MDM2 Group Project 2: Infections on Networks, Group 12

Network Based Epidemic Modelling

Kieran Clarke
hr19845@bristol.ac.uk

Theo Le Vay
xh19142@bristol.ac.uk

Yuyang Liu
gf19229@bristol.ac.uk

George Purvis
gp17954@bristol.ac.uk

Christopher White
gd19031@bristol.ac.uk

March 11, 2021

Abstract

Studying the spread of infectious diseases using network-based modelling extends the standard SIR method, which assumes the population to be uniform and well-mixed and therefore cannot assess the response of an epidemic to changes in population density or structure. Such network-based models are therefore more applicable to real-world scenarios, and provide epidemiologists with useful insights into more complex disease behaviours. This report demonstrates that simulating disease spread in homogeneous networks can produce similar results and is therefore related to traditional SIR models. It then determines that non-homogeneous networks differ to homogeneous networks, and how varying structure and properties of such networks further influences the spread. Network clustering and the cross-links between them are found to be key factor in disease propagation, and may be related to the global spread of diseases such as COVID-19.

1 Introduction

The current COVID-19 pandemic has brought the risk of infectious diseases to the centre of public attention. Now more than ever there is a global emphasis on researching the spread of diseases and investing in methods to prevent and mitigate future disease outbreaks. Modelling diseases is a crucial aspect of such research as it can be used to project potential future disease outbreaks or predict the risk of a new outbreak evolving into an epidemic. The original epidemiology studies showed disease spread can be modelled for populations using differential equations, however similar but more complex models provide epidemiologists with a deeper understanding of their behaviour.

The emphasis of this paper is on modelling the relationship between human networks and disease spread; diseases spread from human to human and so can be represented as spreading through human networks. An advantage of using network based modelling is that it can be used to represent different human networks of varying sizes and with varying properties, which is how real human networks exist. Human networks are especially complex and so they cannot be well described with any one network type.

2 Compartmental Epidemic Modelling and SIR

Epidemiologists model the progress of an epidemic through a compartmentalised mathematical model, most commonly realised as a system of ordinary differential equations.The classic model, produced by Kermack and McKendrick in 1927 [1], models the epidemic as the population exposed to the disease experiencing three successive states, which are “Susceptible”, “Infectious”, and “Removed” (SIR). However, some varieties of the model will change the states of population, including “Susceptible,” “Infectious” and “Susceptible" (SIS), and “Susceptible”, “Infectious”, “Removed”, and “Dead" (SIRD) [2].

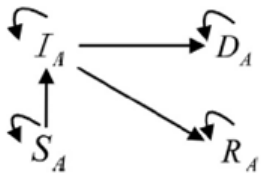

Figure 1: *The state transition diagram of one individual among the states Susceptible, Infectious, Removed and Deceased of a basic, compartmentalised epidemic model.*

Figure 1 shows the alteration of the basic SIR model introducing the “Dead” state to the model. However, the mechanism of how the members of the population transfer from their original susceptible state to others remain the same. Individuals in the population either stay in their current physical condition or perform a one-way movement, such as from susceptible to infected or infected to dead or removed. In this example, removed population members are considered immune forever, and the deceased will not transmit after death.

The SIR model captures two key characteristics of an epidemic. The first is the development of an infectious population due to social interactions between infected and susceptible individuals. The latter is the termination of the epidemic, caused by the “exhaustion of susceptible

members" [1] of the population. The epidemic may also be terminated once all the infected individuals are either removed or deceased.

The basic compartmentalised SIR model of disease transmission can be encoded as a series of ordinary differential equations that describe the rate of change of the different states that individuals in the population can inhabit.

$$\begin{cases} \dfrac{\mathrm{d}S}{\mathrm{d}t} = -\dfrac{\beta SI}{N}, \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \dfrac{\beta SI}{N} - \gamma I, \\ \dfrac{\mathrm{d}R}{\mathrm{d}t} = \gamma I. \end{cases} \tag{1}$$

$$\begin{cases} \dfrac{\mathrm{d}S}{\mathrm{d}t} = -\dfrac{\beta SI}{N}, \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \dfrac{\beta SI}{N} - \gamma I - \mu I, \\ \dfrac{\mathrm{d}R}{\mathrm{d}t} = \gamma I, \\ \dfrac{\mathrm{d}D}{\mathrm{d}t} = \mu I. \end{cases} \tag{2}$$

Equation (1) describes the spread of an epidemic throughout a population of fixed size N, where $N = S(t) + I(t) + R(t)$. The number of individuals in the states "Susceptible," "Infectious," or "Removed" change, with respect to time, according to β and γ. $\beta = ab$ where a is the probability of transmission (per contact between a susceptible and infected individual) and b is the number of contacts with another individual that an infected individual has per unit time. γ is the reciprocal of the infectious period, which is the average amount of time that an individual is infectious after themselves becoming infected.

Equation 2 shows the SIRD variation of the model [2], including $D(t)$ for deceased. In this case, $N = S(t) + I(t) + R(t) + D(t)$ and the parameter μ is the mortality rate of infected individuals.

The overall rate at which an epidemic grows can be expressed through the R_0 number, known as the basic reproduction number. This can be calculated as shown in equation 3, where

$$R_0 = \frac{\beta}{\gamma}. \tag{3}$$

However, the SIR model (and its derivatives) suffer from an important limitation. The model assumes that the population has a uniform density and is well mixed. This means that every individual is just as likely to encounter a certain infected individual as anyone else. As stated by Craig et al [3], 'social interactions are not organised in this stylised way' with individuals in a community often forming groups or clusters and interacting with mostly the same people each day.

In order to capture the variations in the spread of a disease caused by the structure of a population, modern epidemiologists introduced network models of disease transmission, providing measures of social cohesion of individuals [4].

3 Introduction to Epidemic Modelling Using Networks

Networks can be used as models to describe a population of interest as a community of socially linked individuals. The individuals are represented by nodes and mutual links between them are represented by the edges. This model of a community can be applied to epidemic modelling by endowing each individual (node) with a state from the SIR model or its derivatives. Rather than infected individuals being able to contact anyone in the population, individuals can now only contact those with which they share an edge.

Figure 2 shows how a simple network can capture the variations in population density within a community. It would be expected that the spread of the disease may be accelerated if it starts from node A rather than node H. Since the former may interact with four others, it exposes two more individuals to potential infection than the latter, H.

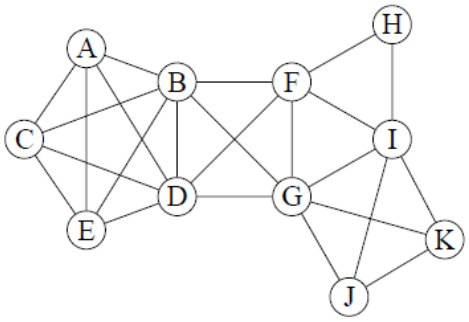

Figure 2: *Example network with 11 nodes, showing potential variations in how connected individuals within a population are.*

Any network can be described through the means of an adjacency matrix. In an adjacency matrix A, elements $A_{ij} = A_{ji} = 1$ if exists a link from node i to j and $A_{ij} = A_{ji} = 0$ if otherwise. For our use of the networks, a disease may spread either way along an edge, simulating human-to-human transfer, therefore our networks will be undirected and symmetric. However, cases such as transfer via contaminated blood or non-direct contacts through an intermediate object (such as vector transmission) also exist for certain pathogens.

