Universitetet i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: **MATHEMATICS AND PHYSICS/MATHEMATICAL STATISTICS**	Spring / Autumn semester, **2020.....** Open/Confidential
Author: **Francis Osei**	 …… (signature of author)
Programme coordinator: **Eli Dragne Vee** Supervisor(s): **Stein Andreas Bethuelsen**	
Title of master's thesis: **Stochastic Epidemic Models on Complex Networks**	
Credits: **60 ECTS**	
Keywords: **stochastic processes, epidemics models, branching processes, complex networks, evoSIR**	Number of pages: ……**62**…… + supplemental material/other: …… Stavanger,…**14/06/2020**…… date/year

Universitetet i Stavanger

Faculty of Science and Technology
Department of Mathematics and Physics

Stochastic Epidemic Models on Complex Network

Master's Thesis in Mathematics and Physics
by

Francis Osei

Supervisor

Stein Andreas Bethuelsen

A thesis submitted in partial fulfillment for the
degree of Master of Science

June 14, 2020

Abstract

The spread of a virus or the outbreak of an epidemic are natural examples of stochastic processes. Classical mathematical descriptions of such phenomenon include various branching processes such as the SIR (Susceptible-Infected-Recovered) model and the SIS (Susceptible-Infected-Susceptible) model. The basis of this thesis consists of giving a comprehensive overview of the mathematical theory behind these models with an emphasis on the SIR model and its evolution on complex networks. Further, following [1],[2],[3], we consider the evoSIR on three network structures (Erdös Rényi Graph (ER graph), configuration model network and the preferential attachment model) in which a susceptible after learning the status of his neighbour breaks that connection at rate ρ and rewire to a randomly chosen individual in the population. We show through simulations that, delSIR can reduce the final size of an outbreak of a diseases with a higher probability. Finally, we show that the network structure crucially influences the measures to control the outbreak of diseases at the population level.

Acknowledgements

It has not been a smooth journey. After all the determination and long hours that were devoted for this project, it is hard to believe that a finished product has finally been produced.

My first and sincere gratitude goes to the ALMIGHTY GOD whose infinite mercy and love has brought me this far. To Him alone be glory and honor.

A big "THANK YOU" goes to my Supervisor, Stein Andreas Bethuelsen for his support, dedication, advice and tolerance in ensuring that my project is successful. This work could not have been complete without you. God bless you.

I wish to express my profound gratitude to Grayson York and Co.. Thank you for your genuine care in making this project a success.

To my parents/guardians for their support and immeasurable help; I am grateful.

Finally, to all those who have contributed in diverse ways to my study and all lecturers of the Department of Physics and Mathematics, UIS, I say GOD BLESS YOU.

Contents

Chapter 1

Introduction

Infectious diseases are responsible for significant health and economic problem in society. Mathematical models use assumptions and statistical inferences in determining parameters for the spread of diseases. Mathematical models in recent years have been used to guide policy makers responding to the emergency of the diseases including Measles[4][5][6], H1N1 influenza[7][8], Hepatitis C Virus (HCV)[9][10], Whooping cough[11], HIV[12][13], Ebola[14][15][16], Coronavirus[17][18][19] and many others. For example, in 1760, Daniel Bernoulli aimed at evaluating the effectiveness of inoculating against smallpox. Ross (1911) modelled the transmission of malaria. The 1920s saw the emergence of the compartmental model (deterministic). Kerneck-Mckendrick in 1927 studied rigorously the general SIR model describing the relationship between susceptible, infection, and immune individual in the population. The early models (deterministic) addresses questions like, how many people in the population will get infected during an outbreak of an epidemic? It is possible for a large proportion of the population to be infected? What is the effect of vaccinating a fraction of the population prior to the arrival of the disease? As more attention was given to this process, assumptions were made in several ways to make it simple and realistic. Another generalization is the stochastic epidemic model which was introduced by Bartlett (1947) who studied the stochastic version of the Kerneck-Mckendrick model and Kendall (1956) who also studies both the deterministic and stochastic epidemics in a closed population. Since then, there have been various reviews in the modelling of stochastic epidemic models. [20] studied the variation of infection rate where he suggested that infection rate acting upon each susceptible individual is independent over time but varies between susceptible. [21] in a collection paper, provides detailed analysis of the stochastic process in epidemics theory, [22] provides a review of some epidemics model quantities providing different ways of combining them and their relationship between epidemic models. [23], described epidemics spread in both homogeneous and inhomogeneous population for deterministic and stochastic models

in continuous and discrete-time. They also provide detailed methods for constructing and analysing mathematical models of viral and bacterial diseases. [24], focused on the epidemic process model on deterministic and stochastic, where model assumptions were mostly stochastic. [25], explain three different types of stochastic epidemic models (stochastic differential, discrete-time and continuous-time Markov chain) and properties special to epidemic model. A very recent book by [26] provides a detailed understanding to stochastic epidemic using various techniques in addressing epidemic spread problems. This generalization (stochastic epidemic modelling) serves as a tool for estimating the probability distribution of the potential outcomes. This model came into existence to address questions like, What is the probability of major outbreak? What is the size of the epidemics? What is the duration of the epidemics? and many more. The early study of epidemiology models was assumed to be model in randomly mixing population where individuals can contact any member in the population. However, in practice, each individual has a finite number of individual (neighbours, family, friends, classmate e.t.c.) he can contact. Diseases like HIV can only spread through a network of individuals having sexual intercourse. In the mid-1980s and 2000s, network science became a well-established and productive research area in modelling real-world phenomena ranging from biological to social phenomena. These models lead to an improved understanding of the significance of network properties on the evolving process on networks, including, the flow of rumours and disease transmission. There is a closed relationship between epidemics spread models and network theory; this is because the connections between the individual that allows infection spread in a population define a network (family, friends, classmates, co-workers e.t.c.) while the network generated provides insight on how epidemic evolves. Improving the prediction of the infection degree distribution and the early growth of infection can be well explained when the structure of the transmission network is well understood. New infectious disease (SARS-CoV-2) have made it clear how the study of a network can relate to the propagation of infectious disease. Contact tracing and breaking the connection with an infected individual (isolation) are the highly effective public health measures which use the underlying dynamics to control the effect infecting the majority of a population[27][28]. In this thesis, we shall study the effect of stochastic modelling of infectious diseases on a random network (closed population) where susceptible individual breaks friendship with a neighbour after learning his status at a constant rate ρ.
This thesis is organized in five chapters. Chapter one is used to present background and an introduction to the work. Chapter two introduces the branching process, which serves as a tool in modelling the early stage of an epidemic. Chapter three explaining the fundamentals of network theory and the relationship between network theory and epidemiology. In Chapter four, we present how epidemic spread evolves on networks and Chapter five laying a valid conclusion as well as capturing future works to the study.

1.1 Why Stochastic Model

Both stochastic and deterministic models play an essential role in better understanding the mechanism of infectious diseases. The deterministic model serves as an introductory model when studying a new phenomenon[29]. One main advantage of the deterministic over the stochastic lies in its simplicity. The model can be more complex but possible to analyse when numerical solutions are available, for analysis of the stochastic model to make possible, the model should be simple and does not entirely need to be realistic. When the analysis is possible, the stochastic is preferred over the deterministic because of the following reasons

1. The stochastic models show what the deterministic model converges to when the size of the population is large[29]. The deterministic model, the spread of infectious diseases is based on the assumption of large numbers.

2. In a large population, epidemic models may lead to either minor outbreak or significant outbreak. The natural way to calculate the probability of the two events is only mathematically possible in the stochastic setting.

3. Questions about uncertainty in estimates requires a stochastic model and an estimate is not of much use without some knowledge of uncertainty.

1.2 Reed-Frost Model

During the 1920s, two researchers at the John's Hopkins University, Lowel Reed and Wade Frost develop a mathematical model to give the exact prediction of how infectious diseases spread through a small population. The Reed-Frost model is one of the simplest epidemic models which is usually specified using discrete time dynamics. Despite its simplicity, it is sufficiently complicated that more tractable approximations are still needed. The Reed-Frost model $(S(t), I(t), R(t),\ t \geq 0)$ is a discrete-time SIR model which means that at time $t = 0$, an individual is either at the susceptible stage or infection stage or recovered/removed stage. A susceptible individual at time $t + 1$ after contacting an infectious individual (infective) will develop the infection and will be infectious to others within a period and become immune and recovered. In the discrete time, we usually think of the infectious period as being short whiles the latent period (the period between infection with a virus and the onset of symptoms) is assumed to be long. Each individual in the population has a uniform probability of coming into contact with any other specified individual in the group with time $t + 1$ depend on the previous step t (i.e. the process evolve to a Markovian recursive) and the binomial probability specifies the

events.

Let $S(t)$ and $I(t)$ denote the number of susceptible and infective individuals in the population at time $t > 0$ respectively. Then the chain-binomial Reed-Frost model in [30] is given as

$$P(I(t+1) = i_{t+1} \mid S(0) = s_0, I(0) = i_0, \cdots, S(t) = s_t, I(t) = i_t) \tag{1.1}$$

$$= P(I(t+1) = i_{t+1} \mid S(t) = s_0, I(t) = i_t,) \tag{1.2}$$

$$= \binom{S(t)}{I(t+1)} (1 - q^{i_t})^{i_{t+1}} (q^{i_t})^{s_i + i_{t+1}} \tag{1.3}$$

and $S(t+1) = S(t) - I(t+1)$. This can be explained as given a susceptible individual at time t will remain susceptible at time $t+1$ if he escapes from all infectives at time t where q is the event probability which is independent of each other. Now lets assume that at the initial state, $S(0) = n$ and $I(0) = m$ then we can compute the probability of the complete chain $i_1, \cdots, i_{k+1} = 0$ by conditioning sequentially and using the Markov property of the chain. Then we have;

$$\begin{aligned} &P(I(1) = i_1, \cdots, I(k) = i_k, I(k+1) = 0 \mid S(0) = n, I(0) = m) \\ &= P(I(1) = i_1 \mid S(0) = n, I(0) = m) \times \cdots \times P(I(k+1) = 0 \mid S(k) = s_k, I(k) = i_k) \\ &= \binom{n}{i_1} (1 - q^m)^{i_1} (q^m)^{n - i_1} \times \cdots \times \binom{s_k}{0} (1 - q^{i_k})^0 (q_k^i)^{s_k} \end{aligned} \tag{1.4}$$

With probability $1 - q$, each individual who becomes infected has infectious contact with any other individual in the population. Denote the final number of infected individual by Z, then we can compute $P(Z = z \mid S(0) = n, I(0) = m)$ by summing the probabilities of all chains for which $\mid i \mid = \sum_{t \geq 1} i_t = z$. It is seen from the equations that new infection may only occur whenever there exist some infectious individual, thus, the length of a chain cannot be longer than the total number of infected, which makes the number of possible chain finite. Based on our findings, we can write the probability function for the final number of infected as

$$\begin{aligned} &P(Z = z \mid S(0) = n, I(0) = m) \\ &= \sum_{i : \mid i \mid = z} P(I(1) = i_1, \cdots, I_k = i_k, I_{k+1} = 0 \mid S(0) = n, I(0) = m) \end{aligned} \tag{1.5}$$

A typical example is given in [30] and [26] for $S(0) = 1, 2, 3$ and $I(0) = 1$, we can see that computing number of infected becomes very complicated when the number of susceptible individuals become very large.

1.3 The Standard SIR Model

In this section, we define the stochastic SIR epidemic model. The stochastic SIR is tough to analyse when the model is complicated. Several assumptions were made to make the model manageable and realistic. The population is assumed to be closed and homogeneously mixing, the effect of the latent period changes the behaviour, constant/-exponential infective period and no partial immunity.
Consider a closed population $n+m$, where n is the total number of susceptible individuals in the population and m is the initial infectives. At any given time t, each individual is either susceptible (S), or infected (I), or Removed/Recovered(R). Let $S(t)$, $I(t)$ and $R(t)$ denote the number of individuals in the population such that $S(t) + I(t) + R(t) = n + m$ at any given time t. In this thesis, we shall assume that $m= 1$ (i.e. at time $t= 0$ we have only one infected individual). While an individual is still susceptible, he contacts an individual in the population according to a poison process with parameter μ. If the individual is still susceptible, he becomes infected; otherwise, the infectious contact does not affect the epidemic process (this will be referred to ghost contact throughout the thesis). An individual becomes infectious for a duration F (independent and identically distributed) with distribution F_L after which they become recovered and immune for the remaining time in our epidemic process. The epidemic continues to exist until time t_f when $I(t_f) = 0$. At this time, no individual can be infected. At the final stage, the population will consist of susceptible and removed individuals. Then the final size $Z=n - S(t_f)$ The process described above is referred to as the standard SIR process, which will be denoted as $B_{n>1,1}(F, \mu)$. Here we considered the two most studied cases of the epidemic process infectious individual becomes active (infectious) for exponentially distributed with mean one and random case where an individual becomes infectious at a constant time 1 (i.e. $F \approx \exp(1)$ or $F \approx 1$).
Two key quantities that appear in the study of the spread of infectious disease is the basic reproduction number R_0 and the escaping probability which will be analyse in detailed in chapter 4 and chapter 5. R_0 denotes the mean number of infectious contact a typical infected has during the early stage of an outbreak[26]. This quantity agrees with the number of infection caused by a typical infectious individual when the population becomes large.

