NETWORK MODELLING OF INFECTIOUS DISEASES

by

Cameline Nafula

A thesis submitted in partial fulfilment of the requirements for the degree of Master Of Science in Applied Mathematics

DEPARTMENT OF PURE AND APPLIED MATHEMATICS

MASENO UNIVERSITY

ABSTRACT

Mathematical modelling of infectious diseases is an important tool for assessing disease dynamics. This branch of mathematics has provided many significant insights concerning the epidemiology of infectious diseases. Previous disease models studied normally took the deterministic approach in modelling. This form of modelling, even though it has helped to bring out some powerful epidemiological insights; has been criticized to be less realistic because it oversimplifies the biology of real world disease dynamics. In recent years, where machines with more computing power have been introduced; more complex modelling scenarios have been developed. Much of this complexity can be incorporated within the population level framework provided by compartmental models. This form modelling of diseases has heavily borrowed techniques from network science to come up with complex epidemiological scenarios that do not make the unrealistic assumptions of homogenous populations. These complex modelling scenarios are believed to give a more realistic picture of epidemiological dynamics thus providing important insights in disease control and prevention. In this study we investigated how network modelling can be incorporated in existing compartmental models. This was done with an aim of gaining a better understanding of different disease properties and more importantly to predict how various disease mitigation and prevention measures can work in different disease scenarios. A comparison of compartmental modelling scenarios and network modelling scenarios is also done using an interactive modelling software called NetLogo.

Chapter 1

Introduction

1.1 Background of the Problem

Many infectious diseases are spread by contact between susceptibles and infectives, this contact may be direct or indirect. Direct transmission maybe spread directly through human contact networks while indirect transmission can be spread through other vectors, that is, environment, water, insects e.t.c.

Mathematical modelling of infectious diseases came in the limelight during the 18th century. Many scholars notably Bernoulli studied how these infectious diseases spread in the populations and came up with models of the said diseases. This arose predominantly from their large public health importance. In 1760 Daniel Bernoulli formulated and solved a small pox model that evaluates the effectiveness of variolation of healthy people with the smallpox virus [1]. However; deterministic epidemiology modeling seems to have started in the 20th century. Hamer; formulated and analyzed a discrete time model in 1906 in his attempt to understand

the recurrence of measles epidemics [30]. In 1911 Ross developed differential equation models for malaria as a host- vector disease. He was interested in the incidence and control of malaria [37]. Other deterministic epidemiology models were then developed in papers by Ross, Ross and Hudson, [5, ?]. Starting in 1926 Kermack and McKendrick published papers on epidemic models and obtained the epidemic threshold result that the density of susceptibles must exceed a critical value in order for an epidemic outbreak to occur [15, ?, ?]. All these papers gave very powerful insights on disease prevention and control. Various targeted public health measures were developed towards this end. These measures were very insightful in development of physiological, biotic, abiotic and social-dynamical tools and techniques of disease prevention and control.

Despite improved measures in disease prevention and control as a result of insights from disease modelling; infectious diseases still continue to be major causes of suffering and mortality. Moreover, infectious disease agents adapt and evolve, so that new infectious diseases have emerged and some existing diseases have re-emerged [39]. Popular books have given us exciting accounts of the emergence and detection of new diseases [41, 42, 43, 48]. The Human Immunodeficiency Virus (HIV), which is the etiological agent for Acquired Immunodeficiency Syndrome (AIDS), emerged in 1981 and has become an important sexually-transmitted disease throughout the world. Drug and antibiotic resistance have become serious issues for diseases such as tuberculosis, malaria, and gonorrhea. Malaria, dengue, and yellow fever have re-emerged and are spreading into new regions as climate changes occur. Diseases such as plague, cholera, enteric fever and hemorrhagic fevers (Bolivian, Ebola, Lassa, Marburg,

etc) continue to erupt occasionally. Surprisingly, new infectious agents called prions have recently joined the previously known agents: viruses, bacteria, protozoa, and helminths (worms). There is strong evidence that prions are the cause of spongiform encephalopathies, e.g. bovine spongiform encephalopathy (BSE, mad cow disease), CreutzfeldtJakob disease (CJD), kuru, and scrapie in sheep [42]. Biological terrorism with diseases such as smallpox or plague has become a new threat. In the 21st century, we have already encountered Severe Acute Respiratory Syndrome (SARS), a disease that emerged from a mutation of a wild animal coronavirus. An outbreak of foot and mouth disease in the United Kingdom in 2001 caused great economic hardships there [52, 53]. Avian influenza has devastated bird populations in SE Asia; moreover, there is the concern that a new human influenza strain will arise by recombination of an avian influenza strain with a human influenza strain. In the future we will undoubtedly face more new infectious disease challenges. It is clear that human or animal invasions of new ecosystems, global warming, environmental degradation, increased international travel, and changes in economic and social patterns will continue to provide opportunities for new and existing infectious diseases [53] The emergence and reemergence of novel and deadly forms of infectious diseases, global climate change, and population growth have increased the need for sound quantitative methods to guide disease interventions.

Due to the ever expanding complexities in infectious disease spread on the population, there is also more urgent need to develop new novel ways of modelling infectious diseases. The contact network approach, originally developed in the field of statistical physics has recently gained much pop-

ularity in the field of epidemiology. A very appealing property of network modelling is its ability to easily depict the complexity of the real world. In particular the degree of distribution captures heterogencity in transmission among hosts, allowing disproportionate role of highly connected individuals super spreaders to be easily investigated [20]. In this study we examined how network modelling of diseases could be integrated into the over-simplistic approach of compartmental modelling.

1.2 Statement of the problem

Most previous modelling initiatives on infectious disease are mostly classical in their approach and thus assume homogenous mixing and perfect mixing of individuals within the disease compartments. This picture is not realistic enough in some instances because as we all know human interactions cannot be assumed to be homogeneously distributed. Compartmental modelling of diseases does not account for the complexity of the social network. Hence in this study we sought to study how network models can be used in modelling of infectious diseases.

1.3 Objectives of the study

The main objective of this study was to investigate how network models could be integrated into compartmental modelling scenarios. The specific objectives were:

- To examine how compartmental disease models are used in epidemiology.
- To examine the limitations of compartmental disease modelling in epidemiology.
- To investigate how network modelling can be used in modelling of diseases.

1.4 Significance of the study

The main significance of this study is to bring out a better understanding of how network models could be utilised to examine communicable disease properties, formulate policies and provide targeted interventions amongst disease populations; this can be very fundamental in reducing the high disease prevalence in developing nations like Kenya.

Chapter 2

Literature Review

Mathematical epidemiology has a long history going back to the small pox model of Daniel Bernoulli in [1]. Daniel Bernoulli used census data and statistical methods to study the advantages of variolation or variolization as inoculation against smallpox. From the year 1960 to 2002, we can see the documentation of the literature and development of mathematical epidemiology [2, 3, 4]. These historical literatures clearly show us that the diseases that motivated the development of epidemiology theory are arguably those due to infectious diseases and in light of the previous literature we note that mathematical models have proved to be useful tools to analyze the spread and control of infectious diseases.

Most of these previous epidemic models incorporate a homogeneous mixing assumption, sometimes called the law of mass action, see [5, 2, 6], whereby the rate of increase in epidemic incidence is proportional to the product of the number of infectious and susceptible individuals. This assumption has been relaxed in some compartmental models by Driessche et al [7] and metapopulation models, see [2, 11, 12, 14], but not eliminated. The *mass-action assumption* is robust in the sense that it is consistent

with several scenarios for the individual-toindividual transmission of disease. In particular, it is equivalent to a model in which all individuals in a population make contact at an identical rate and have identical probabilities of disease transmission to those contacts per unit of time.

On the other hand, the compartmental approach models the behaviour of an infectious disease in a large population of people by considering each individual as being in a particular state. These states are often called compartments, and the corresponding models are called compartment models. The simplest compartment models assume a person can be in one of only two states, either susceptible (S) or infectious (I) [15].

We see that in these classical deterministic compartmental models, mainly based on Kermack and McKendricks SIR model [15] have the main advantages of transparency and simplicity. They are easy to develop and fast to solve, allowing for rigorous sensitivity analysis to explore the dependence of model output on uncertainty of the parameters. They can be used, either in simple and/in structured form; considering age structure [12] and/or geographic component [18], for describing the temporal dynamics of an epidemic and for the assessment of some containment or mitigation strategies, such as mass vaccination [19] or border restrictions [12]. Although these assumptions are sometimes unrealistic, they facilitate mathematical analysis and, in some cases, offer a reasonable approximation.

Populations can be quite heterogeneous with respect to susceptibility, infectiousness, contact rates or number of partners and simple homogeneous mixing models do not allow for extreme variation in host

parameters. On the other hand, the evaluation of realistic, individually targeted, public health intervention strategies, such as household or school/workplace contacts of index cases, in turn requires highly detailed models. Spatially explicit models provide a plausible system in which the precise spatial location of individuals and movement patterns can be employed to evaluate the disease properties hence come up with intervention options [16].

The newer approach to epidemiological modelling is to also model the underlying social network and population interactions with the modern methods of Network Theory [16]. The contact network approach, originally developed for applications in the field of statistical physics, has only recently gained in popularity. In network terminology, individuals, or groups of individuals, are defined as nodes, connections between those nodes are edges, and the number of edges from one node to another is the degree. In network epidemiology, diseases spread from node to node following the edges. If the transmission probability along edges is high enough, an epidemic can occur. The network approach is not inherently different from the other modelling tools. It is simply a more general way of representing epidemiological systems. In fact, most alternative models can be considered as particular cases of network model. For example, modelling an epidemic using an SIR compartmental model is equivalent to using a complete network model in which all the nodes are connected to each other [20].

Although Network Models still require some fitting parameters, they generally provide broader models of epidemic outbreaks than simple equation-based models (i.e. they have richer and more flexible capability for sce-

nario representation [17]. A very appealing property of networks is their ability to easily depict the complexity of the real world. In particular, the degree distribution captures heterogeneity in transmission among hosts, allowing the disproportionate role of highly connected individualssuper spreadersto be easily investigated [21]. Networks also often include lists of attributes to nodes or edges that describe between-edge variation in disease transmission or between-host variation in infectiveness or pathogen excretion patterns.

For example, from the Network Theory point of view an epidemic outbreak simply corresponds to a percolation threshold on an underlying social network. Such an approach to Epidemiological Modelling provides a new theoretical foundation for consistent data analysis, parameter estimation, modelling and forecasting. It is important to recognize that disease spread through population is controlled by two groups of parameters: the disease properties (infection and incubation period, symptoms timing, and infection mechanisms) and the structure of the social network of the population (contact rate, social clustering, and migration). These groups of parameters are completely unrelated and can be studied independently. For instance, the structure of the social network of the community can be studied long before an outbreak occurs within it (if ever).Only by combining together a given disease and a given social network; an epidemic can be created, but this depends on threshold conditions and the relationship between the network connectivity and infection and recovery rates, otherwise the disease dies out [20, 21]. Network Models, sometimes called Agent Based Models, can be used in such instances.

In a nutshell we see that in traditional modelling, the population under

consideration is divided into disjointed classes whose sizes change with time and is assumed to have a constant size. Individual properties of persons within the population classes are not considered. These models assume that all susceptible people in the population are equally at risk of infection from any infected individuals (that is homogenous mixing) and that all infected individuals have a constant and equal infectiousness. Therefore; in light of the previous classical models we have come across in this study, we can see network models represent the most obvious way to relax assumptions that were considered mandatory in traditional mathematical models of diseases, such as the assumption of homogeneous mixing, which most mathematical previous models are based on.

In this study we present an overview of compartmental approach of modelling, parameterization of infectious disease models and some limitations of the classical modelling approach. An example of these limitations is the Severe Acute Respiratory Syndrome (SARS) outbreak experienced in 2002-2003. Estimates of R_0 (basic reproduction number/expected number of new infections created by an infected individual) based on the initial outbreak of SARS ranged between 2.2 and 3.6 [50, 52]. The case fatality ratio was estimated to be between 11% and 13% [52, 53]. For comparison, the U.S. Department of Health and Human Services assigns the greatest pandemic severity ranking to pandemics with a case fatality ratio of 2%; pandemics with this ranking would require the strictest national response strategies [54, 59]. Based on the estimates of R_0, SARS should have caused a great world pandemic with cases numbering easily in the millions. However, for the entire SARS outbreak (from November 1,2002 to July 31,2003), only 8,096 cases were reported with 774 deaths [57].

Certainly, one explanation for the limited spread of SARS is the quick response by world public health agencies, who imposed strict quarantines on infected individuals. Another likely explanation for the discrepancy is that the estimates for R_0 were based on data involving large numbers of transmissions in hospitals, where people have unusually high rates of contact in comparison to other population groupings. Compartmental models assume a fully mixed, homogeneous population; the mass-action assumptionin which each individual has the same amount of contacts as every other individual. Thus, simple compartmental models do not accurately model the increased rate of contact at hospitals and the decreased rate of contact of quarantined individuals. If the population at large had as many contacts as the population within a hospital, perhaps the estimates of R_0 would have been more accurate, and SARS would have infected many more people. Although compartmental models have proven to be quite useful in modelling epidemics, they do no properly model some important aspects of disease spread. From giving an overview of compartmental modelling we will then proceed to discuss and analyze network modelling of diseases, the new approach of modelling infectious diseases that explicitly considers complex human population structures. For many infectious diseases, the transmission occurs in diverse populations, so a complex epidemiological model must divide the heterogeneous population into subpopulations, groups or units, in which then members have similar characteristics i.e. in the SARS example given above, the population can be divided into hospital units, household units, schools units and workplace units etc. These divisions into groups can be based not only the mode of transmission, contact patterns, behavioural patterns, latent pe-

riod, infectious period, genetic susceptibility or resistance, and amount of vaccination or chemotherapy, but also on social, cultural, economic, demographic or geographic factors. Network epidemiological models capture the diverse host interactions that underlie disease transmission [8, 9] thus providing a more realistic picture of the disease dynamics.

2.1 Basic Concepts

2.1.1 Compartmental Modelling of Infectious Diseases

This form of modelling divides the population under consideration into disjointed classes whose sizes change with time t and has constant size N. They are usually classified by a string of letters that provide information about the model structure. If the model is to include vital basic dynamics, then it is assumed that births and natural deaths occur at equal rates and that all newborns are susceptible.

Individuals in the population are classified into the following basic class disjoints of the population;

- **Susceptible (S)** : Initially individual/host is susceptible to infection; no pathogen is present; just a low-level nonspecific immunity within the individual/host. These individuals are naive/ susceptible to the disease.

- **Infectious (I)** : host encounters infectious individual and becomes infected with a micro parasite; abundance of the parasite grows with time. These individuals are able to transmit the parasite to others.
- **Recovered (R)** : The host is neither no longer infectious or is removed (dead). These are immune or removed (dead) individuals that do not contribute to further transmission.

