Modelling Epidemic Spread in a Dynamic Network

Connor PJ. Tynan
MTH3022 Project

1 Introduction and Background

The modelling of COVID-19 has played a key role in understanding and planning against the epidemic and as a result is of major importance to planners and people alike. There are several ways of modelling an epidemic, one of which is as a *dynamical system* on a graph in which vertices represent individuals and edges describe contact between said individuals, hence two adjacent vertices $i \sim j$ represent a contact between individuals i and j.
The focus of this investigation is to gain insight into how network structure can affect the outcome of SIR epidemic simulations, i.e. how does specification of our network parameters affect the short/long-term behaviour of simulations and whether or not an epidemic arises for some given infectivity parameters. In order to study such, an R-based package called EpiModel (1) is utilised and used as a primary means of data collection and analysis.

2 EpiModel and Methodology

As mentioned, EpiModel is our primary tool for data collection and analysis for this investigation. The package provides the means for modelling how infectious disease transmission affects population dynamics using R.

EpiModel supports the following classes of epidemic model: Deterministic Compartmental models (DCM), Stochastic Individual Contact models (ICM) and finally Network models. For this project EpiModel is used to build a network model. Network-type models are arguably the most flexible of the three model types, enabling the user to model independent contacts between individuals over time. The underlying network model will be considered as being: probabilistic (subject to randomness) and non-static (dynamic/ever changing), furthermore EpiModel enables us to model viral transmission in a 2 step process: Firstly, generate a network using EpiModel, in which network parameters must be specified.

Network parameters requiring specification include:

- Number of Nodes (Number of individuals/vertices)
- Mean Degree (Mean number of adjacent edges with other vertices for a given vertex - represents the mean number of people an individual is in contact with)
- Mean Partnership Duration (Average amount of time spent with a contact)
- Concurrency (The proportion of nodes with multiple connections i.e. vertices with at least two adjacent vertices)

Once a network has been generated, simulations can be run in order to analyse how our network parameters affect population dynamics. In order to run simulations, additional initial conditions and *infectivity parameters* need to be specified. The objective with specifying infectivity parameters is firstly, to enable us to easily represent the effects of modifying our network parameters and secondly, to make our infectivity parameters as realistic as possible, that is making them as reflective of what they would roughly be in real life. Making our

parameters completely realistic (completely representative of what they would be in real life) is extremely difficult due to the complexities involved in obtaining said values. The aim is to achieve a balance between these two objectives.

Infectivity parameters include:

- Act rate (The mean number of (potentially-transmitting) acts occurring between connected individuals during each time step e.g. sneezing, coughing, kissing, etc..)
- Transmission probability per act (The probability of an infected individual passing on the virus to a susceptible person through an act)
- Recovery rate (The probability of an infected individual recovering per time step)

In addition to infectivity parameters, a number of initial conditions also need specifying for the simulations to run, of which include:

- Starting number of infected (The number of infected individuals at $t = 0$ of our simulations)
- Starting number of recovered (The number of recovered individuals at $t = 0$ of our simulations - important because recovered individuals take-away from the set of susceptible people)
- Number of time steps (Time steps per each simulation - Used to determine the longevity of a simulation. In general, let 1 timestep represent 1 day in real-life)
- Number of simulations (Running an increased number of simulations will improve the reliability of our results but demands more computational power in exchange)

After defining necessary parameters, simulations can be run through either EpiModel's web GUI, where data is automatically formatted into a table and a series of auto-generated graphs are produced, or by using the R-Studio command line, in which simulated data can be stored into a data frame and analysed manually. Using the command line requires additional effort but grants extra freedom when analysing and investigating collated data. For this reason, simulations will be primarily be run through the R-Studio command line.

Parameter	Chosen Value	Justification
Act Rate	9.6	A mean value calculated from data collected investigating the spread of Influenza [2]
Transmission Probability	0.05	Heuristic value derived from literature [1]
Recovery Rate	0.05	Based on the average recovery time of a range of patients from various cultures [3,4]
Infected at $t = 0$	1	Disease transmission typically begins with a single infected individual
Recovered at $t = 0$	0	Recovery not possible until infection transmission has begun
Number of Time Steps	1000	Ideal given that Covid-19 has been around for nearly 2 and a half years now
Number of Simulations	10	A nice balance between managing computational times and maintaining the reliability of our results

Table 1: Table of chosen infectivity parameters and initial conditions.