1.3.1 Sellke Construction

The final size distribution of an epidemic in a closed population can be derived from the Sellke construction. The process keeps track of the total infection pressure generated by the infectious individual[30]. Let $1, 2, \cdots, n$ and $-(m-1), -(m-2), \cdots, 0$ denotes the

initial susceptible and the initially infected individuals in the population, respectively. $F_{(m-1)}, F_{(m-2)}, \cdots, F_n$ are independent and identically distributed random variable, each following the distribution of F. Let $F_{(m-1)}, F_{(m-2)}, \cdots, F_0$ denote the length of the infection period for the m initial infected. Again, let $Q_1, Q_2, \cdots, Q_n$ be an independent random variable, each exponentially distributed with mean 1. This represents the threshold of the initial susceptible. The total infection pressure (cumulative force of infection) exerted on a given susceptible up to time t is given as

$$A(t) = \mu \int_0^t I(u)du \tag{1.6}$$

where $I(u)$ is the number of infected at time u. The infection pressure $A(t) > 0$ and is increasing or constant. A susceptible individual i ever becomes infected at time t when $A(t) = Q_i$. The jth person to get infected will remain infected for the time period F_j (i.e. $F_1, F_2, \cdots, F_n$ is used sequentially).

Proposition 1. The Sellke construction gives a process equivalent to the Standard SIR epidemic

Proof. We have to show that the time when infection happens is the same for both models. The proof was taken from [30]. Let $I(t) = y$ and the ith individual is susceptible at time t and becomes infected during $(t, t + \Delta t)$ with the probability $\Delta y \Delta t + o(\Delta t)$, the probability of the complementary event (two outcomes of an event that are the only possible outcome) is given by

$$\begin{aligned}
P(\text{susceptible } i \text{ infected at } t + \Delta t \mid \text{not infected at } t) &= P(Q_i < A(t + \Delta t) \mid Q_i > A(t)) \\
&= \frac{P(A(t) < Q_i < A(t + \Delta t)}{P(Q_i > A(t))} \\
&= \frac{\left(1 - e^{-A(t+\Delta t)}\right) - \left(1 - e^{-A(t)}\right)}{e^{-A(t)}} \\
&= 1 - e^{\mu y \Delta t + o(\Delta t)} \\
&= 1 - (1 - \mu y \Delta t + o(\Delta t)) \\
&= \mu y \Delta t + o(\Delta t) \approx \mu I(t) dt
\end{aligned}$$

□

Now, let's compute the exact result of the final size of the epidemic using the Sellke's construction. Let Z be the final size of the epidemic and $A = A(\infty) = \mu \int_0^\infty I(u)du$ be the total infection pressure. The integral is finite since the number of infected is finite, and each recovers at a finite time. Then we can write the total infection pressure in terms of

the infection period as

$$A = \sum_{j=-(m-1)}^{Z} F_j [30] \tag{1.7}$$

We can also express the final size in terms of the infection period and the individual threshold as

$$Z = \min \left\{ i \;\middle|\; Q_{(i+1)} > \mu \sum_{j=-(m-1)}^{Z} F_j \right\} [30] \tag{1.8}$$

Where $Q_{(1)}, Q_{(2)}, \cdots, Q_{(n)}$ are the order statics (smallest value) of $Q_1, Q_2, \cdots, Q_n$ because epidemic ceases when infection pressure generated by a previously infected individual is insufficient to infect any susceptible in the population.

1.3.2 Early Stage of an outbreak of infectious diseases

Now let us consider the standard SIR epidemic process $B_{n>1,1}(F, \mu)$ in a closed population with n individuals (a population with no birth, death, immigration and emigration). We study the epidemic process where less than $k{=}k(n)$ individuals has been infected. We know from the previous section that infectious contact between two individual in the population happens independently. Future contact with an already infected individual has no impact on the epidemic process, and in such situations, contact between individual becomes dependent. In the early stage of an epidemic in a large population, it is unlikely for two already infected individuals to have contact (i.e. at the beginning of an epidemic, infectious contact occurs between infected and susceptible individuals in the population). This suggests that the number of infected at the beginning of an epidemic can be approximated by the Branching process where "giving birth" refers to infecting an individual in the population and "being born" corresponds to be infected. Branching processes are a class of stochastic processes that model the growth of populations. In the next chapter, we describe the discrete-time and continuous-time Branching process approximation and obtain an asymptotic result for the epidemic for a major and minor outbreak.

Chapter 2

Branching processes

2.1 Introduction

In this section, we study single type branching processes which serve as toy models to determine the number of infectious individuals in a large population and also as a tool to determine the final behaviour of an epidemics (explode/extinct). Branching processes play a central role in the theory of mathematical epidemiology. Branching processes are stochastic process (The model becomes Markovian if F is exponentially distributed) in which the size of a generation only depends on the size of the previous generation and the number of their offspring. Following [31] branching processes can be define as follows: Let $\{I(t)|t \in [0,\infty)\}$ taking the set of values $\{1, 2, 3, \cdots\}$ be discrete random variable which has the Markov property. The process starts with a single individual $m= 1$ at time $t= 0$. The time before an individual in the population dies out is i.i.d according to a continuous random variable F with expectation τ and variance σ^2. Denote the branching process by $B_1(\mu, F)$ where an active individual gives birth at a time point of a Poison process with intensity μ.

2.2 Discrete Time Branching Process

The number of newly infected persons j from a previously infected individual i at time $t \geq 0$ is a stochastic variable denoted as $Y_i(t)$. Then $I(t)$ follows the recursive relation

$$I_j(t+1) = \sum_{i=1}^{I_t} Y_i(t) \tag{2.1}$$

$Y_i(t)$ are independent and identically distributed random variable according to the distribution $\mathbb{P}(Y_i(t) =j) =p_j$. We assume that individuals give birth independently of all

others and the transitional probability must satisfy.

$$\sum_{j=0}^{\infty} \mathbb{P}_{i,j}(t)s^j = \left(\sum_{j=0}^{\infty} \mathbb{P}_{1,j}(t)s^j\right)^i, \quad s \in [0,1][31] \tag{2.2}$$

Equation 2.2 implies the process starting from state $I(0) = i$ is equivalent to the sum of i independent processes beginning from state $I(0) = 1$. From equation 2.2, if $I(0) = \sum_{l=0}^{i} I_l(0)$ for $I_l(0) = 1$, then the probability generating function of $I(t)$ must satisfies

$$E(s^{I(t)}) = \Pi_{l=1}^{i} E(s^{I_l t})[31] \tag{2.3}$$

2.2.1 Generating Function and Extinction Probability

Generating functions (GF's) predict several properties about the initial phase of the spread of epidemics. At the same time, the population is still effectively infinite, including the probability of large epidemic, the size distribution after some number of generations, and the cumulative size distribution of non-epidemic outbreaks[32]. Let $G(s)$ denote the generating function of the offspring random variable.

$$G(s) = \sum_{j=0}^{\infty} p_j s^j \tag{2.4}$$

Now lets consider the event that the progeny of our process goes extinct (i.e. $\mathbb{P}(I(t) \to 0) = \lim_t \mathbb{P}(I(t) = 0 \mid I(0) = 1)$. Let $\{I(t) \mid t \geq 0\}$ be the number of active individuals in the population at time $t > 0$. As shown later in this section, the process will be extinct if $\mathrm{E}(I(t)) \leq 1$ since there is an average $\mathrm{E}(Y_1(t))^t$ at any given discrete-time point t. Considering the supercritical cases where $\mathrm{E}(Y_1(t)) > 1$, from [30], the extinction probability q of the branching process is given as

$$q = \sum_{k=0}^{\infty} \mathbb{P}(extinction|Y^*(0))\mathbb{P}(Y^*(0) = k)[30] \tag{2.5}$$

where $Y^*(0)$ is the number of children of an ancestor, thus $B_1(\mu, F)$ dies out if and only if all offspring generated by these children become extinct, $q = \sum_{k=0}^{\infty} q^k \mathbb{P}(Y^*(0) = k)$. The prove of theorem 2.1 and 2.2 below are taken from [31] and [33]

Theorem 2.1 (uniqueness of extinction probability)**.** *If $\mu \leq 1$ then 1 is the only root of $G(s)$ in [0,1] and if $\mu > 1$, then the extinction probability is a unique solution less than 1 of $G(s)$ in [0,1].*

Proof. From equation 2.4, it is easy to show that $s = 1$ is a fixed point thus $G(1) = \sum_{j=0}^{\infty} p_j s^j = \sum_{j=0}^{\infty} p_j = 1$. Then 1 is a root of $G(s)$

Now assuming $\mu \leq 1$, then $p_0 \neq 0$ taking $s = 0$ to be a fixed point of equation 2.4 which implies $G(0) = p_0 = 0$ but $G(0) = p_0 > 0$ whenever $\mu \leq 1$. Thus 0 is not a fixed point when $\mu \leq 1$. Now assume $s \in (0, 1)$ is a root of G(s), then the inequality $G'(s) < G'(1)$ since $f''(s) > 0$ for any s. We know from the Mean Value theorem that for some $c \in (s, 1)$, $G'(c) = \dfrac{G(1) - G(s)}{1 - s} = 1$ but this contradict the fact that for some $c \in (0, 1)$ $G'(c) < 1$ thus 1 is the only root of $G(x)$ whenever $\mu \leq 1$.
Next, $\mu > 1$ then $p_0 = 0$, $G(0) = p_0 = 0$ clearly it can be seen that 0 is the smallest root on $[0, 1]$. Now taking $p_0 > 0$, $G(s) < s$ when s is sufficiently close to 1 thus there is a at least one solution in $[0, 1)$ whenever $p \geq 0$. Assuming s_0 and s_1 are the roots of $G(s)$, then the Rolle's theorem implies that there exist ξ_0 and x_1, $s_0 < \xi_0 < s_1 < \xi_1$ such that $G'(\xi_0) = G'(\xi_1)$. But we know from above that G is strictly convex thus $G'(\xi_0) = G'(\xi_1)$ is not possible. Therefore there is a unique root $s^* \in [0, 1)$ of $G(s)$. □

Theorem 2.2. *For any finite value m^*, if $\mathrm{E}(I_1(t)) = m^*$ then $\lim_{t\to\infty} \mathbb{P}(I_1(t) = k) = 0$ for $k \in \mathbb{N}$. Again, $I_1(t)$ becomes extinct with probability q where q is the smallest root of $G(s) = s$ in $[0, 1]$ or explode with probability $1 - q$.*

Proof. Now assuming $R_k = \mathbb{P}(I(t + \Delta t) = k | I(t = k) = 0$ for some $k = 1, 2, 3, \cdots$, then we can show that at any given time $t + \Delta t$ the space k is transient thus $R_k = 0$. If $p_0 = 0$, then $R_k = \mathbb{P}(I(t) = 1) = p_1^k < 1$. Again if $p_0 > 0$, $R_k = 1 - \mathbb{P}_k, 0 = 1 - p_0^k < 1$. Thus for each value of $k \in \mathbb{Z}$, $\lim_{t\to\infty} \mathbb{P}(I_1(t) = k) = 0$ and $\lim_{t\to\infty} \mathbb{P}(I_1(t) = k$, for infinitely many values of $k) = 0$. $I(t)$ is either 0 or ∞ since it does not take same values of $k \in \mathbb{Z}$ but we know from equation 2.5 that $I(t) \to 0$ with probability q. □

Definition 2.3 (Martingales in Discrete Time)**.** Let $(\Omega, \mathcal{F}, \mathcal{P})$ be the probability space and $\mathcal{F}_n$, $n \geq 0$ be an increasing sequence of the sub-σ-algebras of $\mathcal{F}$ A sequence $\{I_n,\ n \geq 0\}$ of random variables is called Martingale if

1. For all $n \geq 0$, I_n if $\mathcal{F}$-measurable and integrable
2. For all $n \geq 0$, $\mathbb{E}(I_{n+1} \mid \mathcal{F}_n) \leq I_n$

A more detailed explanation of the Martingale can be found in [34] and [26]. The following theorem follows from the Martingale convergence from [35]

Theorem 2.4 (Martingale convergence Theorem)**.** *For a branching process with i.i.d. offspring $(Y_i(t))$, $\lim_t I(t)\mathrm{E}(Y_i(t))^{-1} \to W$ almost surely, for some integrable random variable W, where $\mathbb{P}(W > 0) > 0$ iff $\mathrm{E}(Y_i(t)) > 1$*

2.3 Coupling Between Branching Process and SIR

Consider a sequence of Epidemics $B_{n>1,1}(\mu, F)$. Denote $\{Y_n(t) \mid t \geq 0\}$ the number of susceptible individuals to have been infected at time $t \geq 0$ in the nth epidemics and $\{Y(t) \mid t \geq 0\}$ is the number of individuals alive in the branching process $B_1(\mu, F)$. This section gives an intuitive explanation of why the limiting process of the $\{Y_n(t); t \geq 0\}$ is $\{Y(t); t \geq 0\}$. Denote the numbering of individuals born into the population in the branching process by $-(m-1), -(m-1), \cdots, 0$. Assuming the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ holds for the individual life histories $\mathcal{H}_{-(m-1)}, \mathcal{H}_{-(m-2)}, \cdots, \mathcal{H}_0$ of the m ancestors and let H_i, $i \geq 1$, be the life history of the ith individual born. Let U_N be i.d.d defined on $(\Omega, \mathcal{F}, \mathbb{P})$ to be uniformly distributed on $\{1, 2, 3, 4, \cdots, n, N | N = 1, 2, 3, 4, \cdots, n, n+1, \cdots\}$. For a fixed n label the initial susceptible of the process $B_{n>1,1}(\mu, F)$ as $1, 2, 3, \cdots, n$. Birth and death of non ghost individual in the branching process correspond to contact and removal in the SIR epidemic process, respectively. The individual contacted at the ith contact has the label $C_i \in \{1, 2, 3, 4, \cdots, n, N \mid N = 1, 2, 3, 4, \cdots, n, n+1, \cdots\}$. A contacted individual becomes infected when still susceptible in the epidemic process; otherwise, the individual and the descendants in the branching process is ignored[30]. The two processes agree until the first ghost at time $T' > 0$. It is shown in [33] that as $n \to \infty$, the probability that there will be a ghost contact at any given time interval $[t, t + \Delta]$ is 0.