Compartments with labels such as **S**, **I**, and **R** are often used for the epidemiological classes. **SIR** is the mostly widely used compartmental model. Others models i.e. **SEIR** , **SI** , **SIS** are mostly formulated from it. Below is an example of a general disease framework for an **SIR** model.

Figure 2.1: SIR model flow framework

Where;

β = rate of progression to infectious state
= 1/ latent period
γ = rate of recovery
= 1/ infectious period

2.1.2 Network modelling of Infectious diseases.

In network terminology, individuals or groups of individuals are defined as nodes, connections between those nodes are edges, and the number

of *edges* from one node to another is the *degree*. In network terms, disease spread from one node to another following the edges. If the transmission probability along edges is high enough, an epidemic can occur. Network also often include lists of attributes to nodes or edges that describe between- edge variations in disease transmission or between host variations in infectiveness or pathogen excretion patterns.

Below is an example of a network structure of individuals within a population

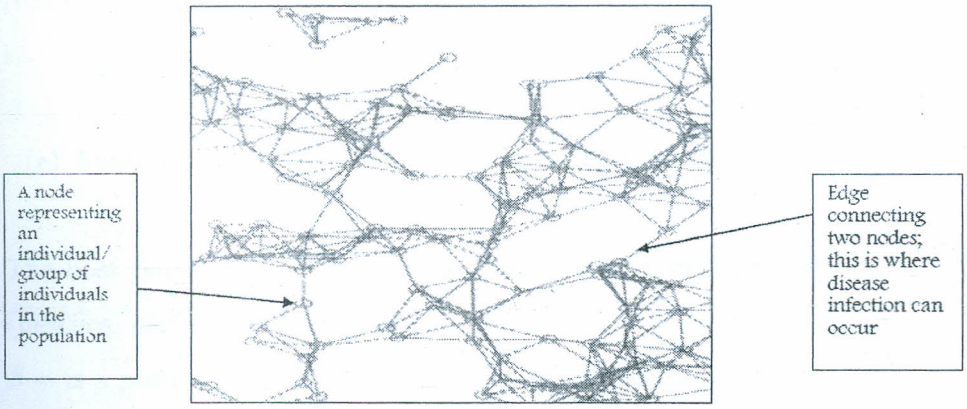


Figure 2.2: Network diagram showing connection of individuals within a population

2.1.3 Definition of concepts in epidemiology

(i) β = **rate of infection (transmission parameter)**

Each susceptible individual is drawn either uniformly random (homogeneity) or through contact network techniques (heterogeneity) from the population. If the person drawn is infected then the susceptible individual

changes his state to infected with probability β. This parameter captures the ability of the disease to be transmitted from one person to another (Contact rate).

(ii) λ = **force of infection**

This is the per capita rate at which susceptible individuals contract the infection. Rate at which new infected are produced = λS where S = Number of susceptibles. There are different kinds of transmissions, that is:

(a) **Frequency- dependent or mass- action transmission**

The host population must be large enough to support the infection i.e. (measles, pertussis)

β = is a product of contact rate (c) and transmission

(b) **Density dependent transmission**

The rate of transmission of these infections depends on the proportion of infectives in the population.

Contact rate (c) increases linearly with population size

(iii) **Basic reproduction number-** R_0

The basic reproduction number R_0 is defined as the average number of secondary infections that occur when one infective is introduced into a

completely susceptible host population [40]. This epidemiological quantity is very fundamental in disease dynamics as it helps in defining which disease become epidemics and also in disease control. In general, for any disease in any host population, the disease can become an epidemic only if $R_0 > 1$. The mathematical condition $R_0 > 1$ can be intuitively interpreted as saying that there exist some conditions under which the disease can grow. These conditions vary with each kind of model as we are going to discuss in the following sections. Clearly, when $R_0 < 1$ each successive infection generation is smaller than its predecessor, and the infection cannot persist. Conversely, when $R_0 > 1$ successive infection generations are larger than their predecessors, and the number of cases in the population will initially increase. This increase does not continue indefinitely. The infection process reduces the pool of *susceptibles*, and hence reduces the probability that an infectious individual contacts a susceptible within its period of infectiousness. This non-linear effect can only be neglected at the beginning of an epidemic. The *basic reproduction ratio* depends on the rate of *contact* between individuals, the probability of transmission given contact, and the time for which an *infected* remains able to transmit the infection. These components are all the subjects of disease control method; isolating those with the infection from the rest of the community, for example in hospital or at home, reduces their *rate of contact* with others; hygiene measures reduce either the contact rate or the probability of transmission given contact; and drug treatment reduces the probability of transmission and/or the length of the infectious period. Vaccination or prophylactic treatment is also used in disease control to reduce the number of susceptibles. Examples of R_0 : 34 for influenza [14]. 10 12 for

chickenpox and 16 18 for measles [34, 35].

(iv) **Network properties**

(a) **Connected network** - A network is connected if it is possible to travel between any pair of individuals by moving along edges of the network. An epidemiologic interpretation of connectedness is that a single individual can transmit infection to any other individual in the population, typically via a number of intermediates.

(b) **The degree or connectivity** of a node, often written as k is equal to the number of neighbours that an individual has on the graph (that is, the number of people to whom our individual is directly connected). Since different individuals may have different numbers of neighbours, we talk about the degree distribution, often written as $P(k)$ of the network.

(c) The average degree, written as $\bar{k}$ or $\langle k \rangle$, can be calculated as $\sum kp_k$. The variance of the degree distribution is given by $\sigma^2 = \sum \left(k - \bar{k}\right)^2 p_k$. This variance equals zero if every individual has the same number of neighbours, in which case we say the network is homogeneous. Otherwise, the network is said to be *heterogeneous*.

(d) **The size of the network**- the *distance* between two nodes is the length of the shortest path that connects them. The *diameter* of a graph is the largest of these values when all pairs of nodes are examined. The *average path* length can be calculated and provides some idea of the typical number of steps between individuals on the network [51].

(c) **Mixing pattern of the network**, [61, 75, 76]. Mixing is usually described with respect to one or more relevant attributes (such as spatial location or an individuals age) and can be summarized by the *mixing matrix*. If the values that can be taken by the attribute(s) are labelled by the subscript i, j then the entries of the mixing matrix, $p_{i,j}$ depict the probabilities that a given contact of an individual of type i is with an individual of type j. In order to describe mixing patterns, the relevant attributes of both an individual and those to whom they are connected must be known. *Assortative mixing* describes situations in which individuals are more likely to interact with other individuals who are similar to themselves in some respect [75, 76]. *Disassortative mixing* describes the opposite situation, in which individuals tend to interact with dissimilar individuals. *Proportionate mixing* (also known as random mixing) occurs when interactions have no particular preference. In order to define proportionate mixing, we imagine the process of constructing a network with a given connectivity distribution p_k. An individual of connectivity k will make k connections in the network. Listing all the connections to be made gives us a set, C, which we call the *"connection pool"*. Since each edge of the network involves a connection between two individuals, the set $\mathbf{C}$ has twice as many elements as there are edges in the network. If there are N individuals in the population, then the Np_k individuals of type k contribute kNp_k connections to $\mathbf{C}$. Consequently, we have that $\mathbf{C}$ has $\sum kNp_k$ elements. Proportionate mixing assumes that connections are made at random from the connection pool. Consequently,

the fraction of connections that are made to individuals of type $\acute{k}$ is given by $\acute{k}Np_{\acute{k}} \sum jNp_j$, regardless of the connectivity of the first individual. Notice that connections are not made at random from the population of individuals (which has connectivity distribution p_k), but rather from the connection pool (which has distribution $kp_k / \sum jp_j$). An interesting consequence of proportionate mixing is that the average connectivity of the neighbours of individuals exceeds the average connectivity of individuals in the population. The former quantity can be shown to equal $\langle k \rangle + Var \langle k \rangle / \langle k \rangle$ which is clearly greater than $\langle k \rangle$ if the network is heterogeneous [60].

Connectivity-based mixing patterns have commonly been used within the STI setting. Here, connectivity equates to the number of sexual partners (or, more likely, to the total number of partners over some period of time). Assortative mixing means that highly sexually active individuals tend to pair up with other highly active individuals and that individuals with few partners tend to be involved with similarly poorly connected individuals.

(f) Another important property of networks is the degree to which they exhibit *local clustering*, also known as *cliquishness, mutuality* or *transitivity* [25, 49, 51]. Considering pairs of connected individuals, we consider how many of their neighbours are common to both of them. The existence of common neighbours leads to the appearance of triangles in the graph (i.e. paths from A to B to C and back to A, where A, B and C are vertices). The clustering coefficient Co is defined as the average fraction of pairs of neighbours

who are also neighbours of each other. That is, suppose that the node i has k_i edges that link it to other nodes. If all the neighbours of i were neighbours to each other, there would be $k_i(k_i - 1) = 2$ edges between them. Suppose that, instead, there are E_i. Then the clusterization around is defined as: $Co_i = \frac{2E_i}{k_i(k_i-1)}$. The clustering coefficient of the network is then defined as the average of Co_i over the network. In a fully connected network, $Co = 1$. In probabilistic terms, Co is the probability that nodes that are neighbours to a particular node, are also neighbours of each other [63].

Betweenness and *centrality* attempt to quantify the importance of different individuals in terms of the population-level properties of the network [63, 65, 66]. More precisely, they provide information about the numbers of paths between pairs of nodes that pass through a given node. Clearly, these properties are global properties of the network. Betweenness (also called betweenness centrality) measures the fraction of shortest paths in a connected component that contain the node of interest [66]. Let $b(j,k)$ represent all of the shortest paths between nodes j and k and $b_i(j,k)$ represent the number of those paths that pass through node i. The betweenness of node i is then given by summing the fractions $g_i(j,k) = \frac{b_i(j,k)}{b(j,k)}$ over all pairs of nodes in the network [65, 63]. Another measure of centrality, information centrality, is similar to betweenness but investigates all paths between nodes that include some other node, not just the shortest paths. The various paths are weighted according to the inverse of their lengths, thus assigning greater importance

to the shorter paths which are likely to be more significant in the spread of infection [65].

(g) *Dynamic network* properties should also be considered i.e. Most notably, sexual partnership networks change as partnerships are formed and break up. They have another notable property in that most individuals tend to be monogamous, so a large fraction of the partnership network consists of isolated nodes (singletons) who are not involved in a partnership and isolated connected pairs of nodes. Any further connections between nodes involve individuals who are involved in several simultaneous partnerships (concurrency) [67].

(h) **Classes of contact network representations (properties) used in disease modelling**. Modelling of different activity contact patterns that usually underlie transmission of diseases usually results in contact networks with different network relationships i.e. *undirected* network, *bipartite* network, *semi-directed* network and weighted network. Undirected networks is used to model person person contacts meaning transmission may occur in either direction along an edge, bipartite networks represent transmission in a hospital between caregivers and wards, the semi-directed network has both directed and undirected edges and can be used to model one-way contacts(flow of disease) from the general population to health-care workers. The weighted network is used to model travel patterns between cities, the edges of networks are weighted by travel influx and disease spread within cities [13, 25, 26, 32, 40, 86].

2.1.4 Epidemiological softwares

Epidemiological programmes allow epidemiologists to build disease models of various diseases alongside the disease dynamics propagating their spread and also allow the programmers to introduce various control measures into the model. These sofware are very useful in simulating disease dynamics especially for network models where analytical models may be very complex to be developed. Some Network analysis software can perform predictive analysis. This includes using network phenomena such as a tie to predict individual level outcomes (often called peer influence or contagion modeling), using individual-level phenomena to predict network outcomes such as the formation of a tie/edge (often called homophily models or particular type of triad, or using network phenomena to predict other network phenomena, such as using a triad formation at time 0 to predict tie formation at time 1. These softwares models are able to illustrate how if there are a certain number of infectious persons in the population; they can spread the infection to the susceptible persons and then the infected individual may recover or not. The results of this simulation can be viewed in monitors and graph plots. Epidemiological analysis software generally consists of either packages based on graphical user interfaces (GUIs), or packages built for scripting/programming languages. Commonly used and welldocumented scripting tools used for network analysis include: NetMiner with Python scripting engine, the statnet suit of packages for the R-statistical programming language, igraph, which has packages for R and Python, the NetworkX library for Python, and the SNAP package for large-scale network analysis in C++. Visual representations of social networks are important to understand network data

and convey the result of the analysis. Visualization often also facilitates qualitative interpretation of network data. With respect to visualization, network analysis tools are used to change the layout, colors, size and other properties of the network representation. All of the tools above contain visualization capabilities, NetMiner, igraph, Cytoscape, NetworkX have the highest level of functionality in terms of producing high-quality graphics. This software makes epidemic problems possible to be explored and analyzed in populations and the patterns that make interaction of many persons [61, 93, 94, 95].

Chapter 3

Methodology

3.1 Compartmental Modelling

A compartmental model is one for which the individuals in a population are classified into compartments depending on their status with regard to the infection under study. They are usually classified by a string of letters that provides information about the model structure. Below we give examples of different compartmental models.

3.1.1 Basic SIR

A basic SIR model without vital dynamics is described by the following set of ordinary differential equations.

$$\frac{dS}{dt} = -\beta SI \tag{3.1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{3.2}$$

$$\frac{dR}{dt} = \gamma I \tag{3.3}$$

The total size of the population, $N = S + I + R$

$S(0) = S_0 > 0$, $I(0) = I_0 > 0$, $R(0) = 0$

Any steady state solution ($S(t) = S^*$), $I(t) = I^*$), $R(t) = R^*$) $\forall t$) of equations (3.1), (3.2) and (3.3) require $I^* = 0$,

$$\frac{dI}{dt}_{t=0} = I_0(\beta S_0 - \gamma) > 0$$

if

$$S_0 > \frac{\beta}{\gamma} = \rho$$

or

$$S_0 < \frac{\beta}{\gamma} = \rho$$

We have that when $\frac{dS}{dt} \leq 0$ then we have always $S < S_0$.

This model is very simplistic in nature, but very important and general conclusion can be derived from it. Some of the questions answered from this model is whether the infection will spread or not? And if it does, how will it develop in time? And when will it start to decline?

Case one

$S_0 < \rho$, $\frac{dI}{dt} \leq 0$, $\forall t \geq 0$ In this case the number of infected remain I_0 and goes to zero as $t \to \infty$. Hence this model does not admit an endemic equilibrium with infection present.

