Modeling Diffusion Phenomena : Homogeneous and Heterogeneous Approaches

A *Thesis Submitted*

in Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

by

Kumar Gaurav

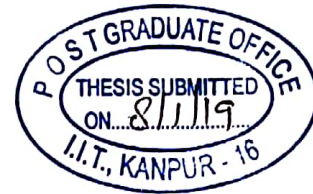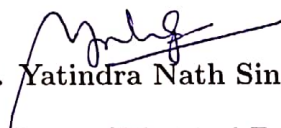*to the*

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

January, 2019

CERTIFICATE

It is certified that the work contained in the thesis entitled **"Modeling Diffusion Phenomena : Homogeneous and Heterogeneous Approaches"** being submitted by **Mr. Kumar Gaurav** has been carried out under my supervision. In my opinion, the thesis has reached the standard fulfilling the requirement of regulation of the Ph.D. degree. The results embodied in this thesis have not been submitted elsewhere for the award of any degree or diploma.

Prof. Yatindra Nath Singh

Department of Electrical Engineering

Indian Institute of Technology Kanpur

Kanpur, INDIA

08th January, 2019

Dedicated to

All My Teachers

For their lessons that crafted my life

Acknowledgements

First of all, I would like to thank my thesis advisor, Prof. Yatindra Nath Singh for providing me an exceptional environment of trust, freedom, and support during all these years. Under his guidance, I have developed all my interest and vision for research in network science. Most importantly, apart from technical knowledge, I was also able to explore the field of value-based education only because of the space he provided me during my entire Ph.D.

No word is sufficient to describe the unconditional support that I received from my entire family. I would always be indebted for their patience, trust, and encouragement. The ways Sapna, my younger sister, single-handedly took care of the entire family in this duration were often unimaginable. Since last year, she and her husband Rakesh ji had released me from almost all of my duties so that I can only concentrate on my studies.

I would like to thank Saumik. He was there with me since the first day of the programme and played all the roles, whenever and whatever was required − an affectionate friend, an ideological mentor, and an academic collaborator. Beyond academics, we often talked on topics like objective, necessity, and direction of research that always motivated me to move forward.

I have never met Sayantari in person, but she was there whenever I had any technical

confusion. She had promptly solved the critical issues whenever required the most. She provided constant support so that I can work on my topics with sufficient analytical background.

All my friends, Manjeer, Amrita, Ankit, Panchjanya, Nandini, Aakash are like my second family in the campus. The colorful moments that I have shared with them during this journey are true assets of my life.

My labmates, especially Sateeshkrishna helped me a lot to pursue network science as my area of research. His constant encouragement and suggestions were extremely helpful during my entire stay. All past and present lab mates Ruchir sir, Sateesh Awasthi ji, Amit munjal, Rameshwar ji, Anupam, Nitin, Varsha, Anjali, Amit vishvakarma, Rashmi, Rahul and Gaurav were always there to help and support me.

Being a senior Ph.D. student, Shyam Bhaiya helped me at all the important stages of this long journey. Apart from that, he introduced me to the activities of Jeevan Vidya which becomes an integral part of my life. I would also like to thank Dushyant bhaiya, Nimisha di, D D Pal ji, Pramod ji, Manoranjan, Aasif, Bhuvan, Amandeep, Mahendra and many other friends for their wholehearted affection, care, and support that I received while participating in various Jeevan Vidya activities.

I would like to thank all the members of Manaviya Shiksha Sanskar Sansthan (MS3). I can not imagine this journey without their presence and contributions. They helped whenever I was in need in all possible ways.

Finally, I would want to take this opportunity to thank Ganesh Bagaria sir, Archana didi and Rajul bhaiya for their constant guidance in all aspects of my life. Their directions has helped me to understand the life in a more meaningful manner.

