Epidemic Modeling with Contact Heterogeneity  
and  
Multiple Routes of Transmission  

AN UNDERGRADUATE STUDY OF THE MODEL FIRST PUBLISHED BY KISS ET AL. IN 2006, "THE EFFECT OF CONTACT HETEROGENEITY AND MULTIPLE ROUTES OF TRANSMISSION ON FINAL EPIDEMIC SIZE".

By:  
SAMA SHRESTHA  

Faculty Advisor:  
PROF. JENNIFER MILLER  

TRINITY COLLEGE  
MATHEMATICS DEPARTMENT 2013-2014
Preamble

This thesis is an undergraduate study of the work of Istvan Z. Kiss, Darren M. Green and Rowland R. Kao in the field of mathematical epidemiology which was published in the journal of Mathematical Biosciences in 2006 [8]. They analyzed the effect of incorporating contact heterogeneity and considering multiple routes of transmission on the epidemic dynamics of a disease.

The purpose of my thesis was to perform an in-depth analysis of their disease transmission model as well as to replicate their analytical and numerical results. In addition, I looked at the effects of altering fixed parameters in their model to get a deeper understanding of the disease dynamics in the model. In this study, I briefly introduce standard stochastic and deterministic epidemic theories used in the field of mathematical epidemiology. The first half of this thesis is focused on discussing the authors’ disease transmission model and understanding their analytical results for calculating the threshold criterion and the final epidemic size. The second half focuses on the simulation of their numerical results and delves deep into the algorithm behind generating networks to run the simulations. Lastly, as an extension to the paper, results from varying a parameter in the simulation are discussed.

Front page picture source: http://regan.med.harvard.edu/pictures/Hier/3Dhier.jpg
I would like to thank my thesis advisor, Professor Jennifer Miller, for her constant guidance and support throughout the academic year as I worked my way through decoding this journal paper word by word. This thesis would not have been possible without all her help. I am very grateful to her for keeping me on track towards completing this undergraduate study even when I wandered off course from time to time. I would also like to thank Professor Peter Yoon in the Computer Science Department for letting me run my simulations on the powerful computing server ‘pine’ which saved me weeks of simulation time. Additionally, I would like to thank the Mathematics Department at Trinity for giving me a chance to work on a senior thesis and fostering an environment for academic growth. Lastly, I would like to thank my family, Bicky, my lovely roommates as well as friends for providing me with constant love and support as always.

- Sama Shrestha
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preamble</td>
<td>1</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>2</td>
</tr>
<tr>
<td><strong>1 Introduction to Epidemic Models</strong></td>
<td>4</td>
</tr>
<tr>
<td>1.1 Deterministic SIR Models for Closed Population</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Network Models</td>
<td>7</td>
</tr>
<tr>
<td><strong>2 Disease Transmission Model from the Paper</strong></td>
<td>9</td>
</tr>
<tr>
<td>2.1 Calculation of the Threshold Criterion</td>
<td>12</td>
</tr>
<tr>
<td>2.2 Calculation of the Final Epidemic Size, $r(\infty)$</td>
<td>17</td>
</tr>
<tr>
<td><strong>3 Numerical Simulations to Validate the Model</strong></td>
<td>25</td>
</tr>
<tr>
<td>3.1 Building a Scale-free Network</td>
<td>25</td>
</tr>
<tr>
<td>3.2 Building a Homogeneous Network</td>
<td>27</td>
</tr>
<tr>
<td>3.3 Spread of Infection</td>
<td>29</td>
</tr>
<tr>
<td><strong>4 Comparison of Simulation Results and Analytical Predictions</strong></td>
<td>31</td>
</tr>
<tr>
<td><strong>5 Discussion</strong></td>
<td>35</td>
</tr>
<tr>
<td><strong>6 Further Analysis</strong></td>
<td>36</td>
</tr>
<tr>
<td>6.1 Effect of Varying the Recovery Rate $(g)$</td>
<td>36</td>
</tr>
<tr>
<td><strong>7 Conclusion</strong></td>
<td>39</td>
</tr>
<tr>
<td><strong>8 Appendix</strong></td>
<td>40</td>
</tr>
<tr>
<td>References</td>
<td>45</td>
</tr>
</tbody>
</table>
1 Introduction to Epidemic Models

Modeling in mathematics is a method of finding good analytical representations of different phenomena that occur in the world. Consequently modeling has a broad scope which includes modeling disease propagation and epidemic outbreaks. In epidemiology, mathematical models help in quantifying the structure through which a disease propagates and the potential of the disease to cause an outbreak. By validating models against previous data of disease prevalence, valuable information can be extracted from these models to mitigate or prevent future outbreaks. The history of mathematical modeling in epidemiology goes back to the eighteenth century when Daniel Bernoulli made a simple model of the spread of small pox in 1760 [7]. The early 1900s saw an increase in dynamical system approaches to epidemic models with the introduction of the first simple deterministic compartment model by A.G. McKendrick and W.O. Kermack in 1927 [6].

Although there are a variety of different types of models that mathematicians have created in order to successfully model specific diseases and interactions in various species, epidemic models can still be categorized into two broad categories:

**Deterministic Models** The word ‘deterministic’ suggests a fixed structure which does not have any random variation in it and such a model has outcomes that are determined through defined relationships and parameters. Examples of deterministic models are compartmental models such as the simple SIR model. Compartmental models have different categories of states that represent different stages of the disease and the propagation of the disease cause individuals to move within these compartment governed by a set of relationships containing relevant parameters. Deterministic models are good models for large populations where the randomness of parameters has less of an effect in the steady state of the solution. However, when the population size is small, ignoring the aspect of randomness may introduce a large error in the outcome. For such a situation, we consider stochastic models.

**Stochastic Models** Stochastic models incorporate the randomness that we see in our daily life. This could be crucial in cases where the number of people spreading the infection is very small due to small population size, recent invasion of disease or successful control measures of disease [6]. In these cases, not accounting for stochasticity might significantly deviate the
results. Stochastic models have random variables in them, hence different simulations of these model will give rise to different results. This is different from deterministic models which will always give the same results given the same initial conditions and parameters.

1.1 Deterministic SIR Models for Closed Population

SIR models have been successfully used for representing spread of acute infections since the 1900s. This model was initially developed by Kermack and McKendrick in 1927 [7]. It assumes that pathogens cause infection for a certain period of time and when individuals recover, they are immune throughout the rest of their lives. Thus this model applies to acute infections rather than chronic conditions. Individuals within a population are categorized into three major compartments: Susceptible, individuals who have not yet acquired the infection; Infected, individuals who have currently acquired the infection; and Recovered, individuals who have successfully recovered from the infection.

Let us define $\beta$ as the transmission rate. It is the number of infectious contacts per individual per unit time. Thus, the term $\beta SI$ represents the rate at which susceptible individuals are infected. Additionally, $g$ is defined as the removal or recovery rate. It is the number of infected individuals that recover per unit time. The average infectious period is thus represented by $\frac{1}{g}$.

![SIR Model Diagram](image)

Figure 1: Visualization of the compartment model for SIR where the arrows represent the rate at which individuals move from one compartment to the other. The direction represents the only movements that are assumed in the simple model.
The following system of differential equations represents the SIR model for a closed population \((N)\) where \(N = S + I + R\). Under closed population, we assume that there is no birth, death or migration. Thus the total population remains constant throughout time.

\[
\frac{dS}{dt} = -\beta SI, \quad (1.1.1a) \\
\frac{dI}{dt} = \beta SI - gI, \quad (1.1.1b) \\
\frac{dR}{dt} = gI \quad (1.1.1c)
\]

This is an example of a compartment model where each category \((S, I, R)\) represents a compartment and different groups of people move within these compartments at rates represented by the differential equations given above. One of the major assumptions of this model is that there is random mixing of hosts or individuals such that each individual has a small and equal chance of interacting with any other individual [6]. Thus, the rate at which people move out of the susceptible compartment is given by \(\beta SI\) with the negative sign representing out flow. Similarly, the rate at which people move into the infected compartment is the same as the rate at which people move out of the susceptible compartment \(\beta SI\) except the positive sign which represents in flow. The term \(-gI\) represents the number of people who are recovering and leaving the infected compartment. Finally, the rate at which people are entering the recovered compartment is equal to \(gI\).