Case two

On the other hand, if we consider the passage of an epidemic through the population i.e. $S_0 > \rho$. From equation (3.1) and (3.2) we get

$$\frac{dI}{ds} = \frac{I(\beta S - \gamma)}{\beta SI} = -1 + \frac{Y}{BS} = -1 + \frac{\rho}{S}, \ (I \neq 0) \tag{3.4}$$

Integrating this equation gives

$$I + S - \rho \, lnS = \kappa(Constant)I_0 + S_0 - \rho \, lnS_0 \tag{3.5}$$

Another important issue that can be calculated is how severe an epidemic will be from (3.4) we can find the maximum of I, I_{max} and it lies at $S = \rho$. Then we have

$$I_{max} = N - \rho + \rho ln\left(\frac{\rho}{S_0}\right)$$

For any values of I_0 and $S_0 > \rho$, I increases from I_0 to I_{max} and an epidemic takes place. Also from (3.1) and (3.3) we have:

$$\frac{dS}{dR} = -\frac{BSI}{YI} = -\frac{S}{\rho}$$

$$S = S_0 e^{[-\frac{S}{\rho}]} \geq S_0 e^{[-\frac{N}{\rho}]} > 0$$

$$0 < S(\infty) \leq N \tag{3.6}$$

We see that $0 < S(\infty) \leq \rho$ hence $R(\infty) = 1 - S(\infty)$. So from (3.6)

$$S(\infty) = S_0 e^{[\frac{R(\infty)}{\rho}]} = S_0 e^{[-\frac{N-S(\infty)}{\rho}]} \tag{3.7}$$

On solving equation (3.7) $S(\infty)$ can be found as its positive root.

Now from (3.4) we can calculate the total number of susceptibles who get the disease in the full course of the epidemic:

$$I_{total} = I_0 + S_0 - S(\infty)$$

When $I(t) \to 0$ and $S(t) \to S(\infty) > 0$ the infection dies out due to lack of infectives and not from lack of susceptibles. The epidemic does not grow unlimited to infect the whole population. There will always be some susceptibles that did not get the disease.

Epidemic behaviour is directly related to the relative removal rate ρ. For a given disease the relative removal rate varies with the community; and it determines why an epidemic of a certain disease can occur in a certain community and not another. For example if the density of susceptibles is high ($S_0 \to \infty$) and the removal rate γ is small (for ignorance, lack of adequate medical care, poverty etc), then an epidemic is likely to occur. In some cases all other functions constant, then γ be high if the disease is very serious and kills the infected very fast e.g. Ebola, in this case an epidemic is unlikely to occur.

Since in real life it is difficult to know the number of infectives each day, it is easier to count the removed, then applying SIR model , we need to get the number of removed per unit time i.e.

$$\frac{dR}{dt} = \gamma I = \gamma(N - R - S) = \gamma I = \gamma(N - R - S_0 e^{\frac{R}{\rho}}) \tag{3.8}$$

Knowing the parameters, we can compute the solution numerically. But since the parameters are not easily known, an approximation is usually made assuming the epidemic is well described by the SIR model i.e. the model fits the disease dynamics in the population.

If an epidemic is not large, $\frac{R}{\rho}$ is small, and if it is large we can expand the exponential in (3.8) to get

$$\frac{dR}{dt} = \gamma(N - S_0 e + (\frac{S_0}{\rho} - 1)R - \frac{S_0 R^2}{2\rho^2}) \tag{3.9}$$

We integrate to get the solution

$$R(t) = \frac{\rho^2}{S_0}\left[\left(\frac{S_0}{\rho} - 1\right) + \alpha tanh\left(\frac{\alpha\gamma t}{2} - \phi\right)\right] \tag{3.10}$$

where

$$\alpha = \left[\left(\frac{S_0}{\rho} - 1\right) + \frac{2S_0(N - S_0)}{\rho^2}\right]^{\frac{1}{2}}$$

and

$$\phi = \frac{tanh^{-1}\left(\frac{S_0}{\rho} - 1\right)}{\alpha}$$

From this the removal rate is found to be:

$$\frac{dR}{dt} = \frac{\gamma\alpha^2\rho^2}{2S_0} Sech^2\left(\frac{\alpha\gamma t}{2} - \phi\right) \tag{3.11}$$

Which has only three parameters, $\frac{\gamma\alpha^2\rho^2}{2S_0} Sec\, h^2$, γ, α and ϕ.

A small introduction of an infection in the population would create an epidemic if the basic reproduction ratio $R_0 > 1$.

Below are examples showing the SIR model, given scenarios and parameters are fitted. The resulting diagrams are simulated from net logo software [58].

N Represents total population

I represents number of infectious persons in the population

β - Is the infection chance - probability of disease transmission from one individual to another

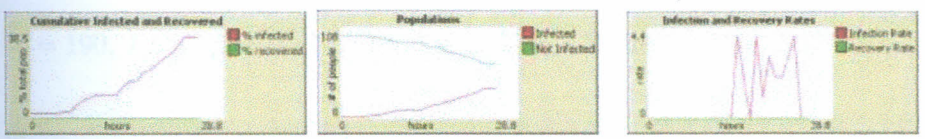
γ - is the recovery rate - probability of an infected individual to recover.

3.1.2 Results of SIR deterministic model

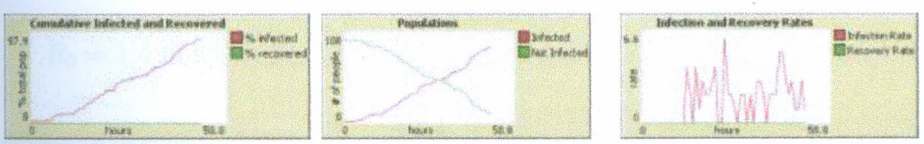
a)Case One

$N = 100$, $I = 2$, $\beta = 0.3$ and $\gamma = 0.6$ $R_0 = 0.7$ and $t = 10$ we have the following graphs

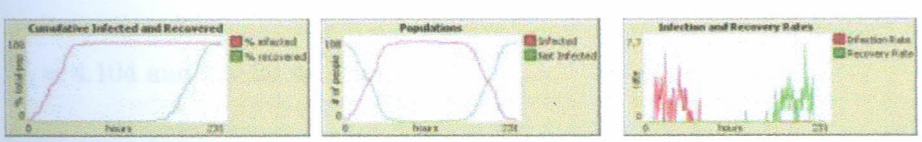
At $R_0 = 1.15$ and $time = 25$



At $R_0 = 2.495$ and $time = 50$, we have

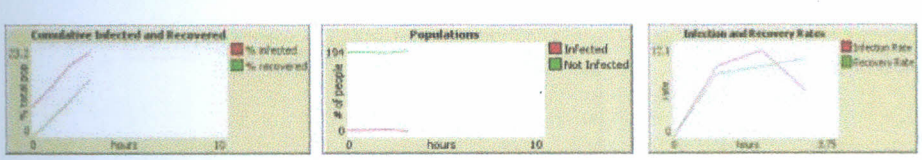


At full simulation where $R_0 = 3.89$ and time is 225 we have

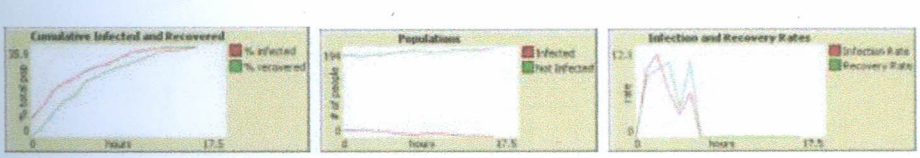


b)Case Two

$N = 190$, $I = 14$, $\beta = 0.25$ and $\gamma = 0.5$ $R_0 = 0.7$ and $t = 3.75$ we have



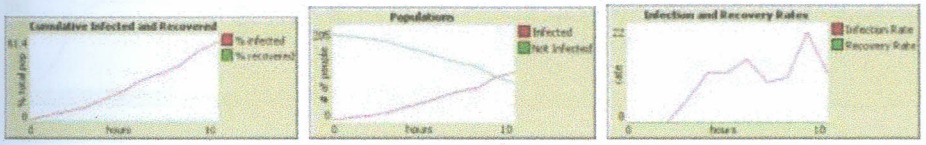
$R_0 = 0.94$ and $t = 15$ we have the full simulations with the graphs above



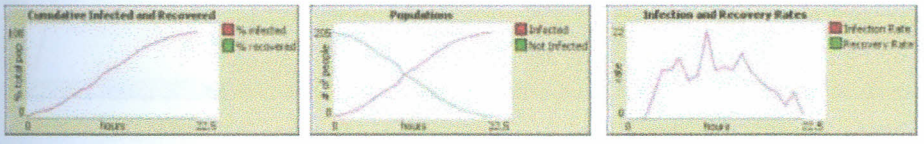
c)Case Three

$N = 190$, $I = 4$, $\beta = 0.45$ and $\gamma = 0.5$ $R_0 = 0.7$ and $t = 25$

$R_0 = 1.4$ and $t = 10$

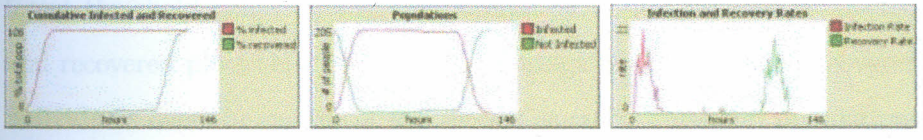


$R_0 = 4.104$ and $t = 20$ we have



$R_0 = 5.14$ and $t = 40$ we have

$R_0 = 5.14$ and $t = 146$ we have



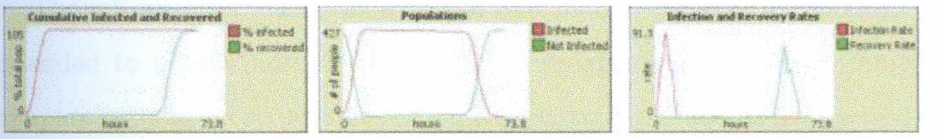
d)Case Four

$N = 400$, $I = 12$, $\beta = 0.45$ and $\gamma = 0.5$ $R_0 = 1.34$ and $t = 5$ we have

$R_0 = 3.17$ and $t = 22.5$ we have

$R_0 = 3.17$ and $t = 46.3$ we have

$R_0 = 3.17$ and $t = 73.8$ we have



From the above results we are able to deduce threshold behaviour of an epidemic in various population sizes. Different diseases have different threshold behaviours in the population. The Cumulative infected and recovered plots show the total percentage of infected and recovered individuals over the course of the disease spread, while the populations

graphs plot the total number of people with or without the disease over time. Lastly the infection and recovery rate graphs plot the estimated rates at which the disease is spreading. βN is the rate at which the cumulative infected changes, and γN is the rate at which the cumulative recovered changes. In the different case scenarios we are able to observe that the cumulative number of people recovering over time, in the event of an epidemic traces out an exponential graph. As stated earlier in the study, an epidemic occurs when $R_0 > 1$, so we can observe from the graphs that as the number of infectious persons increases an epidemic behaviour is observed from the graphs. From the graphs we are also able to see that despite the different disease parameters,in all the simulated SIR graphs we are able to observe in the long run the disease dies out(epidemic burnout). We established that $S(t) \geq e^{-R(0)}$ and since $R(t)$ is obviously less than one ($R(t) < 1$), hence always a fraction of susceptibles in the population, then the chain of transmission eventually breaks down due to lack of number of infected, not lack of number of susceptibles.

Though we have discussed a simple SIR model, the approach can be extended to model more complex disease progressions as well as more complex population structures. Other compartmental models are derived from extending this basic SIR model to represent their disease behaviour and also include more vital dynamics to represent the complex nature of population structures.

passively immune state M to the susceptible state S. Infants who do not have any passive immunity, because their mothers were neither infected nor vaccinated, also enter the class S of susceptible individuals, who can become infected. When there is an adequate contact of a susceptible with an infective so that transmission occurs, then the susceptible enters the exposed class E of those in the latent period, who are infected, but not yet infectious. The incubation period is defined as the period from initial exposure to the appearance of symptoms. Since a person may become infectious before or after symptoms appear, the incubation period is often different from the latent period. After the latent period ends, the individual enters the class I of infectives, who are infectious in the sense that they are capable of transmitting the infection. Models with the M and E epidemiological states have behaviours that are analogous to the models without these states. Examples of models with different epidemiological states are MSEIR, MSEIRS, SEIR,SEIRS, SIR, SIRS, SEI, SEIS, SI, and SIS. Models such as MSEIRS, SEIRS, SIRS, SEIS, and SIS with a flow back into the susceptible class S are always endemic models and never epidemic models, since there is always a threshold above which the disease remains endemic. Models of the types MSEIR, SEIR, SIR, SEI, and SI without return flow into the susceptible class S are endemic if they have a birth and death process, and are epidemic if they are for a short time period without a vital dynamics process.

Traditional compartmental or metapopulation epidemiological models assume that individuals constituting an epidemiological system can be pooled in a small number of functional groups within which the disease incidence rate is simply proportional to the number of susceptible

and infectious individuals. These models are often qualified as "mass-action models ". Within these functional groups, all the individuals are therefore assumed to be epidemiologically identical. The simplicity of these models facilitates the use of analytic techniques to gain general understanding, but at the cost of oversimplifying the biology of real-world disease processes. The weaknesses of simple models have long been clear, particularly when model behaviour has been compared to epidemiologic data, an example is the Severe Acute Respiratory Syndrome (SARS) outbreak experienced in 2002 - 2003 Estimates of R_0 (basic reproduction number/expected number of new infections created by an infected individual) based on the initial outbreak of SARS ranged between 2.2 and 3.6 [50, 52]. The case fatality ratio was estimated to be between 11% and 13% [52, 53]. For comparison, the U.S. Department of Health and Human Services assigns the greatest pandemic severity ranking to pandemics with a case fatality ratio of 2%; pandemics with this ranking would require the strictest national response strategies [54, 59]. Based on the estimates of R_0, SARS should have caused a great world pandemic with cases numbering easily in the millions. However, for the entire SARS outbreak (from November 1, 2002 to July 31,2003),only 8,096 cases were reported with 774 deaths [57]. Certainly, one explanation for the limited spread of SARS is the quick response by world public health agencies, who imposed strict quarantines on infected individuals. Another likely explanation for the discrepancy is that the estimates for R_0 were based on data involving large numbers of transmissions in hospitals, where people have unusually high rates of contact in comparison to other population groupings. Compartmental models assume a fully mixed, homogeneous

their ability to take into account interindividual or intergroup (i.e., internode) variations in epidemiological properties (e.g., degree, infectiveness, recovery rate). With this high resolution, the role played by each individual in the network can be assessed. Since network models capture more heterogeneity among nodes than traditional models, fitted network models can be used to predict the impact of interventions targeting individuals that are critical for disease percolation.