List of Symbols

A Adjacency matrix of a network

N Total number of nodes in the network

k Degree of the node

p_k degree distribution of network

$\langle k \rangle$ Average degree of the network

$\langle k^2 \rangle$ Second moment of degree

C_i Local clustering coefficient of ith node

Θ_k Density function of nodes around a node of degree k

Θ_{b_k} Density function of class B nodes around a node of degree k

μ Birth and death rate of nodes

T Total population size

U Population size of Unaware Class

B Population size of Believer / Broadcater / Bootlegger class

I Population size of Inert class

A Population size of Aware class

u Fractional population of Unaware class

b Fractional population of Believer / Broadcater / Bootlegger class

i Fractional population of Inert class

a Fractional population of Aware class

u_k Fractional population of Unaware class having degree k

b_k Fractional population of Believer / Broadcater / Bootlegger class

i_k Fractional population of Inert class having degree k

a_k Fractional population of Aware class having degree k

ρ Rate of conversion from class U to B to believer

σ Linear conversion rate from class B to I or A

α Non-linear conversion rate from I or A to B

λ Linear conversion rate from I or A to B

γ Conversion rate from class B to A due to effect of media

ψ Effect of media on conversion parameter ρ

$\mathcal{R}$ Reproduction Number

E Equilibrium point

J Jacobian

Contents

List of Figures

List of Tables

Abstract

Diffusion is a widely used concept to define any spreading phenomenon. Differential equation based compartmental models of disease diffusion, originated in the field of epidemiology have been extensively used in many other disciplines. Analysis of these models is based on the assumption of homogeneous mixing which means that any person can interact with everyone else in the population. But, our society is not homogeneous in nature. To include the social heterogeneity, several network-based models have been proposed in the last two decades. So, we have now two different approaches to understand the processes of diffusion. Significant work has been done using both the approaches, but their comparison on the same platform is still missing. This thesis proposes new compartmental models for spreading phenomena involved in viral marketing and online media piracy and compares the similarities and dissimilarities of both homogeneous and heterogeneous approaches.

In the first part of the thesis, we have considered the propagation of rumor in the population based on already existing SIR model. Explaining the basic settings of homogeneous as well as heterogeneous approach, we have focused on the similarities and dissimilarities between both the approaches along with advantages and appropriateness of each of them to understand different aspects of a physical phenomenon. Simulations have been carried out on Random as well as Scale-free networks to understand the effect of network structure on the diffusion process.

In the second part of the thesis, we have considered the scenario of Viral Marketing which exploits the social connections of existing users by prompting them to share the advertisements or campaign by providing rewards in return. Extending the existing models in the literature, we have proposed two new models of viral marketing based on

the consumer's mindset retrieved from the analysis of the extensive survey conducted by us. We have also shown how bistability can be exploited favorably for sustaining the campaign even in the adverse conditions. Apart from Random and Scale-free network, spread over some real networks has also been analyzed to gain the insights of the real scenario. Finally, we have been able to suggest the key factors that need to be considered while designing any viral campaign.

In the third part of the thesis, online media piracy has been discussed. Model is based on the fact that the habit of using pirated content is germinated in a person from their friends or family members. It is a sort of addiction because even after knowing that it is not right, people are involved in piracy. Diffusion of this habit has been modeled as an epidemic. We have later extended our work by including the effect of awareness created by an external agency like media promotions or public awareness programs. Media has been modeled as a time-varying entity with the rate of change being proportional to the fraction of people involved in piracy at that time. By introducing media, we are able to get rid of the steady state where the fraction of aware class was becoming zero.

All these applications have been investigated by both homogeneous as well as heterogeneous approaches. Similarities and dissimilarities of both the approaches have been pointed out along with the identification of the key parameters which are affecting the process most.

Chapter 1

Introduction

Contagious diseases have always been a threat to the existence of any species living on the earth. Around 25 million people, one-fourth of the entire population of Europe, died of Bubonic Plague (Black Death) in 14th century [1]. From 1918 to 1921, Russia observed more than 20 million cases of Typhus where the death rate was as high as 10% [2]. Recent Ebola virus spread of 2013-16, took the life of around 12 thousand people of West Africa [3]. According to WHO report, since 1981, when the first case of HIV was reported, around 70 million people have been infected with HIV with a mortality rate of 50% [4].

Although numerous efforts have been made to fight against infectious diseases during the entire history of human civilization, great progress had been accomplished amid 20$^{\text{th}}$ century. The worldwide vaccination program for smallpox has successfully eradicated the disease which was in existence from thousands of years. Similarly, after 1988, wild poliovirus cases have decreased by over 99% since the launch of the Global Polio Eradication Initiative [5]. There are some other diseases, for example, measles, diphtheria, pertussis and tetanus that can be severe and potentially life-threatening, which have been mitigated in many places. Though major efforts have been given to

develop medicines and vaccines against contagious disease and some good results have also been achieved, analyses of spreading of the infectious disease have also become a prime concern to the researchers.