Theorem 1. Consider a sequence of Epidemics $B_{n>1,1}(\mu, F)$. Denote $\{Y_n(t) \mid t \geq 0\}$ the number of susceptible individuals to have been infected at time $t \geq 0$ in the nth epidemics and $Y_n(\infty)$ the total size of an epidemics, then for any fixed time t', $Y_n(t') \to Y(t')$ and $Y_n(\infty) \to Y(\infty)$ a.s. (almost surely), where $\{Y(t) \mid t \geq 0\}$ is the number of individuals alive in the branching process $B_1(\mu, F)$.

2.4 The Threshold Limit Theorem of Epidemics

In this section, we study the asymptotic distribution of the final size of the general stochastic epidemic model as $n \to \infty$. We have noticed that the final size Z_n of the epidemic process $B_{n>1,1}(\mu, F)$ converges a.s. to the branching process $B_1(\mu, F)$ when the population is large. Let $Z_n = \lim_{t\to\infty} R(t) - m$ be the size of the epidemics with probability $\mathbb{P}_z = \mathbb{P}(Z = z)$ where, $\{S(t)$, I(t) and $R(t)\}$ the number of susceptible, infected and recovered/removed individuals respectively at any given time t. let μ and γ denoted the rate of infection and recovery respectively, then the theorem below follows from [36] which gives a general case of William's threshold theorem.

Theorem 2.5 (William's threshold theorem)**.** *[37] Let* $\theta = \frac{\gamma}{n\mu}$ *denotes the relative removal rate per initial susceptible, then the probability of true epidemic (i.e. infinitely many susceptible are infected when* n *is sufficiently large) is 0 if* $\theta \geq 1$ *whilst if* $\theta < 1$ *a true epidemic occurs with the probability* $1 - \theta^m$*.*

Now we follow [30] to find the final size of an epidemic using the Infection process and the threshold process. Let $\mathcal{F}(t) = \frac{\lambda}{n} \sum_{j=-(m-1)}^{k-m} F_j$, $k \in [0, n+m]$ and $\mathcal{Q}(t) = \sum_{j=1}^{n} 1_{Q_J<k}$, $k > 0$ be the infection pressure process and the threshold process where F_j and Q_j are the infectious periods and individual threshold respectively. From the Selke's construction, A susceptible individual i becomes infected if the total infection pressure reaches Q_i then we have the final size of the epidemic as

$$Z = \min\{i \geq 0 \mid Q_{i+1} > \frac{\mu}{n} \sum_{j=-(m-1)}^{i} F_j\} \tag{2.6}$$

We can see from equation 2.6 that $\mathcal{I}(i-1+m)$ can contact at least i individuals thus expressing the threshold process in terms of the infection process we get

$$Z = \min\{k \geq 0 \mid \mathcal{Q}(\mathcal{F}(k+m)) = k\} \tag{2.7}$$

Now consider the case $m_n = m$ for all n. let $Z'_n = Z_n + m$ and let z_0 be a non-trivial solution of

$$1 - \exp(-\mu\tau z_0) = z_0 \tag{2.8}$$

where τ is the mean of the infection period, then we have the theorem below from [30].

Theorem 2.6. *Consider the sequence of epidemic processes* $B_{n>1,1}(\mu, F)$ *then*

1. *if* $\mu \leq 1$ *then* $Z_n \to Z$ *a.s., where* $\mathbb{P}(Z=k) = 1$ *for a finite value* k*, where* Z *is the total offspring in a continuous-time branching process with* m *initial ancestors.*
2. *If* $\mu > 1$ *then* $Z_n \to Z$ *with probability* $\mathbb{P}(Z=k) = q^m$ *for a finite value* k *and with probability* $1 - q^m$ *the sequence* $\sqrt{n}\left(\frac{Z'}{n} - z_0\right)$ *converges to the normally distributed random variable with mean 0 and variance* $\frac{\rho_b z_0 + \mu^2\sigma^2 z_0 \rho_b^2}{(1-\mu\tau\rho_b)^2}$ *where* $\rho_b = 1 - z_0$
3. *For a large value of* n *the final size of our process falls in the range* $[nz_0 - c_b\sqrt{n}, nz_0 + c_b\sqrt{n}]$ *if the branching process approximation becomes extinct with a higher probability for some fixed large value* c_b

Chapter 3

Random Graph Epidemic models

Real contact network data in the recent years has provide a strong case for the use of networks in the modeling of disease epidemiology[38]. In situations where, the infective profile is complicated, random network may be used to derived many results of the SIR epidemic including the distribution of the final epidemic size and the basic reproduction number[26]. A network consist of discrete elements (nodes) and connections (edges)[39]. Attention has been given to the modeling of stochastic epidemics through networks[39][40][34][41]. Understanding of how real-world network evolve and emerge has led to the improvement of network models. Network theory enters epidemiology as an attempt to relax the assumption that infection happens between each susceptible and infection individual in the population at a constant rate (mass action)[34]. In network epidemiology, each individual in the population is assigned to a neighbourhood and can then contact her neighbour at a normal rate. Graph theory provides the tools and mechanism for describing the application of epidemiology in networks[38].

3.1 Network Representation

A real world network can suitably be described by the means of a set of points together with lines joining certain pair of the points[42]. A graph G is a pair (V, E) of set satisfying $E \subseteq [V]^2$ (i.e. the element of E are 2 elements subset of V). The elements in the set V are called vertices or nodes and elements in E are called edges or lines. We denote the set vertices and a the edges by $V(G)$ and $E(G)$ respectively. In this thesis we consider undirected graphs thus all edges are bidirectional. For example figure 3.1 shows a simple undirected graph with $V = \{1, 2, 3, 4, 5, 6\}$ and $E = \{\{1, 2\}, \{1, 3\}, \{1, 5\}, \{1, 6\}, \{2, 5\}, \{3, 5\}, \{4\}\}$. The order of a graph G is the number of vertices denoted as $|G| = n$ and the number of edges denoted as $||G||$ (i.e.$n = 6$ and $||G|| = 7$). A graph can either be finite or infinite

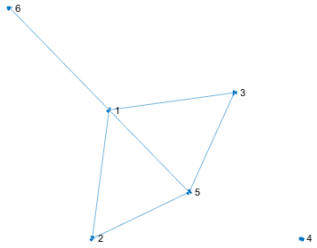


Figure 3.1: Simple undirected random graph

according to the order. The set of all edges in the set E is denoted as (u, v). The set E can be represented in an adjacency matrix where G_{uv} is 1 if $(u, v) \in E$ and 0 otherwise. An graph is said to be undirected if any edge from u_1 to u_2 corresponds an edge from u_2 to u_1 (i.e. a graph G is said to be undirected if the adjacency matrix is symmetric). If all the vertices are pairwise adjacent, the graph G is said to be complete.

3.1.1 Definition of terms

1. **Subgraph:** A graph H is a subgraph of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ (i.e. $H \subseteq G$). A spanning subgraph of G is a subgraph H with $V(H) = V(G)$. A subgraph induced by H' is a subgraph of G whose vertex set H' and edge set is the edges of G that have both ends in H' where H' is a non empty subset of V[42]. The neighbourhood of a vertex V is a subgraph of G induced by all vertices adjacent to V.

2. **Degree and Degree Distribution:** The degree $d(v)$ of a vertex is the number of edges at v. The degree of $deg(v)$ correspond to the number of neighbours of v. The number of vertices of the graph G that lead to v in a single step is given as $deg(v)^{in} = \sum_{u \in V} G_{uv}$ and the number of vertices that can be reached out in one step from v is given as $deg(v)^{out} = \sum_{u \in V} G_{vu}$. For undirected graph $deg(v)^{in} = deg(v)^{out}$. The part of a edge that is attached to a vertex is called a stub: there are two stubs per edge and each vertex is attached to a number of stubs equal to its degree. If the degree occurring in a graph denoted by $d_1, d_2, d_3, \cdots, d_n$ and n_l is the vertices with degree d_l, then the degree distribution is given as $\mathbb{P}_l = n_l/n$[42]. A more general and detailed degree distribution of a finite graph can be found in [26] on page *244*.

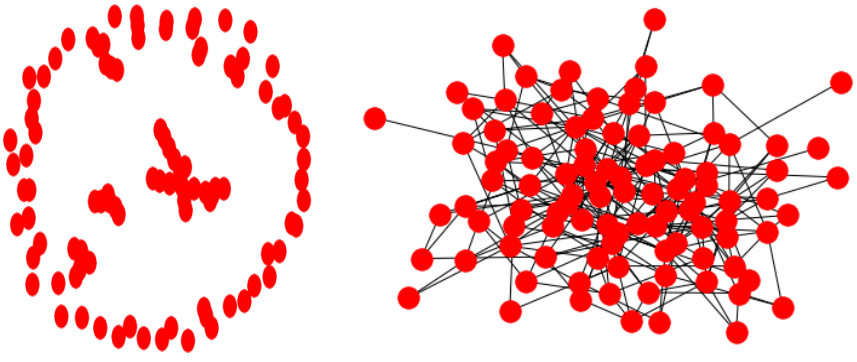
3. **Clustering and Higher Order Structure:** These measure the probability that two neighbors of a randomly chosen node share an edge to form a triangle (i.e. Clustering form a complete subgraph). Clustering leads to a quicker depletion of susceptible vertex around infected ones in network epidemiology[38].

4 **Paths and Connectivity:** The path is a non empty graph $P_{V,E}$ of the form $V = \{v_0, v_1, \cdots, v_m\}$ and $E = \{v_0v_1, v_1v_2, \cdots, v_{m-1}v_m\}$. The length of a path is the number of edges. The path of a graph G is referred to as natural sequence of its vertices. Two vertices v_1 and v_2 are connected if there is a path in G linking v_1 to v_2.

3.1.2 Erdös Rényi Graph

Let $G_{m,n}$ denote the set of all graphs with vertices $\{v_1, v_2, \cdots, v_m\}$ and n edges. A graph belonging to $G_{m,n}$ is obtain by choosing n out of $\binom{m}{2}$ possible edges such that $|G_{m,n}| = \binom{\binom{m}{2}}{n}$. To avoid dependency, we picked $\binom{m}{2}$ independent vertice between the nodes with probability p[43]. Erdös Rényi graph is social network consisting of n vertices where each pair of vertices $(u, v) \in V^2$ is independently connected with a fixed probability p. We denote the Erdös Rényi graph with probability p as $G_{n,p}$ where its distribution is defined as $(G_{uv} : u, v \in V, u < v)$ of i.i.d. random variables with Bernoulli distribution $Ber(p)$,$p \in [0, 1]$[26]. One of the properties that emerged in the modelling epidemics is the existence of giant component. The giant component refers to the largest component of the network.

- If $p = c/n$, $c < 1$, then when n is large, most of the connected nodes/edges of the graph are small with the largest having only $\mathcal{O}(\log n)$ vertice with high probability[44].
- if $c > 1$, there is a constant $\theta(c) > 0$ such that, the largest component has $\sim \theta(c)n$ vertices and the second largest is $\mathcal{O}(\log n)$ for a large n with high probability[44].



(a) (b)

Figure 3.2: Erdös Rényi Graph with (a) c = 0.9, (b) c = 5 using $n = 100$ vertice

Let $p\prime$ be the probability that a random chosen vertex $v \in V$ does not belong to the giant component. Then the probability that $v \in V$ is not in the giant component of a very

large Erdös Rényi graph is given as

$$p\prime = e^{(p\prime-1)c}[45] \tag{3.1}$$

Erdös Rényi Graph can be well approximated by the branching process[44]. We initially starts with a single vertex (m = 1) and is to be connected to Binomial$(n-1, c/n)$ number of neighbors which converges to a poison(c) as $n \to \infty$. Suppose E_k is the number of vertices at distance k then for a small k, E_k behaves like a branching process in which each individual has independent of mean c number of offspring[44].

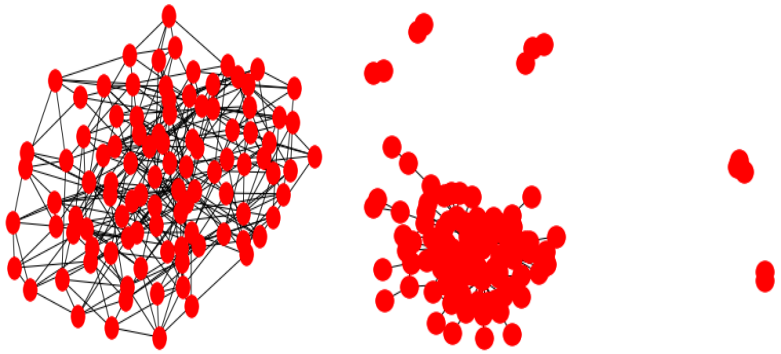
The Dynamics of Erdös Rényi Graph

The random variable $\{G_{n,p}(t), t \geq 0\}$ of the Erdös Rényi Graph $G_{n,p}$ is a stochastic process that evolves according to the dynamics below for any positive integer α_1 and α_2.

1. There is a fixed number of vertex n
2. Independently for any pair of vertices $u, v \in V$, an edge is added after Exp$(\alpha_1/n - 1)$ distributed time if no edge was present and Exp$(-\alpha_2)$ distributed time to remove an edge.