Here, \(\frac{\beta}{g} = R_0\), the basic reproductive ratio, is the average number of secondary cases arising from an average primary case in an entirely susceptible population. It is the measure of maximum reproductive potential for an infectious disease. For the SIR model, epidemics arise when \(R_0 > 1\). This means that the rate of transmission of the disease \(\beta\) is greater than the recovery rate \(g\). Thus, people are getting infected faster than they are recovering which gives rise to an epidemic outbreak. This can also be thought of in the ‘reproductive’ sense where a disease will not spread unless it is successfully transmitted to at least one more host. This is famously known as the ‘threshold phenomenon’ [6].
1.2 Network Models

The assumption of homogeneous mixing in the simple SIR model limits the use of these models when it is applied to small finite networks. In these cases, we can no longer ignore the topology of the network when spreading the infection [9]. This is where the recently recognized network science comes into play. Network science is the application of mathematical graph theory to find solutions to problems in a variety of fields including the natural and the social sciences. Since the late 1900s, social scientists have been using graph theory to study human interactions in groups [9]. Although they are more interested in understanding the reasons behind these interactions, the structure of the network becomes very useful in quantifying contact heterogeneity for epidemic models [6]. The ‘nodes’ and ‘edges’ used in traditional graph theory can directly be used to represent ‘hosts’ and ‘contacts’ of an epidemic model. Hence, a set of nodes can represent a group of susceptible population. Each edge between these nodes can be imagined as an interaction between two individuals through which disease transmission might occur. The number of neighbors that each individual has is their ‘degree’. Thus, each individual interacts with a finite group of individuals which is more accurate of a representation of human interaction than random mixing.

The knowledge of a precise contact network in a finite population is crucial in understanding the disease dynamics of the population. However, mapping such real life networks and collecting enough data to be able to do so is no easy task. In case of human beings, the complexities of human interactions and lack of enough data makes it harder to pinpoint exact networks through which disease may spread. In addition, contact networks also differ from disease to disease. For example, an HIV transmission contact network where the disease spreads mostly through sexual contacts might look very different from a malaria transmission contact network. Nevertheless, the continued progress in this field and several known networks have contributed a lot in understanding the dynamics of network epidemics.

In order to be able to conduct such sensitive analysis, scientist use simulated networks which best represent the real world networks via estimation of parameters and stochastic simulations. The three major class of idealized networks that are extensively used in network epidemiology are random, scale-free and small world. Random networks are formed when nodes are allowed to connect at random to any other nodes until the desired degree distribution is obtained. There are different variations of random graphs, the
most famous one being the Erdös and Rényi graph which will later be used in the simulations [5]. It has been shown that the standard SIR disease transmission differential equations can be solved on a random graph [4]. Small world networks are a variation of the random graphs and these networks are usually based on the model proposed by Duncan Watts and Steve Strogatz in 1998 which involves ‘rewiring’ a regular network into a random network. The ‘rewiring’ involves randomly adding shortcut paths between nodes in a uniformly connected network until a random network is obtained. They found that such networks showed enhanced infection dynamics resulting in an epidemic faster than a regular network [15].

Scale-free networks, used in this paper, are networks whose degree distributions follow a power law distribution with an exponent between 2 and 3 [3]. A power law distribution means that the probability varies as a power of the random variable being observed. Such highly heterogeneous networks are seen in real life such as the world wide web where the highly connected nodes (websites or hubs) attract even more connections due to their popularity [1,3]. This is applicable to infectious disease transmission where the disease is transmitted though sexual contacts. Various studies conducted in Sweden, Britain, Zimbabwe and Burkina Faso have reported that the distribution of the number of sexual partners in different population is well approximated by a power law distribution [10,13,14] . In this case, sex workers can act as major ‘hubs’ having a large number of sexual contacts compared to the general population. Thus, scale-free networks are useful in dealing with the disease dynamics of STDs such as HIV, gonorrhea, chlamydia. Scale-free networks can be generated by many different algorithms of preferred attachment. This paper uses the Barabási-Albert preferential attachment model that was first published in 1999 by Albert-László Barabási and Réka Albert [3]. Their algorithm is used to generate a scale-free network for simulation purposes later in the paper.
2 Disease Transmission Model from the Paper

As discussed earlier, the SIR model in (1.1.1) assumes that all contacts have homogenous networks which means that everybody has an equal chance of encountering everyone in the population. However, using well-defined contact structures, we can now get rid of the simplifying yet unrealistic assumption of random mixing. When each member of a population is considered to be a node in a network, each member will potentially have a group of contacts within the population. The size of this group of contacts is known as the node’s degree. Let us assume that the population is divided into $n$ distinct groups of size $N_k$, $1 \leq k \leq n$, and each individual in group $k$ has exactly $k$ contacts. Thus, the probability that a random person has $k$ contacts is $p(k) = \frac{N_k}{N}$. Thus, if we assume that any node can have a connection with another node, we get an undirected network of size $N$ with node degree distribution $p(k)$. Here, $N = N_1 + N_2 + \ldots + N_n$. Let us define $\langle k \rangle$ as the average number of contacts per node. Thus, $\langle k \rangle = \sum_l lp(l)$ and in general, $\langle f(k) \rangle = \sum_l f(l)p(l)$.

In this paper, the authors bring up the issue of network structures not capturing all transmission between individuals. This might be due to imprecisely defined networks or the deficiency of network structures in capturing transmission through different routes. An example of this is transmission of HIV virus which has several known means of transmissions including sexual contact and blood transfusion. In order to explore this issue, the authors consider additional transmission using approximations by the mean-field terms (homogeneous SIR model) and investigate the effect of both types of transmission on the final epidemic size.

Within group $k$ in the population, there are $S_k$ susceptible individuals, $I_k$ infected individuals and $R_k$ recovered individuals where $S_k + I_k + R_k = N_k$. The authors present the following system of differential equations to capture disease spread for arbitrarily large networks ($N \to \infty$), which accounts for transmission of disease through both the network and the mean-field type transmission. For $k = 1, \ldots, n$,
\[
\frac{dS_k}{dt} = -(1-\lambda)\tau kS_k(t) \sum_l \frac{l-1}{l} p(l|k) \frac{I_l(t)}{N_l} - \lambda \beta \frac{S_k(t)}{N} \sum_l I_l(t) \quad (2.0.1a)
\]

\[
\frac{dI_k}{dt} = (1-\lambda)\tau kS_k(t) \sum_l \frac{l-1}{l} p(l|k) \frac{I_l(t)}{N_l} + \lambda \beta \frac{S_k(t)}{N} \sum_l I_l(t) - gI_k(t) \quad (2.0.1b)
\]

where

- \(\beta\) = mean-field transmission rate,
- \(g\) = removal or recovery rate,
- \(\tau\) = transmission rate across a network contact between an infected and a susceptible node,
- \(\lambda\) = parameter that varies the contribution of the two transmission mechanisms to the overall transmission,
- \(p(l|k)\) = probability that a node of degree \(k\) is connected to a node of degree \(l\).
- \(\sum_l p(l|k) = 1\)

The first term in the model which is multiplied by \(1-\lambda\) represents the transmission through the network where infections are transmitted at a rate of \(\tau\). The strength of the infection also depends on the degree \(k\) of the susceptible nodes being considered and the probability that any given contact or neighbor of a susceptible node with \(k\) connections is infected \((p(l|k)\) term). The sum \(\sum_l p(l|k) \frac{I_l(t)}{N_l}\) represents the potential interaction with the infected neighbors which might cause disease transmission. The second term in the model which is multiplied by \(\lambda\) represents the simple mean-field transmission where the \(\sum_l I_l(t)\) term accounts for the assumption that an individual can get infected from all the nodes in the network. This term is analogous to the simple SIR model that was discussed earlier with an added division of the population into different groups.