Table 1 outlines the investigations chosen values for required inputs with a small summary justification. These values, as mentioned in part above, were chosen based on the reading of relevant literature and personal judgement, considering also both of the previously mentioned objectives. Unless specified otherwise, these parameters will be used throughout for all simulations.

Before running simulations it is important to properly define *short-term* and *long-term*. Some preliminary simulations reveal that the peak number of infections always occurs towards the start of a simulation (at the first peak). This makes sense considering that the number of susceptible people is greatest at the start, $t = 0$. Hence define short-term as being the time of peak infections. This way the short term is fairly reflected regardless of any variability in results/parameter choices etc. Define long-term as being the final time, $t = 1000$. This provides a good contrast to our short-term definition and enables us to compare the two without expecting too-similar results (which could happen if our times for short and long-term are too close together).

Finally, to properly classify the outcomes of our simulations we must define what is meant by an *epidemic* in the context of this report.

- **Epidemic -** When an infectious disease has a large prevalence within a community at a given time (a notable proportion of a population have or had have the disease at time t)
- **Short-term Epidemic -** The peak number of infections exceeds 20% of the total nodes n.
$$\max(\text{i.num}) > \frac{n}{5} \implies \text{Short-term epidemic}$$
- **Long-term Epidemic -** The sum of total infected and recovered nodes exceeds the number of remaining susceptible nodes at time $t = 1000$ (the majority of the population has been in contact with the disease)
$$(\text{i.num} + \text{r.num} > \text{s.num: } t = 1000) \implies \text{Long-term epidemic}$$

Given there is no exact definition for what an epidemic is statistically, these definitions are merely estimates based on personal judgement and wider reading.

3 Simulations - Results and Findings

With the required parameters detailed and key definitions outlined, simulations can now be run. To measure the effects of varying our network parameters, run 10 simulations for each set of network parameters and record the number of short/long-term epidemic occurrences based on the above definitions.

Mean.Degree	Mean.Duration	Nodes	Act.Rate	Trans.Prob.	Recovery.Rate	Long.Term.Prop.	Short.Term.Prop.	Peak.Infected	Concurrency
1	10	100	9.6	0.05	0.05	0.7	0.7	45	25%
1	20	100	9.6	0.05	0.05	0.4	0.2	24	25%
1	10	200	9.6	0.05	0.05	0.7	0.7	75	25%

Figure 1: Demonstration of the data collected from simulations.

Fig. 1 demonstrates the recording of some simulation results alongside their respective network parameters and initial conditions. To vary the network parameters, we conduct a sensitivity analysis. It can be difficult to represent the variation of 3 parameters at once, let alone 4, so we fix 2 network parameters and vary the remaining 2 with time.

3.1 Sensitivity Plots - Results

Consider the following ranges for our network parameters:[†]

$0.5 \leq \text{Deg} \leq 1.5$	$10 \leq \text{Dur} \leq 100$	$100 \leq \text{Nodes} \leq 1000$	$0\% \leq \text{Con} \leq 40\%$

These parameter ranges, and increments (see plot axes below), were chosen to best represent the effects of varying two parameters together whilst also managing the computational expense of our simulations.