3.1.3 Configuration model graph

This is a social network that is generated from a given degree sequence. The network consist of vertices labeled $1, 2, 3, \cdots, n$ and a asymptotic degree of vertice denoted as D. Let the probability that a given vertex has a degree k be denoted as $p_k = \mathbb{P}(D = k), \quad k = 0, , 2, \cdots$ with mean μ_D and variance σ_D^2 of D. The Configuration graph model comes in two ways, Newmann-Strogatz-Watt(NSW) and Molley-Reed(MR)[2]. In the NSW model, a sequence of networks indexed by n vertices is constructed from the sequence of degree $D_1, D_2, \cdots$ of i.i.d. copies of D which uses the first n random variable for the network on n edges. We will often use the MR network model which is used to fit an observed degree distribution. Here we used the power law distribution as the degree distribution. The power law degree sequence is quite used density degree didtribution of random networks though it has many drawbacks when the network under consideration is not well studied. In the MR model, each vertex $v \in V$ is associated with an independent random variable X_v from the degree distribution which represent the number of stubs attached to v such that $\sum_{v \in V} X_v$ is even[26]. Let $d_1, d_2, \cdots, d_l$ and $n_1, n_2, \cdots, n_l$ denote the degrees of the graph and the number of nodes for each degree, then the average degree is given as $\bar{D} = 1/n \sum_{l=1}^{L} n_l d_l = \sum_{k \in Z_+} k p_k$[38]. In this network model, each



(a) NSW (b) MR

Figure 3.3: Configuration Graphs using $n = 100$ vertice

vertex $v_1, v_2, \cdots, v_n \in V$ gets a numbers of half edges(stubs) to its degree such that the stubs are joined into pairs at random to form an edge. The defects (self loop and multiple edges) associated with generating the configuration model becomes a sparse as $n \to \infty$ provides $\sigma_D^2 < \infty$[44]. The probability of a random chosen vertex $v \in V$ will have a self loop is $1/n$ as $n \to \infty$. When a stub of a vertex $v \in V$ is joined to a neighbor, the probability that it connect to a vertex $u \in V$ of degree k is proportional to the total number of stubs degree of edges. The probability of vertex $v \in V$ joining to $u \in V$ of degree k is given as

$$\mathbb{P}(D = k) = \frac{kp_k}{\bar{D}} \text{[38][26]} \tag{3.2}$$

which is refer to as the size degree distribution in [26] which plays a an essential roles in disease epidemiology. The configuration model looks likes the Erdös Rényi Graph with self loop when the degree distribution is Binomial with parameters n and p_k/n.

3.1.4 Preferential attachment models

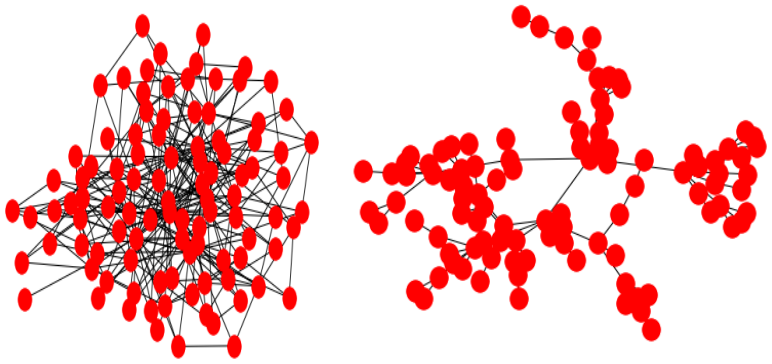
The Preferential attachment is a growing network such that the number of vertices increase throughout the life time of the network[46]. This type of network expand by adding new vertices. This type of network was first identified by Undy Yule[47] who studied the increase in the number of species within genus using the power-law distribution. Unlike the Erdös–Rényi graph and the configuration network model which is static, the preferential model network sequentially add vertices having a fixed number of edges to the network. The preferential attachment is a self organizing network with high degree[46] which is formed by continuously adding new vertices to the existing network. In this thesis we consider the Barabasi and Albert explanation to the preferential attachment in which the probability that a vertex in the network interact with other vertices decay as a power law following $P(C) \approx C^{-\alpha}$ where the exponent ranges between

2.01 to 4[46]. Examples of network that poses this network properties include, the actor network, WWW network and the citation network. There has been some authors which suggest that the preferential attachment combined with growth to produce log–normal in–degree distribution instead of power law[48].

In this network, we start with small number of vertices m_0 at every time step, a new vertex is added to v different vertex already present in the network[46]. Let Π be the probability that a new vertex will connect to vertex $v \in V$ depending on the connectivity k_v of the vertex, then

$$\Pi(k_v) = \frac{k_v}{\sum_{v \in V} k_v}[46] \tag{3.3}$$

At any given time step the model leads to a random network with $t + m_0$ vertices and vt



(a) m_0=3 (b) $m_0 = 1$

Figure 3.4: Preferential attachment model with $n = 100$ individuals

edges[46]. The rate at which a vertices acquired edges is $\partial k_v/\partial t = k_v/2t$ which integrates to give $k_v = v\,(t/t_v)$ where t_v is the time at which $v \in V$ was added to the network[46]. This network is normally referred to as the "Richer get Rich" network since edges with the highest degree will get the highest connections.

3.2 SIR Epidemic Model on graphs

In modelling epidemic in networks, individuals are represented by vertices and the patterns at which individuals get contact are represented by edges. At any given time $t > 0$ a vertex $v \in V$ is either susceptible (S), infected (I) or recovered (R). The state $(3^n states)$ of the network is given by the status of the n vertices. The rate at which an individual becomes infected and recovered is independent of the status of any other vertex $v \in V$. We assume the following;

- An infected individual infect each of its susceptible neighbor at the independent poison process with rate μ.

- An infectious individual recovers and becomes immune (independent of each other) at rate γ

Individual base stochastic simulation is the very first step in studying the stochastic process of disease spread in networks. The process involves keeping track of all 3^n possible events in the network and the rate at which each event occurs. One way to do this simulation is the Gillespie Algorithm sometimes called Gillespie's Stochastic Simulation Algorithm.

3.2.1 Gillespie's Stochastic Simulation Algorithm

This approach is a Markovian process[38] which assumes that all the 3^n events occurs independently with time. The algorithm computes the time until the next event(waiting time) and then calculate the status of the event either susceptible, infective or recovery[38]. This process helps in determining the final epidemic size since at any given time t the status of all the vertices are recorded. In this thesis our discussion and notations are taken from [49]. Let $\Omega(t)$ be the set of transition process which changes independently with time and t be the current time of our epidemic process. Here we assume a smaller step time $t + T = t_1$ of the next 3^n event which is exponentially distributed and each transition happen with a probability proportional to its rate. Let m^* be the transition of a vertice then the probability that there is no transition of vertice $v \in V$ from the state m^* after time T since the last transition is given as $S^*_m(T) = \exp(-\lambda^*_m T)$. Since transition of each vertice happens independently of each other $S^*_m(T)$ can be written as;

$$S(T) = \exp\left(-\sum_{m*=1}^{M} \lambda^*_m T\right) \text{[49]} \tag{3.4}$$

where $\lambda^*_m \in \{\beta, \gamma\}$ is the transition rate from of the state $m*$ and $\sum_{m*=1}^{M} \lambda^*_m$ is the cumulative transition rate. In the SIR model where a susceptible vertex $v \in V$ ever becomes infected only when he comes in contact with an infected vertex $u \in V$, then equation 3.4 becomes.

$$S(T, t) = \exp\left(-\int_t^{t_1} \Lambda(t)dt\right) \text{[49]} \tag{3.5}$$

where $\Lambda(t) = \sum_{m*\in\Omega} \mathbf{1}_m * (t)\lambda_m$ and $\mathbf{1}_m * (t) = 1$ when the process $m*$ takes place (i.e. when there is a transition from one state to the other.) and 0 otherwise. Detailed explanation of the Gillespie algorithm can be found in [49]. Figure 3.5 shows the Gillespie simulation algorithm for the three networks described above with $S = 1000$, $I = 1$ and $R = 0$. In the configuration model, we used the power law degree distribution with exponent 3 for the preferential attachment model. A probability value of 0.004 was used for the Erdös Rényi graph (i.e. $G_{1000,0.004}$).

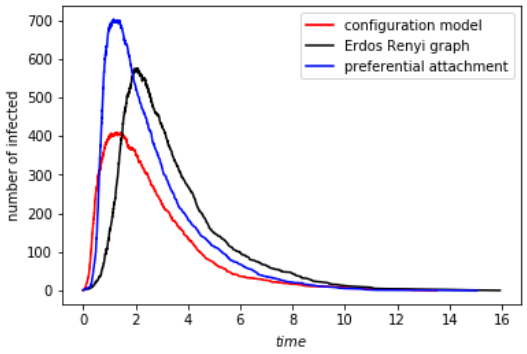


Figure 3.5: Gillespie stochastic algorithm with $\mu = 1.5$ and $\gamma = 0.5$

3.2.2 SIR on Erdös Rényi graph

Consider the sequence of the standard epidemic $B_{n>1,1}(\mu, F)$. If the infectious period F is constant then if an individual v ever gets infected he will then contact his neighbour with probability $p_r = 1 - exp(-\mu F/n)$. Let $n = n^* + m$ where n^* is the number of susceptible individuals in the population and m is the initial invectives at $t = 0$ then we get the Erdös Rényi random network $G(n, p)$. The giant component above can be used to derive results corresponding to the epidemic process. When $c > 1$ the asymptotic probability of large outbreak is equal to the size of the giant component.
Lets consider the construction of epidemic on graph defined by [50]. Let $S(0) = \{1, 2, \cdots, n\}$ and $I(0) = \{-(m-1), -(m-2), \cdots, 0\}$ be the set of initial susceptible and infected individuals in the population. Let $I(t+1)$ denotes the number of susceptible individuals who are infected by the infective(s) $I(t)$ at time $t \geq 0$. At any given time point $t \geq 0$, $I(t+1) \subseteq S(t)$. The set of susceptible individuals who remain susceptible at least time $t+1$ is denoted as $S(t+1) = S(t)/I(t+1)$. At time $\mathcal{T}$, the epidemic ceases, where $\mathcal{T} = \{\inf\{t\} \mid I(t) = 0\}$. For any given individual $v \in I(t)$ and $w \in S(t)$, where $0 \leq t \leq \mathcal{T}$, v infects u with the probability $p_c p_r$ where p_c denotes the probability that there u and v are connected (neighbours). Then if follows that

$$S(t+1) \sim \text{Binomial}(S(t), \mathcal{D}I(t)) \qquad 0 \leq t \leq \mathcal{T} - 1 \tag{3.6}$$

where $\mathcal{D} = 1 - p_c p_r$. Thus for any given individual $-(m-1) \leq v \leq n, \quad v \in I(t)$ has the sampling distribution $1 - \mathcal{D}$ of making infectious contact with individual $u \in S(t)$. This process defines the Reed-Frost model described in chapter 1 since $\mathcal{D}$ is independently and identically distributed according to the distribution $\mathcal{D}_L = 1 - p_c(1 - e^{-\mu F})$.
Now consider the case where $v \in I(t)$ and $u \in S(t)$, $t \geq 0$, and assume that there exist an edge between v and u (i.e. $G_{uv} = 1$) else set $G_{uv} = 1$ with the probability

$$\frac{p_c e^{-\mu F}}{1 - p_c + p_c e^{-\mu F}} [50] \tag{3.7}$$

Equation 3.7 generates graph between infected individuals in the population in different discrete time point $t+1$. We can confirm that $G_{uv}=1$ with probability p_c for $v,u \in I(t), \quad 0 \leq t \leq t\mathcal{T}$ or $v,u \in S(\mathcal{T})$ thus

$$P(G_{uv}=1) = p(1-p_c) + (1-p_c(1-p_c))\frac{p_c e^{-\mu F}}{1-p_c(1-p_c)} = p_c[50] \tag{3.8}$$

We generate exact results from the findings above following [30] and [50]. Consider the standard SIR epidemic $B_{n>1,1}(\mu, F)$. Let $q_k, \quad k=0,1,2,\cdots,n$ denotes the probability that an infective individual whiles infectious fails to contact the set of k susceptible. Let $\phi(\theta) = E(\exp(-\theta F)), \theta \geq 0$ denotes the Laplace transform of F, we have the

$$\begin{aligned} q_k &= E(\mathcal{D}^k) \\ &= E(((p_c e^{-\mu F}) + (1-p_c))^k) \\ &= \sum_{i=0}^{k} k \binom{k}{i} p_c^i \phi(i)(1-p_c)^{k-i}, \quad k=0,1,\cdots,n \end{aligned} \tag{3.9}$$

Let the probability that the final size of our epidemic process j be denoted $P_j^n \quad 0 \leq j \leq n$ then we have

$$\sum_{k=0}^{j} \binom{n-k}{j-k} \frac{P_j^n}{(q_{n-j})^{m+k}} = \binom{n}{j} \qquad 0 \leq j \leq n[30][50] \tag{3.10}$$

If we set $p_c = 1$, equation 3.10, the term q_{n-j} reduces to $\phi(\mu(n-j/n))$

Transmission Dynamics of the Epidemic Process

In general the state of a individual v is either S,I or R. We begin with m infected individual(s) in the population. During the time period F, an infectious individual contacted his neighbour according to the poison process with intensity λ. The contacted individual becomes infected if he is still susceptible. The infected individual recovers and become immune after his infectious period.
Let $A_v = 1$ if the vth individual is of status A and $A_v = 0$ otherwise. Using the above notation to represent the number of individual connected pairs and connected triple, we have

$$[A] = \sum_u A_v, \quad [AB] = \sum_{u,v} A_u G_{uv} B_v \text{ and } [ABC] = \sum_{u,v,w} A_u G_{uv} B_v G_{v,w} C_w$$

From 3.2.2, $[AB] = [BA]$ since the studied network is undirected. The quantities above are $\mathcal{O}(n)$ (proportional to the number of individual) when the total number of a given individual is bounded[30]. If random chosen individual has c neighbours then the basic

reproduction number is given as

$$R_0 = \frac{c\mu}{\mu+\gamma}[30] \tag{3.11}$$

where $\mu/\mu+\gamma$ is the probability of contacting an infected individual before he recovers. Now describing the time dynamics of our model, let $[SI]$ denotes the number of individual u, v where u is susceptible and v is infectious at a given time t. The epidemic on the network $G(n,p)$ obeys exact but unclosed system of equation

$$\begin{cases} \left[\dot{S}\right] = -\mu\left[SI\right], \quad \left[\dot{I}\right] = \mu\left[SI\right] - \gamma I, \quad \left[\dot{R}\right] = \gamma\left[I\right] \\ \left[\dot{SS}=\right] = -2\mu\left[SSI\right] \\ \left[\dot{SI}\right] = \mu\left(\left[SSI\right] - \left[ISI\right] - \left[SI\right]\right) - \gamma\left[SI\right] \\ \left[\dot{SR}\right] = \mu\left[ISR\right] + \gamma\left[SI\right] \\ \left[\dot{II}\right] = 2\left(\left[ISR\right] + \left[SI\right]\right) - 2\gamma\left[II\right] \\ \left[\dot{IR}\right] = \mu\left[ISR\right] + \left(\left[II\right] - \left[IR\right]\right) \\ \left[\dot{RR}\right] = \gamma\left[IR\right] \end{cases} \quad [30][45] \tag{3.12}$$

Writing the system above in triple in terms of higher order structure but allow the equation to be close at the level of pairs lead to $[ABC] = \frac{[AB]\,[BC]}{[B]}$[30]. Manipulating the closed system, the number of infectives in the early epidemics and the final proportion of the population susceptible can be written as

$$I(t) \propto \exp(rt)[45] \tag{3.13}$$

$$s = \left(1 - \frac{R_0}{c} + \frac{R_0}{c}s^{(n-1)/n}\right)^n [45] \tag{3.14}$$

respectively where $r = (n-2)\mu - \gamma$ is the exponential growth rate of infectious individual. The factor -2 existed because at the early stage of an epidemic in a large population, the average infectious individual has already infected exactly one of it contacted individual and also any individual at the susceptible state who becomes infected must have an infectious contact with a individual in infectious state. In a complete graph, $r = \mu - \gamma$[45].