Since there are no correlations between the degrees of connected nodes, \(p(l|k)\) is dependent only on the degree \(l\) and \(p(l)\). Then, the probability of
having a neighbor with degree \( l \) will be equal to the expected number of nodes having degree \( l \) in the network divided by the average degree in the network. Thus, \( p(l|k) = \frac{\binom{l}{k}}{\langle k \rangle} \). Let \( s_k = \frac{S_k}{N} \) and \( i_k = \frac{I_k}{N} \) in order to facilitate simpler analysis. Then the above equations can be written as:

\[
\frac{ds_k}{dt} = -s_k(t) \sum_l \left( (1 - \lambda)\tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) i_l(t), \tag{2.0.2a}
\]

\[
\frac{di_k}{dt} = s_k(t) \sum_l \left( (1 - \lambda)\tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) i_l(t) - gi_k(t) \tag{2.0.2b}
\]
2.1 Calculation of the Threshold Criterion

In the beginning stages of an infection, it can be assumed that the proportion of infected individuals in the population is rather low compared to the susceptible population. Similarly, we can also assume that the proportion of susceptible individuals with $k$ contacts is approximately equal to the proportion of total number of individuals with $k$ contacts in the population. Thus, $i_k \ll s_k$ and $s_k \approx p(k)$. Let us also introduce the Dirac delta function as follows:

$$\delta_{lk} = \begin{cases} 
0 & \text{if } l \neq k, \\
1 & \text{if } l = k.
\end{cases} \quad (2.1.1)$$

Thus, $i_k = \sum_l \delta_{lk} i_l$ and thus this term can be incorporated into the summation in (2.0.2b). Substituting $s_k$ with $p(k)$ we get,

$$\frac{di_k}{dt} = p(k) \sum_l \left( (1 - \lambda) \tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta - \frac{g \delta_{lk}}{p(k)} \right) i_l(t) \quad (2.1.2)$$

Now let us define,

$$I(t) = \sum_k i_k(t) \quad (2.1.3a)$$

$$J(t) = \sum_k ki_k(t) \quad (2.1.3b)$$

Here $I$ represents the total proportion of infected people in the population and $J$ represents the expected value of the number of contacts that an infected node might have in the network. We use these expressions in order to calculate the epidemic threshold for both the homogeneous and heterogeneous transmission case.
Summing (2.1.2) we get,

\[
\frac{dI}{dt} = \sum_k p(k) \sum_l \left( (1 - \lambda)\tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta - \frac{g \delta_{lk}}{p(k)} \right) i_l(t)
\]

(2.1.4)

\[
= \frac{(1 - \lambda)\tau}{\langle k \rangle} \sum_k \sum_l p(k)k(l - 1)i_l(t) + \sum_l \sum_k p(k)\lambda \beta i_l(t) - \sum_k \sum_l g \delta_{lk} i_l(t)
\]

\[
= \frac{(1 - \lambda)\tau}{\langle k \rangle} \sum_k p(k)k(J(t) - I(t)) + \lambda \beta I(t) - g I(t)
\]

\[
= (1 - \lambda)\tau(J(t) - I(t)) + \lambda \beta I(t) - g I(t)
\]

\[
= [\lambda \beta - g - (1 - \lambda)\tau] I(t) + (1 - \lambda)\tau J(t)
\]

(2.1.5)

Similarly, we get an expression for \( J'(t) \) from 2.1.3b and 2.1.2.

\[
\frac{dJ}{dt} = \sum_k kp(k) \sum_l \left( (1 - \lambda)\tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta - \frac{g \delta_{lk}}{p(k)} \right) i_l(t)
\]

(2.1.6)

\[
= \frac{(1 - \lambda)\tau}{\langle k \rangle} \sum_k \sum_l kp(k)k(l - 1)i_l(t) + \sum_k \sum_l kp(k)\lambda \beta i_l(t) - \sum_k \sum_l g \delta_{lk} i_l(t)
\]

\[
= \frac{(1 - \lambda)\tau}{\langle k \rangle} \sum_k p(k)k \left( \sum_l i_l(t) - \sum_l i_l(t) \right) + \lambda \beta \sum_k p(k) \sum_l i_l(t) - \sum_k g k i_k(t)
\]
\[ \frac{(1-\lambda)\tau}{\langle k \rangle} \sum_k p(k) k^2 (J(t) - I(t)) + \lambda \beta < k > I(t) - gJ(t) \]

\[ = \frac{(1-\lambda)\tau\langle k^2 \rangle}{\langle k \rangle} (J(t) - I(t)) + \lambda \beta < k > I(t) - gJ(t) \]

\[ = \left[ \lambda \beta \langle k \rangle - \frac{(1-\lambda)\tau\langle k^2 \rangle}{\langle k \rangle} \right] I(t) + \left[ \frac{(1-\lambda)\tau\langle k^2 \rangle}{\langle k \rangle} - g \right] J(t) \quad (2.1.7) \]

A linear stability analysis is performed of the disease free state \((I, J) = (0, 0)\). Linear stability analysis helps us determine the stability of a steady state by analyzing small changes from the equilibrium state. Analysis of the eigenvalues of the Jacobian matrix provides us information about the stability of this steady state. A negative eigenvalue translates to a stable equilibrium while a positive eigenvalue suggests an unstable equilibrium. In this case, we can use the eigenvalues of the Jacobian to get an expression for the epidemic threshold at which the sign of the eigenvalue should change thereby representing a shift in the stability. Let \(I'(t) = f(I, J)\) and \(J'(t) = g(I, J)\). The Jacobian matrix at \((0, 0)\) is

\[ J_o = \left(\begin{array}{cc} \frac{\partial f(0,0)}{\partial I} & \frac{\partial f(0,0)}{\partial J} \\ \frac{\partial g(0,0)}{\partial I} & \frac{\partial g(0,0)}{\partial J} \end{array}\right) \]

\[ = \left(\begin{array}{cc} \lambda \beta - g - (1-\lambda)\tau \langle k \rangle & (1-\lambda)\tau \langle k^2 \rangle \langle k \rangle - g \\ \lambda \beta \langle k \rangle - (1-\lambda)\tau \langle k^2 \rangle \langle k \rangle & (1-\lambda)\tau \langle k^2 \rangle \langle k \rangle - g \end{array}\right) \quad (2.1.8) \]
Thus, eigenvalues $\lambda$ are calculated as follows:

\[
\det \begin{pmatrix}
\lambda - g - (1 - \lambda)\tau - \Lambda & (1 - \lambda)\tau \\
\lambda\langle k \rangle - (1 - \lambda)\tau\frac{\langle k^2 \rangle}{\langle k \rangle} & (1 - \lambda)\tau\frac{\langle k^2 \rangle}{\langle k \rangle} - g - \Lambda
\end{pmatrix} = 0
\]

\[
(\lambda - g - (1 - \lambda)\tau - \Lambda)\left((1 - \lambda)\tau\frac{\langle k^2 \rangle}{\langle k \rangle} - g - \Lambda\right) - ((1 - \lambda)\tau)\left(\lambda\langle k \rangle - (1 - \lambda)\tau\frac{\langle k^2 \rangle}{\langle k \rangle}\right) = 0
\]

