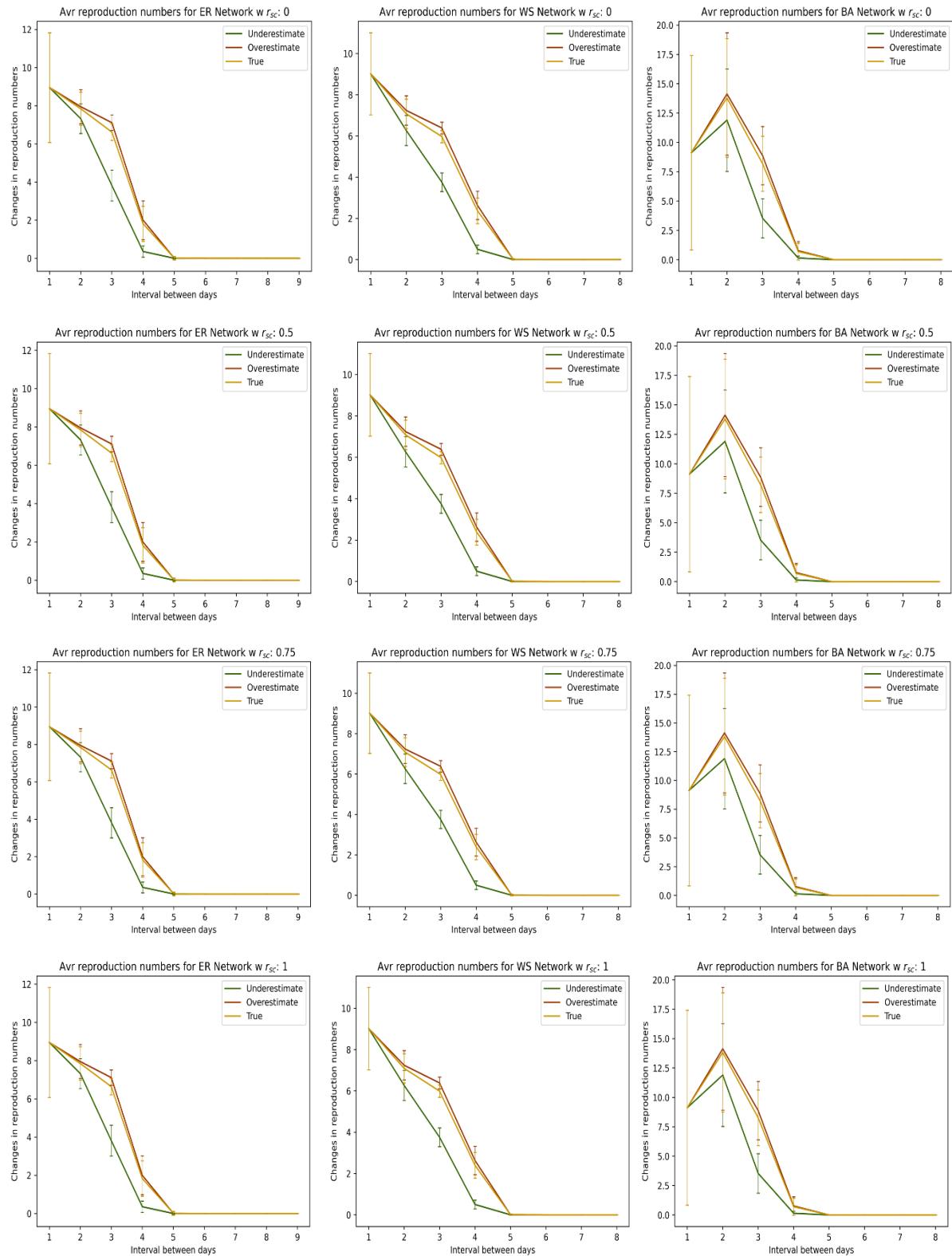
Data used for plots are available at: <https://github.com/Lelion9865/Annex-B-Data-availability.git>

Annex C: Detailed plots for 2-simplex

- (a) Detailed plots of under (r_U) and overestimate (r_O) reproduction for $r_t = 3$, $D_{\text{recover}} = 3$ and $I_0 = 1$ (section 4.4.1)



(b) Detailed plots of under (r_U) and overestimate (r_O) reproduction for $r_t = 9$, $D_{\text{recover}} = 3$ and $I_0 = 1$ (section 4.4.1)



(c) Detailed plots of under (r_U) and overestimate (r_O) reproduction for $r_t = 1$, $D_{\text{recover}} = 3$ and $I_0 = 1$ (section 4.4.2)