Chapter 4

SIR Dynamics on Graph

Consider the sequence of the standard SIR epidemic $B_{n>1,1}(\mu, F)$ where an infected individual continues to be active (infectious) until time F with the distribution given by F_L.

We now describe an adaptation of the standard SIR model with dropping of $S-I$ connections. We begin with m initially infected individual (I_t= m) in the population. During his active period, an infected individual will continue to spread the disease until he recovers and become immune after a random time F with cumulative distribution function given by F_L. A susceptible neighbour of an active individual after learning the status of his neighbour breaks their connection at rate ρ, and with probability p_d he reconnects with a randomly chosen individual in the population. The rate at which a susceptible individual forms new edges in place of the previously deleted edge is ρp_d. The process described with the parameters μ and ρ will be referred to as $evoSIR$. We shall also consider the case where after a susceptible individual breaks the connection and never rewires with any individual in the population (i.e. ρ= 0). This process will be referred to as delSIR. We focus on two (2) exceptional cases in the next section where infected individual lasted for a constant time period and exponentially distributed time with mean 1.

In this thesis, we focus on the dynamics of the Erdös Rényi graph, configuration graph model and the Preferential Attachment model.

4.1 Critical Value

This quantity plays an exceptional role in disease modelling. Denote the number of infected individual at time t during an epidemic by I_t, Then the final size of the epidemic

can be written as

$$I_\infty = \lim_{t\to\infty} I_t \tag{4.1}$$

From equation 4.1, we have that a large epidemic occurs if I_∞ is $\mathcal{O}(n)$. Then we can say for any positive number ϵ a large epidemic occurs if

$$\lim_{n\to\infty} \sup \mathbb{P}_1 \left(\frac{I_\infty}{n} > \epsilon \right) > 0[51] \tag{4.2}$$

where $\mathbb{P}_1$ is when the process begins with one initially infected individual in the population (i.e. (i.e. $I_0 = 1$). Now fixing all parameters in our SIR process (i.e. μ, γ and ρ), the critical value μ_c is the smallest value of μ for which we have a large epidemic. From 4.2 if $\mu < \mu_c$, then for any positive real number ϵ

$$\lim_{n\to\infty} \sup \mathbb{P}_1 \left(\frac{I_\infty}{n} > \epsilon \right) = 0[51] \tag{4.3}$$

4.2 SIR with fixed infection period on ER–graph

From the previous chapter, we know that if the infection period of a randomly chosen individual in the population is constant, then the final set of infected individual becomes the Erdös Rényi graph. Consider the case where an edge becomes $S - I$ only once (i.e. friendship between susceptible individual only last for time one) An infected individual will transfer the infection to his neighbour with the probability

$$p = 1 - e^{-\mu} \tag{4.4}$$

a susceptible individual will escape from an infection with the probability $e^{-\mu}$. The reduced graph after edges have been deleted gives the final epidemic size when a member of the cluster is infected with mean degree cp where c is the mean degree of the original graph. Therefore there is a positive probability of a large epidemic if $cp > 1$. Since the number of infected at the early stage of the epidemic can be approximated by the branching process, representing the generating function in equation 2.4 by

$$G(s) = e^{-cp(1-s)} \tag{4.5}$$

then the probability that an individual will start a large epidemic is $1 - q$ where q is the smallest root of equation 4.5 .[1] described a more detailed approach by Martin Löf by proofing the central limit theorem for the number of infected individual in a large epidemics. The critical value of this process must satisfy the equation

$$cp = 1 \implies c\left(1 - e^{-\mu_c}\right) = 1 \tag{4.6}$$

Now let us consider the case where infection can occur before an individual breaks connection with an infected individual at rate ρ. Consider the process $B_{n>1,1}(\mu, \rho, F)$ where $S-I$ edges break the connection at rate ρ. For infection to cross an edge at a constant rate 1, the infection must happen before rewiring and before an infected individual gets immune or dies. $S-I$ edge becomes $I-I$ is exponentially distributed with parameter μ (i.e. $\exp(\mu)$), and $S-I$ breaks the connection and rewired with a random individual in the population is exponentially distributed with parameter ρ (i.e. $\exp(\rho)$), then minimum becomes $\exp(\mu+\rho)$. Let T be the time until infection spreads across an edge, and let Q be the time until the edge is removed, then the probability that infection is transferred from an active individual to a susceptible before rewiring is given by,

$$\mathbb{P}(\text{infection spread}) = \mathbb{P}(T < 1, T < Q) = p_c^r = \frac{\mu}{\mu+\rho}\left(1 - e^{-(\mu+\rho)}\right) \tag{4.7}$$

Let $\tilde{G}(n, p_c^r)$ be the graph generated after rewiring during the epidemic process. The large epidemic occurs with positive probability if the reduced graph $\tilde{G}(n, p_c^r)$ has a giant component.

Theorem 4.1 ([1])**.** *Consider the evoSIR with parameters μ, γ and ρ where an $S-I$ edge becomes $S-I$ only once,*

1. *If p_c^r gives the transmission probability after rewiring, then the critical value of the total infection rate must satisfy $p_c^r = 1$.*
2. *As $n \to \infty$, the ratio of the expected epidemics in delSIR to the size in evoSIR in the sub-critical case (when $cp < 1$) converges to 1.*

Point two of theorem 4.1 can be explained as follows as given in [1]: From coupling techniques, the final set of recovered/removed individuals in delSIR corresponds to the set in evoSIR when same parameters are defined on each process thus

$$\mu_c(\text{evoSIR}) \leq \mu_c(\text{delSIR}) \tag{4.8}$$

To prove that the two are equal, we have to show that if $\mu \leq \mu_c(\text{delSIR})$, then the evoSIR will dies out. Also we can compare the evolution of the two processes by

1. Run the the delSIR to completion.
2. We randomly rewired all the deleted edges in the delSIR until no infection is created after rewiring.

Now Consider the Erdös Rényi graph $\mathcal{G} = G(n, cp/n)$, where p is the probability that infection is transferred from an infected individual to a susceptible neighbour. Let the probability of eliminating a successful infection due to rewiring be denoted as

$$\alpha = 1 - \frac{p_c^r}{p} \tag{4.9}$$

We also interpreted equation 4.9 as; the probability of deleting edges from the graph $\mathcal{G}$. Edges are rewired in the evoSIR with the same probability value α. Let R' and R denotes the set of sites that are infected in the delSIR at $t = \infty$ and the set of sites infected at time $t = \infty$ in the evoSIR respectively. From lemma 4 in [1], If $cp\alpha < 1$, then there exist some positive constant C_1 and C_2 so that

$$P(R > C_1 \log n) \leq C_2 n^{-3/2} \quad and \quad E(R) = E(R') + o(n) \tag{4.10}$$

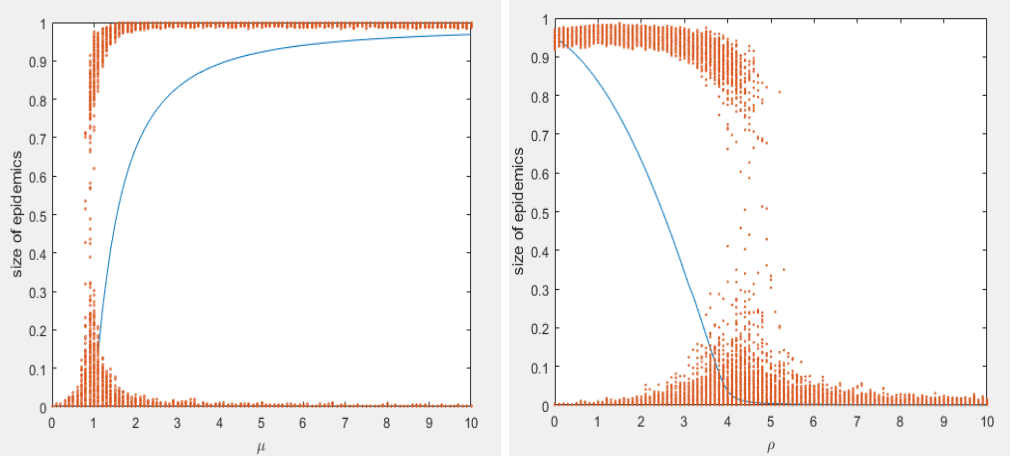
We can see from figure 4.4 below that if $\mu > \mu_c$, then the delSIR will have a large epidemic with positive probability. Since the epidemic size in evoSIR couples to be large, we will also have a large epidemic in the evoSIR. If $\mu < \mu_c$, the summable bond of equation 4.10 implies $R \leq C_1 \log n$ for a large enough n with higher probability. Thus the probability of large epidemic converges to 0. [52] shows the proof of theorem 4.1 but this time they considered the dynamics on the configuration model graph.
The equality of the two critical values for delSIR and evoSIR hold since the sub-critical epidemics, delSIR dies out quickly, so rewiring has no significant effect. When the size of the population is large, the degree distribution of the Erdös Rényi graph with parameter $n - 1$ and c/n is approximately poison distribution with mean c. Then the limiting generating function after rewiring can be written as

$$\tilde{G(s)} = e^{\mu p_c^r (1-s)} \tag{4.11}$$

where μp_c^r is the mean of the asymptotic poison distribution.

Theorem 4.2 ([1]). *If q is the fixed point of $\tilde{G}(s) < 1$, then $1 - q$ gives the probability of the large evoSIR and delSIR epidemics.*

When $c_c = 1/p_c^r = 1$, the proportion of the fraction of infected individual in the large epidemics is zero[1].
The left-hand side of figure 4 ((a)) shows a varying infection rate μ while the rate at which a susceptible individual breaks friendship with an infected individual $\rho = 4$. The red plot shows a simulation of the fraction of individuals infected during the epidemic in the evoSIR. The blue curve shows the delSIR with the same parameters. The delSIR shows a continuous transition since the final epidemic size is the size of the giant component of the subgraph of the ER-graph obtained by deleting edges randomly. It can be seen



(a) (b)

Figure 4.1: simulation of the constant time Evosir on the Erdös Rényi graph with $c = 5$ and $n = 1000$

from the figure that the final size of the epidemic increases as ρ increases. [1] also shown that with the same parameters, the fraction of infected individual in the population is seemingly discontinuous at the critical value (i. e. as the infection rate μ decreases to μ_c, the final fraction of infected individuals does not converge to 0.). In right-hand side ((b)), μ= 1 while ρ varies. As ρ increases, the final size of the epidemic approaches 0. Deleting edges reduces the ratio of infection by each active individual during epidemics.