3.2 Network modelling of diseases

The contact network approach, originally developed for applications in the field of statistical physics, has only recently gained in popularity. In network terminology, individuals, or groups of individuals, are defined as *nodes*, connections between those nodes are *edges*, and the number of edges from one node to another is the *degree*. The distribution of the number of such contacts within a population, is called the *degree of distribution*, fundamentally it influences the spread of pathogens through the population. In network epidemiology, diseases spread from node to node following the edges. If the transmission probability along edges is high enough, an epidemic can occur. A very appealing property of networks is their ability to easily depict the complexity of the real world. In particular, the degree of distribution captures heterogeneity in transmission among hosts, allowing the disproportionate role of highly connected individuals superspreaders to be easily investigated [18, 65]. Networks also often include lists of attributes to nodes or edges that describe between-edge variation in disease transmission or between-host variation in infec-

a) Canonical network types (Erdös-Renyi random graph and the Lattice models)

Regular lattices and random graphs both have a long history of use in network theory and as models for the structure of populations. A classic example of a lattice model is provided by Harriss contact process model [72]. Lattice models assume that individuals are located at the sites of a regular lattice and connections are made to some collection of the nearest neighbours of each site. As an example, in one dimension, individuals may be regularly spaced along a line and each is assumed to interact with their k nearest neighbours. In two dimensions, individuals might be cited on a regular square grid, with connections made to their four nearest neighbours (up, down, left and right: the so-called *von Neumann neighbourhood*) or their eight nearest neighbourhood (up, down, left, right, and the four diagonals: the *Moore neighbourhood*). In order to avoid edge effects, periodic boundary conditions are often imposed. The diagram below (fig 3.1) shows a regular lattice network (*von Neumann neighbourhood lattice*).

The random graph [67], most commonly associated with Erdös and Renyi from their graph theoretical treatment of this setting, assumes that each pair of individuals has some probability p, of being connected. These connections are made randomly and independently between the pairs. Below is a diagram showing a random network with eight nodes with a 0.55 degree of connectivity.

Pairs of nodes in an N node network are independently connected at random, with per pair connection probability p. This leads to a bino-

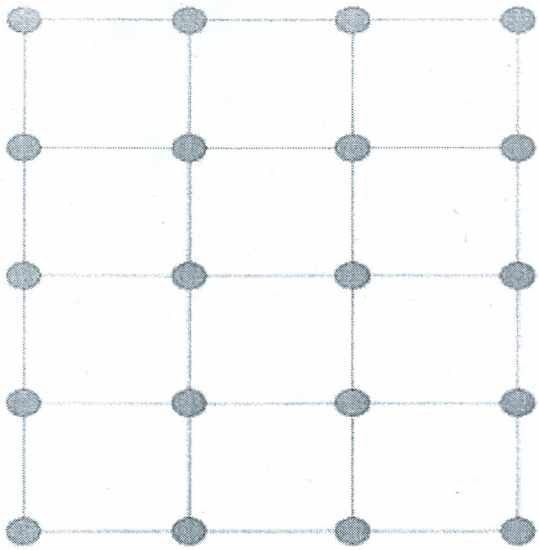


Figure 3.1: Von Neumann neighbourhood lattice *Constructed using Mathematica 9.0 [95]*

mially distributed connectivity distribution, with mean $(N-1)p$. If is sufficiently large, this distribution can be well approximated by a Poisson distribution with mean Np. This connectivity distribution is fairly closely *centered* about its mean: most individuals have a similar number of neighbours. The connectedness of the graph depends on the value of Np if this quantity is small then the graph consists of a large number of disconnected components, but when Np is large most sites are found to form a connected component of the graph. This component is known as the giant component of the graph. A theorem stated by Diekmann and others [58] makes this statement more precise, stating that (for large) the random graph has a (single) giant component if and only if $\Phi = NP$ is greater than one. This component then contains a proportion z of the population, where z is the greatest root of the equation $z = 1 - exp(-\Phi z)$.

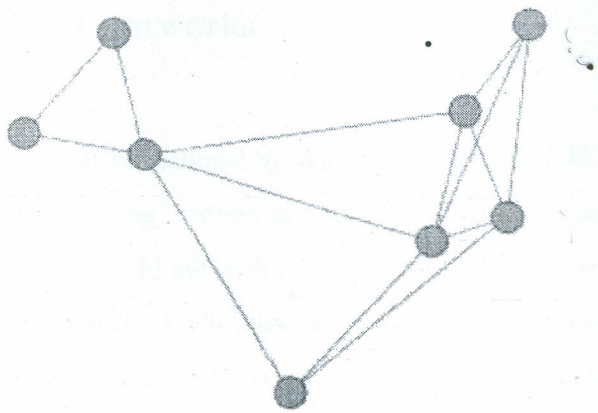


Figure 3.2: Random network with eight nodes and 0.55 degree of connectivity *Constructed using Mathematica 9.0 [95]*

The random nature of connections means that such graphs have little local structure, so exhibit low levels of clustering [67]. On the other hand, path lengths in random networks are relatively short. No individual is especially important in terms of the global structure of the network: since there are no preferred individuals, measures of betweenness and centrality tend to be low. In marked contrast, connections in lattice models tend to be highly localized. Individuals are assumed to be situated on a regular lattice and are only connected to some local neighbourhood. All individuals on a regular lattice (ignoring potential edge effects) have the same number of neighbours. Path lengths in lattices tend to be relatively long: one typically has to pass through a large number of intermediates in order to travel between any pair of nodes. In a one dimensional lattice, path lengths scale linearly with the network size N. Since connections are localized, lattices exhibit high values of the clustering coefficient [67]. As in the case of random graphs, there are no preferred nodes in the network so betweenness and centrality are low.

b) Small-World networks

Small- World networks introduced by Watts and Strogatz [58] have played a major role in stimulating interest in network modeling. Starting from a lattice model, a small world network can be generated by rewiring existing edges within the network. Each edge is examined in turn and is rewired with probability ψ: if it is to be rewired, then one of its ends is left in place and the other is reconnected to a randomly chosen node. (In an alternative formulation, connections are added between randomly chosen pairs of nodes with some probability [69]). This leads to a network that is, in some sense, intermediate between the regular lattice and the random graph. If equals zero, we have a regular lattice and if equals one (all edges are rewired) then we have a random graph.

When $0 < \varphi << 1$, the majority of the connections are local in nature but there are a small number of long-range connections. The surprising result of Watts and Strogatz is that it only takes a relatively small number of these long-range links to give the small world network many of the properties of the random graph. In particular, path lengths in the network rapidly decrease as φ increases. In the small world regime, the network exhibits short path lengths (like the random graph) while still being highly locally clustered (like the lattice) [58, 69]. The connectivity distribution of the small world network remains fairly tightly centered around its mean. Below is an example diagram of a small- world network with a 0.55 degree of connectivity.

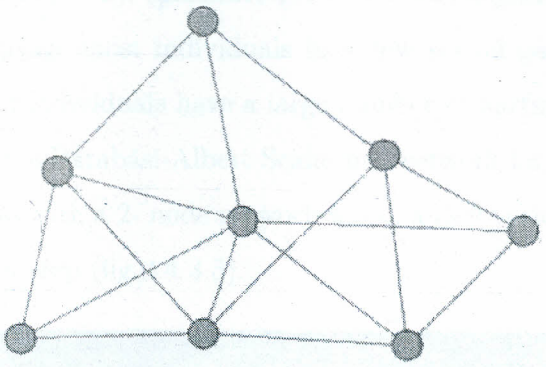


Figure 3.3: Small world network with a 0.55 degree of connectivity *Constructed using Mathematica 9.0 [95]*

c) Scale -Free networks

Scale free networks, as proposed by Barabási and Albert, put the emphasis on the evolution in the construction of the networks. Their original model, just as many real networks, does not start with a given set of N nodes and some links between them. Instead, starting from a small nucleus, the network is grown by the addition of new nodes and links according to some rules that, presumably, mimic the mechanisms by which real networks grow [20, 21].

For the Barabási and Albert scale free network, the connectivity distribution can be shown to follow a power law, with $P\langle k\rangle \sim k^{-3}$ [22]. An important observation is that this distiteribution has infinite variance [22, 23]. The main rule in this model is that of preferential attachment, such that the likelihood of connecting to a node depends on the node's degree, for example according to a probability: $\prod k_i = \frac{k_i}{\sum_j k_j}$. The highly heterogeneous nature of scale-free networks echoes an observation that

has often been made by epidemiologists and sociologists in the sexual partnership setting: most individuals have few sexual partners, while a small number of individuals have a large number of partners [40]. Below is an example of a Barabási-Albert Scale- free network for 220- node network that began with a 2- node network where a new vertex with 2 edges is added at each step (fig 3.4 3.5).

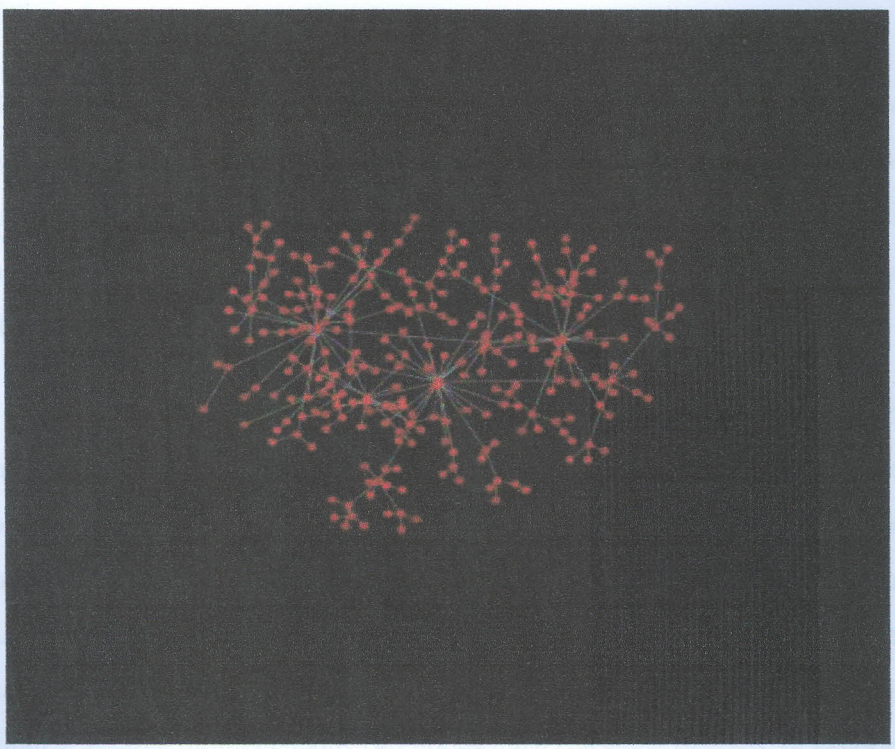


Figure 3.4: Barabási-Albert Scale free network for 220- node network *Constructed using NetLogo 5.0.1 [61]*

You can see the degree distribution of the network in this model by looking at the plot diagrams below the network diagram. The first plot is a histogram of the degree of each node. The second plot shows the same

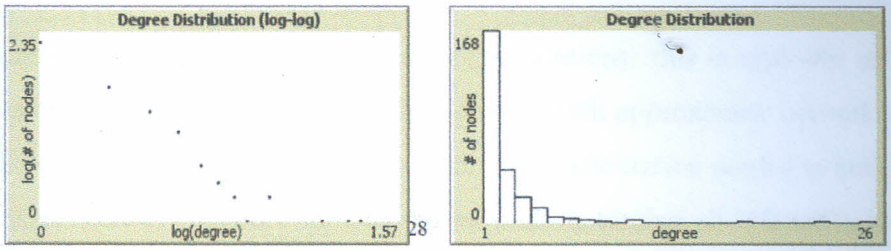


Figure 3.5: Histograms showing the degree of each node in Barabasi-Albert Scale- free network with 220- nodes.

data, but both axes are on a logarithmic scale. When degree distribution follows a power law, it appears as a straight line on the log-log plot. One simple way to think about power laws is that if there is one node with a degree distribution of 1000, then there will be ten nodes with a degree distribution of 100, and 100 nodes with a degree distribution of 10.

Several epidemiologically relevant networks including sexual contact networks have been characterized as scale free [80, 82]; however, researchers have found that realistic contact networks do not always exhibit these well studied structural properties [10, 52]. For example, contact networks underlying the spread of respiratory and airborne diseases tend to have degree distributions that appear more exponential in shape than scale-free, homogeneous (all nodes have the same degree), or random (Poisson degree) distributions.

To construct a contact network for any particular disease or class of diseases, you have to first define an epidemiological contact. An exact contact network model requires knowledge of every individual in a population and every disease causing contact between individuals (e.g. sneezing in the case of airborne diseases or sexual contact in the case of sexually

transmitted diseases). For even small populations, this is typically unfeasible, and thus researchers typically work with approximate networks. There are several techniques for gathering the information needed to build realistic contact network models. These include tracing all infected individuals and their contacts during or following an outbreak [57], sociological surveys of individuals in populations and using census information [12, 29, 28] or the distributions can be inferred from general information about activity patterns, using statistical tools or computer simulations that generate explicit networks from such patterns [35, 23]. Characterizing network structure has become a multidisciplinary cottage industry, with researchers across epidemiology, sociology, biology, computer science and physics searching volumes of data for meaningful patterns [9]. These transmission patterns can be modelled on small scale(households, schools, workplaces, hospitals etc), intermediate scale(in a city, an urban or rural setup) or on large scale(large cities or countries) [11, 23, 24, 31, 38, 87].

3.2.2 Spread of disease on networks

As stated in the previous sections the parameter R_0 is very crucial in disease dynamics. Calculation of R_0 is more involved in the network setting and typically requires simplifying assumptions to be made. As an example, the presence of loops in the network is usually ignored. This enables analysis to be undertaken, albeit at the cost of neglecting some aspects of network structuresuch as cliques that may impact upon the spread of infection. For a static network, each individual has a fixed set of contacts and so an important quantity T [26, 27] is the probability of transmission

from an infective node to a susceptible node along a given edge over the entire duration of their infection. Newman calls this the "transmissibility "of infection [25]. T summarizes core aspects of disease transmission including the rate at which contacts take place between individuals, the likelihood that a contact will lead to transmission, the duration of the infectious period, and the susceptibility of individuals to the disease. In the SIR model described in the previous section whereby the infection is transmitted at a rate β along a given edge and the duration of infectiousness is exponentially distributed with mean $\frac{1}{\gamma}$, $T = \frac{\beta}{\beta+\gamma}$. β can be thought of as a composite of three factors: $\beta = \alpha i_t c$, where α is the number of individuals(nodes) with which a susceptible individual(node) has effective contact(link); $i_t = \frac{I(t)}{N}$ is the proportion of infectious contacts(nodes); and c is the per contact rate at which disease is transmitted between an infectious individual and susceptible individual [49]. The compartmental model can be approximately mapped to a network model in which all individuals have identical numbers of contactsa regular random network. Considering a regular random network with homogeneous degree k. Suppose that disease spreads from infected nodes to susceptible contacts with a rate of $\bar{c}$. Because contacts are random, we can assume that the fraction of contacts that are infected is equal to the fraction of infected individuals in the network as a whole, or $i_t = \frac{I(t)}{N}$. For most populations, however, it is unrealistic to imagine that individuals are in constant contact with all other individuals. Bansal and others [9] thus stated that it is more realistic to interpret the edges in the complete network as 'possible 'contacts and the transmission term $\acute{\beta}$ as a combined effective contact and transmission rate $\acute{\beta} = kT$. In compartmental mod-

els, the identity of a contact is determined randomly and instantaneously for each transmission even. For a homogeneous network, in which every individual has k neighbours, the basic reproductive number (R_0) is $R_0 = T(k-1)$. The average number of secondary infections is proportional to the average number of neighbours minus one [27]. The minus one accounts for the fact that every infectious individual, except for the initial infective (origin), must have acquired infection from one of their neighbours.