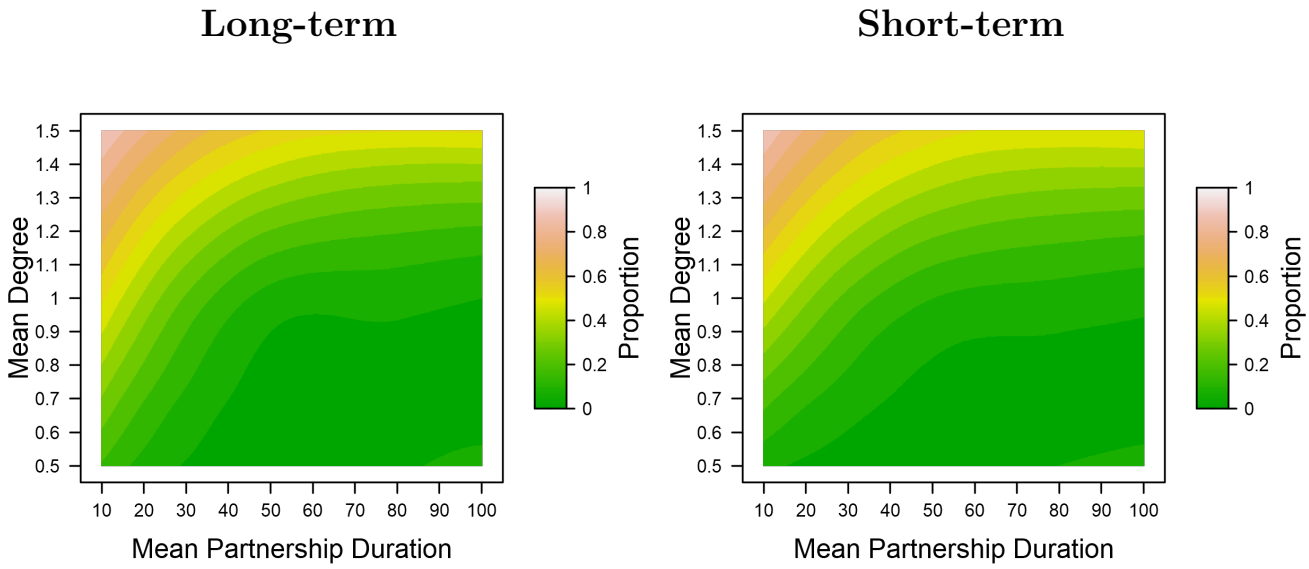

Figure 2: Varying mean degree and mean partnership duration. Concurrency: 0% - Nodes: 100

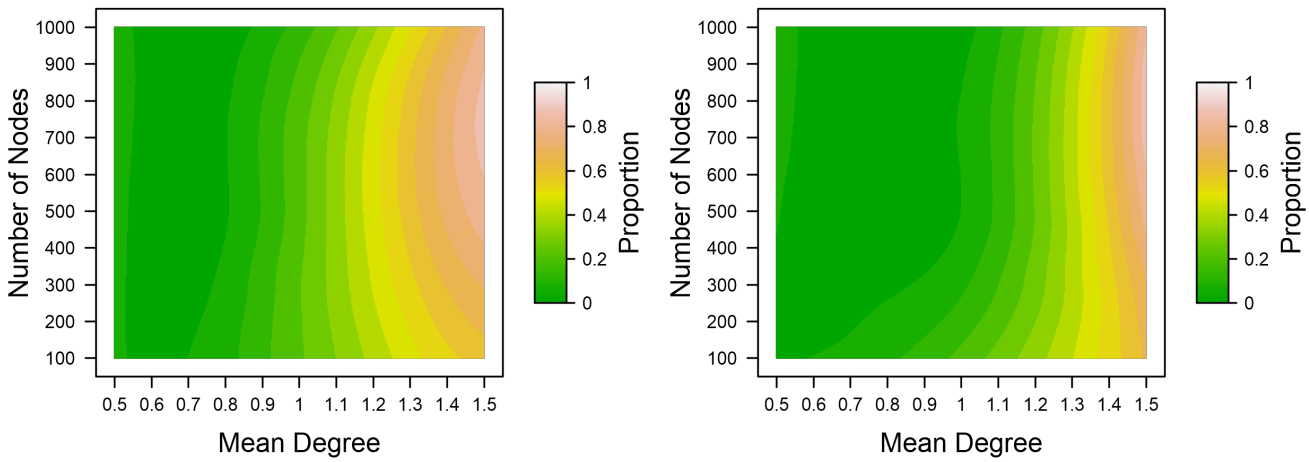

Figure 3: Varying mean degree and number of nodes. Concurrency: 0% - Duration: 35

Now introduce concurrency to our simulations. Introduction of concurrency requires additional computational power due to utilisation of a Markov chain Monte Carlo method, in addition, the effect of introducing concurrency was rather minimal and so we only explore 3 key values from the range (see below: Fig. 4-6).

[†]Mean Degree (Deg), Mean Partnship Duration (Dur), Concurrency (Con).