1.1   Background

Mathematical modeling of the epidemic spread can be traced back to the models for smallpox given by Daniel Bernoulli in 1760 [6], but the Modern mathematical epidemiology is considered to begin with the inclusion of *Mass Action Principle* for a discrete-time deterministic epidemic model for the measles spread in 1906 by Hamer [7]. Soon after the work of Hamer, Dr. Ronald Ross followed a simple differential equation model to explain the transmission of malaria in 1911 [8]. Kermack and McKendrick, in 1927, generalized the differential equation approach in their popular SIR (Susceptible−Infective−Recovered) model [9]. Extending their work for infections that do not confer any permanent immunity like common cold and influenza, they formulated the SIS (Susceptible−Infective−Susceptible) compartment model in 1932 [10]. The notion of *threshold* was formally introduced through this model. Threshold provides the boundary condition for a disease to spread in a given population. The concept of threshold helped a lot to establish the foundation of the theory of epidemiology.

More exhaustive researches on epidemic spreading occurred at the end of the 20th century. New models were developed taking into consideration various factors like latent period, isolation, migration, vaccination with or without immunity loss, age structure, spatial structure, genetic heterogeneity, etc. [11]. Recently, models have been created for vector-borne diseases like dengue, chikungunya, yellow fever, etc. [12]. These diseases are caused by vectors, i.e., living organisms that can transmit infection between

humans or from animals to humans like mosquitoes, ticks, flies, sandflies, fleas, and some freshwater aquatic snails. Apart from all the other parameters, these models include time delays to account for the incubation period needed by vectors to become infectious [13, 14]. Specific models have also been created for diseases like Ebola hemorrhagic fever (caused by Ebola Virus) and AIDS/HIV (caused by sexually transmitted infections) [15]. Adding more social and biological factors in these models bring them closer to reality, but their complexity increases simultaneously. Hence, along with the use of high-speed computers, required for complicated simulations, more advanced mathematical analyses such as degree, chaos, theories of bifurcation and semigroup have been comprehensively applied in the model investigation [16, 17, 18].

Although most of these models have been borrowed by various other disciplines like Computer science, Social Science and Network science, but due to their origin from epidemiology, these are still termed as epidemic models [18, 19, 20]. In fact, all these disciplines share a common interest in studying diffusion phenomenon and rely on very similar models.

A computer virus is a potentially malicious software program or script that diffuse from computer to computer by replicating itself to another program. When broke out, it can cause significant damage by modifying sensitive information, file encryption, file deletion, formatting disks or slowing down the system. In the past, outbreaks of computer viruses have caused notable economic damages [21]. Factors such as the mutation of existing viruses and the sudden breakthrough in the antivirus technology are akin to the appearance of new infectious disease and invention of its medicines and vaccines. Motivated by the close resemblances between computer viruses and their analogous biological counterparts, various models have been borrowed from epidemiology to comprehend the spread of computer viruses on computer networks.

Similarly, internet-based spreading mechanisms such as tweeting and sharing of online content are contagious by nature [22]. Like a disease's behavior, an Internet-based outbreak also starts with a few individuals who pass on the information to their contact and eventually the information reaches to a significant number of people. One of the frequent events in social media is the spread of rumors, gossips, and hoaxes. Depending on its malicious content, they sometimes create threatening situations [23, 24]. With the increasing influence of social media, viral marketing techniques are being used to advertise a product which exploits social contacts of existing users [25]. Political parties are using it as an important communication channel to propagate their agenda. Even the extreme opinions or ideologies that were once restricted to small groups are easy to spread via web. Such activities are responsible for opinion formation and adaptation among the population and impact the decisions of vital importance. Researchers are using epidemic modeling as a tool to understand the mechanism of these information diffusion phenomena.