(2.1.9)

Setting $\Lambda = 0$, the transition from a negative to a positive eigenvalue is expressed by solving for $g$ in (2.1.9),

\[
g^2 - g \left[\lambda - (1 - \lambda)\tau\left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1\right)\right] + \left[\lambda(1 - \lambda)\tau\frac{\langle k^2 \rangle}{\langle k \rangle} - (1 - \lambda)\tau\lambda\langle k \rangle\right] = 0
\]

\[
g = \frac{1}{2} \left[\lambda - (1 - \lambda)\tau\left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1\right)\right] + \frac{1}{2} \left\{\lambda(1 - \lambda)\tau\left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1\right)^2 + 4\lambda(1 - \lambda)\beta\tau\langle\langle k \rangle \rangle\right\}^{1/2}
\]

(2.1.10)

In order to find an expression for $R_0$, the threshold criterion, we can rewrite (2.1.10) as functions of $\rho_r = \frac{\lambda\beta}{g}$ and $\rho_0 = \frac{(1-\lambda)\tau\langle k \rangle}{g}$.

\[
R_0 = \frac{1}{2} \left[\rho_r + \rho_0 \left(\frac{\langle k^2 \rangle}{\langle k \rangle^2} - \frac{1}{\langle k \rangle}\right)\right] + \frac{1}{2} \left\{\rho_r - \rho_0 \left(\frac{\langle k^2 \rangle}{\langle k \rangle^2} - \frac{1}{\langle k \rangle}\right)\right\}^2 + 4\rho_r\rho_0 \left(1 - \frac{1}{\langle k \rangle}\right)^{1/2}
\]

(2.1.11)

Here, $\rho_r$ represents the transmission potential through the mean-field transmission mechanism and $\rho_0$ represents the transmission potential through the heterogeneous transmission mechanism. These two transmission potentials are the number of secondary infections that a randomly chosen infectious node will generate in its entire lifetime of being infected within the given susceptible population. The reproductive potential, $R_0$ of this model also represents the number of secondary infections generated by a node but this node is not randomly chosen. If the value of $R_0 < 1$, the epidemic does not spread since a node is not infecting enough nodes to replace itself.
when it recovers. Thus, $R_0$ is the threshold criterion for epidemic spreading. Rather, as the expression suggests, it is chosen based on the number of infectious neighbors that the node has and the probability of the node contracting the infection from its neighbor. As seen above, $R_0 = 1$ was the assumption as the epidemic threshold, similar to the traditional SIR models. Highly heterogenous networks such as the scale-free network lack an epidemic threshold [11]. Thus, it is interesting to see that multiple transmission networks result in the existence of an epidemic threshold similar to a homogeneous network.
2.2 Calculation of the Final Epidemic Size, \( r(\infty) \)

In order to look at the behavior of the disease in the long run, we can look at the final size of the recovered population after letting time go to infinity. Thus, the final epidemic size gives us information about the steady state of the disease propagation. This section deals with calculating the final epidemic size using our differential equations.

We have the following equations 2.0.2a and 2.0.2b from our disease transmission model.

\[
\frac{ds_k}{dt} = -s_k(t) \sum_l \left( (1 - \lambda) \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) i_l(t), \quad (2.2.1a)
\]

\[
\frac{di_k}{dt} = s_k(t) \sum_l \left( (1 - \lambda) \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) i_l(t) - gi_k(t) \quad (2.2.1b)
\]

Let us now assume that

\[
\lambda_k = \sum_l \left( (1 - \lambda) \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) i_l(t) \quad (2.2.2)
\]

Thus we get,

\[
\frac{ds_k}{dt} = -s_k(t)\lambda_k \quad (2.2.3)
\]

Solving equation 2.2.3, we get

\[
\int_0^T \frac{ds_k}{dt} \, dt = \int_0^T -s_k(t)\lambda_k \, dt
\]

\[
s_k(T) = s_k(0) \exp(-\Phi_k(T)) \quad (2.2.4)
\]

where

\[
s_k(0) = \frac{N_k}{N} \quad \text{and} \quad \Phi_k(T) = \int_0^T \lambda_k(a) \, da
\]
Adding equations 2.2.1a and 2.2.1b, and integrating from 0 to \( T \), with boundary condition \( T \rightarrow \infty \), we get

\[
\int_0^T \frac{ds_k}{dt} dt + \int_0^T \frac{di_k}{dt} dt = \int_0^T -gi_k(t)dt
\]  

(2.2.5)

Using the Fundamental Theorem of Calculus,

\[
s_k(T) - s_k(0) + i_k(T) - i_k(0) = \int_0^T -gi_k(t)dt
\]  

(2.2.6)

Going back to the definition, \( s_k(T) = \frac{S_k(T)}{N} \), we can substitute it back in equation 2.2.6. Additionally, at \( t = 0 \), the susceptible population having \( k \) contacts is approximated by the entire fraction of the population having \( k \) contacts. Thus, there is no infection present at \( t = 0 \). Equation 2.2.6 becomes

\[
\frac{S_k(T)}{N} - \frac{N_k(T)}{N} + \frac{I_k(T)}{N} - 0 = \int_0^T -gi_k(t)dt
\]

\[-\frac{R_k(T)}{N} = \int_0^T -gi_k(t)dt\]

\[-r_k(T) = \int_0^T -gi_k(t)dt\]

\[
\frac{r_k(\infty)}{g} = \int_0^\infty i_k(t)dt
\]  

(2.2.7)

In 2.2.7, we are letting \( T \rightarrow \infty \). From our previous definition of \( \Phi_k(T) = \)}
\[
\int_0^T \lambda_k(a)da \text{ in equation 2.2.4 and } \lambda_k \text{ from equation 2.2.2,}
\]

\[
\Phi_k(\infty) = \lim_{T \to \infty} \int_0^T \sum_l \left( (1 - \lambda) \tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) i_l(t) dt
\]

\[
\Phi_k(\infty) = \sum_l \left( (1 - \lambda) \tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) \lim_{T \to \infty} \int_0^T i_l(t) dt
\]

\[
\Phi_k(\infty) = \sum_l \left( (1 - \lambda) \tau \frac{k(l - 1)}{\langle k \rangle} + \lambda \beta \right) \frac{r_l(\infty)}{g}
\]

(2.2.8)

However, \( s_k(\infty) = \frac{N_k}{N} - r_k \) since all the infected have recovered in the long run. We also know from equation 2.2.4 that \( s_k(\infty) = \left( \frac{N_k}{N} \right) \exp(-\Phi(\infty)) \). Thus, we get

\[
r_k(\infty) = \left( \frac{N_k}{N} \right) (1 - \exp(-\Phi(\infty)))
\]

(2.2.9)

Therefore,

\[
\Phi_k(\infty) = \sum_l \left( (1 - \lambda) \tau \frac{k(l - 1)}{\langle k \rangle g} + \frac{\lambda \beta}{g} \right) \left( \frac{N_l}{N} \right) (1 - \exp(-\Phi(\infty)))
\]

(2.2.10)

The paper expands this sum into two parts. Thus, \( \Phi(\infty) = k\alpha + \alpha_r \) where

\[
\alpha = \sum_l \left( \frac{(1 - \lambda) \tau}{\langle k \rangle g} \right) (1 - \exp(-l\alpha - \alpha_r))(l - 1) \frac{N_l}{N}
\]

(2.2.11)

\[
\alpha_r = \sum_l \frac{\lambda \beta}{g} (1 - \exp(-l\alpha - \alpha_r)) \frac{N_l}{N}
\]