This matrix can be used in the calculation of graph properties that will help to show and explain the variation that can arise from a model with a network. For example, the local clustering coefficient for a node i can be calculated as

$$C_i = \frac{\sum_{j,k} A_{ij} A_{jk} A_{ki}}{k_i(k_i - 1)} \quad \text{where} \quad k_i = \sum_{j,k} A_{ij} \tag{4}$$

and A is the adjacency matrix of a network [5]. The local clustering coefficient quantifies whether or not a node and its neighbours are close to being a clique. A clique is a subset of a graph such that every pair of nodes in the clique are joined by an edge. A clique could represent a household of closely interacting family members or, perhaps, a work environment.

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i \tag{5}$$

A network average clustering coefficient can be obtained using equation 5, where n is the number of nodes [6].

In nearly all cases populations are distributed in clusters, within these clusters nodes are highly interconnected. The clusters are connected but not to the same degree as the interconnections within the clusters. This property is true at most scales of population. For example individuals cluster to form "bubbles", families and friend groups. families cluster to form towns or communities, which when clustered form countries.

Using an epidemic model with a network comes at a cost. The simulations are much harder to run promptly and the the results will always be limited by the fact that creating a truly representative graph for a community of individuals is likely impossible [3]. Therefore all the networks we will be working with can only approximate the real structure of a community.

However, by applying a network approach to epidemic modelling, we can see how factors such as heterogeneity and clustering can effect the course of an epidemic. Furthermore, by creating an individual-based model we are also able to monitor quantities such as the number of nodes a certain node infects. This could provide useful insight into so called 'super-spreaders'.

In order to model contagion spread across a network, we must generate networks to represent population distributions. Creating one network that is structurally similar in all properties to the natural population distribution in a given location is nearly impossible. However, we can generate networks to have some structural similarities to population distribution, then study specifically the properties within that network that are accurate to population distributions.

4 Initial Modelling with Homogeneous Networks

For our initial modelling we will be using homogeneous networks. These networks contain a node for each individual in the population and each node will have the same number of edges, all with randomly assigned destinations.

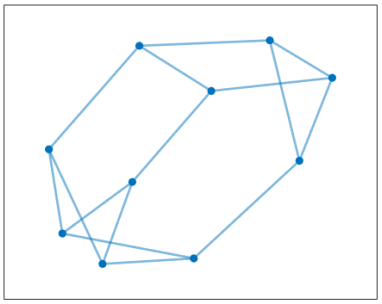

Figure 3: *Example homogeneous network containing 20 nodes, each with 4 randomly assigned edges*

Transferring the SIR model of an epidemic to a homogeneous network is the very beginning of developing a model with a network since it provides a baseline for comparisons. The key limitation of the standard SIR model is that the population is assumed uniform and well mixed and these properties are mirrored in the homogeneous network, since all the individuals have the same number of connections (uniformity) and the random nature of these connections ensure that the network will be well-mixed [3].

Figure 3 shows a small homogeneous network. In order to

use this network to model an epidemic, each node must have a state (described in figure 1). Most nodes start as susceptible but a selection of nodes start as infected to begin the epidemic. Using the same parameters described for equation 2, we can simulate the state transitions for each node in the network using the algorithm described in figure 4.

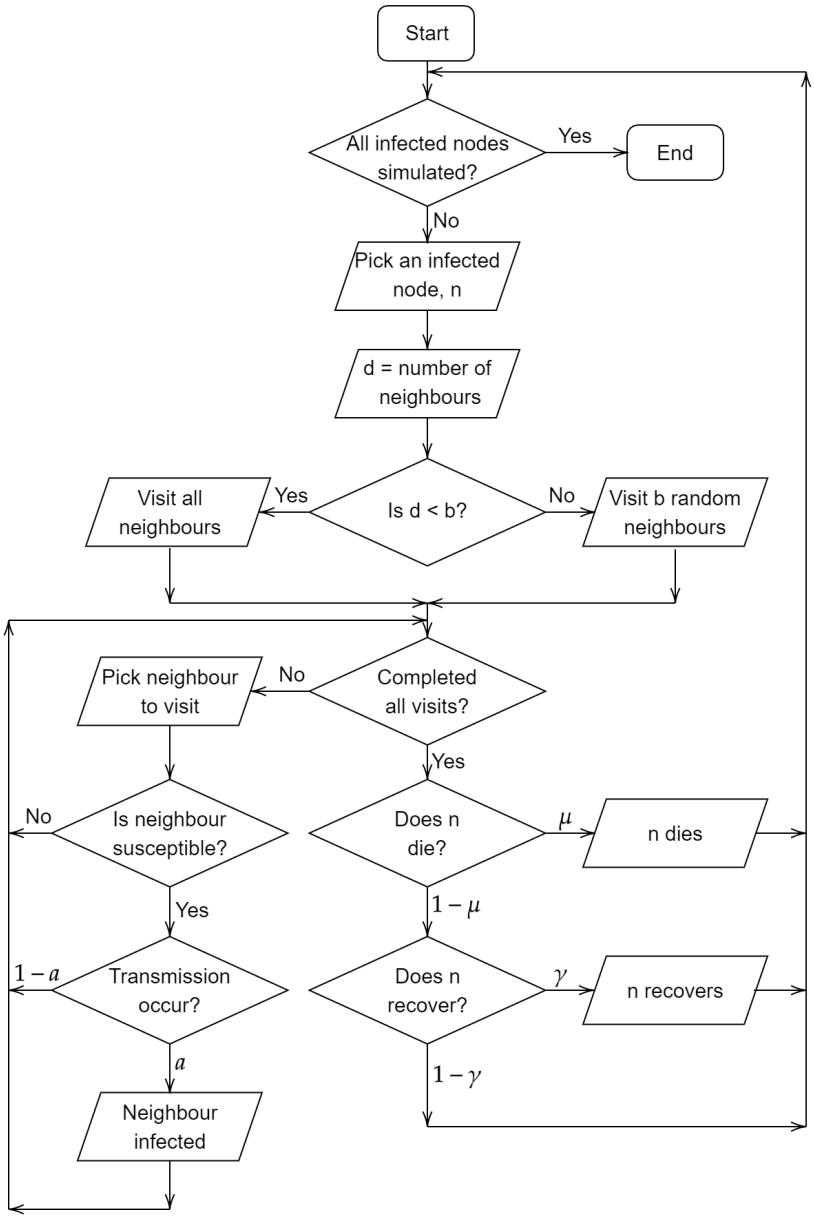

Figure 4: *Flowchart displaying the algorithm used by the model for each day that the simulation is run over. The algorithm loops through all the infected nodes in the network and, using the parameters described in equation 2, determines any further infections, deaths and recoveries.*

A key difference when moving from a differential equation approach to a network approach is that a node may only contact those it shares an edge with. This property interacts with the number of contacts per day parameter (b) since if $b < degree(n)$, the node n will only make

contact with a random selection of b of its neighbouring nodes. Furthermore, if $b > degree(n)$, the node n will make contact with all of its neighbours but this number of contacts will be smaller than the effective number of contacts simulated by the standard SIR model. For this initial stage in the modelling, we will ensure that $b < degree(n)$ for all nodes to produce results more comparable with the standard SIR model.