4.3 SIR with exponential infection period on ER–graph

Now let's consider the sequence of the standard SIR epidemic $B_{n>1,1}(\mu, F)$ where the infected individual recovers and becomes immune with an exponential distribution with parameter 1 (i. e. exponential(1)). Then infection is transferred from an infected individual during his time of infection with probability

$$p_e = 1 - \int_0^\infty e^{-t} e^{-\mu t} dt = \frac{\mu}{\mu + 1} [1] \tag{4.12}$$

The critical infection rate must satisfy the equation $cp_e = 1$ which gives

$$\mu_c = \frac{1}{c - 1} \tag{4.13}$$

If the infection times are assumed to exponential(α_0) for any $\alpha_0 > 0$, the infection status of the edges going out of the vertex are correlated which makes it not easy to reduce to percolation (the behaviour of connected clusters in the graph)[1]. The generating

function of the number of infections directly caused by one infected individual is

$$G(s) = E\left(G\left(e^{-\mu F} + \left[1 - e^{-\mu F}\right]\right)\right) \text{ [1]} \tag{4.14}$$

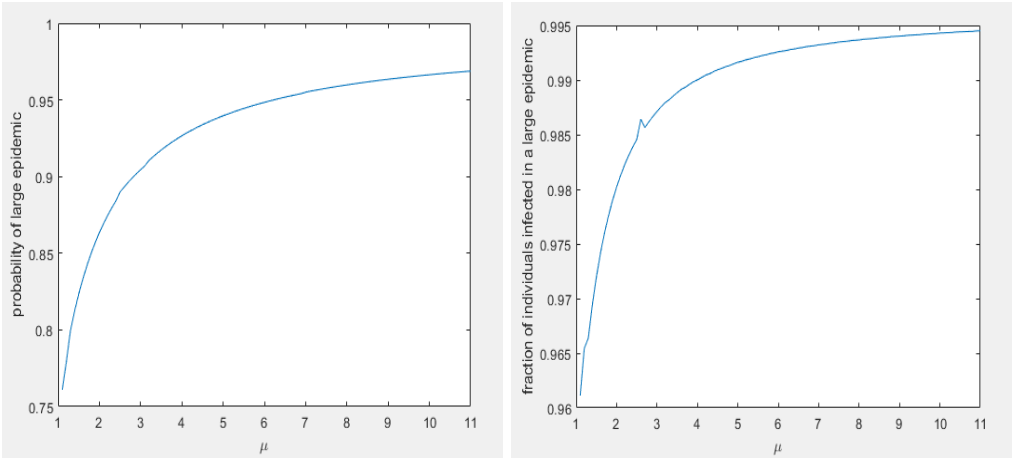
Since when n is large, the degree distribution is approximately poison distributed with the generating function $G(s) = e^{(c(1-s))}$ then we have

$$\hat{G}(s) = e^{-c(1-s)} \int_0^\infty e^{-t} e^{c(1-s)e^{-\mu t}} \tag{4.15}$$

Theorem 4.3 ([1]). *Consider the evoSIR with parameters μ and ρ where infection times are exponentially distributed with mean 1, then*

1. *If $q_1 < 1$ is a fixed point of equation 4.15, then $1 - q_1$ gives the probability of a large epidemic.*
2. *If $q_2 < 1$ is a fixed point of the the equation $\exp(-cp(1 - q_2)) = q_2$, the $1 - q_2$ gives the fraction of individual infected in a large evoSIR.*

From theorem 4.3, we can see that the probability of large epidemic differs between the delSIR and evoSIR. Figure 4.2 shows a pictorial view of 4.3. Here we can see that both



(a) (b)

Figure 4.2: Comparing the probability of large epidemic and fraction of individuals infected in a large epidemic using $c = 5$ on the ER–graph

quantities increase when the infection rate μ increase in a giant ER-graph. For any giving value μ, the fraction of individuals infected has a more substantial value compared to the probability of large epidemics for the case of our chosen parameters.
Assuming that infected individual recovers and becomes immune with exp(1) and after a susceptible individual learning the status of infected neighbour rewired with a randomly

chosen individual in the population with $\exp(\rho)$. The minimum of the two exponential distribution is $\exp(1+\rho)$. The transition probability of this process is

$$p_e^r = 1 - \int_0^\infty (1+\rho)e^{-(1+\rho)t}e^{(-\mu t)}dt = \frac{\mu}{\mu+\rho+1}[1] \tag{4.16}$$

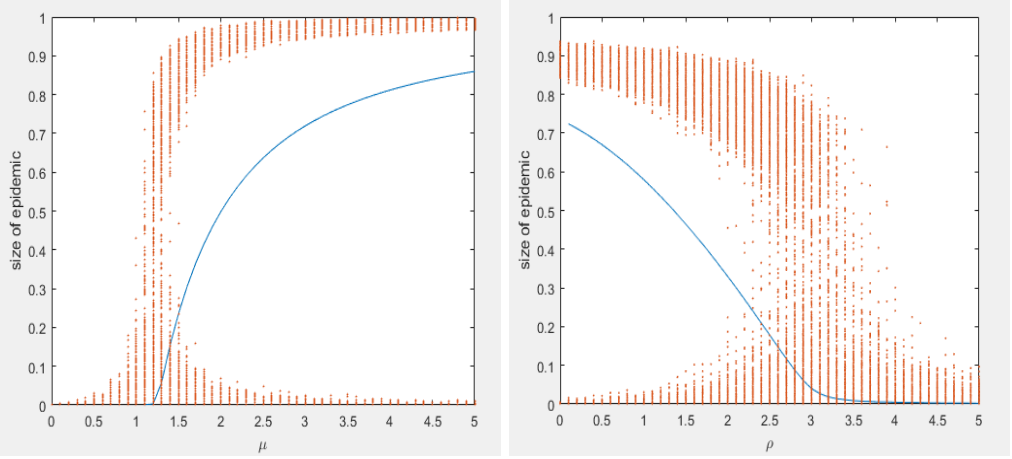
From the same reasoning from the previous sections, the critical value μ_c must satisfy the equation $cp_e^r = 1$. [1] proved that if the value of$T$ is conditioned (i.e. infection occurs before rewiring and infection and rewiring times occurs before time T), then the two (delSIR and evoSIR) processes have equal critical value using the transmission probability and the generating function below respectively

$$\tau(T) = \int_0^\infty \mu e^{-(\mu+\rho t}dt = \frac{\mu}{\mu+\rho}\left(1 - e^{-(\mu+\rho)T}\right)[1] \tag{4.17}$$

$$\hat{G}(s) = e^{-\mu_r(1-s)}\int_0^\infty \mu e^{-t}\exp(\mu_r(1-s)e^{-(\mu+\rho)t})dt[1] \tag{4.18}$$

where $\mu_r = \dfrac{c}{\mu+\rho}$

Theorem 4.4 ([1]). *If q is the fixed of $\hat{G}$, $1-q$ gives the probability of the large epidemics for delSIR and evoSIR when infection period is exponentially distributed with mean 1.*



(a) (b)

Figure 4.3: simulation of the exponential time evoSIR on the Erdös Rényi graph with $c = 5$ and $n = 1000$

The figure shows the fraction of individuals infected in a SIR epidemics. Here we consider the case of a giant component graph $c > 1$. For the left-hand side (a) of figure 4.3 , μ varies and $\rho = 4$. The red plot shows a simulation of the fraction of individuals infected during a SIR epidemic. The blue curve shows the delSIR with the same parameters. It can be seen from the figure that the final size of the epidemic increases as μ increases

but the fraction of individuals infected is continuous at the critical value. In (b), μ= 1 while ρ varies. As ρ increases, the density decreases for a large μ

4.3.1 The basic reproduction number (R_0)

This is the expected number of secondary infections caused by a newly infected individual at the early stage of the epidemic. This is a threshold value that determines whether there will be a major outbreak of an epidemic or dies out quickly. If R_0> 1, there is a positive probability of a major outbreak and if $R_0 \leq 1$, the epidemic dies out when introducing in a finite population. [3][2] gives the general formal formulation for R_0 on a configuration graph model where the degree distribution has the parameters mean μ_D and variance σ^2 for preventing dropping edges as

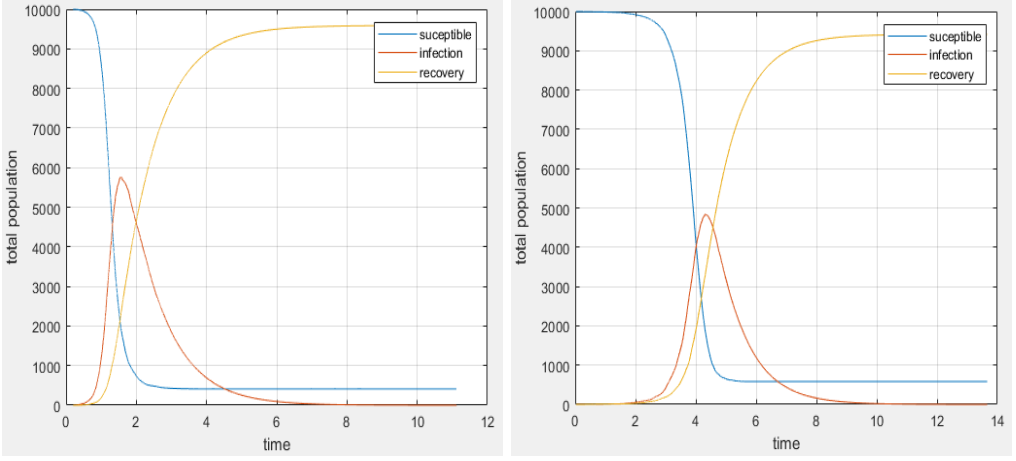
$$R_0 = \frac{\beta}{\beta + \omega + \gamma}\left(\mu_D + \sigma^2/\mu_D - 1\right) \tag{4.19}$$

For our case β=μ, ω=ρ. Since the degree distribution of the Erdös-Rényi is approximately poison(μ_D), which implies σ_D=μ_D thus, we have the following as our reproduction number

$$R_0 = \frac{\mu}{\mu + \rho + \gamma}(2\mu_D - 1) \tag{4.20}$$

The equation is independent of p_d. This is because, at the early stage of the epidemic, a susceptible individual after deleting his connection with an infected individual will connect to a susceptible individual.

Figure 4.4 shows the plots of the delSIR dynamic on a giant component. We assumed



(a) $\rho = 0$ (b) $\rho = 4$

Figure 4.4: SIR dynamics on a giant component with $\mu = 2$, $c = 4$ and $n = 10000$

that the epidemic started with 1 initial infected individual and the length at which infected individual becomes infectious to his neighbour is exponentially distributed with

mean 1. We see from our figure that, the fraction of epidemics that result in major outbreaks decreases as the rate at which edges are deleted increases. The results show that delSIR reduces the final size of the epidemic but takes a longer period for the epidemic to dies out based on our chosen parameters.

4.4 SIR dynamics on Configuration graph model

Recall that a vertice $v_1 \in V$ connect with another vertex $v_2 \in V$ with probability proportional to the degree, then the number of stubs attached to v_1 is given as $kp_k/\overline{D}$ where $\overline{D}$ is the mean degree distribution. During an early stage of an epidemics, individuals that get infected at the time of infection, have a sized biased degree distribution of neighbours $D^* \sim \{\tilde{p_k}\}$ where $\tilde{p_k} = kp_k/\mu_D$. Since the connection is made from a single edge $e_1 \in E$, the number of edges outgoing from all neighbours of v_1 has the degree distribution $D^* - 1$. Assuming a neighbour of v_1 is infected, and the remaining neighbour (s), with positive probability during the early stage of an outbreak is (are) susceptible. Since an infected individual cannot infect his infector, $D^* - 1$ susceptible are of interest. The mean number of susceptible neighbours a typical newly infected individual will have during the early stage of an epidemic is given as

$$E(D-1) = \mu_D - 1 + \sigma_D/\mu_D \tag{4.21}$$

where μ_D and σ_D^2 are the mean and variance of D, respectively. Since a susceptible individual may break the connection and rewire with a randomly chosen individual in the population at rate ρ, we know from the previous section that with probability $\mu/\mu + \gamma + \rho$, a newly infected individual can infect his neighbour. Now assuming a susceptible neighbour after breaking the connection with an infected individual, do not rewire (ρ= 0) then the basic reproduction number in equation 4.19 is given as

$$R_{\rho_0} = E(D-1)\frac{\mu}{\mu+\gamma} \tag{4.22}$$

From equation 4.19 and 4.22, deleting an edge at a higher rate reduces the proportion of the number of number of individual effected during an epidemic. At the early stage of an epidemic, the number of infectives will asymptotically behave like the branching process. The expected birth rate plays an essential role in determining the exponential growth of an epidemics. To achieve the transmission from a newly infected individual to his neighbour, the following conditions must be satisfied

- The newly infected individual must still be infectious.

- The should be no rewiring from a susceptible neighbour away from the infectious person..

Let $\mu(t)$ denote the expected birth rate of our branching process at time t, then from [53] we have

$$\mu(t) = \left(\mu_D - 1 + \frac{\sigma_D^2}{\mu_D}\right)\mu e^{-(\mu+\gamma+\rho)t} \tag{4.23}$$

The expected birth rate $\mu(t)$ determines the exponential growth of the epidemics and also can be used in calculating the average number of births (i.e. R_0) during an epidemic by computing $\int_0^\infty \mu(t)dt$. For equation 3.13, $r = (n-2)\mu - \gamma$, which is referred to as Malthusian parameter for growth. In the case of evoSIR and delSIR, r is given by the solution of the Lotka-Voltera equation $1 = \int_0^\infty e^{-r}\mu(t)dt$. Then after integrating and performing some algebra yields,

$$r = \mu E(\bar{D} - 2) - \gamma - \beta = \mu\left(\mu_D - 2 + \frac{\sigma_D^2}{\mu_D}\right) - \gamma - \rho[53] \tag{4.24}$$

The equation 4.24 is also independent of p_d.