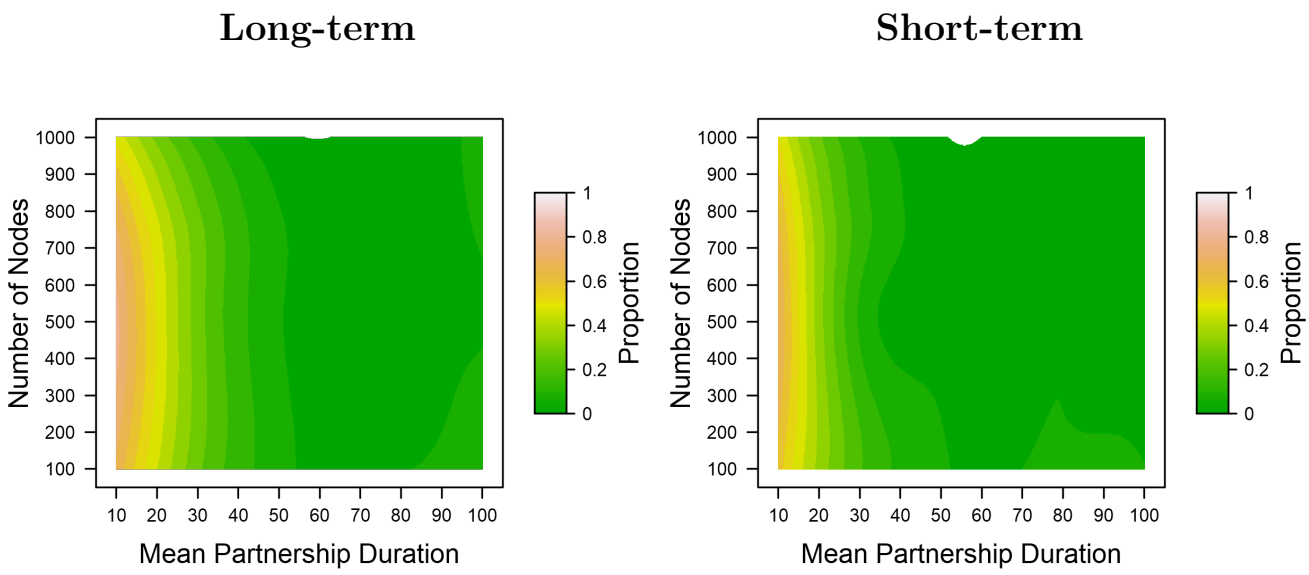

Figure 4: Varying mean partnership duration and number of nodes.
Concurrency: 0% - Mean Degree: 1