It is apparent, for the network described in figure 3, that the size of the population is tiny, with individuals connected to 1/5 of the entire population. Such a low population with such a high number of connections would not be representative of a real community. Furthermore, running our algorithm on a network this small would provide limited results which would have a high level of variance due to the probabilistic nature of the model. Therefore, the homogeneous structure shown in figure 3 can be scaled up to a network of 100,000 nodes, each with 50 randomly assigned connections. Using a network of this size will provide much more consistent results and, as stated in the literature [3], this number of connections being randomly assigned is comparable to the uniform and well-mixed approach of the standard SIR model.

30 of the 100,000 individuals are randomly chosen to be initially infected and then the algorithm described in figure 4 is run over a period of 100 days. The results are shown in figure 5.

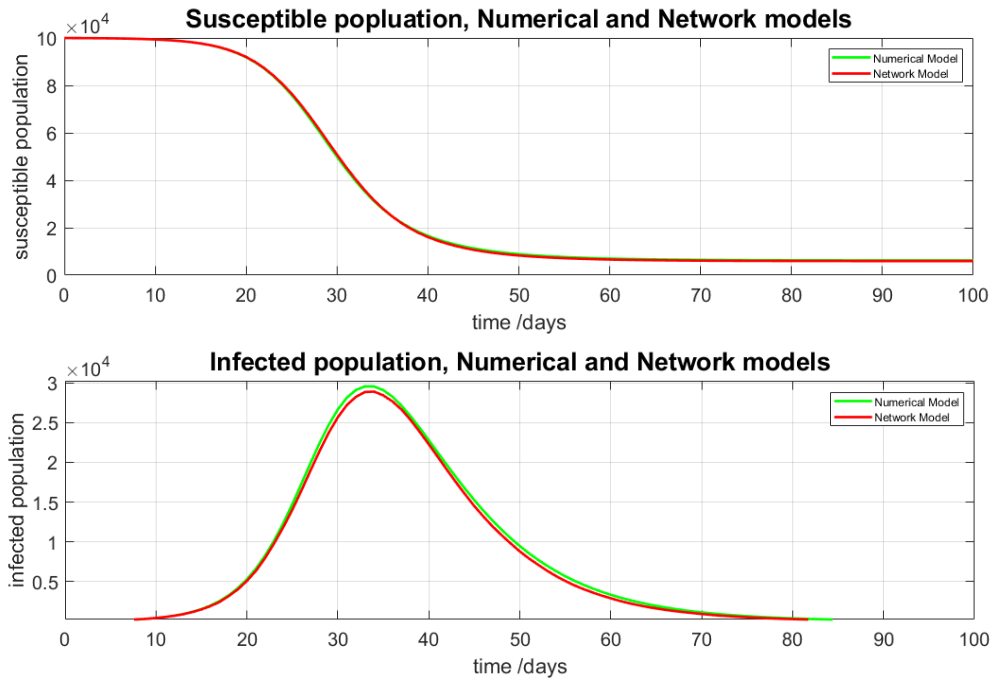

Figure 5: *Number of susceptible and infected people in the population against time, calculated by the standard ODE model (Numerical Model) and our model on a homogeneous network (Network Model).* $N = 100,000$, $a = 0.1$, $b = 4$, $\gamma = 0.125$ *and* $\mu = 0.01$ *for both models.*

In both graphs of figure 5, the simulated trends of change have identical characteristics. In the graph for the susceptible population, both models show an identical curve, decreasing until until the 60^{th} day before leveling off. Both models predict approximately $5,000$ people

never get infected.

In the infections graph, the peak, although slightly different in height, happened on about the 33^{rd} day after the starting point. Furthermore, both models introduced the termination of the epidemic after around 80 days, when no new infection happens.

The similarity between the curves shown in figure 5 shows that our algorithm on a homogeneous network produces results very close to the standard ODE model. This will provide us with a basis for further comparison as new network types are introduced that are not homogeneous or well-mixed. Therefore, in the next section, a variety of non-homogeneous network types will be introduced before epidemics will be modelling using them.

5 Non-Homogeneous Networks

From the basic non-homogeneous network, more complex network structures are developed by altering the data contained in the adjacency matrix. Additionally, changing the construction process of the network, such as the scale-free network in subsection 5.3, can also generate very different results.

Therefore, the following types of non-homogeneous networks are chosen to identify the varying characteristics of population in epidemic model.

5.1 Random Networks

For random graphs, the set of edges between the nodes is randomly assigned. Our model's random network, visualised in figure 6, is constructed using Steger and Wormald's algorithm [7], which generates a network with a binomial distribution of its nodes' degrees. This functions by setting the number of nodes, n, and the probability of edge creation, p. Some of the first network based epidemiology papers were based on random networks, however their usefulness in modelling the spread of diseases is limited since human networks tend to be more complex, and there have been many papers expanding upon them.

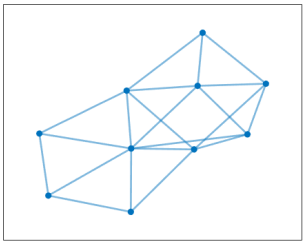

Figure 6: *Network of 10 binomially distributed degree nodes.*

5.2 Small-World Networks

The construction of a small-world network, also known as a Watts-Strogatz graph [6], begins with a ring network of n nodes, and then each node is connected to its k nearest neighbouring nodes. Cross-links are then implemented by applying a probability, p, of each of a node's connections being replaced with an edge that connects it to any random node in the network. This p value is also known as the rewiring probability. This network is useful to model the significance of cross-links, which exist in human networks, in altering the spread of diseases.

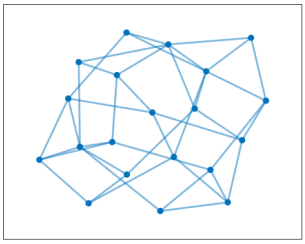

Figure 7: *20 node small-world network.*

5.3 Scale-Free Networks

Scale-free networks are a type of random graph whose construction is based upon the Barabási-Albert preferential attachment model, which produces a network with a power-law degree distribution [8]. The network of n nodes is created by preferentially connecting new nodes, of degree m, to high degree nodes, where the probability, Π, of a new node being connected to a node i is given by

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}. \tag{6}$$

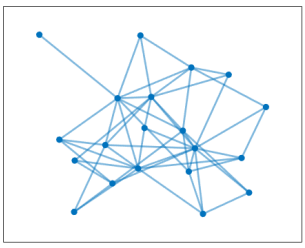

Figure 8: *20 node scale-free network.*

Real human networks possess similar properties to scale-free networks, as they are constantly 'growing', and newly introduced people are more likely to connect to more highly connected people. These network properties are an advancement on the random network model and make it useful in modelling real-world scenarios.