As for the case of heterogeneous networks with proportionate (random) mixing, the basic reproductive number R_0 is given by the following formula $R_0 = \left(\langle k \rangle - 1 + \frac{Var(k)}{\langle k \rangle}\right)$ [7, 96]. This expression contains an extra term, involving the variance of the connectivity distribution, that leads to the value of R_0 being inflated in heterogeneous settings. This result was not unexpected, since similar "mean and variance "formulae for R_0 had earlier appeared in a wide number of epidemiological settings [46, 58]. This formula is similar to the formula of average connectivity of individuals neighbours under proportionate mixing [81]. It should be noted that the value of R_0 no longer simply reflects the arithmetic mean of the numbers of secondary infections: in heterogeneous settings, one must adopt a more appropriate notion of the word "average "in the verbal definition of the basic reproductive number. The appearance of the variance in the above formula for R_0 has a surprising impact on the spread of infections in scale free networks [90, 97]. The basic reproductive number is infinite whenever the transmissibility is non-zero: infection can spread on a scale free network whenever there is some possibility of transmission. This result reflects the infinite variance of the connectivity distribution

of the scale-free network. It should be noted that this result only applies in the limit as the number of nodes becomes infinite: for a finite network, the variance will be large but can only be finite. Any real world scale-free network can only have a finite number of nodes and so there would be an epidemic threshold, albeit for a much smaller transmissibility than would be the case in the corresponding homogeneous network (by which we mean a network with the same value $\langle k \rangle$. The impact of heterogeneity has long been recognized in the setting of sexually transmitted infections. Epidemiologists had realized that certain sections of the community, for instance highly sexually active individuals such as sex workers, were at much greater risk of infection than the general population. Such "core groups "are responsible for a large fraction of the cases and transmission events [46, 98]. The prevalence of infection is high within the core group, but low in the general population. In many cases the infection could not spread or persist without the core group: the heterogeneity in the population leads to the basic reproductive number being greater than one. This effect is often given as an explanation of why many infections are able persist at low levels in a population. Heterogeneity in proportionate mixing settings, therefore, promotes the spread of infection compared to the corresponding homogeneous setting. Comparing two settings with the same value of R_0 heterogeneity leads to less severe outbreaks or lower prevalences of infection at endemic equilibrium, because infection tends to be concentrated amongst the highly connected individuals.

There are several approaches used to mathematically analyse the spread of a disease on networks [90, 99, 100, 101, 102, 103]. One method mostly used is the moment generating function methods adapted from an area of

statistical physics called percolation theory [?]. In the Bond Percolation method of disease modelling [?], one can imagine that a parasite initially appears at a random node in a contact network. The disease propagates through the network similarly as in an SIR compartmental model, except that the spread is guided by the structure of the contact network instead of the uniformly random contact patterns of a compartmental model. The initial node remains infectious for some period of time, during which it has the potential to transmit disease to each of its contacts. The secondary cases likewise can transmit disease to their contacts during their infectious periods, and so on [40]. This process resembles simple bond percolation from statistical physics, which models, for example, the flow of a liquid through a porous material [?]. Just as the liquid traverses gaps in the porous material with a characteristic viscosity, a disease spreads from person to person with a characteristic level of infectiousness [40]. In general, percolation theory describes connectivity in random graphs and thus can be applied to predict the size of the infected cluster, that is, the number of nodes reached via parasite transmission along the edges in the network. This approach was initially suggested by Grassberger and Newmann [?, 96] and extended by Meyers [?, 25, 24]. These methods allow us to make predictions for infinite networks with a specified degree distribution. To use it, we must assume that;

a) The contact network is infinite (or quite large)

b) The epidemiologically relevant structure of the network is adequately summarized in its degree distribution.

The second assumption means that we ignore additional structure, like

local clustering, beyond what is expected in an infinite network with the given degree distribution. These assumptions can be tested for their reasonability with comparisons to simulations of disease spread through a finite-sized contact network [40].

Network theory makes a technical distinction between outbreaks and epidemics. An outbreak is a causally connected cluster of cases that, by chance or because the transmission probability is low, dies out before spreading to the population at large. In an epidemic, on the other hand, the infection escapes the initial group of cases into the community at large and results in population-wide incidence of the disease. The crucial difference is that the size of an outbreak is determined by the spontaneous dying out of the infection, whereas the size of an epidemic is limited only by the size of the population through which it spreads. The fate of an outbreak depends on both the level of contagion and the structure of the underlying network. To model disease transmission, every edge in a network is given a probability of disease transmission along it (T_{ij}), that is, the probability that individual i , if infected, will transmit disease to individual j during his or her infectious period. It is reasonable to assume that the T_{ij}s are independently and identically distributed (*iid*) random variables, then we can make calculations based solely on the average of these probabilities $\bar{T}$ or T as defined above [25, 96]. This value summarizes core aspects of disease transmission including the rate at which contacts take place between individuals, the likelihood that an encounter will lead to transmission, the duration of the infectious period, and individual susceptibility. In these calculations, we use the degree distribution of a network $P(k)$ to indicate its structure. Probability generating functions (pgfs) are

functions that completely describe discrete probability distributions. For infectious disease modelling, the pgf of a contact networks degree distribution summarizes useful information about the structure of the contact network. The pgf for a degree distribution is

$$G_0(x) = \sum_{k=1}^{\infty} P_k x^k \tag{3.16}$$

where P_k is the relative frequency of nodes of degree k in the network. If we choose a random edge and follow it to the nearest vertex, then the pgf for the "excess degree ", that is, number of edges emanating from that vertex other the one along which we arrived at;

$$G_1(x) = \frac{\sum_{k=1}^{\infty} k P_k x^{k-1}}{\sum_{k=1}^{\infty} k P^k} \tag{3.17}$$

The average degree and the average excess degree equals the derivative of these equations at $x = 1$ i.e. $\langle k \rangle$ and $\langle k_e \rangle$ respectively;

$$\langle k \rangle = \sum_{k=1}^{\infty} k P^k and \langle k_e \rangle = \frac{\sum_{k=1}^{\infty} k(k-1) P^k}{\sum_{k=1}^{\infty} k P^k} = \frac{\langle k \rangle^2}{\langle k \rangle - 1} \tag{3.18}$$

Using straight forward probabilistic arguments in [24, 40, 96], we sequentially derive pgfs for the distributions of;

a) The number of edges emanating from a node reached along a randomly chosen edge

$$G_1(x) = \frac{G_0\prime(x)}{\langle k \rangle} \tag{3.19}$$

above where $\langle k \rangle = G_0{}'(1)$ as described in (3.18) above.

b) The number of edges along which disease transmits from an infected node;

$$G_0(x, \bar{T}) = G_0(1 + (x - 1))\bar{T} \tag{3.20}$$

Where $\bar{T}$ is the average probability of transmission) and;

c) The size of outbreaks stemming from a single introduction of disease

$$H_0(x; \bar{T}) = xG_0(H_1(x; \bar{T})\bar{T}) \tag{3.21}$$

Where H_1 is defined as the self-referential pgh

$$H_0(x; \bar{T}) = xG_1(H_1(x; \bar{T})\bar{T}) \tag{3.22}$$

For a random network with a given degree distribution, there typically exists a threshold transmission rate below which small, finite-sized outbreaks occur and above which large scale epidemics , comparable to the size of the network, are possible. This epidemic threshold is analogous to the well-studied percolation threshold [?] and it depends on the network structure. In an undirected random network with a given degree distribution, for example, the epidemic threshold is $\bar{T}_c = \frac{1}{G_1{}'(1)}$. Intuitively $G_1{}'(1)$ can be interpreted as the pgf for the number of susceptible contacts for each infected individual. So, as in similarity with the epidemiology of the basic reproductive number (R_0), where an epidemic occurs when $R_0 > 1$ i.e. if $\bar{T} \bullet G_1{}'(1) > 1$. $\bar{T}$ is related to the traditional (R_0) according to $R_0 = \frac{\bar{T}\langle k \rangle^2}{\langle k \rangle - 1} = \bar{T}\langle k_e \rangle$, where $\langle k \rangle$ and $\langle k \rangle^2$ are the mean degree and mean square of the network respectively.

Highly connected networks, with ample opportunities for transmission, have low epidemic thresholds. In such networks, even mildly infectious individuals will be able to cause epidemics. Less connected networks will have higher epidemic thresholds. The value of the epidemic threshold $\bar{T}_c$ the predicted average size of an outbreak $\langle s \rangle$ and probability of an epidemic P were first derived by Newmann [?]. By nesting pgfs for the number of new infections emanating from an infected vertex one can construct a pgf for the size of an outbreak, and hence derive the average size of an outbreak:

$$1 + \frac{\bar{T}\langle k \rangle}{1 - \bar{T}\langle k_e \rangle} = \frac{1 - \bar{T}G_0(1)}{1 - \bar{T} \bullet G_0\prime(1)\prime} \tag{3.23}$$

This expression diverges when an outbreak becomes a large-scale epidemic. The epidemic threshold $\bar{T}_c$ (above) is the transmissibility value that marks this point. The probability of a full-blown epidemic S is derived by first calculating the likelihood that a single infection will lead to only an outbreak instead of a full-blown epidemic, and then subtracting that value from one:

$$S = 1 - \sum_{k=1}^{\infty} p_k (1 + (u-1)\bar{T})^k = 1 - G_0(u, \bar{T}) \tag{3.24}$$

where u is the probability that the person at the end of an edge does not have the disease and is the solution to the equation;

$$u = \frac{\sum_{k=1}^{\infty} k p_k (1 + (u-1)\bar{T})^{k-1}}{\sum_{k=1}^{\infty} p_k} \tag{3.25}$$

u can be solved using the numerical root finding methods, while S is the expected proportion to solve for is also the expected proportion of the population that will be infected should an epidemic occur [40, 96].

Ideally, an epidemic model would incorporate the following realities of human-to-human contacts:

- A given individual has contact with only a finite number of other individuals in the population at any one time, and contacts that can result in disease transmission are usually short and repeated events.
- The number and frequency of contacts between individuals can be very heterogeneous.
- The numbers and identities of an individual 's contacts will change as time goes by.

The bond percolation theory described above captures the first two points mentioned above i.e. it considers contact network the contact network to be static and it predicts the final state of an outbreak but not the temporal progression of the disease. Although this assumption may be reasonable for rapidly spreading diseases, there are many situations in which the underlying network will change considerably during an outbreak. For example, concurrent and serial contacts are known to strongly influence the spread of sexually transmitted infections like HIV [?, ?, ?]. Epidemics in dynamics have been modelled using high- dimensional pair- approximations methods by Altmann and Eams [?, ?] and moment closure methods on dynamic contact networks by Fergusson and Bauch [?, ?].

Volz [?] recently developed a low-dimensional system of nonlinear ordinary differential equations to model the dynamical progression of a disease spreading through static random networks with arbitrary degree distributions. This model improves on the bond percolation approach in that it

both predicts the temporal progression of disease and allows for changing structure in the underlying network. Specifically, the model considers a simple class of dynamic networks in which pairs of edges are randomly chosen and swapped. As stated above human contact patterns are potentially complex, as the numbers and intensity of contacts can vary considerably across a population. Furthermore, contacts are transitory events such that the identities of ones contacts change in time. To capture such heterogeneity, Volz and Meyers [?] came up with a neighbour exchange (NE) model as a simple extension of a static contact network model. In this model, an individuals number of concurrent contacts remains fixed while the composition of those contacts changes at a specified rate. The model assumes that at any given time, an individual will be in contact with an individual-specific number of neighbours with whom disease transmission is possible. Each contact is temporary, lasting a variable amount of time before coming to end, at which point the neighbour is replaced by a different individual. Let the population of interest consist of n individuals, each of which falls into one of three exclusive states: susceptible, infectious or recovered. At some time , an individual will have contacts with other individuals (i.e. φS): $(\sigma, \varphi_1), (\sigma, \varphi_2), \ldots, (\sigma, \varphi_k)$. Only undirected contact networks will be considered such that if there exists a contact (σ, φ) there will also be a contact (φ, σ). In network terminology, a directed link, denoted (σ, φ), is called an *arc*. An undirected link, denoted $\{\sigma, \varphi\}$, is called an *edge*. The *degree* of a node σ is the number of edges connected to the node. The term *contact* will be used specifically to denote a directed *arc* in the network, where two arcs correspond to each undirected edge. The k - *degree* of a node will be the number of concur-

rent contacts to/from the node. The NE model by Volz and Meyer [?] [?] assumes that the identities of a nodes neighbours will continually change, while the total number of current neighbours remains constant. This occurs through an exchange mechanism in which the destination nodes of two edges are swapped. For example, two nodes σ and $\sigma\prime$ with distinct contacts $(\sigma\prime, \varphi\prime)$ may exchange contacts such that these are replaced with $(\sigma, \varphi\prime)$ and $(\sigma\prime, \varphi)$. There are two edges and four contacts involved in each edge swap. The fate of each edge and contact is summarized in the following pseudochemical equation:

$$\begin{aligned} \{\sigma, \varphi\} + \{\sigma\prime, \varphi\prime\} &\dashrightarrow \{\sigma, \varphi\prime\} + \{\sigma\prime, \varphi\} \\ \{\sigma, \varphi\} + \{\sigma\prime, \varphi\prime\} &\dashrightarrow \{\sigma\prime, \varphi\} + \{\sigma, \varphi\prime\} \\ \{\varphi, \sigma\} + \{\varphi\prime, \sigma\prime\} &\dashrightarrow \{\varphi, \sigma\prime\} + \{\varphi\prime, \sigma\} \end{aligned} \tag{3.26}$$

Any given contact (σ, φ) will be reassigned to $(\sigma, \varphi\prime)$ at a constant rate ρ. Equivalently, a given edge is swapped at a rate $\rho/2$. The model is given the following equation;

$$\begin{aligned} \frac{d\theta}{dt} &= -\theta_r \frac{F_{SI}}{F_S}, \\ \frac{dF_{SI}}{dt} &= rF_{SI}\left(-1 - \frac{\delta}{F}\frac{F_{SI}}{F_S} + \frac{\delta}{F}\frac{F_{SS}}{F_S}\right) + F_{SI}\mu - \rho(F_{SI} - F_I F_S) \\ \frac{dF_{SS}}{dt} &= -2rF_{SI}\left(\frac{\delta}{F}\frac{F_{SS}}{F_S}\right) + F_{SI}\mu - \rho(F_{SS} - F_S F_S) \\ \frac{dF_I}{dt} &= rF_{SI}\left(\frac{\delta + 1}{F}\frac{F_{SS}}{F_S}\right) - \mu F_I \end{aligned} \tag{3.27}$$

The model consists of four core dynamic variables: θ is the fraction of degree one nodes that are still susceptible, F_{SI} is the fraction of edges in the network connecting a susceptible node and an infected node. F_{SS}

is the fraction of edges in the network connecting two susceptible nodes, and F_I is the fraction of edges in the network adjacent on an infected node, regardless of the state of the node at the other end of the edge. There are also four fixed parameters: r is the transmission rate, μ is the recovery rate,$g(x)$ is the pgf for the network degree distribution, and ρ is the neighbour exchange rate. To simplify the equations, we also use three helper values: $F = g\prime(1)$ is the total number of edges in the network, $F_S = \theta g\prime(\theta)$ is the fraction of edges adjacent on a susceptible node, and $\delta = \frac{\theta g\prime\prime(\theta)}{g\prime(\theta)}$ is the average excess degree for a susceptible node selected by following a random chosen $I \leftrightarrow S$ edge. *Excess degree* is defined as the degree of the node minus one. The last two of these helper values vary as the epidemic progresses through the network. Finally, the equations highlight the commonly appearing term rF_{SI}, which is the rate of transmission events in the network per unit time. This model tracks the state of each edge and each stub (one end of an edge) as disease spreads through the network.