As discussed earlier, Kermack and McKendrick's SIR and SIS model [9, 10] are the foundation of epidemic spread analysis. Under this scheme, the whole population is compartmentalized in different classes, and transition from one class to another is modeled by a system of ordinary differential equations. It is very true that it is difficult to identify all the factors affecting a real-world phenomenon and its accurate quantification and modeling is perhaps impossible. Therefore, simplifying assumptions are made to gain insights into epidemic dynamics. One of the fundamental assumptions is homogenous mixing according to which every individual in the population has the same chance of coming into contact with every other individual [8]. The second assumption is referred as mass-action approximation according to which the rate of change of individuals in a compartment at the next time step is assumed to be proportional to the

number of individuals in the compartment at the current time step [26]. The third crucial assumption is about the rate of occurrence of events. It is assumed that an infected person spreads the infection at a constant rate or recovery rate of an infected person is constant.

Mass action assumption appears to hold in most of the cases, but homogeneous mixing is no longer valid in many cases. HIV/AIDS is one of the obvious examples. Moreover, societies are not random and well-mixed instead they have their unique community structures. For example, someone is much more likely to be infected by a co-worker, friend, or family member than by a random individual in the population because (s)he spends much more time and has close contact with the former group of people. Compartmental models do not account for these social structures. Homogeneous mixing hypothesis eliminates the need to know the precise contact network on which the disease spreads, replacing it with the assumption that anyone can interact with anyone else. However, with all its limitation, the homogeneous approach of modeling is proved to be efficient in several fields including public health planning and policy making.

Although, these models are efficient in providing answer to the questions like how many people are infected at any given time; how many will remain infected in steady state; what is the nature of bifurcation, etc., but they are silent on the questions like who are expected to be infected at a particular time; who are the critical persons contributing largest number of infections. In short, all population level statistics is provided by these models, but information related to a particular individual is missing [27]. To understand the dynamics of the individuals, researchers have used simulation-based network analysis where individuals are depicted as nodes, and intercommunications between them as links connecting the respective nodes [28, 29, 30]. The *heterogeneous* network representation to model interactions makes a number of points intuitive. First,

each node has degree much smaller than the size of the network in general, which easily represents the real social structure where a person interacts only with a small fraction of the society. In this way, various kinds of heterogeneity can be efficiently represented by network structure, which is very difficult to be accommodated in compartmental models. Second, while a social network has a dynamic structure as new links are formed through new contacts, and some existing links get deleted as old associations disappear, most of the existing links will remain unaltered for substantial time duration. Therefore without loss of generality, networks are often considered to be static within a timescale of interest (such as the duration of an epidemic) [31].

1.2 Motivation

Though several attempts are made to analyze outbreaks using homogeneous and heterogeneous approaches independently, it is not well-explored what are the similarities and dissimilarities between these two approaches. Upon first inspection, both these approaches seem quite different. Homogeneous modeling is based on differential equations and relies on strict assumptions about homogeneity, whereas heterogeneous analysis is based on simulations carried over a heterogeneous network structure. Conversely, homogeneous analysis is expected to run fast whereas heterogeneous analysis may take hours to run on a computer (depending on the size of the network and complexity of the interactions). However, researchers have begun to realize that these frameworks are much closer than initially thought [32, 33, 34].

In this thesis, we aim to model different social phenomena that exhibit epidemic like diffusion such as rumor propagation, viral marketing, multimedia piracy, etc. To make the models more realistic, extensive surveys have been carried out, wherever re-

quired. Survey results helped us to understand the behavioral transition between differ-
ent classes based on the individual's psychology. We have also analyzed the similarities
and dissimilarities that we can achieve by using homogeneous and heterogeneous ap-
proaches. Apart from qualitative analysis, similarities and dissimilarities between both
the approaches have also been explicitly quantified. We observe that while there are
certain physical information that can be achieved from both approaches, each of them
explicitly reveals specific characteristics about the flow in the society which can be
crucial to develop promotional or inhibiting activities.

1.3   Thesis Outline

The rest of the chapters are organized as follows. Chapter 2 discusses the fundamen-
tal concepts of diffusion dynamics for both homogeneous and heterogeneous approaches.
In this chapter, the discussions are restricted to SIS model which are further extended
for more complex models in the following chapters, and also the concepts of random,
scale-free and real networks are introduced to understand diffusion phenomena over
heterogeneous structures.

Chapter 3 discusses spreading of rumor or hoax in the society as an epidemic spread.
Three subpopulations of the society, typically called as Unaware, Believer, and Inert
migrate from one category to another with specific conversion rates. We analyze the
rumor spreading for both homogeneous and heterogeneous society structures.