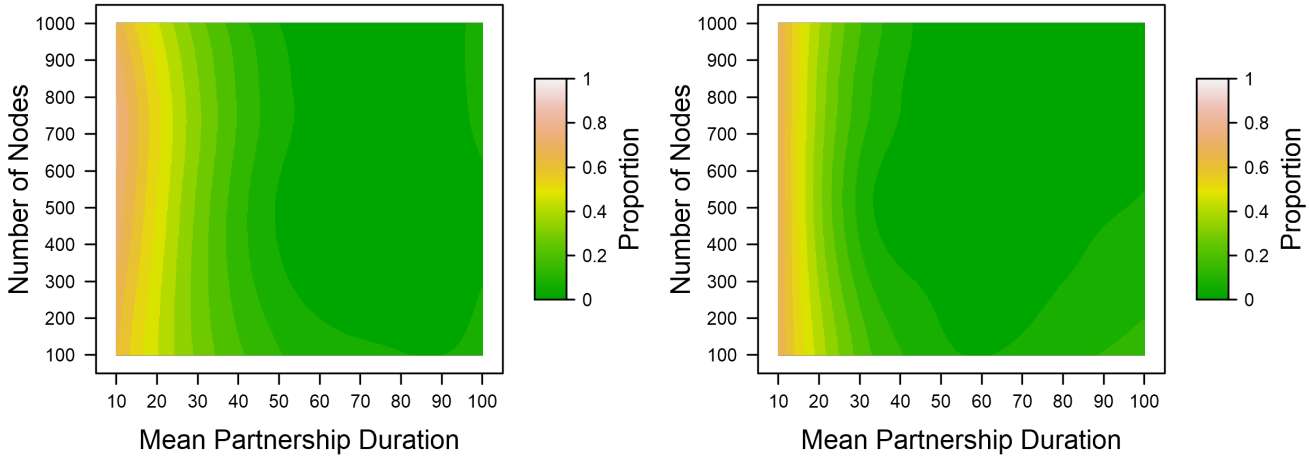

Figure 5: Concurrency: 25% - Mean Degree: 1

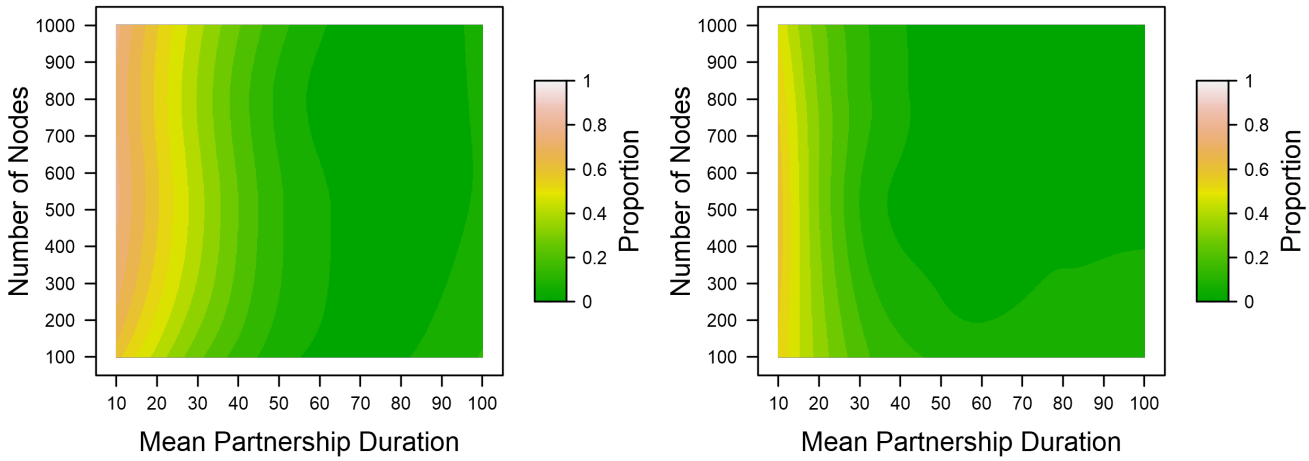

Figure 6: Concurrency: 40% - Mean Degree: 1

3.2 Analysis - Findings

Figures 2 through 6 are the result of thousands of simulations exploring different variations of network parameters. Heatmaps are an ideal way of representing and drawing conclusions from our results. All of the plots reveal that the majority of parameter combinations used result in an epidemic being avoided in both the short and long term. The most common parameters associated with avoiding an epidemic can be seen to be low mean degrees ($\sim 0.5 - 1.0$) and medium-high mean partnership durations ($\sim 40 - 100$) (Fig. 2). We can hence investigate in more detail the effects of a high mean degree and low partnership duration (see Fig. 7 below). The number of nodes appears to have a rather minimal effect on whether or not a short/long-term epidemic will occur, with patterns on Fig. 3-6 being rather linear and varying primarily with the x-axis. Because of this, we arbitrarily choose $n = 1000$ for the next plot:

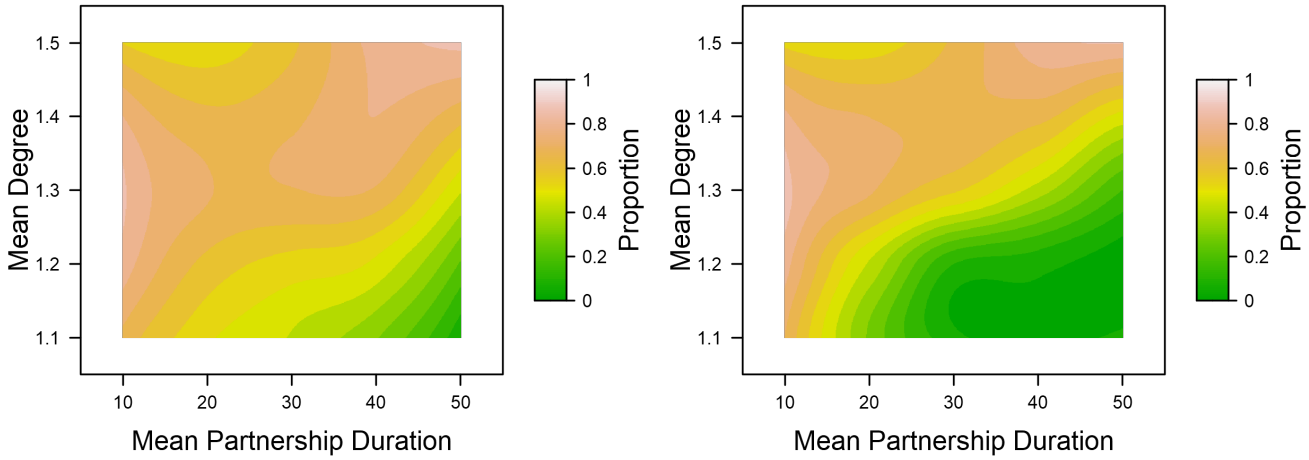

Figure 7: Varying high mean degree and low mean partnership duration. Concurrency: 0% - Nodes: 1000. Long term (left) - Short term (right)