(2.2.12)
Thus the final epidemic size for the population is given by

\[ r(\infty) = \sum_k r_k(\infty) \]

\[ = \sum_k \left( \frac{N_k}{N} \right) (1 - \exp(-k\alpha - \alpha_r)) \]

\[ = \sum_k p(k)(1 - \exp(-k\alpha - \alpha_r)) \quad (2.2.13) \]

In general when \( \langle f(k) \rangle = \sum_k f(k)p(k) \), we get the following parametric equations:

\[ r(\infty) = \langle 1 - \exp(-k\alpha - \alpha_r) \rangle \]

\[ \rho_r = \alpha_r / \langle 1 - \exp(-k\alpha - \alpha_r) \rangle \quad (2.2.14a) \]

\[ \rho_0 = \alpha \langle k \rangle^2 / \langle (k - 1)(1 - \exp(-k\alpha - \alpha_r)) \rangle \quad (2.2.14b) \]

Thus the final epidemic size is expressed using the above average function consisting of the two transmission potentials, \( \rho_r \) and \( \rho_0 \) that are also expressed as average functions. In order to evaluate the final epidemic size for various values of \( \alpha \) and \( \alpha_r \), we can approximate the sums with integrals. The authors in the paper base their calculations on the assumption of a scale-free network which corresponds to the preferential attachment model of Barabási and Albert. This model has a probability density given by \( p(k) = \frac{2m^2}{k^3} \) for continuous \( k \), and an average contacts per node, \( \langle k \rangle = 2m \). We talk of a continuous \( k \) here because we are using it in the probability density distribution function.

We can express equation (2.2.14) as,

\[ r(\infty) = \sum_l (1 - \exp(-l\alpha - \alpha_r))p(l) \quad (2.2.15) \]

We will now apply the transformations \( x = \frac{k}{m} \) and \( \Phi = m\alpha \) in equations (2.2.15). Thus, \( p(l) \) can be written as,
\[ p(l) = \frac{2m^2}{k^3} = \frac{2}{x^3} \frac{1}{m} \]

Since \( \Delta k = 1 \) in our discrete sum and \( m \) is a constant,

\[ p(l) = \frac{2 \Delta k}{x^3} m = \frac{2}{x^3} \Delta x \] (2.2.16)

Therefore (2.2.15) becomes,

\[ r(\infty) = \int_{1}^{\infty} (1 - \exp(-\Phi x - \alpha_r))^2 \frac{2}{x^3} dx \] (2.2.17)

Thus the following family of parametric equations are obtained.

\[ r(\infty) = 2 \int_{1}^{\infty} \frac{1 - \exp(-\Phi x - \alpha_r)}{x^3} dx \] (2.2.18a)

\[ \rho_r = \left(\frac{\alpha_r}{2}\right) / \int_{1}^{\infty} \frac{1 - \exp(-\Phi x - \alpha_r)}{x^3} dx \] (2.2.18b)

\[ \rho_0 = 2 \Phi / \int_{1}^{\infty} \frac{mx - 1}{mx^3} [1 - \exp(-\Phi x - \alpha_r)] dx \] (2.2.18c)

It is interesting note that the qualitative behavior of \( r(\infty) \) does not change with \( m \) which we shall see later in our simulations as well. Additionally, as \( m \to \infty \), \( r(\infty) \) converges towards the case where

\[ \rho_0 = 2 \Phi / \int_{1}^{\infty} \frac{1 - \exp(-\Phi x - \alpha_r)}{x^2} dx \] (2.2.19)

The authors present an interesting contour plot of the final epidemic size, \( r(\infty) \) as a function of the two transmission potential: the mean-field type transmission, \( \rho_r \), and network based transmission, \( \rho_0 \), in Figure 2.
In Figure 2, the fixed overall transmission potential ($\rho_{\text{total}}$) illustrates how the varying contributions of the two types of transmission affects the final epidemic size. At low values of $\rho_{\text{total}}$, the final epidemic size contour lines intersect the $\rho_{\text{total}}$ lines in such a way that increased contributions from network transmissions, $\rho_r$, will lead to a larger epidemic size while increased contribution of mean-field type transmission, $\rho_0$, will lead to a smaller epidemic size. However, at higher values of $\rho_{\text{total}}$, the $r(\infty)$ contour lines intersect $\rho_{\text{total}}$ lines in a way that the opposite phenomena occurs. The value of $\rho_{\text{total}}$ where this transition occurs is the critical value, $\rho_{\text{crit}} \approx 1.4$. Additionally, the figure also demonstrates that for a fixed $\rho_{\text{total}}$, the highest final epidemic sizes are observed at the extreme cases of $\lambda = 1$ or 0. This can be observed by looking at the contour plot edges before and after the critical point. For values of $\rho_{\text{total}}$ less than the critical point, the largest epidemic size for a fixed value of $\rho_{\text{total}}$ occurs at $\lambda = 1$ (purely mean field). However, for values of $\rho_{\text{total}}$ greater than the critical point, the largest epidemic size for a fixed value of $\rho_{\text{total}}$ occurs at $\lambda = 0$ (purely scale-free network).

The authors consider the two extreme cases of $\lambda$ in further detail by plot-
ting the final epidemic sizes for each of the two extreme cases. As seen in Figure 3, the two final epidemic size curves intersect at the critical point of $\rho_{\text{crit}} \approx 1.4$.

When $\lambda = 1$, equation 2.2.14 becomes

$$r(\infty) = 1 - \exp(-\rho r(\infty))$$  \hspace{1cm} (2.2.20)

When $\lambda = 0$, equation 2.2.14 becomes

$$r(\infty) = \langle 1 - \exp(-k\alpha) \rangle$$  \hspace{1cm} (2.2.21a)

$$\rho_0 = \alpha(k)^2 / \langle (k - 1)(1 - \exp(-k\alpha)) \rangle$$  \hspace{1cm} (2.2.21b)
Figure 3: This figure shows the analytical predictions using the equation (2.2.20) and equation (2.2.21) with $m = \langle k \rangle / 2 \to \infty$. The final epidemic size is plotted as a function of the transmission potential where $\rho = \rho_r = \rho_0$ for each of the different modes of transmission. This figure was directly borrowed from Kiss et al. 2006.
3 Numerical Simulations to Validate the Model

The analytical results obtained for the final epidemic size in (2.2.18) using the disease transmission model were compared to simulated final epidemic sizes on computer-generated homogeneous and scale-free network. We are mostly interested in the maximum epidemic sizes here since we want to see how final epidemic sizes vary with different transmission potential. Since earlier analysis (Figure 2) showed highest epidemic sizes at the extreme cases of $\lambda$, the simulations will be centered around approximating the two extreme cases using ideal networks. The homogeneous network used in our simulation is assumed to approximate the mean-field transmission case very well i.e. $\lambda = 1$ and this will be discussed in detail later. A scale-free network is used in simulating the heterogeneous network. Thus, we assume that transmission occurs only through the connections of these networks i.e. $\lambda = 0$. Transmission through individual homogeneous and scale-free networks should approximate the cases of the extreme values of $\lambda$ well.

3.1 Building a Scale-free Network

A scale-free network was generated using the Barabási-Albert preferential attachment model. Their algorithm to generate the model is as follows [2,8]:

- Start with a small ($m_0$) number of nodes.
- **Growth**: At every time step, add a new node with $m \leq m_0$ edges linking the new node to $m$ different nodes already present in the system. This $m$ is appropriately chosen and remains constant throughout the network building.
- **Preferential Attachment**: Assume that the probability $\Pi$ that a new node will be connected to an existing node $i$ depends on the degree $k_i$ of this node such that
  \[
  \Pi(i) = \frac{k_i}{\sum_l k_l}
  \]
  Here $l$ runs through all the existing nodes in the network and $\sum_l k_l$ represents the total number of connections present in the network. Thus, the probability of being connected to a node $i$ depends on the number of connections $i$ in comparison to the total connections in the network.
- After $t$ timesteps, this algorithm results in a network with a total of $N = t + m_0$ nodes and $mt$ edges.
Numerical simulations were implemented in MATLAB with the following pseudo code [12]. This code creates a scale-free graph with $N$ nodes and $d$ average degree. It outputs an adjacency matrix where the entry $1$ in $a_{ij}$ represents a connection between nodes $i$ and $j$. A lack of connection is represented by the entry $0$. Each node connects to $d/2$ other nodes at each timestep in order to maintain the average number of contacts for the network at $\langle k \rangle = d$.