5.4 Clustered Networks

To further analyse the effect of clustering on epidemics, we will be generating our own graphs, referred to as 'clustered networks'. This includes our standard clustered networks and clustered networks with central clusters with more connections.

We generate clusters in a stochastic way, to capture the differing size and inter-connectivity of different clusters. Clusters are generated with a varying number of nodes but the same average population throughout the network. This excludes any central clusters that have an optional base population multiplier. Once clusters have been generated they can be interconnected to produce the intended network.

Standard clustered graphs are generated by adding random interconnections between the clusters. They all have the same base number of connections and random variation is optional. In a real world setting, these graphs could represent a society that is split into 'bubbles' where the cluster population is the allowed bubble size and the number of interconnections between the clusters is a function of both the willingness to follow social distancing rules and stick to interactions within clusters across the community. These can be changed to show that the interconnections between bubbles significantly affect the rate of spread throughout a network. This is parallels the findings of [9].

We also generated a structurally different type of clustered network. This network uses the same cluster generation but generates 1 or more central cluster with increased population and guaranteed connections to every outer cluster. Interconnections between clusters are optional but included by default. This is an attempt to represent the friendship paradox [10]. The friendship paradox is the observed mathematical phenomenon in social networks that, connections between nodes are not equally distributed, and a subset of less than half of the nodes have the majority of the connections. This is important because these more connected nodes will have a larger effect on the spread of the virus.

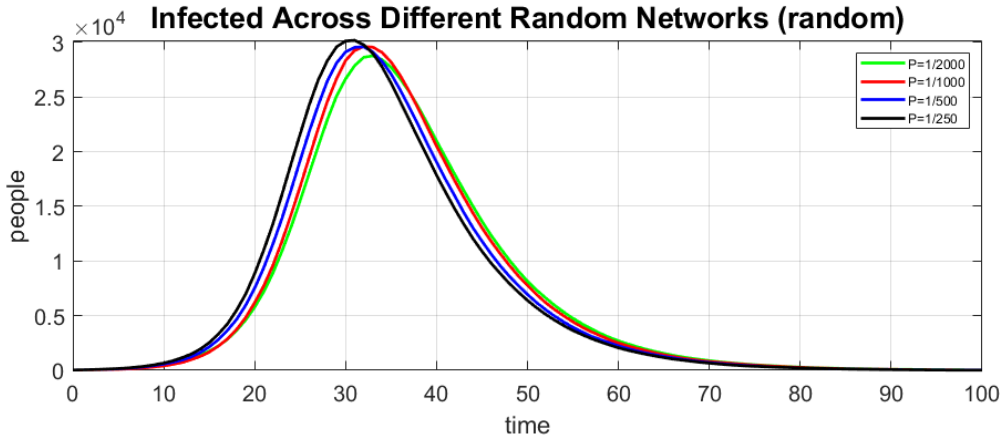

Figure 9: *Graphs showing the number of infected individuals on 4 random networks, each with a population of 100,000 and each with a different probability of edge creation, against time in days. The disease parameters are $a = 0.01$, $b = 40$, $\mu = 0.01$, $\gamma = 0.125$ and there are 30 initially infected individuals. These initially infected are, for the top graph, randomly placed, for the middle graph, placed on the 30 highest degree nodes and for the bottom graph, placed on the 30 lowest degree nodes.*

To compare these networks, the ratio of nodes to connections has been kept very similar, differing between generations but averaging about 0.001 difference. This as well as other connectivity metrics were kept the same. meaning, the significant changes in contagion behaviour and spread is due to the structural changes.

6 Results

By applying the algorithm described in figure 4 to these new, non-homogeneous network types, we can analyse the effects of relaxing the constraints of well-mixing and uniformity present in the standard compartmental model. Furthermore, we can analyse the impact of varying the parameters used to generate the networks to see how a disease would respond.

6.1 Epidemics on Random Networks

The results shown in figure 9 show that as the probability of edge creation decreases the peak of the infections gets delayed and its height is reduced. This behaviour is expected since as the probability of edge creation decreases the network becomes less connected, giving the disease less paths on which to spread. In fact, for random networks, as the population size $N \to \infty$, the probability of edge creation, $p \to \bar{C}$, the network average clustering coefficient [11]. Therefore, networks with a higher level of clustering will increase the speed of the spread of a disease and increase the peak number of infections.

The graphs in figure 16 in appendix A.3 show the variation that arises when the initially infected individuals are more or less connected. When the nodes of highest degree are infected first the peak number of infections is raised and it occurs earlier in the simulation and when the nodes of least degree are infected first, the peak occurs later and is not as high.

6.2 Epidemics on Small-World Networks

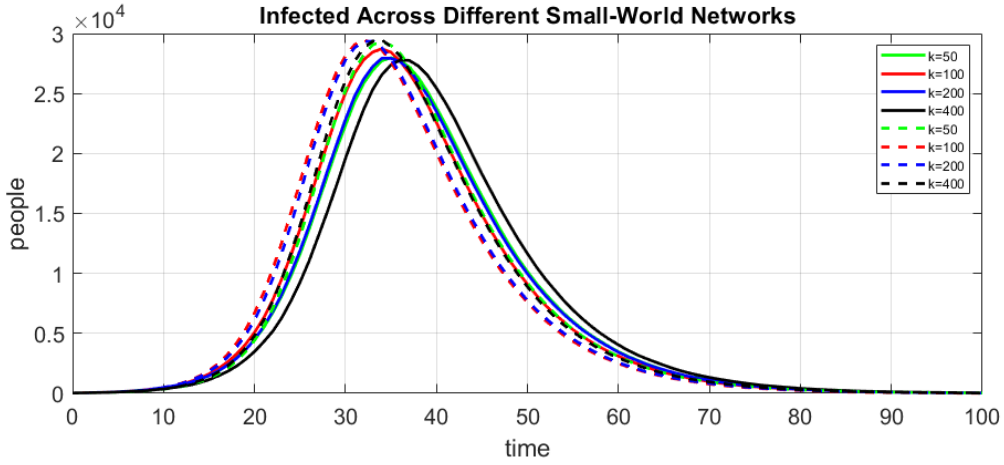

Figure 10: *Graph showing the number of infected individuals on 8 small-world networks, each with a population of 100,000, against time in days. The disease parameters are* $a = 0.01$, $b = 40$, $\mu = 0.01$, $\gamma = 0.125$ *and there are 30 initially infected individuals. These initially infected are randomly placed. The dashed lines show the networks generated with an edge rewiring probability of 0.75 and the solid lines show networks with an edge rewiring probability of 0.25. K is the number of neighbours used to form each initial cluster.*

Figure 10 shows, firstly that increasing the probability of edge rewiring will increase the peak height of infections and make it occur sooner. Increasing this value will increase the number of connections between the different 'small worlds' present in the network, therefore, making it easier for a disease to transmit between clusters of people.