Another major conclusion to draw from our results is that the proportion of recorded epidemic occurrences is relatively similar for both short and long-term cases. Fig. 7 (which highlights the particularly high number of epidemic cases for high mean degree and low mean duration) is testament to this, with the only major difference being that there were less epidemics in the short-term plot (right), for mean degrees ($\sim 1.1 - 1.3$) and for mean durations ($\sim 30 - 50$), than there were in our long-term plot (left). The trends of Fig. 7 support the trends of Fig. 2, once again suggesting that changing the number of nodes has little effect on the outcome of our simulations.

Concurrency, as mentioned above, was explored rather sparsely due to the limited variation in results (as seen in Fig. 4-6) and the computational expense, which comes with EpiModel needing to account for an additional network parameter during network generation and simulations. There are some variations in results however, with most noticeably being that epidemic cases were higher in the higher-concurrency cases (ever-so-slightly), which can be inferred from the exaggerated orange/yellow regions of Fig. 6 when compared to say Fig. 4.

We conclude our analysis with a mini-summary:

- Mean degree and mean partnership duration are the most influential variables in epidemic occurrence.
- High mean degree and low partnership duration sport the highest chance of an epidemic occurring.
- Effects of varying the number of nodes n and concurrency are limited.

4 Extending the Model

To further our our investigation of network parameters and their effects, we extend our model in a number of ways.

4.1 Compartmentalisation

Firstly, compartmentalising the population enables us to assign specific network parameters and infectivity conditions to unique groups of people. In the absence of realism, this can simply be used to compare the effects of varying network parameters. In some pursuit of realism however, we try to reflect a real life situation in which some members of the population did obey the lockdown, social distancing and mask mandates, and others did not. The simplest way of doing this is to assign the disobedient group a higher mean degree, assuming that they will be in contact with more people given tha they are breaking the rules.

With our two compartments (EpiModel best suited for 'dual compartmentalisation' [6]) we run simulations and compare the effects of varied mean degree in real time.

4.2 Time-varying Parameters

EpiModel supports an *updater function* which allows for the updating of infectivity variables at specific times t (whether it be random or specified change). We can use this functionality to improve the realism of our simulations, altering infectivity parameters for both compartments to mirror a situation reflective of what has happened in the UK over the last 2-3 years. Below is an example of some (but not all) key events we incorporate into our model in hopes of improving realism.

Event	Time (t)	Effect
Lockdown + Mask Mandate	100	Reduced Infection Probability & Act Rate
Christmas (2020)	355	Massively Increased Act rate
Vaccines (1, 2 & Booster)	440, 560, 770	Reduced Infection Probability
Christmas (2021)	725	Massively Increased Act Rate
Variants (Delta and Omicron)	660, 760	Massively Increased Infection Probability

Table 2: Example of key events, their respective times and respective effects on the model.

4.3 Extending to SIRS

Our final extension to the model is an additional *susceptible state* which follows from our already defined, recovered state. According to research, reinfections can occur as early as 47 days after initial infection [7], meaning that at this stage, individuals are already susceptible again. There is however quite a lot of variability in results surrounding rates of reinfection and an all round lack of data. We hence define a recovered-to-susceptible rate based on a more general benchmark: Immunity after initial infection lasts for 90 days $\implies$ r.s.rate $= \frac{1}{90}$

Following these extensions to our investigative model, we want to once again investigate the effects of varying network parameters. As per our previous model's findings, choose the number of nodes and concurrency arbitrarily ($n = 1000$ and 0% respectively), instead we focus on the effects of changing mean degree and mean partnership duration. A visualisation of simulation flows (changes to SIRS populations) is a suitable way of comparing the effects of varying our network parameters but also a great way to highlight the effect our extensions have on the model.