The first equation describes the decline in the number of degree one nodes that are susceptible. If a degree one individual is susceptible, then F_{SI}/F_S is the probability that his or her single edge is connected to an infected node, and r is the probability of transmission along that edge. Thus, $r\frac{F_{SI}}{F_S}$ is the rate at which such nodes become infected. The second equation describes the change in the fraction of edges connecting susceptible nodes to infected nodes. Considering one such edge, connecting a susceptible node attached to an infected node. The first term in the equation performs the accounting required due to transmission events. When a transmission event occurs, turning a susceptible node to an infected

node, we must;

1. Remove the single $S \leftrightarrow I$ edge carrying the infection.

2. Remove any other $S \leftrightarrow I$ edges adjacent on the newly infected node and,

3. Add any $S \leftrightarrow S$ edges adjacent on the newly infected node.

The second term accounts for recovery events, which convert edges from $S \leftrightarrow I$ to $S \leftrightarrow R$. The final term corresponds to the impact of neighbour exchanges on the fraction of $S \leftrightarrow I$ edges. By randomly mixing the network edges, neighbour exchanges slowly bring the fraction of $S \leftrightarrow I$ edges to the expected fraction of such edges, $F_S F_I$. The value $F_S F_I$ corresponds to the expected number $S \leftrightarrow I$ edges in a network that has the same fraction of stubs connected to susceptibles and infecteds as the original network, but has edges redistributed randomly between nodes.

The third equation describes the change in the fraction of edges connecting susceptible nodes to other susceptible nodes. Considering one such edge connecting a susceptible with another a susceptible node; the first term in the equation corresponds to loss of $S \leftrightarrow S$ edges following an infection transmission event. The $S \leftrightarrow S$ edges that are turned into $S \leftrightarrow I$ edges, added in the previous equation, must be subtracted here. The rate of change on $S \leftrightarrow S$ edges is doubled, because both their endpoints become infected at the same rate. As before, the final term corresponds to the impact of neighbour exchanges on the fraction of $S \leftrightarrow S$ edges. Neighbour exchange acts like a spring with tension ρ

slowly bringing the fraction of $S \leftrightarrow S$ edges to the expected value $F_S F_S$ for a comparable random network.

The final equation describes the change in the fraction of stubs that are adjacent on an infected node. The first term performs accounting due to newly created infections. When a susceptible node becomes infected all of the edges emanating from the node add to the class of F_I stubs, including the single edge involved in the transmission. The second term in the equation governs the loss of infected stubs through the recovery of infected nodes. Below are examples showing the spread of an infection in a network structure, the scenarios and parameters used are similar to the ones used in the SIR compartmental modelling examples in section 3.1 above. i.e. for the four cases discussed above. The four cases discussed below have similar number of individuals (nodes), the probability of disease spread chance, and the recovery chance as the four cases discussed above. The resulting diagrams are simulated from net logo software [61].

Other parameters included in all these simulation case scenarios are;

- The average degree node for all this simulations the average degree node is 12 for all the simulations except the last scenario (case) five whereby we have varied the degree node to 6 while all the other factors are held constant to case one (for comparison purposes i.e. the see the effect of degree variance on disease spread on a network) and (averaged on highly connected, least connected and medium connected individuals). The average degree node can be derived from all the social contacts an individual encounters in the network i.e. at home, at the workplace, at school etc.

- The disease originates from one infectious individual in the network (patient zero)- shown as a red node in the network diagrams) and spreads to other susceptible individuals who become infectious and spread on the disease to other susceptible nodes(secondary infections.) **Initial outbreak size = 1**

- If a node (individual) does recover, there is some probability that it will become resistant to the disease in the future. In all these simulations the **probability of gaining resistance to the disease is 0.3**

- In the network diagram and the plots; red infectious individual(s), green susceptible individual(s) and grey are resistant individuals.

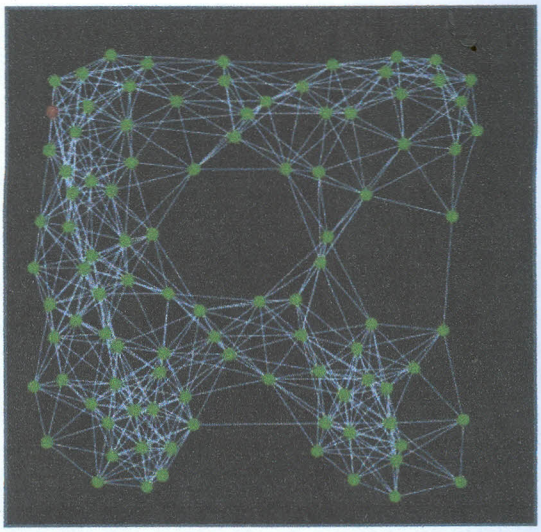
- The plots are drawn at time $t = 52, 105, 168, 211, 333, 418, 656, 1290, 2010$ and 3140 for all the plots, though in some cases the disease dies out before the final time step of 3140.

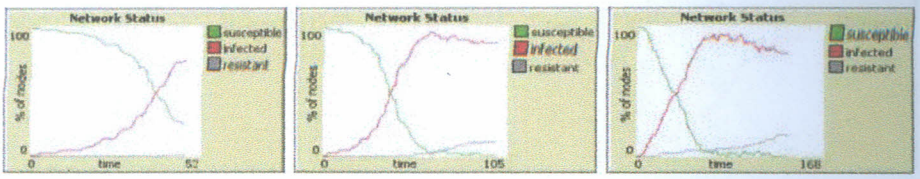
- The different cases have different parameters for infectiousness and recovery, implying they infer different diseases, but in these cases the parameters are not representing any disease per- se but chosen randomly and in comparison to the previous compartmental SIR models so as to understand network settings on disease setups.

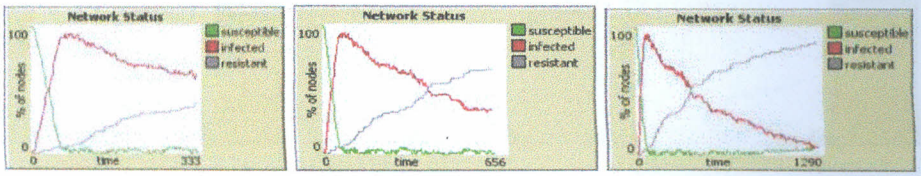
3.2.3 Results of Network modelling Cases

a) **Case One**

$N = 100$, probability of disease spread = 0.3, probabilty of recovery is 0.6

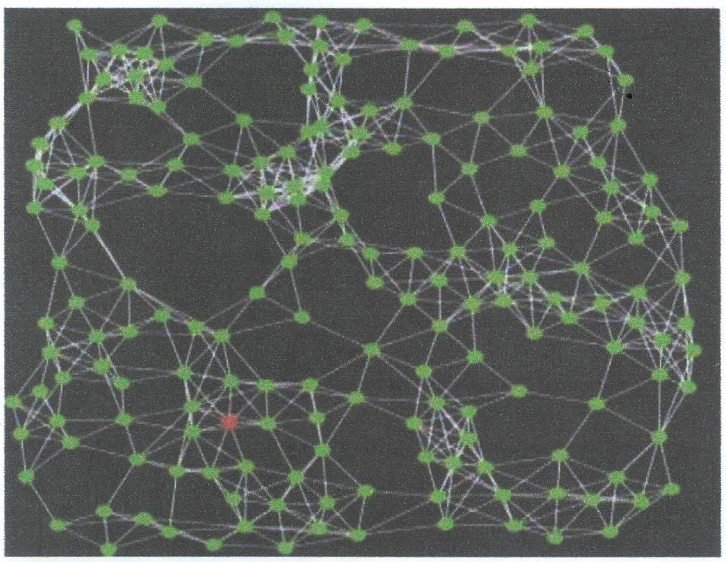


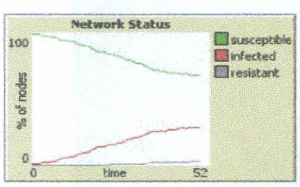


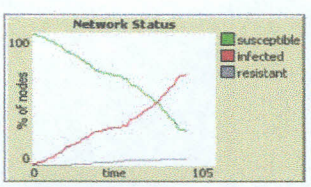


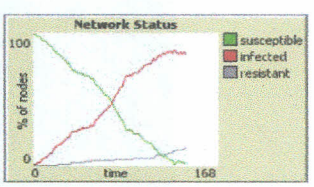
b) **Case two**

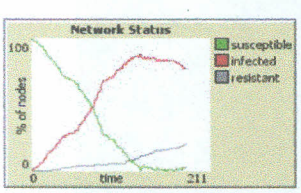
$N = 190$, probability of disease spread $= 0.25$, probabilty of recovery is 0.5