For the case with lots of rewiring (the dashed lines), the networks with cluster sizes of 100 and 200 hundred individuals produce similar results as do the networks of cluster sizes of 50 and 400. This goes against the pattern demonstrated by the random networks, as a greater amount of clustering does not necessarily result in faster transmission. An explanation for this inconsistency may be that, for a small cluster, the high rewiring probability breaks down the clustered structure enough so that this trend cannot be observed. For the case with clusters of many individuals, the larger cluster size may be reducing the effect that it has on the spread since the larger the clusters get, each individual will become less isolated from many other individuals, resulting in smaller effective clusters.

Upon calculating the network average clustering coefficients for these networks we observe a different trend to that shown in the random networks. The small-world networks with a higher rewiring probability have lower clustering coefficients yet the disease spread faster and infects more people. This suggests that the inter-cluster connections are pivotal in the spread of the disease.

6.3 Epidemics on Scale-Free Networks

Figure 11 shows the variation on the spread of a disease across different scale-free networks. It can be observed that, generally, the networks constructed with a greater value for m make the disease spread faster and with a higher peak number of infections.

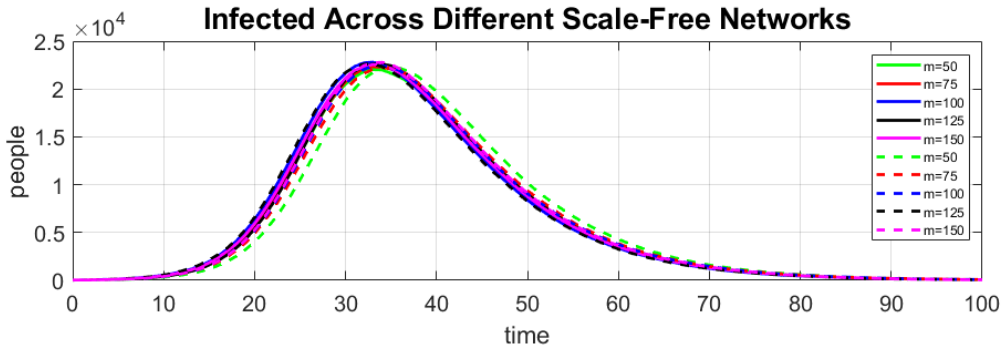

Figure 11: *Graph showing the number of infected individuals on 10 scale-free networks, each with a population of 100,000, against time in days. The disease parameters are* $a = 0.01$, $b = 40$, $\mu = 0.01$, $\gamma = 0.125$ *and there are 30 initially infected individuals. These initially infected are randomly placed.* m *is the number of edges to be preferentially attached, each time a node is added. The dashed lines represent networks where the 100 most connected (highest degree) nodes have been removed.*

Furthermore, it can be seen that removing the 100 most connected nodes will delay the peak of infections by around a day or two. Somewhat unexpectedly, removing these nodes did not have a consistent effect on the peak number of infections. For some networks the peak is higher when the nodes are removed and for some the peak is lower. This may be due to variations that arise from the random assignment of the initially infected or just the stochastic nature of the model.

6.4 Epidemics on Our Clustered Networks

From figures 12 and 13 you can see that, despite identical disease parameters, similar connectivity metrics and population, the network with a central cluster is much more vulnerable. This is due to the many connections from the central cluster to each other cluster. This results in any node being only 4-6 connections away from any other node meaning contagion spreads faster and more reliably. Whereas, in the standard clustered network, if the disease does not spread outside its starting cluster quickly, the ratio of susceptible to non-susceptible people in said cluster will fall until the disease can no longer spread. The top sequence of figure 12 shows an epidemic with the same initial conditions as figure 13, however, it only manages to spread to 1 cluster outside the starting cluster before the disease dies out.

These results imply that the number of connections between clusters is instrumental in the propagation of a disease. To test this, we ran our simulation on a standard clustered network with a higher average inter-cluster connectivity of 7 (average 14 connections per cluster). This is significantly higher than the centralised network, which had 2 connections to randomly chosen clusters per cluster plus 2 connections to the central cluster(average 6 connections per cluster). Despite this, the epidemic still spread much faster in the centralised network. This implies that the destination of a connection is relevant as well as the frequency of them. For example a connection to a isolated node or cluster will do little but infect that cluster but a connection to a well connected cluster can significantly accelerate spread. implying a better predictor for network venerability, be a sum of edges weighted by there centrality or "significance". These finding could be used to model which connections need to be removed

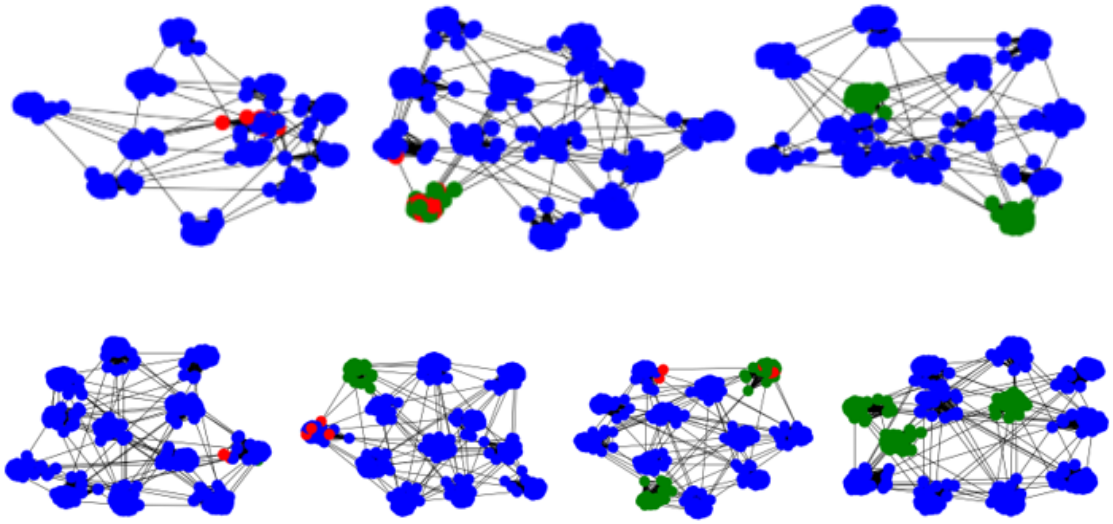

Figure 12: *Sequence of snapshots showing an epidemic in two standard clustered networks. The top sequence shows an epidemic in a network with average inter-cluster connectivity of 2, the bottom sequence shows an epidemic in a network with an increased average connectivity of 7, between the clusters. Blue nodes represent susceptible individuals with red representing infected and green representing removed. Further information on the visualisations can be found in appendix A.1.*

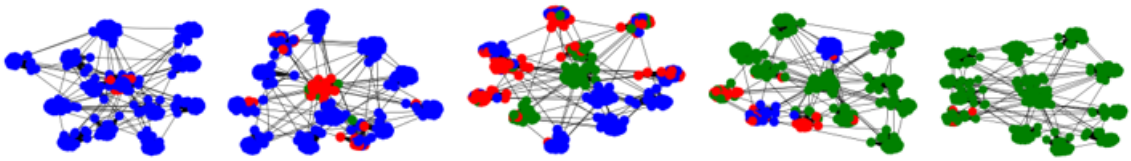

Figure 13: *Sequence of snapshots showing an epidemic in a clustered network with a central cluster. Blue nodes represent susceptible individuals with red representing infected and green representing removed. Further information on the visualisations can be found in appendix A.1.*

from real life networks to have the greatest impact on disease spread.