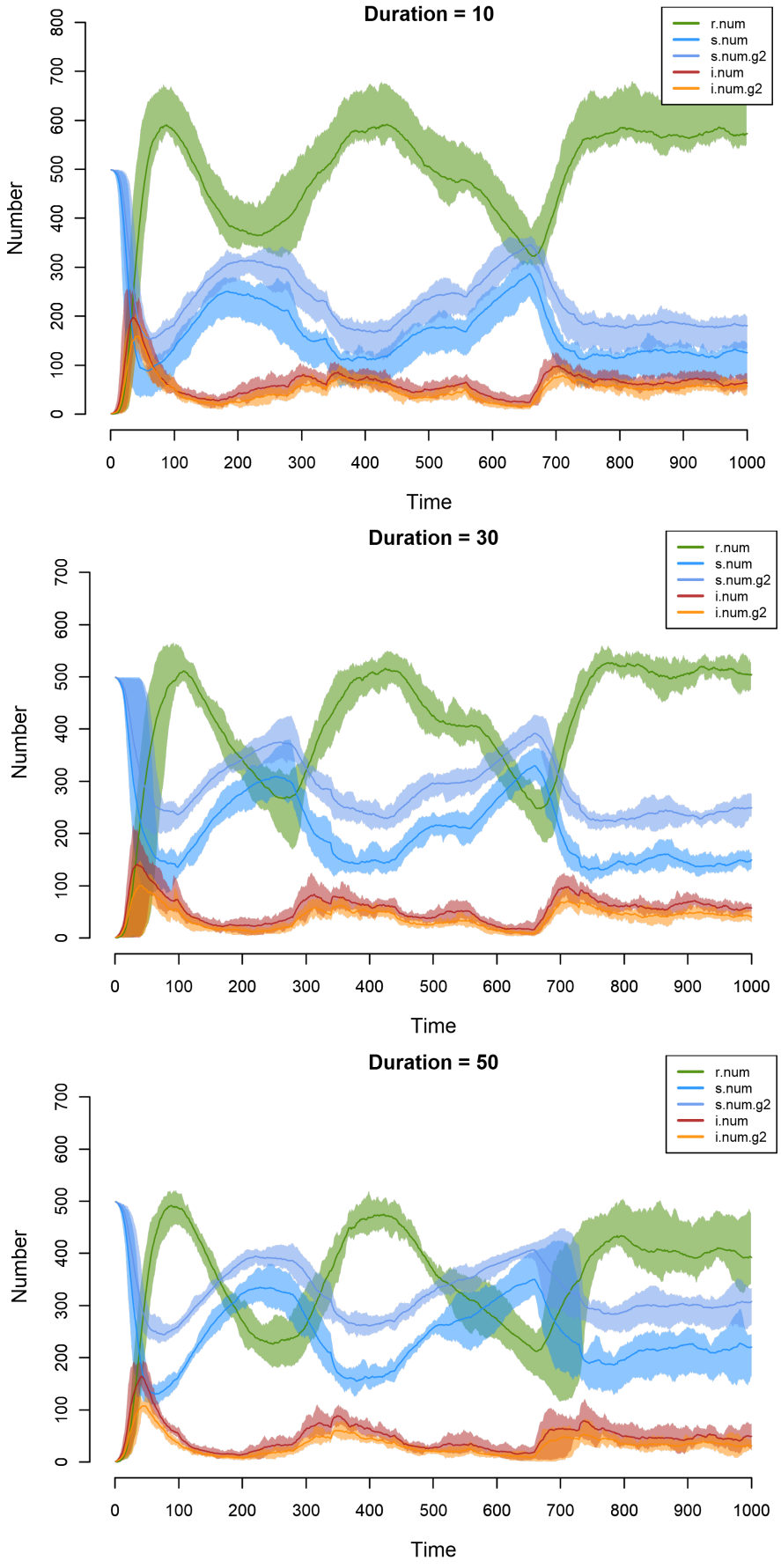

Figure 8: Plotting SIRS populations with time. Group 1 represents rule breakers (Mean Degree: 1.5) and group 2 (.g2) represents the contrary (Mean Degree: 1)

Figure 8 incorporates our previously discussed extensions into the model and demonstrates how varying mean degree and mean partnership duration influences the outcome of epidemic simulations. The effects of varying mean degree are demonstrated through the phase difference between the rule-breaker and rule-follower's susceptible populations (s.num & s.num.g2). Increasing mean degree resulted in an almost-linear increase in infections and hence a reduction in susceptible peoples. This overall trend is in line with figures 2, 3 and 7. Additionally, increasing the mean partnership duration had an overall positive effect, decreasing infections all around and hence less people were affected by the disease spread. Such a result reinforces our previous findings as seen in figures 3-7.