- Create a $N \times N$ zero matrix
- Fully connect the first $(d + 1)$ nodes by setting the first $(d + 1)$ rows and $(d + 1)$ columns except the diagonal elements to $1$.
- For loop through the rest of the rows of the matrix starting from $(d + 2)$
  
  # While the new node has less than $d/2$ neighbors
  - Generate a random number $z$ between $0$ and $1$
  - Sum rows of matrix to a vector $L$ to count the total neighbors
  - Calculate $\Pi_i$ for each node $\left(\frac{L_i}{\text{sum}(L)}\right)$ where $L_i$ is the $i$th entry of $L$
  - Consider probabilities as subintervals of the interval from $0$ to $1$
  - Choose a node to connect to based on $z$ and the subinterval containing $z$
  - Enter the connection in the matrix
  - Repeat until while condition is met
  
  # End While loop
- End For Loop

The implementation of this code (Appendix A) produced networks which showed a scale-free distribution i.e. a power-law distribution of the number of connections in the network represented by $p(k) \propto k^{-\gamma}$ where $\gamma = 2.9 \pm 0.1$. The figure shows the regression line for only the upper part of the graph since the bottom part of the graph skews the line with a lot of nodes having the same number of few contacts.
3.2 Building a Homogeneous Network

The authors briefly talk about generating a homogeneous network such that each node has the same number of links in the network. Thus, they chose to use one of the most prevalent methods of generating random networks characterized by a lack of clustering. For my simulations, I chose to use an alternative formulation of a random network which produces networks with a probability distribution that approximately follows the Poisson distribution. This random graph is also famously known as the Erdös and Rényi graph [5]. It is the original random-graph model and has been used extensively due its simplicity. Barbour and Mollison showed that the epidemic dynamics on this particular random network are analogous to an SIR model epidemic in a randomly mixing population [4,6]. Additionally, I also wanted to compare the results from this graph with those produced by the authors.

The following pseudo code [12] was used to produce the random graph with $N$ nodes and an average degree distribution of $d$.

- Create an empty $N \times N$ matrix.
- Run a For loop with $d \ast N/2$ loops.
- Generate two random numbers \( x \) and \( y \) and multiply each number by \( N \).
- Round the two numbers up to whole numbers.
- While \( x = y \) or \( x^{th} \) and \( y^{th} \) node are connected:
  
  # Repeat steps 1 and 2.
- End While loop
- Connect \( x^{th} \) and \( y^{th} \) node.
  
- End For loop

Here, we let the loop run \( d \times N/2 \) times because \( d \times N \) represents \( d \) connections for all \( N \) nodes. However, since two nodes connect to each other, \( d \times N \) counts this connection twice. Thus, the division by 2 takes care of this issue, giving us a random graph with \( d \) average connections. The Erdös and Rényi graph generated by this code (Appendix B) has a Poisson distribution as seen in Figure 5.

![Graph](image)

Figure 5: Results of the numerical simulation with \( N = 1000 \) and \( d = 20 \). Degree distribution follows Poisson distribution as desired.
3.3 Spread of Infection

The epidemics are seeded into the network with 10 initial randomly chosen nodes for transmission of the disease. The probability \( P_j \) of a susceptible node \( j \) with \( k_j \) infectious neighbors acquiring infection is calculated using the following relationship,

\[
P_j = 1 - (1 - \tau \Delta t)^{k_j}
\]  

(3.3.1)

Equation 3.3.1 comes from analyzing the relationship of node \( j \) with its \( k_j \) neighbors. Node \( j \) can contract the infection from any of its \( k \) neighbors and each neighbor has the transmission rate \( \tau \) of spreading the infection. So, \( j \) has a \( (1 - \tau \Delta t)^{k_j} \) probability of not getting infected from any of its neighbors. Therefore, the probability that \( j \) gets infected from any of its \( k_j \) neighbors is \( 1 - (1 - \tau \Delta t)^{k_j} \).

This is slightly different from what the authors of the paper used in their own simulations. In the paper, they describe the probability of infection as being directly related to \( \tau k \Delta t \). After some analysis, we concluded that the authors most likely expanded equation 3.3.1 using Taylor Series expansion and simply took the first two terms. To demonstrate this, let \( h(\tau \Delta t) = (1 - \tau \Delta t)^{k_j} \). Taking the first two terms of the Taylor Series expansion of \( h \) around \( \tau \Delta t = 0 \) gives us the following:

\[
(1 - \tau \Delta t)^{k_j} \approx 1 - k_j \tau \Delta t
\]  

(3.3.2)

Therefore, \( P_j \approx 1 - 1 + k_j \tau \Delta t \approx k_j \tau \Delta t \) which is equivalent to what the authors mention in the paper. Thus, our probabilities should be more accurate in comparison since they do not seem to be accounting for the error terms of the Taylor Series in their expansion.

An infectious node recovers at a rate of \( g \) and \( g = 1 \) is used for the simulations in the paper. Hence, all infected nodes recover in the next time step and move to the recovered state remaining immune throughout the rest of the simulation. The following pseudo code was used to spread the infection on given network (Appendix C).

- Randomly choose ten numbers (nodes) less than \( N \) to seed the infection and store it in a \( N \times 1 \) vector.

- While the number of infected nodes in the previous time step > 0:
• Multiply the network and infection vector to get a vector of the number of infectious neighbors

• Run a For loop through each node
  – Generate a random number \( z \) between 0 and 1
  – Calculate the probability \( (P_i) \) in 3.3.1 for node \( i \)
  – If \( P_i > z \) and node \( i \) is not in the recovered vector:
    * Node \( i \) gets infected

• End For loop

• Move all infected nodes to recovered vector and update the infection vector

  – End While loop

  – Calculate final epidemic size by adding the entries of the recovered vector

The simulations (Appendix E) were then averaged over 50 different network realizations with 50 simulations of epidemic spreading on each network.
4 Comparison of Simulation Results and Analytical Predictions

Here, we compare the theoretical prediction for the final epidemic size in the homogeneous network based on the random graph that the authors use in the paper to the simulated results that we obtained using the Erdős and Rényi graph. The graph that the authors use is characterized by $N$ nodes having the same number of contacts, $d$. Thus, the analytical expression for the final epidemic size can be obtained from 2.2.14 by using the probability density function consisting of the Dirac delta function, $\delta$, where $p(k) = \delta(k - \langle k \rangle)$. Here,

$$\delta(k - \langle k \rangle) = \begin{cases} 0 & \text{if } k \neq \langle k \rangle, \\ 1 & \text{if } k = \langle k \rangle. \end{cases} \quad (4.0.3)$$