To understand the intrinsic correlations between both the approaches, a different diffu-
sion dynamics, known as viral marketing, is further modeled and analyzed in Chapter
4. A survey was conducted to understand the key-parameters and the dominant inter-
actions in the dynamics. Our findings indicate that people who are not participating in

a campaign, may regain his/her interest, and such transitions help in the sustainability of the campaign. We first introduced a minimal model to understand the diffusion dynamics of viral marketing campaigns. Further, we modify the interactions between the subpopulations according to the survey results to come up with a more realistic model to understand a viral campaigning spread.

Opposite to the beneficial effect of diffusion, we model a complicated adverse habit of online piracy in Chapter 5. The spread of the habit of online piracy, in the presence of social contacts, is often compared with an epidemic. We proposed two different awareness models- word-of-mouth and media induced awareness. The models help to understand the benefit of one-to-one awareness programs as well as the effect of mass awareness programs. We also analyzed the effect of heterogeneity on both of these awareness schemes.

Chapter 6 concludes the thesis and discusses the future scopes of this work.

Chapter 2

Fundamentals of Epidemiology for SIS Model

In this chapter, we will present a short account of the concepts that appear to be of importance in discussing mathematical models of epidemiology. We have used basic SIS (Susceptible−Infective−Susceptible) model to explain the essential terminologies of the epidemiology.

2.1 Homogeneous Approach

SIS model was proposed by Kermack and Mckendrick in 1932 [10]. It is a compartmental model in which the total population is compartmentalized into two classes: susceptible class (S) in which all the individuals are susceptible to the infection if they come in contact; and an infected class (I) in which all the individuals are infected by the disease and spread the infection to susceptible individuals. This model is applicable if there is no permanent immunity from the disease. So, infected individuals, after recovery goes back to the susceptible compartment. Processes of infection, as well as recovery, are characterized as the migration of individuals between compartments.

Treating number of susceptible and infectious individuals as continuous variables, they have described this movement by the following pair of coupled ordinary differential equations:

$$
\begin{aligned}
S' &= -\beta \frac{I}{T} S + \gamma I, \\
I' &= \beta \frac{I}{T} S - \gamma I.
\end{aligned}
\tag{2.1}
$$

Here, S and I indicates the number of susceptible and infectious individuals, respectively and T denotes the overall population. The rate of change of S and I with time is denoted by S' and I' respectively. In rest of the thesis, similar notation has been used to represent the rate of change of a variable. Total population, T, remains constant when the population is considered to be closed while ignoring birth, death, immigration, and emigration. Infection and recovery parameters are denoted by β and γ respectively. The term $\beta \frac{I}{T} S$ denotes the rate at which new infections arise. While formulating this term, authors have assumed population to be well mixed (homogeneous mixing) which means that individuals are equally probable to interact with anyone else in the population. Under this assumption, the law of mass action (a well-established relation for chemical reactions) holds, and the rate at which susceptible acquires infection is proportional to the densities of susceptible and infective both.

The second term γI depicts the number of individuals recovered from the infected compartment per unit time. Since, $S + I = T$, the pair of differential equations can be reduced to the single equation

$$
I' = \beta I (1 - \frac{I}{T} - \frac{\gamma}{\beta}).
\tag{2.2}
$$

$\frac{I}{T}$ appearing in the above equation is nothing but the fraction of infected individuals in the population. Similarly $\frac{S}{T}$ is the fraction of susceptibles. These fractions are denoted

by i and s respectively. In this whole thesis, we have used fractional population while analyzing differential equations. Dividing the eq. 2.2 by T, we get the rate of change of fractional population of infectives as below

$$i' = \beta i(1 - i - \frac{\gamma}{\beta}). \tag{2.3}$$

2.1.1   Steady State Analysis

Once the system reaches to steady state, the fraction of a particular class will become constant and thereafter the rate of change of i or s with time will be zero. Equating eq. 2.3 to 0, we get

$$\beta i(1 - i - \frac{\gamma}{\beta}) = 0. \tag{2.4}$$

As clear from the above equation, there are two possible steady states $-$(a) $i_0^\star = 0$ and (b) $i_1^\star = (1 - \frac{\gamma}{\beta})$. Corresponding values of susceptible class will be $s_0^\star = 1$ and $s_1^\star = \frac{\gamma}{\beta}$ respectively. The solution $i_0^\star = 0$ is always a feasible solution but $i_1^\star = (1 - \frac{\gamma}{\beta})$ is feasible only if it is non-negative as negative fractional population has no physical significance. Therefore, depending upon whether $(1 - \frac{\gamma}{\beta})$ is positive or negative (which implies $\frac{\beta}{\gamma} = \mathcal{R}$ is greater or less than 1) system may lead to one of the two different scenarios:

1. **Case 1**: $\mathcal{R} < 1$

   In this case, only one solution $(s_0^\star = 1, i_0^\star = 0)$ is feasible, and hence the system reaches an endemic free equilibrium state where no infected individual exists in the entire population. Endemic free equilibrium state is represented as $E_0(s_0^\star, i_0^\star)$ i.e., $E_0(1,0)$.

2. **Case 2**: $\mathcal{R} > 1$

In this case, along with epidemic free state, an endemic steady state also exists which is denoted by $E_1(s_1^\star, i_1^\star)$ where $s_1^\star$ and $i_1^\star$ are $\frac{\gamma}{\beta}$ and $(1 - \frac{\gamma}{\beta})$ respectively.

2.1.2   Reproduction Number ($\mathcal{R}$)

It is evident that the quantity $\mathcal{R}$ is exhibiting the *epidemic threshold* which predicts whether a disease will persist or die out with time. In epidemiological literature, $\mathcal{R}$ is called reproduction number, and it is defined as the average number of secondary infections caused by a single infectious individual during their entire infectious lifetime [35]. The parameter β is the rate at which an infected individual creates a new infectious individual by spreading the disease to susceptible class. The parameter γ is the rate at which an individual departs back from infected class to susceptible. Hence, $\frac{1}{\gamma}$ is the average time spent by a person in the infected class and $\frac{\beta}{\gamma}$ represents the Reproduction number of SIS dynamics. It is intuitive that if $\mathcal{R} > 1$, then every individual is causing more than one infection and hence the disease will survive; otherwise, it will die out.

2.1.3   Stability

Apart from finding out the feasible equilibrium point, it is also necessary to investigate whether the equilibrium point is stable or unstable, i.e., whether the system moves towards (stable) or away from (unstable) a given equilibrium point. Local stability of an equilibrium point can be estimated using linear stability analysis. For a generalized set of ordinary differential equations, $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ with an equilibrium point $\mathbf{x}^\star$, we can linearize the equation by Taylor series expansion around the equilibrium as,

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}^\star) + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{x}^\star} (\mathbf{x} - \mathbf{x}^\star) \tag{2.5}$$

Let $\mathbf{x} = \mathbf{x}^\star + \delta\mathbf{x}$, where $\delta\mathbf{x}$ is a small perturbation. The question of stability now translates into the eventual decay (or growth) of $\delta\mathbf{x}$, so that $\mathbf{x}$ comes back to (or moves away from) the steady state $\mathbf{x}^\star$, making the equilibrium point stable (or unstable). To study the behavior of $\delta\mathbf{x}$ with time, we take a time derivative of the expression $\mathbf{x} = \mathbf{x}^\star + \delta\mathbf{x}$ and find that,

$$\delta\dot{\mathbf{x}} = \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{2.6}$$

as $\mathbf{x}^\star$ is a constant. Drawing the equivalence between eq. 2.5 and 2.6, as both express a form of $\mathbf{f}(\mathbf{x})$, we write,

$$\delta\dot{\mathbf{x}} = J^\star \delta\mathbf{x}, \tag{2.7}$$

where $J^\star$ is the Jacobian evaluated at the equilibrium. For equilibrium $\mathbf{x}^\star$ to be *stable*, *all* the eigenvalues of $J^\star$ should have *negative* real part. For a system of N ordinary differential equations, where N variables are coupled with each other, the components of state vector $\mathbf{x}$ are $[x_1, x_2, x_3..., x_N]$ and the components of rate vector $\mathbf{f}$ are $[f_1, f_2, f_3,..., f_N]$. In this case, the Jacobian is

$$J^\star = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_N}{\partial x_1} & \frac{\partial f_N}{\partial x_2} & \cdots & \frac{\partial f_N}{\partial x_N} \end{bmatrix} \tag{2.8}$$

evaluated at $[x_1^\star, x_2^\star, x_3^\star, \ldots, x_N^\star]$. To analyze the stability of the fixed points of the SIS model we have,

$$
\begin{aligned}
f_1 &= -\beta s i + \gamma i \\
f_2 &= \beta s i - \gamma i
\end{aligned}
\tag{2.9}
$$