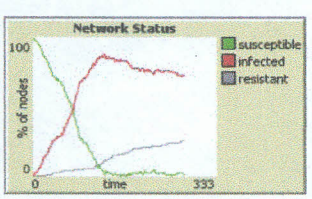


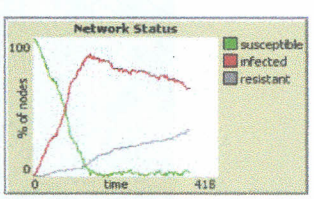

Network Status
100
0
% of nodes
0
time
52
susceptible
infected
resistant

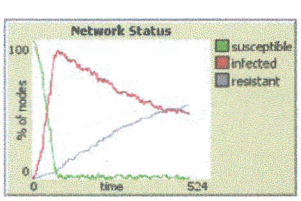

Network Status
100
0
% of nodes
0
time
105
susceptible
infected
resistant

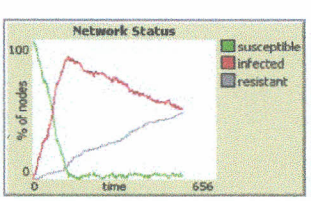

Network Status
100
0
% of nodes
0
time
168
susceptible
infected
resistant

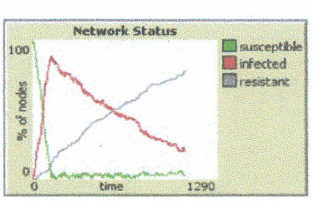

Network Status
100
0
% of nodes
0
time
211
susceptible
infected
resistant

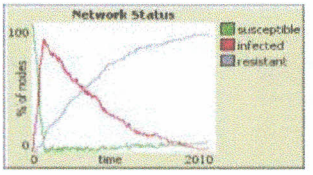

Network Status
100
0
% of nodes
0
time
333
susceptible
infected
resistant


Network Status
100
0
% of nodes
0
time
418
susceptible
infected
resistant


Network Status
100
0
% of nodes
0
time
524
susceptible
infected
resistant

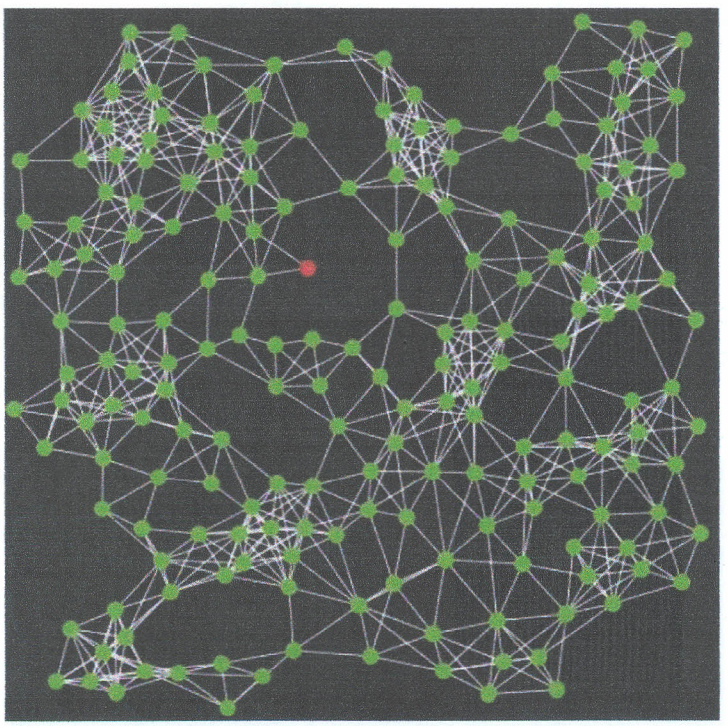

Network Status
100
0
% of nodes
0
time
656
susceptible
infected
resistant

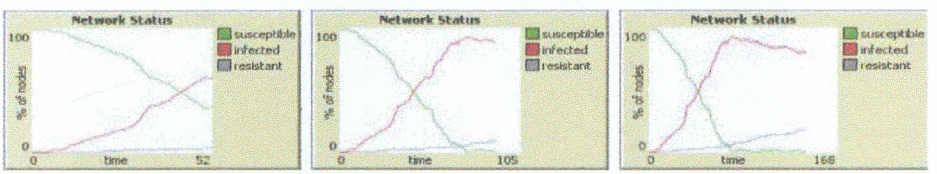

Network Status
100
0
% of nodes
0
time
1290
susceptible
infected
resistant

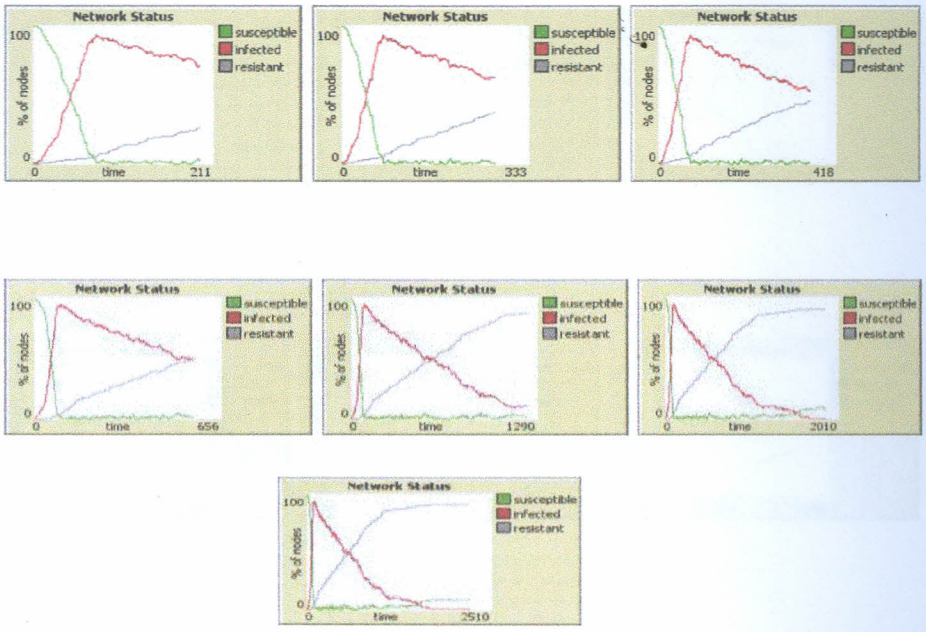


c) **Case three**

$N = 190$, probability of disease spread $= 0.45$, probabilty of recovery is 0.5

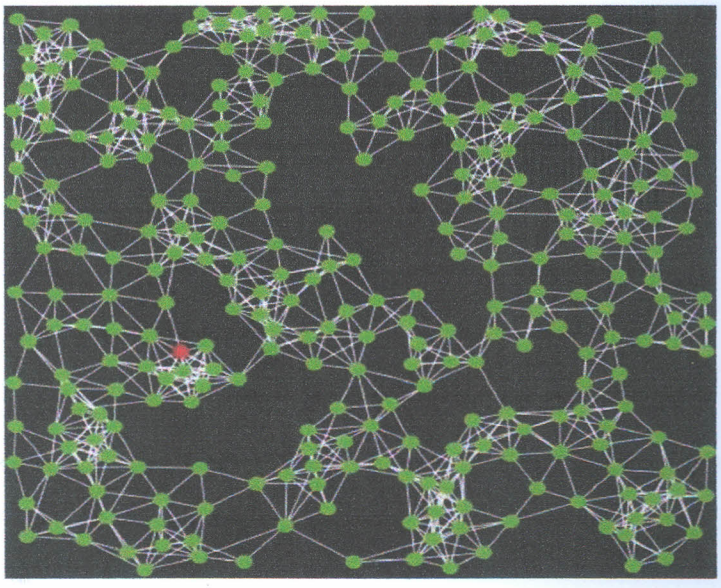


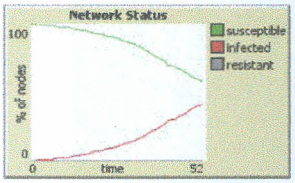


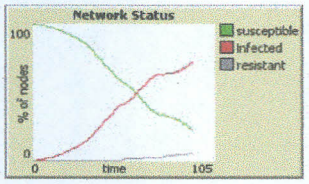


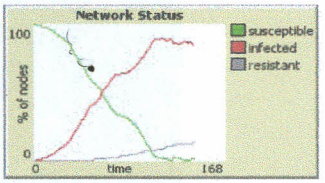
d) **Case four**

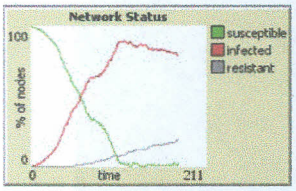
$N = 400$, probability of disease spread $= 0.45$, probabilty of recovery is 0.5

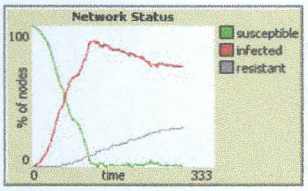


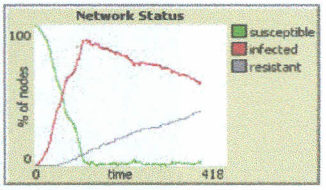


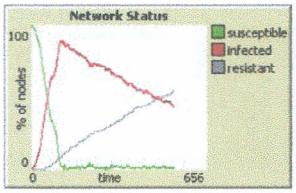


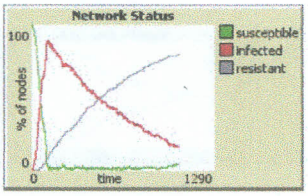


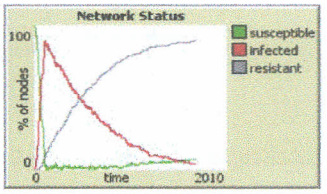


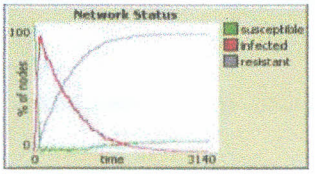






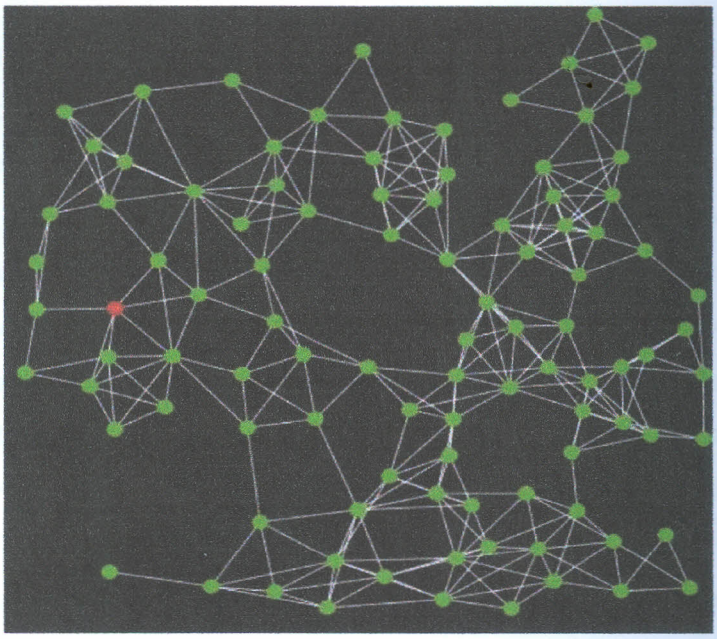


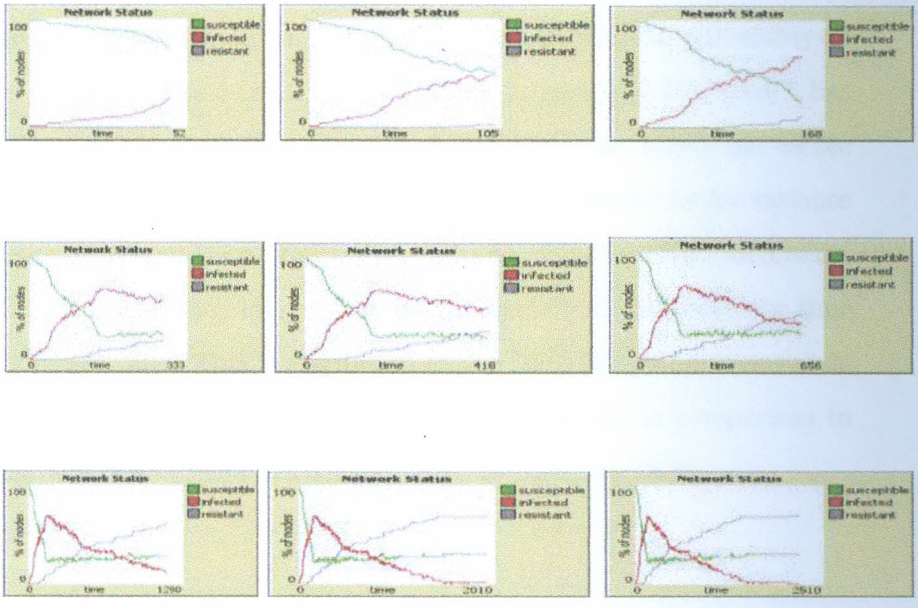




e) **Case five**

$N = 400$, probability of disease spread $= 0.45$, probabilty of recovery is 0.5





The % of infectious nodes (individuals) eventually die out in the network. The disease modelled in this network signify a highly contagious disease

that reaches high levels of infectious very quickly but eventually dies out due to recovery rates and immunity gained by some of the nodes either due to vaccination or gaining resistance to the disease after recovery. It can signify a highly contagious virus where individuals initially susceptible of the virus gain resistance to the disease after infection. There is no transition $R \rightarrow S$, the removed become resistant to the disease. Eventually the disease dies out in the population. The disease has some critical points in the network above which the disease increases its infectiousness or below which the infectiousness slows down and eventually dies out in the network. No of individuals (nodes) in the network also influence the disease dynamics i..e the disease spreads faster in case one(100 nodes) than in case 2(190 nodes). Also comparing case 2 and case 3 which all have the same number of nodes (190 nodes), the disease spreads and dies out faster in case 3 than in case 2. This is because the probability of disease spread in case 3 is higher (0.45) that in case 2(0.25). Case 4 has (400 nodes) but has all other parameters similar to case 3(190) but we see the disease takes much longer to die out in the network. As for variance in average degree node i.e. for case one (12) but 6 for case five, we observe that disease takes much longer time in the network almost twice the time it takes in case one. This is attributed to the fact that an infectious node in case 5 has fewer contacts to spread the disease too in comparison to case one where he or she has almost twice as many contacts.

Chapter 4

Conclusion and recommendations

4.1 Conclusion

Contact network modelling of disease transmission allows us to make quantitative predictions about disease outbreaks and epidemics. From the results observed in the previous chapter and using the methodology of networks, we are able to see effect of different parameters and average node distribution on the disease dynamics. We have also been able to observe that prior immunisation or vaccination of a disease against it can dramatically influence the host network. This is observed through variance in probability of gained resistance. Reducing S, the number of susceptibles in the population, through vaccination is a critical and often long lasting disease control strategy. When susceptible individuals are effectively immunised, they move to resistant compartment without experiencing the disease. if the fraction of susceptibles is reduced to less than the transmission parameter then disease is unable to spread. This

shows that partial vaccination of the population, reducing S to a small but non zero value, is sufficient to protect the population as a whole (herd immunity). From the network we are able to observe connectivity patterns of individuals some are highly connected while others are lowly connected or mediumly connected. From the network we are also able to observe clusters of neighbourhoods or communities. The results shed light on contact network structure to the fate of an epidemic and effective public health strategies on the network. So an effective disease control strategy can target to immunize the highly connected individuals in the network or in a cluster. This will be very effective in reducing the disease spread rates. i.e. bring the average transmissibility below the epidemic threshold. One can also manipulate the network through several ways i.e. quarantining the first patients of the disease i.e.(patient zero and immediate secondary infections from him or her) since they can be located in the network, then they can be quarantined and his or her contacts provided with disease prevention alternatives e.g. quarantine also, vaccination or travel ban to or from the neighbourhood where patient zero is originating from, this can only be done if the outbreak is caught up in the initial stages; these measures will go in a long way to reduce the susceptibility population very rapidly. An interesting observation is the time course of the epidemic in the population echoes a compartmental SIR model albeit with different values of its parameters [?]. Another important observation that we can bring out form the results is that one does not have to posses the full knowledge of the network characteristics to be able to manipulate it but with information on only a few important nodes one can predict the full network properties and hence can easily subject it to manipu-

lation and this makes this study of epidemiology very fundamental and important in study of diseases.

4.2 Recommendations

Although there is an ever growing literature concerning network approaches to diseases, much work remains to be done especially in application of network modelling to diseases prevalent in developing nations, we find that many advances in this area of science have been done in developed nations but is still quite a new and unknown concept in developing nations where diseases have high prevalence rates. With the recent increases in computing power on network models, much study still needs to be done in this sector (network modelling of diseases). This calls for greater effort in the design and operation of all kinds of complex dynamical networks and their coupling with other factors of disease i.e. environment, economics etc; so as to provide better analysis and prediction of various potential issues.

References

[1] D. Bernouilli, *Esai dune nouvelle alanyse de ls mortilate causee par la petite verole et des advantages de lionoculaiton pour la prevenir, in Memoires de Mathemtique et de Physique*, Academie Royale des Sciences,Paris, 1760, pp.1 - 45

[2] Lloyd, A. L. & May, R. M. 1996, *Spatial heterogeneity in epidemic models.*, J. Theoret. Biol.179, 1 11.

[3] Bailey, N. *The Mathematical Theory of Infectious Diseases.* Charles Griffin 1975, London

[4] Castillo-Chavez,C., Z. Feng and W. Huang. *On the Computation of Ro and its role in Global Stability in: Mathematical Approaches for Emerging and Reemerging Infectious Diseases*, Springer, New York pp 2002, 229 - 250.

[5] Ross, R. 1910 *The prevention of malaria.* London, UK: John Murray.

[6] Diekmann.O and Heesterbeek. J. A. P., *Mathematical epidemiology of infectious diseases: model building, analysis and Interpretation.2000*, John Wiley and Son.

[7] Van den Driessche, P. & Watmough, J. 2002, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission.* Math. Biosci.

[8] Lloyd, A. L. & May, R. M. 1996, *Spatial heterogeneity in epidemic models.* J. Theoret. Biol. 179, 1 - 11.

[9] S. Bansal, B. Grenfell, and L. A. Meyers, *When individual behaviour matters: Homogeneous and network models in epidemiology.* Journal of the Royal Society Interface 4(16), 2007.

[10] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk and M. Massari. *Social contacts and mixing patterns relevant to the spread of infectious diseases.* PLoS Medicine 5:e74, 2008.

[11] S. Bansal, B. Pourbohloul, and L. A. Meyers, *Comparative analysis of influenza vaccination programs.* PLoS Medicine 3:e387, 2006.

[12] Finkenstadt, B. & Grenfell, B. 1998, *Empirical determinants of measles metapopulation dynamics in England and Wales.* Proc. R. Soc. B 265, 211 - 220.

[13] Grenfell, B. T., Bjrnstad, O. N. & Kappey, J. 2001, *Travelling waves and spatial hierarchies in measles epidemics.* Nature 414, 716 - 723.

[14] Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. 2005, *Multiscale, resurgent epidemics in a hierarchical metapopulation model.* Proc. Natl Acad. Sci. USA 102, 11 157 - 162.