This function translates into the condition that the probability of having $k$ contacts is 1 when $k$ is equal to the average degree distribution. For all other times, the probability is 0 since every node in the network has the same number of contacts. Thus, 2.2.14 can be written as

$$r(\infty) = 1 - \exp(-\langle k \rangle \alpha) \quad (4.0.4)$$

where

$$\alpha = \frac{\tau}{g(k)} r(\infty)(\langle k \rangle - 1)$$

Then, 4.0.4 can be written as

$$r(\infty) = 1 - \exp \left( -\frac{\langle k \rangle \tau r(\infty)}{g} - \frac{\tau r(\infty)}{g} \right)$$

$$= 1 - \exp \left( \rho_0 r(\infty) - \frac{\rho_0 r(\infty)}{\langle k \rangle} \right)$$

$$= 1 - \exp(-\rho_0(1 - 1/\langle k \rangle)r(\infty)) \quad (4.0.5)$$

The theoretical prediction for the final epidemic size in the scale-free network is obtained from 2.2.18 with $\alpha_r = 0$ since $\lambda = 0$. Thus, we get the
following parametric equations:

\[
r(\infty) = 2 \int_{1}^{\infty} \frac{1 - \exp(-\Phi x)}{x^3} \, dx \quad (4.0.6a)
\]

\[
\rho_0 = 2\Phi \int_{1}^{\infty} \frac{mx - 1}{m^x^3} [1 - \exp(-\Phi x)] \, dx \quad (4.0.6b)
\]

Figure 6: Comparison between the theoretical predictions of final epidemic size obtained by plotting equation 4.0.5 and simulated final epidemic size on homogeneous networks that have a Poisson degree distribution with \( N = 1000 \).
Figure 7: Comparison between the theoretical predictions of final epidemic size obtained by plotting the parametric equation 4.0.6 and simulated final epidemic size on scale-free networks with $N = 1000$.

In Figure 6, we can see that the general behavior of the simulated epidemic sizes for values of $\rho_0 > 1$ matches the theoretical predictions provided by the random graph with equal degree nodes. Thus variations in the randomness of a graph seems to have little effect on its overall epidemic behavior. The simulated results produce almost identical final epidemic sizes for both the values of $\langle k \rangle$. Additionally, the epidemic sizes are larger for corresponding values of $\rho_0$ than the theoretical predictions for both values of $\langle k \rangle$. The theoretical epidemic size for a higher value of $\langle k \rangle$ is larger because each node has more neighbors (going from 6 to 20) to spread the infection at a faster rate. It is interesting to see a lack of this behavior in our simulated results. This could be attributed to the difference between the degree distributions of the two variations of random graphs used.

In Figure 7, the difference between theoretical prediction and the simulated results for the epidemic size when $\langle k \rangle = 20$ is less significant than the difference when $\langle k \rangle = 6$. However, in both cases, the final epidemic sizes for the simulated results were consistently larger than the analytical prediction. This is different from what the authors got and can be attributed to our
alternative method of calculating the probability of infection. Our probabilities should theoretically always be higher than the ones the authors used as discussed earlier, thus contributing to this behavior in the figure.

![Figure 8: The simulated final epidemic sizes for the two network transmissions when $\langle k \rangle = 20$. The critical point is seen in the intersection of the two curves where $\rho_0 = \rho_{\text{crit}} = 1.439$.](image)

In Figure 8, the general behavior of the two final epidemic size curves is similar to the theoretical predictions seen in Figure 3. Similar results were obtained for simulations with $\langle k \rangle = 6$, hence the above figure is representative of both the average degree distributions. The critical value of $\rho_{\text{total}}$ also matches the predicted critical value of approximately 1.4. At this point, the final epidemic sizes is the same for transmission through both the networks.

Therefore, despite the stochastic nature of this simulation, I was able to obtain results that match the general behavior of those produced in the paper. The shift in positions of various curves are accounted by the variations in my method compared to those of the authors'. Nevertheless, the simulated results validate the theoretical predictions made by the authors.
5 Discussion

This paper based its focus on the importance of incorporating heterogeneity into contact networks when looking at the transmission of a disease and the idea of looking at multiple transmission routes. The importance of taking such factors into account is evident in the way human beings interact with each other. Thus, looking at several transmission mechanisms including heterogeneous network seems like a better representation of human interactions. The authors choose scale-free network due to its relative accuracy in portraying real world networks. In the case of diseases, sexually-transmitted disease contact networks show scale-free behavior. The authors also talk about a ‘hierarchical’ spread of infection that is seen in the scale-free network as being representative of heterogeneity in a network. This means that a few highly connected nodes will significantly influence how the infection spreads in the network and this introduces heterogeneity since disease transmission is no longer random. If these highly connected nodes were to rapidly contract the infection, it would effectively alter the general approach taken to prevent an epidemic. The prevention approaches would need to shift the focus from the general behavior of the network to these highly connected individual contacts.

Both the analytical and numerical results demonstrated that for transmission potential values smaller than the transmission potential critical point ($\approx 1.4$), the scale-free network rapidly ensured a larger final epidemic size than the homogeneous network. Thus, although the number of secondary infections generated by each node is small, the presence of a few highly connected ‘super nodes’ ensured a large epidemic size in the scale-free network. The fact that we see an epidemic for values of $\rho_0$ below 1 in the homogeneous network for my simulations in Figure 8 suggests that the random graph that I chose fails to be a good representation for the pure mean-field case for transmission potential below 1. This could also be a result of an error in creating the network itself. For transmission potential values higher than the transmission potential critical point, the scale-free network has a consistently lower final epidemic size in comparison to the homogeneous network. This is most likely due to a rapid spread of infection in the few highly connected nodes leaving a high amount of poorly connected susceptible nodes in the network which causes the epidemic to end faster. Interestingly, homogeneous networks show the opposite behavior where the spread of infection is slow at first meanwhile ensuring a larger epidemic size for high values of transmission potential.
6 Further Analysis

6.1 Effect of Varying the Recovery Rate (g)

The recovery rate in our model, $g$, represents the rate at which infected individuals recover and move into the recovered category where they stay immune for the duration of the infection. Thus, in our discrete simulation of the epidemic, $g$ can also be interpreted as the fraction of people who recover after the spread of infection at one timestep. I was interested in looking at the effects of varying the recovery rate because the authors choose a specific value $g = 1$ for the simulation of the epidemic and argue that a different value of $g$ would not significantly change the result. I wanted to validate their claim by performing simulations with $g = 0.7$ and $g = 0.5$ and comparing the results with $g = 1$ case. We have the relation:

$$\rho_0 = \frac{\tau \langle k \rangle}{g}$$
$$\tau = \frac{g \rho_0}{\langle k \rangle}$$

From equation 6.1.1, we can determine that for a given value of $\rho_0$, decreasing the value of $g$ from 1 will decrease the value of $\tau$ since $\langle k \rangle$ is held constant. With a reduced force of infection or transmission potential, we should see a lower epidemic size for a lower value of $g$ at each value of $\rho_0$.

I altered my code for the spread of infection slightly in order to incorporate for this change in value of $g$ (Appendix D). The major change was that I had to keep track of infected nodes from the previous time step that evaded recovery due to the reduction in the recovery rate and add them to the group of newly infected nodes from the current time step. The results of the simulations are discussed below.
Figure 9: Results from the epidemic simulation with varying values of $g$. Figures (a) and (b) are results for homogeneous networks with $\langle k \rangle = 6$ and $\langle k \rangle = 20$ respectively. Similarly, figures (c) and (d) are results for scale-free networks with $\langle k \rangle = 6$ and $\langle k \rangle = 20$ respectively. The three values of $g$ used were: 0.5, 0.7, 1.
As discussed above, we would expect to see a lower epidemic size for lower values of $g$ for each fixed $\rho_0$ value and we see a similar result in figure 9. Interestingly, for simulations with average degree values of 20, this difference in epidemic sizes seem to be reduced in comparison to those with average degrees of 6. Numerically, this can be explained by equation 6.1.1, where a big increase in the magnitude of the denominator, $\langle k \rangle$, will reduce the effect of a small increase in the numerator, $g$.