4.4.1 Critical values and Generating functions

Let V_m be the number of vertices distance m from $v_1 \in V$; then the process converges to the branching with $V_0 = 1$ where D and D^* represent the distribution of the initial and future degree distribution respectively. Let m_k, $k > 1$ be the kth moment of D, then the limiting branching process will not go extinct if

$$1 < E(D^* - 1) = \frac{m_2 - \mu_D}{\mu_D} \tag{4.25}$$

From section 2.4, we let $G(s) = \sum_{k=0}^{\infty} p_k s^k$ denote the generating function of the degree distribution D, then the distribution of the subsequent generation should satisfy

$$\mathbb{P}(D^* - 1 = k - 1) = \frac{kp_k}{\mu_D} \tag{4.26}$$

From equation 4.26, we can compute the generating function of the $D^* - 1$ as

$$G_p(s) = E(s^{D^*-1}) = \sum_{k=1}^{\infty} \frac{kp_k}{\mu_D} s^{j-1} = \frac{G'(s)}{\mu_D} \tag{4.27}$$

For a fixed time delSIR, we know from the section 4.2 that infection will cross an $S - I$ edge (i.e. $\mathbb{P}(T < 1)$) is given us $\mu/\mu + \rho$, then using theorem 4.1, a large epidemic occurs

if

$$\frac{\mu}{\mu+\rho}\frac{m_2-\mu_D}{\mu_D} > 1 \tag{4.28}$$

Theorem 4.5 ([52]). *Let μ_{cc} denote the critical value of the delSIR and evoSIR, if the third moment (skewness) of the degree distribution if finite ($E(D^3) < \infty$) then*

1. *delSIR and evoSIR have the same critical value given as $\mu_{cc} = \rho\mu_D \dfrac{1}{m_2 - 2\mu_D}$.*
2. *Let $\alpha = \rho\mu_D/\mu$ and $\alpha_c = m_2 - 2\mu_D$, then if $\alpha = \alpha_c$, the probability of a large epidemic is the same for both delSIR and evoSIR.*

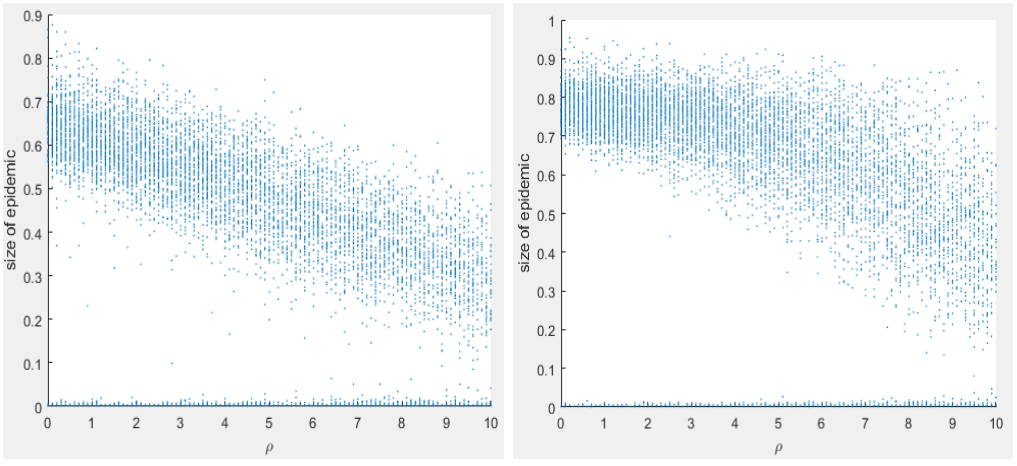
In the case of exponential infection period, there will be a large epidemic for the delSIR if

$$\frac{\mu}{\mu+\rho+1}\frac{\mu_D}{m_2-2\mu_D} > 1 \tag{4.29}$$

then the critical value for exponential infection period is given as

$$\mu_c^e = (1+\rho)\frac{\mu_D}{m_2-2\mu_D} \tag{4.30}$$

Figure 4.5 shows a simulation of the evoSIR on the configuration model network taking



(a) exponential infection period (b) constant infection period

Figure 4.5: Simulation of evoSIR on the configuration model with $n = 10000$

μ= 4 whiles varying the rate at which and $S - I$ breaks the connection and connect to a randomly chosen individual in the population. Here we use the power law as the degree distribution with exponent 2.5. We can see from the figure that there is a positive probability of large epidemic size if $\rho \to 0$ for the two infection period considered (i.e. the fraction of epidemic resulted in a major outbreak decreases with increasing the rate at which edges are deleted). Based on our parameters, the fraction of individual infected during an epidemic of the exponential infection period is less as compared to the

constant infection period as ρ increases. We can see a surprising result that the size of the epidemic in the supercritical case in (b) of figure 4.5 is continuous but discontinuous in 4.1 the critical value. [3] (figure 1) shows similar simulation by assuming that the process starts with 10 initial infected individuals in the population. Another surprising result is the simulation of the exponential infection time in figure 4.5 which needs to be studied further.

Here we will follow [52] to compute the the size of the giant component in the delSIR. Let $\eta = \mu/\mu + \rho$ denotes the probability that infection will cross a $S - I$ edges. The generating function of the reduced graph after edges have been deleted is given as

$$\begin{aligned} G_0(s) &= \sum_{k=0}^{\infty}\sum_{j=k}^{\infty} p_j \binom{j}{k} \eta^k (1-\eta)^{j-k} s^k \\ &= \sum_{j=0}^{\infty} p_j \sum_{k=0}^{j} \binom{j}{k} (\eta s)^k \\ &= G(s\eta + (1-\eta)) \end{aligned} \tag{4.31}$$

Thus, G_0 is the distribution of the degree distribution thinned by flipping coins with probability η of head to see the edges retained. Also we can compute the generating function of the distribution $(D^* - 1)$ (i.e. distribution of the reduced graph of the offspring distribution in the second phase of the branching process) as

$$G_1(s) = G_p(s)(s\eta + (1-\eta)) \tag{4.32}$$

There is the need to also compute the distribution of the deleted edges denoted in [52] as $(D^* - 1)_{1-\eta}$ which is given as

$$G_2(s) = G^*(s(1-\eta) + \eta) \tag{4.33}$$

To compute the size of the delSIR, we follow theorem 4.2 and 4.3. From chapter 2, we know that the probability that a branching process started by a first generation particle dies out is given by the smallest solution $G_1(q) = q$. Also the probability that the two-phase branching process dies out is given as $G_0(q)$. Since there is a unique giant component in the configuration graph model, any vertices that start a supercritical branching process must have the same component, thus giving the size of the giant component in the delSIR as

$$1 - G_0(q) \tag{4.34}$$

4.5 SIR Dynamics on the Preferential Attachment Model

The canonical preferential attachment model discussed in chapter 3 begins with m nodes all connected. This network sequentially adds nodes until there are n. For the preferential attachment, each time a node joins the network, it connects to m existing node selecting the neighbour with probability proportional to the current degree. The order of how individual enters the network is required but not necessarily the same as the epidemic order[54]. Let $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_1 n)$ define a random vector variable whose support (set of indices i such that $\sigma_i = 0$) is all $n!$ possible permutation of the set $\{1, 2, \cdots, n\}$. Labelling the individual $\sigma_i (1 \leq i \leq n)$ by the epidemic order is the ith individual that enters the network. Assume that, initially we have two individuals connected such that when individual $\sigma_i (3 \leq i \leq n)$ enters the network, it connects with X_i existing individual, where X_i followed the censored poison distribution with mean μ_p and support $\{1, 2, \cdots, i-1\}$ that is

$$P(X_i = x) = \begin{cases} e^{-\mu_p(1+\mu_p)} & , x = 1 \\ \dfrac{\mu_p x e^{\mu_p}}{x!} & , 2 \leq x \leq i-2 \\ \sum_{z=i-1}^{\infty} \dfrac{\mu^z e^{-\mu}}{z!} & , x = i-1 \end{cases} \tag{4.35}$$

where X_i, X_j are independent if $i \neq j$. When individual $\sigma_j (1 \leq j < i)$ gets connected to σ_i with a probability proportional to its current degree $\sum_{k=1}^{i-1} G_{\sigma_k \sigma_j}$.

Epidemic studies on the preferential attachment have suggested that the epidemic threshold for a large outbreak is zero[55]. In a large population, an infection can spread and create a large epidemic no matter the value of the transmission probability.

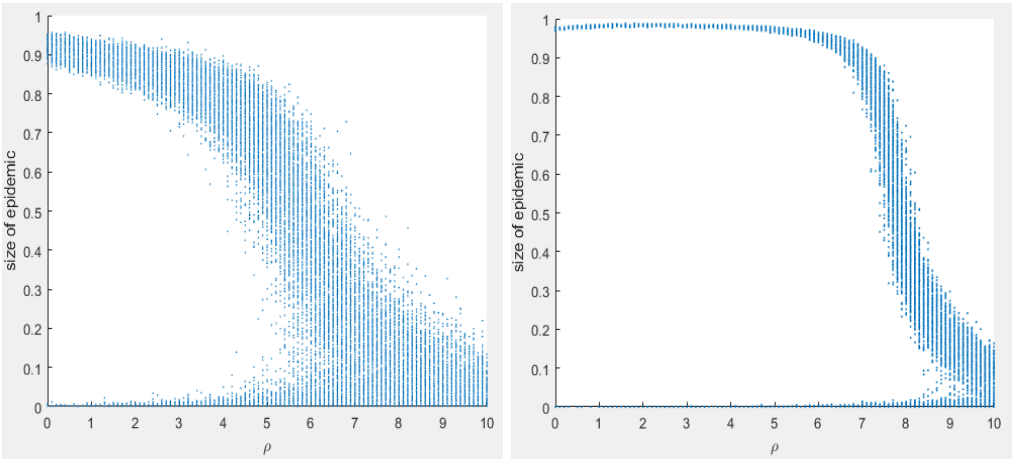
Let's consider epidemic dynamics on the preferential attachment with four individuals $\{v_1, v_2, v_3, v_4\}$ as given in [55] where the infected individual remains active for a constant time h and let R_∞ denote the number of individual recovered at time $t = \infty$. Denote the network by G with probability δ of each individual in the population attached to $I_0 = v_1$ to each i in the population such that

$$\delta = \begin{cases} 0 & \text{if } i \text{ isolated in } G \\ 1/2 & \text{if } i \text{ has one neighbor in } G \\ 1 & \text{if } i \text{ has two neighbors in } G \end{cases}$$

then the probability that infection does not spread from v_1 is given by a transmission probability $p_h = 1 - e^{-\mu h}$

$$\begin{aligned} Prob(R_\infty = \{v_1\}) = \int_0^\infty \rho^{-\rho h} \Big(& (1-\delta)^3 + 3/4(1-\delta)^2\delta + 3/8(1-\delta)\delta^2 \\ & + (1-p_h)(3/2(1-\delta)^2)\delta + 9/8(1-\delta)\delta^2 \\ & + (1-p_h)^2(3/2(1-\delta)^2 + 9/8(1-\delta)\delta^2) \\ & + (1-p_h)^3(3/8(1-\delta)\delta^2 + \delta^3) \Big) dh \qquad (4.36) \\ = \frac{1}{4(\mu+\rho)(2\mu+\rho)(3\mu+\rho)} \Big(& \mu^3(24 - 54\delta + 45\delta^2) \\ & + 2\mu\rho(22 + 27\delta + 21\delta^2 - 12\delta^3) - 3\mu\rho^2(-8 + 4\delta - 5\delta^2 + 5\delta^3) \\ & + \rho(4 + 3\delta^2 - 2\delta^3) \Big) \text{[55]} \end{aligned}$$

Figure 4.6 shows SIR dynamic on the preferential attachment, starting with three (3)



(a) exponential infection period (b) constant infection period

Figure 4.6: evoSIR simulation on the preferential attachment

initial individuals (i.e. the network was built by starting with 3 individuals connected to each other as seen in figure 3.4). Individuals are added until the network had n= 100000 individuals. In both cases, We can see from the simulation that, the preferential attachment for a smaller epidemic size the rate at which edges are deleted should be much larger as compared to the other simulations in Erdös Rényi graph and the configuration model graph. Most notably, the structure of the network plays a very crucial role when determining the final epidemic size as we can see in figure 4.1, 4.3, 4.5 and 4.6 (i.e. individual preventive measures should differ based on the structure of the population).

Chapter 5

Conclusion and Future Works

5.1 Conclusion

One reason for mathematical modelling is to analyze and predict the extent of emerging diseases and develop proposed control measures. In this thesis, we studied the stochastic SIR epidemic model on complex networks. The stochastic model studied captured the randomness in disease transmission observed in a real-life epidemic which serves as a model to influence the outcome of an emerging epidemic. The model studied combined the contact structure and the properties of an infectious individual to give understanding to the epidemic behaviour. We translated the observed population into three networks (Erdös Rényi Graph, configuration model graph and the preferential attachment). We devoted to making rigorous approximation and obtained asymptotic result of an epidemic in regards to having major/minor outbreak. As the size of the population becomes large ($n \to \infty$), the initial phase of the standard SIR can be approximated by the branching process by using coupling technique.

After introducing the random network and describing how infection spread can be modelled from such structure, we explained the dynamics and determining preventive measures to reduce the spread of an epidemic. In the dynamics, changing a contact pattern can have both negative and positive impact on the population during an outbreak. The size of an epidemic can be reduced if an individual breaks connections with an infected individual as seen in figure 4.1, 4.3 4.5 and 4.6. In [3], if a susceptible neighbour is exposed to an infected individual and rewired with an unexposed susceptible, it increases the probability of a larger epidemic. However, in our model, where susceptible immediately breaks connection after getting information about infected neighbour reduces the final size of the epidemic. These models serve as a preventive measure in sticking to the public health advice concerning the SARS-CoV-2 and any other infectious diseases

that spread through contact. We also showed in figure 4.1 and 4.3 that when susceptible individual rewired to a randomly chosen individual in the population (evoSIR) will have a larger epidemic size at the population level when compared to breaking the connection and never rewired (delSIR). Further, from equation 4.20, we realised that deleting edges reduced the basic reproduction number (R_0), we also noticed that rewiring has no effect at the beginning of an outbreak if rewired edges are dropped since R_0 is independent of p_d (i.e. the probability that a susceptible individual will connect to a randomly chosen individual in the population after breaking friendship with an infected person). This is as a result of a susceptible individual connecting to a new susceptible individual after deleting an edge.
Finally, Whether or not the evoSIR will have a larger epidemic size depends on the social structure of the population. The assumption of constant and exponential infection period also influence the fraction of individuals infected during an epidemic. From our simulations for networks considered, we can see that probability of generating a secondary infection is larger for constant infection period than the exponential time. The results highlighted in this thesis shows the importance of modelling individual changes in response to an epidemic and how individuals can play their roles in controlling the spread of the diseases.

5.2 Future works

Future work concerns a more in-depth analysis of particular mechanisms, new proposals to try different approaches/methods. Some essential ideas would have to be add up to improve or extend further in this thesis. The following ideas could be used for more in-depth analysis of infectious diseases:

1. Though the SIR model can serve as a preventive model for controlling infectious diseases in the absence of a vaccine, since individual behaviours are far more complex and not predictable, the SIS epidemic model can also be used in modelling the dynamical system of infectious diseases.