Our clustered networks can produce periodic results, similar to the wave progression of real life contagions. These cases are more likely in our simulation when cluster population is significantly increased. With relatively lower inter-cluster connections. Many real life contagions have a yearly cycle caused mostly by more people being inside and close to each-other over winter, however our model does not account for this but still shows periodic behaviour, this is likely because the disease will quickly spread through a cluster creating a small peak, before spreading to another cluster, and dying out in the previous.

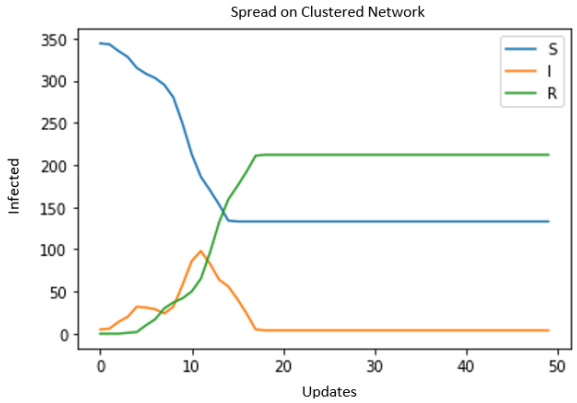

Figure 14: *SIR plot showing an epidemic similar to the one shown in figure 12. It can be seen that not all of the population gets infected since the disease does not spread outside of a few clusters.*

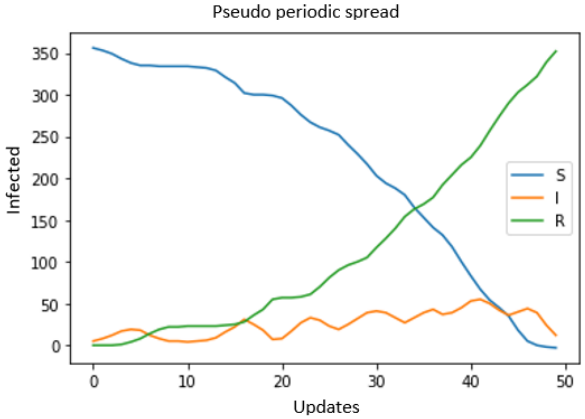

Figure 15: *SIR graph showing periodic changes of infection numbers caused by the disease spreading quickly within clusters before taking longer to spread to other clusters.*

7 Conclusion on our Epidemic Model Using Networks and their Key Characteristics

In conclusion, upon analysing the results, it is clear that an epidemic has a high degree of sensitivity to changes to the network through which it spreads. This suggests that variations in the structure of human communities will have a large impact on the spread of a disease.

For random networks, the results demonstrate a clear relationship between the numbers of social connections made by individuals and the rate of growth of an epidemic. Therefore, a group of individuals with more possibility of making physical contact will experience a peak of infections sooner and higher than those who put much effort in avoiding contact with others. This is shown through the networks with a higher average degree, and therefore clustering coefficient, having a more serious outbreak. Therefore, such analysis suggests that social distancing and reducing the amount of people with whom you have contact with will greatly help to slow the spread of a disease.

On the other hand, it is seen that using small-world networks, a more refined representation of a community, that the number of connections between individuals and the amount of clustering present is not as vital as the existence of long-range, inter-cluster connections. This suggests that a lockdown which places a heavier emphasis on restricting travel, and therefore transmission, between isolated communities would have a greater impact. The impact is significant on curbing the spread of a disease than other strategies, such as heavy restrictions on the number of interactions among nearby individuals.

Our analysis of scale-free networks suggests that removing nodes with a high degree will not have a large impact on the peak number of infections but it will delay the peak. This suggests that individuals who have the potential to encounter and interact with many people are very important. For example, workers in places such as supermarkets would potentially be able to significantly speed up the spread of a disease if they were not to take the proper precautions. Our results on scale-free networks were not completely consistent, suggesting that other factors than the number of high degree nodes would have a more significant effect on the epidemic. The idea of having nodes, or groups of nodes, interacting with many more individuals in the population will be explored further in the analysis of our own clustered networks.

Finally, our clustered networks may be the most complicated among all. Formed by multiple sub-networks or clusters, this method demonstrates how social circle can impact the behaviour of an epidemic. It should be noticed that, compared to self-clustering behaviour in the construction of other models, its cluster structure is more organised to reflect the structure of real-world interconnected communities in cities or such cities in regions. Thus, the structure is controllable and repeatable to form valid results.

Whilst modelling on our clustered networks, we observed many properties of self similarity. This means clusters of nodes in larger networks behaves similarly to a scaled version of a network of singular nodes. Especially in the standard clustered network where the clusters are structurally similar to the whole network that contains them.

In practice, it has been found the clustered model can be very complex and difficult to work with, taking significantly more time then any other model to generate networks then simulate on. It was also found in early models the stochastic nature of our network generation, caused significantly different results from seemingly identical inputs. However the variations between networks was reduced, and certain key values such as ratio of nodes to edges and total pop were kept similar. causing repeatable affects as long as the change to the independent variable was significant.

The key findings are the large differences in spread rate and peak height, between the centralised clustered network and the standard clustered network. We found that even when the ratio of nodes to edges and average edge count (as well as other metrics) were kept constant there was still a significant difference. This confirms that there is a large structural aspect to rate of contagion spread. This implies average centrality or sum/average of connections weighted by their "significance" would be a better single number metric for a network, or a population's "vulnerability".

This model's accuracy in representing disease spread is constrained by the complex nature of

human networks, which cannot be totally modelled. All epidemiology models are based on assumptions and simplifications, which may still provide accurate projections over wider populations. The purpose of this report was primarily to demonstrate the relationship between network theory and disease spread.

8 Extending Our Model

In the future, our modelling may be improved by adding complexity to the mechanism of disease transmission and to the behaviour of the individuals in the network. For example, in [12], researchers established a "two-layered multiplex network" with physical transmission of the infectious and virtual information exchange in social interactions are related. The investigated population in their report, therefore, take epidemic prevention measures. This alters the rate of transmission significantly.

Another area for development may be applying age demographics since, typically, diseases have higher mortality rates in older people; COVID-19 is an example of a disease that has a mortality rate that significantly increases with age, amongst other factors. In future work, the model could be further developed by applying age demographics to the nodes in its networks. This would involve researching the relationship between age and social circle in human networks, as it is unlikely that older people are as socially connected as younger people. Then, applying proportionally greater mortality rates to older nodes should output interesting results that could be the basis for analysing the effectiveness of preventative measures or even who best to vaccinate first.