We can expect variation in concurrency to minimally affect the general dynamics seen above, same with varying 'number of nodes', it is for this reason that they were chosen arbitrarily. It can be expected that mean degrees less than 1 and mean partnership durations greater than 50 will frequently avoid an epidemic from arising in both the short and long term, as per our previous sections results.

5 Discussion

From both our basic and extended models, there are conclusions to be drawn and evidence-supported advice to be taken away in regards to avoiding future pandemic/epidemic situations.

We first note that varying the number of nodes had little effect on epidemic simulations, from this it can be said that epidemic situations, in general, unfold similarly regardless of population size. We note that this claim is based on over-simplified population dynamics, a limitation of our model. In reality, population dynamics are considerably more complex and complexity is expected to increase with population size. Future models can improve on this lack of realism by including the following:

- The addition of an *exposed state* (SIRS $\longrightarrow$ SEIRS) which describes individuals who have come into contact with the disease but are not currently symptomatic/infectious.
- The addition of arrival and departure rates to the population (People entering and leaving the population).
- Additional compartments to facilitate differences in age, vulnerability and attitude towards the virus.

The effects of varying concurrency were also limited with an increase in concurrency barely resulting in a higher number of infections. It is possible however that the effects of concurrency were underappreciated and so would be my advice to aim for reduced concurrency within a population where possible (keeping social bubbles small), despite the limited effect it had on our model's simulations.

The effects of varying mean degree and mean partnership duration were very clearly demonstrated and we concluded from our findings that a high mean degree (greater than 1) and a low mean partnership duration (less than 40) resulted in the worst cases of epidemic spread, both short and long term. This was demonstrated particularly in our extended model (Fig. 8) where a higher mean degree resulted in considerably more infections and a low partnership duration affected both compartments of people, equally as badly. From this we conclude that lowering mean degree and increasing mean partnership duration is definitely desirable for planners and the management of future pandemic/epidemic situations. This is evidence for the effectiveness of epidemic interventions such as self-isolating, bubbling and shielding, which are designed to lower an individuals mean degree and increase mean partnership duration. The effectiveness of vaccines and mask mandates were also loosely demonstrated through our extended model, with the introduction of such resulting in reduced infection numbers and a boost to recovered and susceptible populations.

The dangers of new variants such as Omicron and Delta was also emphasised, with both the rule-follower and rule-breaker groups experiencing a major increase in infections, despite the vaccination program. The rule-following populations remained considerably better off however, despite the introduction of variants which further stresses the importance behind concepts such as sheilding, self-isolating and social-distancing.

Conclusion of major points:

- In general, epidemic situations can arise in any population regardless of size/scale, given the right conditions.
- Individuals should aim to reduce their concurrency and having concurrent partners should be discouraged during a disease-spread situation.
- Response measures such as social distancing, self isolating, bubbling and shielding, in general, are effective and should be employed to manage disease spread within a population. Individuals are encouraged to follow these measures in order to reduce the population-wide spread of disease.
- Vaccines and mask mandates which reduce the probability of infection are effective and should be employed during pandemic/epidemic situations
- Highly infectious variants of a disease pose a major threat to populations, further stressing the importance of lowering one's mean degree and increasing mean partnership duration through following measures such as social distancing, self isolating etc.

References

[1] Churches T, Jorm L. Flexible, Freely Available Stochastic Individual Contact Model for Exploring COVID-19 Intervention and Control Strategies: Development and Simulation. JMIR Public Health Surveillance. 2020; 6(3): e18965.

[2] Eames KTD, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. PLOS Computational Biology 2012; 8(3): e1002425.

[3] Voinsky I, Baristaite G, Gurwitz D. Effects of age and sex on recovery from COVID-19: Analysis of 5769 Israeli patients. The Journal of Infection. 2020; 81(2): e102-e103.

[4] Costantino V, Heslop DJ, MacIntyre CR. The effectiveness of full and partial travel bans against COVID-19 spread in Australia for travellers from China during and after the epidemic peak in China. Journal of Travel Medicine. 2020; 27(5): taaa081.

[5] Dan JM, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. Science. 2021; 371(6529): eabf4063.

[6] Jenness M, et al. EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks. Journal of statistical software vol. 2018; 84(8).

[7] Rahman S, et al. COVID-19 reinfections among naturally infected and vaccinated individuals. Science. 2022; 1438(12).