2. The assumption of exponentially distributed infection period, constant infection period and other parameters being constant in our model are not realistic. These assumptions simplify our model; theoretically, there is the need to extend our model to time-varying parameters.

3. This thesis provides a theoretical framework of the SIR dynamics to identify plans for investigations and interpretation of the suggested findings. There is a need to design deductive reasoning from real-world data to establish grounds for the study.

List of Figures

Appendix A

Simulation of the evoSIR on configuration graph model

This code generates the configuration graph model and run the SIR simulation on evolving graph suggested by [1]. These are all built-in python 'networkx' package. Here, we first construct the structure of the network by assuming a population of 10000 individuals. After constructing the network, we run the evoSIR as described in chapter 4 by assuming a constant infection period. The code below generates the simulation of the right hand side of figure 4.5. Figure 4.6, 4.1 and 4.3 can be done with the same codes by first considering the structure of the network. This code generates an excel CSV file which can be used for analysis in most programming languages.

```
import networkx as nx
import numpy as np
from random import *
from bisect import *
import matplotlib.pyplot as plt
from numpy import array

def inlist(a, x):
i = bisect_left(a, x)
if i != len(a) and a[i] == x:
return 1
return 0
while(1):
n = 10000
mu = 4
rho = round(100*random())/10
time = 0
infected = []
si = []
while True:
z=[]
while len(z)<10000:
```

```
#setting up the degree distribution using the power law
nextval = int(nx.utils.powerlaw_sequence(n, 2.5)[0])
if nextval!=0:
z.append(nextval)
if sum(z)%2 == 0:
break
G = nx.configuration_model(z)
G=nx.Graph(G) # remove parallel edges
G.remove_edges_from(G.selfloop_edges())
for i in range(len(G.nodes)):
G.nodes[i]['state'] = 's'
G.nodes[0]['state'] = 'i'
infected = [(0,time+1)]
si = sorted(list(G.edges(0)))
while(len(si) > 0):
timetoevent = np.random.exponential(1/(len(si)*(mu+rho)))
if(timetoevent+time>infected[0][1]):
time = infected[0][1]
todie = infected[0][0]
G.nodes[todie]['state'] = 'r'
del infected[0]
for node in G.neighbors(todie):
if(inlist(si,(todie,node))):
del si[bisect_left(si,(todie,node))]
if(inlist(si,(node,todie))):
del si[bisect_left(si,(node,todie))]
elif(random()<mu/(rho+mu)):
cross = choice(si)
time += timetoevent
if(G.nodes[cross[0]]['state']=='s'):
G.nodes[cross[0]]['state'] = 'i'
infected.append((cross[0],time+1))
for node in G.neighbors(cross[0]):
if(G.nodes[node]['state']=='i'):
if(inlist(si,(cross[0],node))):
del si[bisect_left(si,(cross[0],node))]
if(inlist(si,(node,cross[0]))):
del si[bisect_left(si,(node,cross[0]))]
if(G.nodes[node]['state']=='s'):
insort(si,(cross[0],node))
elif(G.nodes[cross[1]]['state']=='s'):
G.nodes[cross[1]]['state'] = 'i'
infected.append((cross[1],time+1))
for node in G.neighbors(cross[1]):
if(G.nodes[node]['state']=='i'):
if(inlist(si,(cross[1],node))):
del si[bisect_left(si,(cross[1],node))]
if(inlist(si,(node,cross[1]))):
del si[bisect_left(si,(node,cross[1]))]
if(G.nodes[node]['state']=='s'):
insort(si,(cross[1],node))
else: #rewiring to a neighbour
time += timetoevent
rewire = choice(si)
del si[bisect_left(si,rewire)]
```

```
G.remove_edge(rewire[0],rewire[1])
if(G.nodes[rewire[0]]['state'] == 's'):
svert = rewire[0]
else:
svert = rewire[1]
newvert = svert
while((newvert==svert) or (newvert in G.neighbors(svert))):
newvert = int(n*random())
G.add_edge(newvert,svert)
if(G.nodes[newvert]['state'] == 'i'):
insort(si,(newvert,svert))
survivors = 0
for i in range(len(list(G.nodes))):
if(G.nodes[i]['state']=='s'):
survivors += 1
writefile = open('/Users/lenovo/OneDrive/Desktop/masters theis codes/Data.CM/consttime.csv','a')
writefile.write(str(n) + ',' + str(mu) + ',' + str(rho) + ',' + str(survivors/n) + '\n')
writefile.close()
print(survivors/n)
```

Appendix B

Matlab code to generate delSIR on the ER-graph

This code uses the fixed point method method to find all roots of the generating function in equation 4.18. Theorem 4.4 is then apply to give the blue curves in figure 4.3 depending on whether μ or ρ varies.

```
%rho = Rewiring rate
%mu = Critical value
%c = mean degree
clear;clc;format('short','g')
c = 5;
%rho = 4
mu = 1
zvec = zeros(1,50);
xVec = zeros(1,50);
for l = 1:100
s = 0.1;
rho = l/10;
mu_r = c*mu/(mu+rho)
error = 0.1;
while error >= 0.1
zOld = s;
gs = @(t) exp(-t).*exp(mu_r.*(1-zOld).*exp(-(mu+rho).*t));
intf = integral(gs, 0, inf)
%generatinf function
s = exp(-mu_r*(1-zOld))*intf;
error = abs((s-zOld)/(s))*100;
end
xVec(l) = rho;
zVec(l)=1-s;
end
plot(xVec,zVec);
hold on
data = xlsread('exprhovaries');
x = data(:,4);
```

```
y = data(:,5);
sz = 1;
scatter(x,1-y,sz)
xlabel('\rho')
ylabel('size of epidemic')
xlim([0 5])
hold off
```

Bibliography

[1] Y. Jiang, R. Kassem, G. York, M. Junge, and R. Durrett. Sir epidemics on evolving graphs. 2019.

[2] F. Ball, T. Britton, K. Y. Leung, and D. Sirl. A stochastic sir network epidemic model with preventive dropping of edges. *Journal of Mathematical Biology*, 2019.

[3] K. Y. Leung, F. Ball, D. Sirl, and T. Britton. Individual preventive social distancing during an epidemic may have negative population-level outcomes. *R. Soc. Interface.*, 2018.

[4] C. Macintyre, N. Gay, H. Gidding, B.Y Hull, L. Gilbert, and P. Mcintyre. A mathematical model to measure the impact of the measles control campaign on the potential for measles transmission in australia. *International journal of infectious diseases*, pages 277–82, 2003.

[5] S. Verguet, M. Johri, S. Morris, C. Gauvreau, P. Jha, and M. Jit. Controlling measles using supplemental immunization activities: A mathematical model to inform optimal policy. *Vaccine*, 2014.

[6] D. Sinha, N. Klahn, and S. Pehlivan. Mathematical modeling of the 2019 measles outbreak on us population. *Acta Scientific Microbiology*, pages 209–214, 2020.

[7] S. Kim, J. Lee, and E. Jung. Mathematical model of transmission dynamics and optimal control strategies for 2009 a/h1n1 influenza in the republic of korea. *Journal of Theoretical Biology*, pages 74 – 85.

[8] N. F. Beggs and H. M. Dobrovolny. Determining drug efficacy parameters for mathematical models of influenza. *Journal of Biological Dynamics*, 2015.

[9] S. Burban, L. Buffat, T. Poynard, and A.-J. Valleron. Modeling the hepatitis c virus epidemic in france. *Hepatology (Baltimore, Md.)*, pages 1596–601, 1999.

[10] A. Pitcher, A. Borquez, B. Skaathun, and N. Martin. Mathematical modeling of hepatitis c virus (hcv) prevention among people who inject drugs: a review of the

literature and insights for elimination strategies. *Journal of Theoretical Biology*, 2018.

[11] G. Fabricius, P. Bergero, M. Ormazabal, A. L. Maltz, and D. Hozbor. Modelling pertussis transmission to evaluate the effectiveness of an adolescent booster in argentina. *Epidemiology and infection*, pages 1–17, 2012.

[12] E. Rosenberg, M. Davidian, and H. Banks. Using mathematical modeling and control to develop structured treatment interruption strategies for hiv infection. *Drug and alcohol dependence*, pages S41–51, 2007.

[13] J. Stover. Influence of mathematical modeling of hiv and aids on policies and programs in the developing world, sexually transmitted diseases. 2000.

[14] O.s Deepa, S. Nallamalli, L. Naik, and G. Teja. Mathematical model for transmission of ebola. *Procedia Computer Science*, pages 742–746, 2015.

[15] A. Mhlanga. *Dynamical analysis and control strategies in modelling Ebola virus disease*. 2019.

[16] A. Rachah. A mathematical model with isolation for the dynamics of ebola virus. *Journal of Physics*, 2018.

[17] I. Benjamin, M. R. Ferrández, M. Vela, and A. M. Ramos. Mathematical modeling of the spread of the coronavirus disease 2019 (covid-19) taking into account the undetected infections. the case of china, 2020.

[18] R. Sameni. Mathematical modeling of epidemic diseases; a case study of the covid-19 coronavirus. *Populations and Evolution*, 2020.

[19] K. Prem, Y. Liu, T. W. Russell, A. J Kucharski, R. M. Eggo, and N. Davies. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health*, 2020.

[20] N. Becker and P. Yip. Analysis of variation in an infection rate. *Australian and New Zealand Journal of Statistics*, pages 42–52, 1989.

[21] J. P. Gabriel, C. Lefevre, and P. Picard. *Stochastic Processes in Epidemic Theory: Proceedings of a Conference held in Luminy*. Springer Berlin Heidelberg, 2014.

[22] D. Mollison. The structure of epidemic models. 1995. Preprint of chapter 2 of Epidemic Models: their Structure and Relation to Data, ed. Denis Mollison, Cambridge UP 1995.

[23] D. J. Daley and J. Gani. *Epidemic Modelling: An Introduction*. Cambridge University Press, 2001.

[24] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation.* Wiley, 2000.

[25] L. J. S. Allen. *An Introduction to Stochastic Epidemic Models.* Springer, Berlin, Heidelberg, 2008.

[26] F. Ball, T. Britton, E. Pardoux, C. Larédo, D. Sirl, and V. C. Tran. *Stochastic Epidemic Models with Inference.* Springer International Publishing, 2019.

[27] Case investigation and contact tracing : Part of a multipronged approach to fight the covid-19 pandemic. online, 2020.

[28] L. Feil. The ever-growing importance of contact tracing. online, 2020.

[29] T. Britton. Stochastic epidemic models: a survey. 2009.

[30] H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis.* Springer, 2000.

[31] L. J. S. Allen. *Stochastic Population and Epidemic Models: Persistence and Extinction.* Springer International Publishing, 2015.

[32] J. C. Miller. A primer on the use of probability generating functions in infectious disease modeling. *Infectious Disease Modelling*, 2018.

[33] T. E. Harris. *The Theory of Branching Processes.* Springer-Verlag, 1963.

[34] Martengale convergence theorem. Massachusetts Institute of Tecnnology. 2013.

[35] C. Jacob. Branching processes: their role in epidemiology. *International Journal of Environmental Research and Public Health*, 2010.

[36] M. Rajarshi. Simpler proofs of two threshold theorems for a general stochastic epidemic. *Journal of Applied Probability*, 1981.

[37] F. Ball. The threshold behaviour of epidemic models. *Journal of Applied Probability*, pages 227–241, 1983.

[38] I. Z. Kiss, J. C. Miller, and P. L. Simon. *Mathematics of Epidemics on Networks From Exact to Approximate Models.* Springer, 2017.

[39] X. Song. The spreading of epidemics in complex networks. PHY 536 term paper, Department of Physics, UIUC., 2017.

[40] H. Andersson. Limit theorems for a random graph epidemic model. *The Annals of Applied Probability*, 1998.

[41] T. Kurtz, E. Lebensztayn, A. Leichsenring, and F. Machado. Limit theorems for an epidemic model on the complete graph. *ALEA*, 4, 2008.

[42] J. A. Bondy and U. S. R. Murty. *Graph Theory and their Application.* 1982.

[43] P. Erdös and A. Rényi. On the evolution of random graph. *Institute of Mathematics, Hungarian Academy of Sciences, Hungary.*, 1966.

[44] R. Durrett. *Random Graph Dynamic.* Cambridge University Press, 2017.

[45] T. House. Modelling epidemics on networks, 2011.

[46] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks, 1999.

[47] R. van der Hofstad. *Random Graphs and Complex Networks.* Cambridge University Press, 2017.

[48] P. Sheridan and T. Onodera. A preferential attachment paradox: How preferential attachment combines with growth to produce networks with log-normal in-degree distributions. 2018.

[49] C. L. Vestergaard and M. Génois. Temporal gillespie algorithm: Fast simulation of contagion processes on time-varying networks. *PLoS computational biology*, 2015.

[50] P. Neal. Sir epidemics on a bernoulli random graph. *Journal of Applied Probability*, pages 779–782, 2003.

[51] E. Volz and L. A. Meyers. Susceptible infected recovered epidemics in dynamic contact networks. *R. Soc. Interface.*, 2007.

[52] D. Yao and R. Durrett. Epidemics on evolving graphs. 2020.

[53] T. Britton, D. Juher, and J. Saldana. A network epidemic model with preventive rewiring: comparative analysis of the initial phase. *Bulletin of Mathematical Biology*, 2016.

[54] C. Lee, A. Garbett, and D. J. Wilkinson. A network epidemic model for online community commissioning data. *Stat Comput*, 2017.

[55] V. S. H. Rao and R. Durvasula, editors. *Dynamic Models of Infectious Diseases: Non Vector-Borne Diseases.* Springer-Verlag New York, 2013.