For future modelling it may be appropriate to forgo the clusters of nodes and simply change node properties to better represent clusters. When modelling with a clustered network, contagions naturally spread through clusters in more than 1 update cycle, meaning neighbours' of said cluster will not be infected simultaneously, we are unsure of the effect this has but it does represent real world behaviour closer. Shared connections between non-fully infected clusters can share infections speeding up the contagion's progress and increasing the chance of full spread in both clusters. These properties and a few more would need to be hard coded. However, a network that uses singular nodes instead of clusters could be larger and run faster.

Due to the generality of how we have conducted our research, our models and data could be applicable, with little change, to other contagion based phenomena. Any situation in which connected objects are assigned a property, where the objects with this property are likely to spread or increase this property, based on connection or proximity to objects without said property. This is clearly very general, examples could be spread of a computer virus, wild fire, ideology, traffic through intersections, and more.

References

[1] Kermack, W.O., McKendrick, and A.G. Contributions to the mathematical theory of epidemics—i. *Bltn Mathcal Biology*, 53:33–55, 1991. URL https://doi.org/10.1007/BF02464423.

[2] Norman T. J. Bailey. *The mathematical theory of infectious diseases and its applications*. Griffin, London, 1975.

[3] Ben Craig, Tom Phelan, Jan-Peter Siedlarek, and Jared Steinberg. Improving Epidemic Modeling with Networks. *Economic Commentary*, pages 1–6, 2020. doi: 10.26509/frbc-ec-202023. URL https://www.clevelandfed.org/newsroom-and-events/publications/economic-commentary/2020-economic-commentaries/ec-202023-network-enhanced-sir-models.aspx.

[4] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994. URL https://www.jstor.org/stable/24362148.

[5] Yu Wang, Eshwar Ghumare, Rik Vandenberghe, and Patrick Dupont. Comparison of different generalizations of clustering coefficient and local efficiency for weighted undirected graphs. *Neural Computation*, 29(2):313–331, 2017. doi: 10.1162/NECO_a_00914. URL https://doi.org/10.1162/NECO_a_00914. PMID: 27870616.

[6] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998. doi: 10.1038/30918. URL https://doi.org/10.1038/30918.

[7] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959. URL https://www.jstor.org/stable/2237458.

[8] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999. URL https://www.sciencedirect.com/science/article/pii/S0378437100000182.

[9] Matt J. Keeling and Ken T. D. Eames. Improving Epidemic Modeling with Networks. *Journal Of The Royal Society Interface*, 2, 2005. doi: 10.1098/rsif.2005.0051. URL https://doi.org/10.1098/rsif.2005.0051.

[10] Scott L. Feld. Why your friends have more friends than you do. *American Journal of Sociology*, 1991. URL https://www.journals.uchicago.edu/doi/10.1086/229693.

[11] ChargeShivers (https://math.stackexchange.com/users/175036/chargeshivers). Expected local clustering coefficient for a node in a network. Mathematics Stack Exchange. URL https://math.stackexchange.com/q/2200642. URL:https://math.stackexchange.com/q/2200642 (version: 2017-03-26).

[12] Peican Zhu, Xinyu Wang, Shudong Li, Yangming Guo, and Zhen Wang. Investigation of epidemic spreading process on multiplex networks by incorporating fatal properties.

Applied Mathematics and Computation, 359:512–524, October 2019. doi: 10.1016/j.amc.2019.02.049. URL https://doi.org/10.1016/j.amc.2019.02.049.

[13] Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002. doi: 10.1103/RevModPhys.74.47. URL https://doi.org/10.1103/RevModPhys.74.47.

[14] James Holland Jones. Notes On R0. Technical report, Stanford University Department of Anthropological Sciences, 2007. URL https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf.

[15] Angelika Steger and Nicholas C Wormald. Generating random regular graphs quickly. *Combinatorics, Probability and Computing*, 8(04):377–396, 1999. URL https://mediatum.ub.tum.de/doc/1315533/file.pdf.

[16] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. 1 2008. URL https://www.osti.gov/biblio/960616.

[17] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008. URL http://conference.scipy.org/proceedings/SciPy2008/paper_2/.

A Appendix

A.1 Visualisations

Some of our data is quite challenging to conveniently and accurately represent. Generally, a pandemic is represented with SIR graphs. However, the focus of this report is on how network structure affects disease spread. It is difficult to show the effects of changes to the structure of a network with SIR graphs. This is because the change to the SIR graph will be seemingly small, and difficult to properly connect to the network changes. To represent our more qualitative findings we have created a simple animation tool to represent spread through scaled-down networks. It simply redraws the network at each update color-coding the nodes corresponding to their current SIR state. We have added regularly sampled frames from these where appropriate to demonstrate spread.

These diagrams are taken from smaller simulations with smaller population count so the end image is coherent at a glance. These are diagrams meant to convey qualitative information, not graphs meant to convey quantitative data. There is no way we have found to draw networks at the same orientation as the drawer only has access to a adjacency matrix thus some of the images are rotated between frames.