If we think about this from a biological perspective, we can take a hypothetical example where a particular disease always has a constant transmission potential, $\rho_0$. In this case, a decrease in the recovery rate will also cause a decrease in the transmission rate to account for this change in recovery rate. In the above figure, we are seeing a result of this decrease in the transmission rate which will reduce the size of the epidemic outbreak.

However, the most important observation that I made from these simulations was that the qualitative behavior of the final epidemic size curve does not change for all values of $g$. Thus, changing the recovery rate simply scales the result for lower average degree values and this effect gets weak with an increase in the average degree values. The authors decision of using $g = 1$ without the loss of generality makes sense for these simulations.
7 Conclusion

As an undergraduate, I have been fascinated by the use of traditional SIR models to model disease transmission. My interest in the application of graph theory and network science to these epidemic models was what led me to choose this paper for my thesis study. I was also interested in how they used a combination of analytical and numerical results to analyze their disease transmission model. I have learned a lot about networks and epidemic models from this experience. I believe that the incorporation of stochasticity in disease transmission models is vital to understanding the core of epidemiology. As shown in this paper, accounting for various modes of transmission is important for diseases that have more than one significant ways of spreading. This aspect becomes very important when considering disease management plans where minimizing the epidemic size is a major goal. If mathematical disease models are to influence public health policies, it is crucial to estimate parameters as close to real life as possible. It is evident that a good understanding of the knowledge of disease transmissibility and contact network structures are necessary to come up with a good strategy to mitigate the effects of an epidemic outbreak. In conclusion, the authors successfully showed that it is not enough to estimate a final epidemic size using the transmission potential of a disease without fully comprehending the contact network that is driving this epidemic.
8 Appendix

Appendix A: Code for generating a scale-free network

```matlab
function Graph = create_SFgraph(N,d)

inNodes = ones(d+1,d+1);
inNodes(1:(size(inNodes,1)+1):end)=0;  % Puts 0 in the diagonal elements
% node doesn't connect to itself

reNodes1 = zeros(d+1,N-(d+1));
reNodes2 = zeros(N-(d+1),N);
Nodes = [inNodes reNodes1];

Graph = [Nodes;reNodes2];

for k = (d+2):N

    while sum(Graph(k,:)) < (d/2)

        z = rand;

        Links = sum(Graph,2);

        Total = sum(Links);
        Prob = Links/Total;
        Interv(1) = Prob(1);

        for f = 2:length(Prob)
            Interv(f) = Interv(f-1) + Prob(f);  % Interval is working fine
        end

        j = 1;
        while Interv(j) < z
            j = j + 1;
        end

        Graph(j,k) = 1;
        Graph(k,j) = 1;
    end

end
```

Published with MATLAB® 7.14
Appendix B: Code for generating a homogeneous network

```matlab
function Graph = create_graph_rnd(N,d)
    
    Graph = zeros(N,N);
    
    for i = 1:(d*N/2)
        
        j = ceil(N*rand);%randomly choosing a node
        k = ceil(N*rand);
        
        while (j==k)||(graph(j,k)==1) %conditions that can't happen
            
            j = ceil(N*rand);
            k = ceil(N*rand);
            
        end;
        
        Graph(j,k)=1;
        Graph(k,j)=1;
    
    end;

Published with MATLAB® 7.14
```
Appendix C: Code for spreading an infection on a network with $g = 1$

```matlab
function [epidemic,r_inf] = infection(Graph,N,p)

tau = p(1);         %transmission rate
del_t = p(2);       %delta t timestep

f = zeros(1,10);     %choosing 10 random nodes
for a = 1:10
    f(1,a) = randi(N);
end

old_Inf = zeros(N,1); %Infected matrix, 1 represents infected
Recov = zeros(N,1);   %matrix that keeps track of nodes who have recovered
Prob_Inf = zeros(N,1); %probability of infection

for a = 1:10
    old_Inf(f(1,a),1) = 1;
end

sum_inf = 1;         %to start the loop

while sum_inf > 0     %loop runs until there are no new infections

    Inf_nbr = Graph * old_Inf; %the number of neighbours that are infected
    new_Inf = old_Inf;          %to keep track of previously infected nodes

    for i = 1:N
        ra = rand;
        Prob_Inf(i) = 1 - (1-(tau * del_t))^(Inf_nbr(i));
        if (Prob_Inf(i) > ra) && (Recov(i) == 0) %making sure recovered node
            new_Inf(i) = 1;
        end
    end

    Recov = Recov + old_Inf; %everyone who is infected gets recovered next
    old_Inf = new_Inf - old_Inf; %getting rid of previously infected nodes
    sum_inf = sum(old_Inf);
end

epidemic = sum(Recov);     %final epidemic size
r_inf = epidemic/N;
```

Published with MATLAB® 7.14
Appendix D: Code for spreading an infection on a network with varying $g$

```matlab
function [epidemic, r_inf] = infection2(Graph, N, p)
    tau = p(1);         % transmission rate
    del_t = p(2); % delta t timestep
    g = p(3);

    f = zeros(1,10); % choosing 10 random nodes
    for a = 1:10
        f(1,a) = randi(N);
    end

    old_Inf = zeros(N,1); % Infected matrix, 1 represents infected
    Recov = zeros(N,1); % matrix that keeps track of nodes who have recovered
    Prob_Inf = zeros(N,1); % probability of infection

    for a = 1:10
        old_Inf(f(1,a),1) = 1;
    end

    sum_inf = 1; % to start the loop
    while sum_inf > 0 % loop runs until there are no new infections
        Inf_nbr = Graph * old_Inf; % the number of neighbours that are infected
        new_Inf = old_Inf; % to keep track of previously infected nodes
        Recov2 = zeros(N,1);

        for i = 1:N
            ra = rand; % f = rand;
            Prob_Inf(i) = 1 - (1-(tau * del_t)^(Inf_nbr(i)));
            if (Prob_Inf(i) > ra) && (Recov(i) == 0)
                new_Inf(i) = 1;
            end
        end

        new_Inf = new_Inf - old_Inf;
        ind = find(new_Inf < 0);
        new_Inf(ind) = 0;

        if (old_Inf(i) == 1) && (g > f)
            Recov2(i) = 1;
        end

        sum(Recov2);
        new_Inf2 = old_Inf - Recov2;

        Recov = Recov + Recov2; % who is infected gets recovered in next state
        old_Inf = new_Inf + new_Inf2; % getting rid of previously infected nodes

        sum_inf = sum(old_Inf);
    end

    epidemic = sum(Recov); % final epidemic size
    r_inf = epidemic/N;
end
```
Appendix E: Code for running all simulations

```
clc
clear all
matlabpool open local 12
avg_r = zeros(1,51);
avg_r2 = zeros(1,51);

q = 1;
for ro = 0:0.1:5
    r_inf = zeros(50,50);
    r_inf2 = zeros(50,50);
    N = 1000;
    d = 6;
    tau = ro/d;

    parfor j = 1:50 %runs parallelly in 12 workers
        g = create_SFgraph(N,d);
        h = create_graph_rnd(N,d);

        for i = 1:50
            [epidemic,r] = infection(g,N,[tau 1 0.5]);
            [ep2,r2] = infection(h,N,[tau 1 0.5]);
            r_inf(j,i) = r;
            r_inf2(j,i) = r2;
        end
    end

    avg_r(1,q) = mean(mean(r_inf));
    avg_r2(1,q) = mean(mean(r_inf2));
    q = q+1;
end

r = 0:0.1:5;
A = [r' avg_r' avg_r2'];
dlmwrite('HalfRecovery_6.txt',A,'delimiter','	','precision',5) %output file
plot(r,avg_r,'.')
hold on
plot(r,avg_r2)
matlabpool close
```
References