[15] Kermack, W. O.; McKendrick, A. G. (1927), *Contribution-s to the Mathematical Theory of Epidemics.* Part i, Proceedings of the Royal Societyof Edingburgh A: Mathematical, Physical and Engineering Sciences 115 : 700 - 721

[16] Viboud.C, Bjornstad.O, Smith D.L, Simonsen.L, Miller.M.A, and Grenfell.B.T., *Synchrony,waves, and spatial hierarchies in the spread of influenza.* Science, 2000; 312(5772):447 - 451.

[17] Murray, J. D., *Mathematical biology.* Berlin, Germany: Springer, 1989.

[18] Colizza. V, Barrat.A, Barthelemy.M, Valleron.A.J, and Vespignani.A., *Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions*, PLoS Med, 2007;4(1), e13.

[19] Monto.A.S, *Vaccines and antiviral drugs in pandemic preparedness: Emerging Infectious Diseases.* 2006,555 4)Sing B. Typhoid fever Epidemiology. Journal Indian Academy of Clinical Medicine, Vol 2, No and 2 , Jan - June 2001. SIAM Rev. 45(2), 2003,167 - 256

[20] Newmann M, *The structure and function of complex networks*, 2003 65 - 89

[21] Meggan E. Craft and Damien Caillaud, *Network Models: An Underutilized Tool in Wildlife Epidemiology*, 2011

[22] A. L. Barabasi and R. Albert., *Emergence of scaling in random networks.* Science 286: 1999, 509.

[23] R. Albert and A.L. Barabasi. *Topology of evolving networks: Local events and universality.* Phys. Rev. Lett., 85: 2000, 5234 - 5237.

[24] L.A Meyers, B. Pourbohloul, M.E. Newmann, D.M, Skowronski and R.C. Brunham. *Network Theory and SARS: predicting outbreak diversity* . J. Theor. Biol., 231:71 - 81,2005.

[25] L. A. Meyers, M. E. J. Newman, M. Martin, and S. Schrag. *Applying network theory to epidemics: Control measures for Mycoplasma pneumoniae outbreaks.* Emerging Infectious Diseases: 2003, 9:204 - 210.

[26] O. Diekmann, M. C. M. De Jong, and J. A. J. Metz. *A deterministic epidemic model taking account of repeated contacts between the same individuals.* J. Appl. Prob., 35: 1998,448 - 62.

[27] 37. M. J. Keeling and B. T. Grenfell. *Individualbased perspectives on R(0).* J. Theor. Biol., 203:51 - 61, 2000.

[28] K.M. Carley, D. Fridsma, E. Casman, A. Yahja, N. Altman, L.C. Chen, B. Kaminsky and D.Nave. BIOWar: *Scalable AgentBased Model of Bioattacks, IEEE Trans. on Sys. Man and Cyber.*Part A:, .36 2006; 252 - 265

[29] P.D. Stroud, S. J. Sydoriak, J. M. Riese, J. P. Smith, S.M. Mniszewski, P. R. Romero . *Semi empirical powerlaw scaling of new infection rate to model epidemic dynamics with inhomogeneous.* Mathematical Biosciences, 203 2003; 301 - 318

[30] W. H. Hamer. *Epidemic Disease in England*, Lancet, 1, 1906, 733 - 739.

[31] K. Dietz. *Epidemics and romours: A survey*, J.Roy. Statistics. Soc. Ser. A 130, 1967, 505 - 528

[32] M. E. Craft, E. Volz, C. Packer, and L. A. Meyers. *Distinguishing epidemic waves from disease spillover in a wildlife population.* Proceedings of the Royal Society of London B 276: 1777 - 1785, 2009.

[33] J. M. Read, K. T. D. Eames, and W. J. Edmunds. *Dynamic social networks and the implications for the spread of infectious disease.* Journal of the Royal Society Interface 134:2008,1001 - 1007.

[34] B. Pourbohloul, L. A. Meyers, D. M. Skowronski, M. Krajden, D. M. Patrick, and R. C. Brunham. *Modeling control strategies of respiratory pathogens.* Emerging Infectious Diseases 11(8): 2005, 12 - 49.

[35] M. E. J. Newman. *Egocentered networks and the ripple effect.* Social Networks, 25: 2003 83 - 95.

[36] R.Ross. *The Presentation of Malaria,* Murray, London 2^{nd} (ed)., 1911

[37] A.G. McKendrick, *Applications of mathematics to medical problems,* Proc. Edinburgh Match Soc.,44 1926, 98 - 130

[38] R. Levins, T. Awerbuch, U. Brinkman, et al., *The emergence of new diseases, American Scientist,* 82 1994, 52 - 60.

[39] Lauren Ancel Meyers, Nedialko.B. Dimitrov. *Mathematical Approaches to Infectious Disaese Prediction and Control.* Operation Research journal: 2010

[40] L. Garrett, *The Coming Plague*, Penguin, New York, 1995.

[41] M. B. A. Oldstone, *Viruses, Plagues, and History*, Oxford University Press, New York, 1998.

[42] R. Shilts, *And The Band Played On*, St. Martins Press, New York, 1987, pages 57 - 59

[43] N. M. Ferguson, C. A. Donnelly, R. M. Anderson, *The Footãndmouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions,* Science, 292 , 2001. 1155 - 1160.

[44] Anderson RM & May RM ; *Infectious Diseases of Humans: Dynamics and Control* Oxford, UK: Oxford University Press, 1991.

[45] P. Martens, *How will climate change affect human health?*, American Scientist, 87,1999, 534 - 541.

[46] Keeling, M. J. *The implications of network structure for epidemic dynamics.* Theor. Popul. Biol.67: 2005, 18. (doi:10.1016/j.tpb.2004.08.002)

[47] M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. Chew, C. C. Tan, M. H. Samore, D. Fisman, and M. Murray. *Transmission dynamics and control of severe acute respiratory syndrome.* Science 300:doi: 10.1126 /science.1086616, 2003.

[48] J. Kleinberg and S. Lawrence, *The structure of the Web.* Science 294, 2001, 1849 - 1850

[49] C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. AbuRaddad, et al. *Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong.* The Lancet 361: 2003, 1761 - 1766.

[50] S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. AbuRaddad, A. J. Hedley, G. M. Leung, et al. *Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of Public Health Interventions.* Science doi:10.1126/science.1086478, 2003.

[51] Texas Department of State Health Services. *Planning Guidelines for Nonpharmaceutical Interventions.* Department of State Health Services, Texas, 2007.

[52] World Health Organization. *Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003.* `http://www.who.int/csr/sars/country/table20040421/en/index.html`, May 2004.

[53] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, *On the definition and the The Basic Epidemiology Models diseases in heterogeneous populations*, J. Math. Biol., 28 (1990), pp. 365 - 382.

[54] M. J. Keeling, D. A. Rand, and A. J. Morris. *Correlation models for childhood epidemics.* Proc. R. Soc. Lond. B, 264, 1997, 1149 - 1156.

[55] Klovdahl, A. S., Dhofier, Z., Oddy, G., OHara, J., Stoutjesdijk, S. & Whish, A. *Social networks in an urban area: first Canberra study.* Aust. N.Z.J. Sociol. 13, 1977. 169 - 172. (doi:10.1177/144078337701300215)

[56] Wilensky, U. N e t L o g o model `http://ccl.northwestern.edu/netlogo/models/` epiDem. 1998; *Center for Connected Learning and Computer-Based Modeling*, Northwestern University, Evanston, IL.

[57] H. W. Hethcote and J. A. Yorke, *Gonorrhea Transmission Dynamics and Control*, 1984, vol. 56 of Lecture Notes in Biomathematics, Springer-Verlag, Berlin.

[58] S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. *Modelling disease outbreaks in realistic urban social networks.* 2004 Nature 429:doi:10.1038/nature02541.

[59] J. H. Jones and M. S. Handcock. *Social networks: Sexual contacts and epidemic thresholds.* Nature, 423: 2003, 605 - 606.

[60] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, *Superspreading and the effect of individual variation on disease emergence,* 2005, Nature, vol. 438, no. 7066, pp. 355 - 359.

[61] L. A. Meyers. *Contact network epidemiology: Bond percolation applied to infectious disease prediction and control,* Bulletin of the American Mathematical Society 2007, 44:63 - 86.

[62] M. E. J. Newman. *Spread of epidemic disease on networks,* 2002, Physical Review E 66:016128.

[63] L. A. N. Amaral and J. Ottino. *Complex networksAugmenting the framework for the study of complex systems.* European Physical Journal B 38:2007,147 - 162.

[64] D. J.Watts. *A simple model of global cascades on random networks.* Proceedings of the National Academy of Sciences of the United States of America, 2002., 99:5766 - 5771.

[65] D. J. Watts. *Small Worlds: The Dynamics of Networks Between Order and Randomness.* Princeton University Press, Princeton, NJ, 1999.

[66] B. Bollobs *Random Graphs.* Academic Press, 1985.

[67] M. E. J. Newman and D. J. Watts. *Scaling and percolation in the small-world network model.* Phys. Rev. E, 60: 1999, 7332 - 7342.

[68] D. J.Watts and S. H. Strogatz. *Collective dynamics of small-world networks.* Nature, 393: 1998.

[69] M. E. J. Newman. *Assortative mixing in networks.* Phys. Rev. Lett., 89:208701, 2002.

[70] M. E. J. Newman. *Mixing patterns in networks.* Phys. Rev. E, 67:026126, 2003.

[71] A. C. Ghani and G. P. Garnett. *Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks.* Sex. Trans. Dis:27: 2000, 579 - 587.

[72] Alun L. may and Steve Valeika: *Network Models in Epidemiology: An overview;* WSPC review Vol 18:38, 2005.

[73] M. Kretzschmar and M. Morris. *Measures of concurrency in networks and the spread of infectious disease.* Math. Biosci., 133: 1996, 165 - 95.

[74] F. G. Ball, D. Mollison, and G. Scalia-Tomba. *Epidemics with two levels of mixing.* Ann. Appl. Prob., 7: 1997, 46 - 89.

[75] A. Barrat and M. Weigt. *On the properties of small-world network models.* Eur. Phys. J. B, 13: 2000, 547 - 560.

[76] R. Pastor-Satorras and A. Vespignani. *Epidemic spreading in scale-free networks.* Physical Review Letters 86(3200), 2001.

[77] *Introduction to Social Network Methods: Chapter 1: Social Network Data.* Faculty.ucr.edu. Retrieved 2012-10-24.

[78] Only connect: Felix Grant looks at the application of data analysis software to social networks", *Scientific Computing* World June 2010: pp 910.[2]

[79] Wolfram Research, Inc. Mathematica Edition: Version 9.0 Publisher: Wolfram Research, Inc. Place of publication: Champaign, Illinois Date of publication: 2013

[80] M. E. J. Newman. *Spread of epidemic diseases on networks.* Phys. Rev. E, 66:016128, 2002

[81] R. M. May and A .L. Lloyd. *Infection dynamics on scale-free networks.* Phys. Rev. E, 64:066112, 2001.

[82] J. A. Yorke, H. W. Hethcote, and A. Nold. *Dynamics and control of the transmission of Gonorrhea.* Sex. Trans. Dis., 5: 1978, 51 - 56.

[83] C. Moore and M.E. J. Newmann, *Epidemics and percolation in small world networks.* Phys. Rev. E 61: 2000, 5678 - 5682.

[84] M. Kuperman and G. Abramson, *Small- world effect in an epidemiological model.* Phys. Rev. Lett 86: 2001, 3200 - 3023.

[85] Y. Moreno, R. Pastor Satorass, and A. Vespignani, *Epidemic outbreaks in complex heterogenous networks.* Preprint cond-mat/0107267,2001.

[86] C.P. Warren, L. M. Sander, and I. Solokov, *Firewalls, disorder and percolation in epidemics.* Preprint cond- mat/0106450.2001.

[87] C.P. Warren, L. M. Sander, I. Solokov C. Simon and J. Koopman, *Percolation on disordered networks as a model of epidemics,* Math. Biosci (in press).

[88] G. Grimmett. *Percolation.* Springer, Berlin, 1999.

[89] P. Grassberger. *On the critical behavior of the general epidemic process and dynamical percolation.* Mathematical Biosciences 63:1983, 157 - 172.

[90] L. A. Meyers, M. E. J. Newman, and B. Pourbohloul. *Predicting epidemics on directed contact networks.* Journal of Theoretical Biology 240:2006, 400 - 418.

[91] H. Wilf, Generating functionology. Academic Press, London 2^{nd} Edition. 1994.

[92] A. Adimora, V. Schoenbach, and I. Doherty. *HIV and African Americans in the Southern United States: Sexual networks and social context.* Sexually Transmitted Diseases 33:2006, S39 - S45.

[93] Newman, M. E.J.,2002. *Spread of epidemic disease on networks.* Phys. Rev. E 66, art. no.-016128.

[94] K. Ford, W. Sohn, and J. Lepkowski. *American adolescents: Sexual mixing patterns, bridge partners, and concurrency.* Sexually Transmitted Diseases 29: 2002,13 - 19.

[95] E. Volz. *SIR dynamics in structured populations with heterogeneous connectivity.* Journal of Mathematical Biology 56: 2007,293 - 310.

[96] E. Volz and L. A. Meyers. *Susceptible-infected-recovered epidemics in dynamic contact networks.* Proceeding of the Royal Society B 274:2007,2925 - 2933.

[97] C. Watts and R. May. *The influence of concurrent partnerships on the dynamics of HIV/AIDS.* Mathematical Biosciences 108(1): 1992, 89 - 104.

[98] M Altmann, *Susceptibleinfectedremoved epidemic models with dynamic partnerships.* J. Math. Biol. 33, 1995, 661 - 675. (doi:10.1007/BF00298647)

[99] K.T.D Eames & Keeling, M. J. *Monogamous networks and the spread of sexually transmitted diseases.* Math. Biosci. 189, 2004, 115 - 130. (doi:10.1016/j.mbs 2004.02.003)

[100] C. A Bauch, *Versatile ODE approximation to a network model for the spread of sexually transmitted diseases.* J. Math. Biol. 45, 2002 375 - 395. (doi:10.1007/ s002850200153)

[101] R. Ross & H.P Hudson . *An application of the theory of probabilities to the study of priory pathometry,* I, II. Proc. Roy. Soc. Lond. A (1917) 212 - 225, 225 - 240.

[102] Kermack, W. O.; McKendrick, A. G. (1932). *"Contributions to the Mathematical Theory of Epidemics". Part ii, the problem of endemicity, Proceedings of the Royal Societyof Edingburgh A: Mathematical, Physical and Engineering Sciences* 138 : 55 - 83.

[103] Kermack, W. O.; McKendrick, A. G. (1933). *"Contributions to the Mathematical Theory of Epidemics". Part iii,further studies of the*

problem of endemicity, Proceedings of the Royal Societyof Edingburgh A: Mathematical, Physical and Engineering Sciences 141 : 94 - 122