A.2 Network Generation and Reproduction

All of our networks have been generated using the NetworkX library for Python [17]. This was chosen for the wide variety of network generators, ease of creation of our own generation algorithms as well as public accessibility. All of our standardized networks were generated from NetworkX generators, so should be easily reproducible using this. Our clustered graphs were made using our own code, but were created by creating clusters then later interconnecting them. Code is included below.

```
def createclustered(avpop,popvariance,avconect,connectivityvariance,
    ↪ numclusters,interclusterconnect):
    #central cluster has increased pop
    #degrees = sorted(centralcluster.degree, key=lambda x: x[1], reverse=True
        ↪ )
    #station=degrees[0] #main connection to outerclusters
    clusters=[]

    g=nx.Graph()
    for i in range(numclusters):
        #print("creating intercluster connections \n\n\n")
        (cluster,popcount)=creatcluster(avpop,populationvariance,
            ↪ avconectivity,connectivityvariance,popcount)
        g=nx.compose(g,cluster)
        clusters.append(cluster)
    for p in clusters:
        startpoints=random.choices(list(p.nodes()),k=interclusterconnect)
        endpoints =random.choices(list(g.nodes()),k=interclusterconnect)
        for l in range(interclusterconnect):
```

```
            g.add_edge(startpoints[l],endpoints[l])
    return g

def clustwithstar(avpop,popvar,avconectivity,connectvar,numclusters,
    ↪ interclustconnect,centralconnect):
    #central cluster has increased pop
    #degrees = sorted(centralcluster.degree, key=lambda x: x[1], reverse=True
        ↪ )
    #station=degrees[0] #main connection to outerclusters
    clusters=[]
    (centralcluster,popcount)=creatcluster(centralclusterfactor*avpop,
        ↪ populationvariance,avconectivity,connectivityvariance,popcount)#
        ↪ central cluster has increased pop
    g=nx.Graph()
    g=nx.compose(g,centralcluster)
    for i in range(numclusters):
        print("creating␣intercluster␣connections␣\n\n\n")
        (cluster,popcount)=creatcluster(avpop,populationvariance,
            ↪ avconectivity,connectivityvariance,popcount)
        g=nx.compose(g,cluster)
        clusters.append(cluster)
    for p in clusters:
        startpoints2=random.choices(list(p.nodes()),k=centralclusterconnect)
        endpoints2 =random.choices(list(centralcluster.nodes()),k=
            ↪ centralclusterconnect)

        startpoints=random.choices(list(p.nodes()),k=interclusterconnect)
        endpoints =random.choices(list(g.nodes()),k=interclusterconnect)
        for l in range(interclusterconnect):
            g.add_edge(startpoints[l],endpoints[l])

        for l in range(centralclusterconnect):
            g.add_edge(startpoints2[l],endpoints2[l])#central cluster to out
                ↪ cluster connections
    print(popcount)
    return g

def creatcluster(avpop,popvariance,avconect,connectvariance,popcount):#
    ↪ popcount previous pop before addition of cluster
    print("creating␣cluster\n\n\n")
    pop=random.randrange(avpop,int((avpop*populationvariance)))#population of
        ↪  cluster
    temp = nx.Graph()
    pop =popcount+pop#highest value primary key node of cluster
    #pop=range(pop)
    for i in range(popcount,pop):
```

```
        temp.add_node(i)#adds node for each pop

    """each node connected to a random sample of other nodes"""

    for i in temp.nodes():
        connectivity=random.randrange(avconectivity-connectivityvariance,
            ↪ avconectivity+connectivityvariance)
        numconnection= int(pop*connectivity)#number of connections current
            ↪ node in cluster has
        connections= np.random.choice(temp.nodes(),size=numconnection)#random
            ↪  set of nodes

        for j in connections:
            temp.add_edge(i,j) #https://stackoverflow.com/questions/28488559/
                ↪ networkx-duplicate-edges/51611005
            #apparently wont add repeated edges
    return temp, pop
def update(g,totals,visits, infectivity = 0.09):
    for i in g.nodes(data=True):

        while(0<i[1]["exposures"] and i[1]["sir"] ==0):
            i[1]["exposures"]-=1
            if infectivity>random.random():#for each exposure infection
                ↪ chance
                i[1]["sir"] =1
                totals[0]-=1
                totals[1]+=1
    for i in g.nodes(data=True):
        count=0
        nei = list(g.neighbors(i[0]))#every connection to infected adds
            ↪ exposure
        if visits < len(nei):
            nei= random.sample(nei,visits)
        #print(nei)
        """this is done poorly if run slow start here"""
        for j in g.nodes(data=True):
            if j[1]["sir"]==1 and j[0] in nei:#checks if node is suseptible
                ↪ and a neibor of checked node
                count +=1
        i[1]["exposures"]=count
    for i in g.nodes(data= True):
        if i[1]["sir"] == 1:
            i[1]["timer"]+=1
        if i[1]["sir"] == 1 and i[1]["timer"] > 3:#recovers after 8 days
            i[1]["sir"] =2
```

```
            i[1]["timer"]=0
            totals[1]-=1
            totals[2]+=1
    for l in g.nodes(data= True):
        if l[1]["sir"] ==2:
            l[1]["timer"]=l[1]["timer"]+1
            """
        if l[1]["timer"]>=4 and l[1]["sir"] ==2:
            totals[0]+=1
            l[1]["sir"] =0
            totals[2]-=1
            #print(len(g.nodes))
            #print("lost immunity")
            """
    return g, totals
```

A.3 Further Results for Random Networks

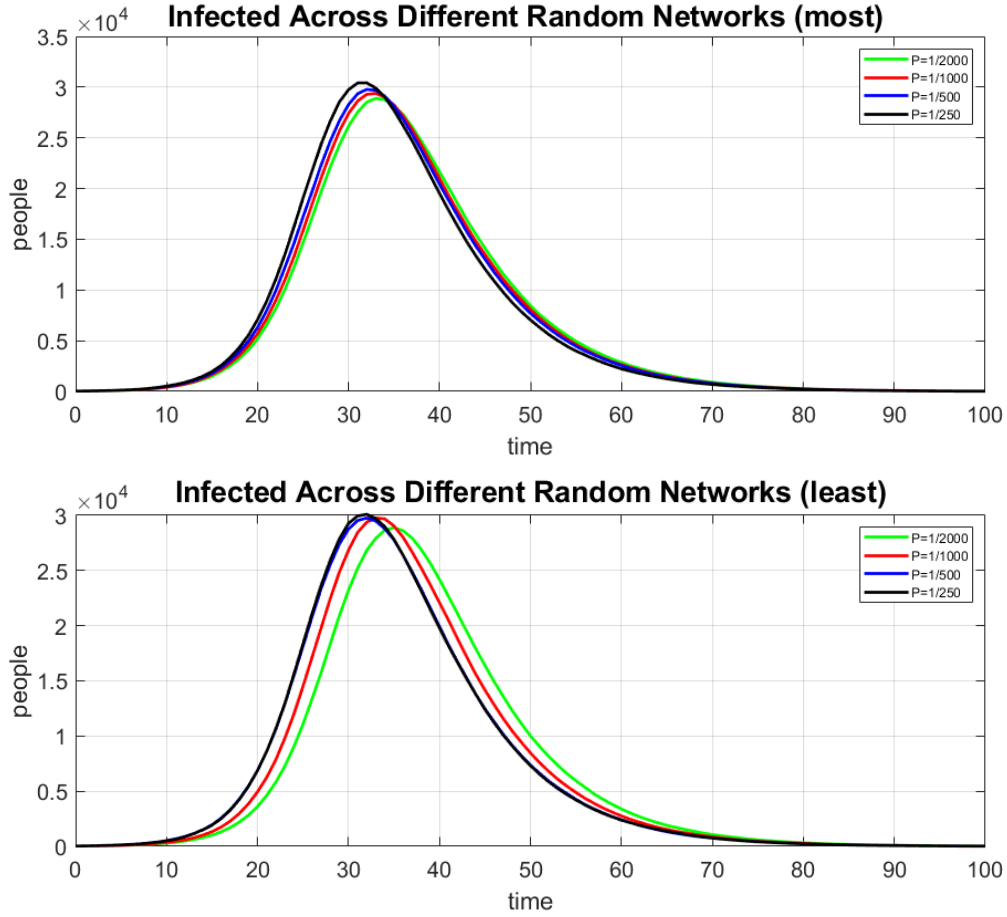

Figure 16: *Graphs showing the number of infected individuals on 4 random networks, each with a population of 100,000 and each with a different probability of edge creation, against time in days. The disease parameters are* $a = 0.01$, $b = 40$, $\mu = 0.01$, $\gamma = 0.125$ *and there are 30 initially infected individuals. These initially infected are, for the top graph, placed on the 30 highest degree nodes and for the bottom graph, placed on the 30 lowest degree nodes.*

Figure 16 shows the results produced by running our model on four random networks. These graphs show the course of the epidemic when the initially infected individuals are the most connected individuals and when they are the least connected individuals.

A.4 Mitigating Circumstances

One of our team members was suffering from illness and could not contribute as a result, bringing our group to 4 people. However we believe this may have already been communicated to the relevant people.