Required jacobian will be

$$J^\star = \begin{bmatrix} \frac{\partial f_1}{\partial s} & \frac{\partial f_1}{\partial i} \\ \frac{\partial f_2}{\partial s} & \frac{\partial f_2}{\partial i} \end{bmatrix} = \begin{bmatrix} -\beta i & -\beta s + \gamma \\ \beta i & \beta s - \gamma \end{bmatrix} \tag{2.10}$$

and two eigenvalues are 0 and $\beta(s - i) - \gamma$. Value of non-zero eiganvalue $\beta(s - i) - \gamma$ at E_0 and E_1 will be $(\beta - \gamma)$ and $(\gamma - \beta)$ respectively. Now again two cases arise:

1. **Case 1:** $\beta < \gamma$

   $\beta < \gamma$ implies $\frac{\beta}{\gamma} = \mathcal{R} < 1$. Under this condition, eigenvalue corresponding to E_0 will be negative and corresponding to E_1 will be positive. Hence, for $\mathcal{R} < 1$ endemic free equilibrium is the stable equilibrium state.

2. **Case 2:** $\beta > \gamma$

   Similar to the analysis in case 1, we can observe that for $\mathcal{R} > 1$, E_0 is unstable and endemic equilibrium E_1 is the stable steady state.

In this way, we can check the stability of the steady state points. For complex models, numerical methods can be helpful to find the nature of eigenvalues of the Jacobian matrix.

2.1.4   Bistability

While exploring the stability of equilibrium points, there might be a scenario when more than one equilibrium point is stable in a given region. Such a condition is referred to as bistability. In such cases, the system leads to different equilibrium points depending on initial conditions, which in our case is the initial number of susceptible or infected individuals.

There are a few more relevant terms like phase transition and bifurcation diagram that we will discuss in Sec. 3.2.3. With this much background of the homogeneous epidemic models, we can now move on to heterogeneous part.

2.2   Heterogeneous Approach

As discussed in the previous chapter, homogeneous modeling assumes the possibility of interaction of an individual to every other individual in the population, but it does not represent the real-life situation where any person can interact only to a limited fraction of the community. Networks are used to model this heterogeneity in the social structure. In the next section, we will discuss the modeling of SIS epidemic spread by the heterogeneous approach, but before that, it is better to get familiarized to few terminologies of network science used in this thesis.

2.2.1   Network Representation

A network is a collection of nodes and links which is most commonly represented as an adjacency matrix. The adjacency matrix $\mathbf{A}$ of a simple network graph is a matrix with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1 & \text{if link exists between nodes } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

A diagonal element $(i = j)$ represents self-loop. For a network with no self-loops, all the diagonal elements of the matrix $\mathbf{A}$ will be zero. The Adjacency matrix for an undirected graph is always symmetric since a link between nodes i and j implies a link between j and i. In case of a directed network, every link has a direction associated with it, pointing from one node to another. The adjacency matrix of such a network is generally asymmetric. If the link of a network represents the strength or frequency of a certain event, number 1 in the adjacency matrix is replaced by various rational numbers representing weights. Such a network is called a weighted network.

2.2.2   Degree Centrality

Various centrality measures have been defined to find out the most important (or central) nodes in the network. Simplest among all of these measures is the number of neighbors it has, which is often called the degree of the node. Higher the degree, more important the node is. Highest degree node is the most central node. If a node has a very high degree compared to most of the other nodes, it is called a *hub*.

There are many other centrality measures like eigenvector centrality, Katz centrality, closeness centrality, betweenness centrality, page rank, etc. These have not been used in our present work, thus have not been elaborated any further.

2.2.3   Degree Distribution

In a network, the fraction of vertices having degree k is denoted by p_k. It can also be thought of as a probability of a randomly chosen node to have degree k. The distribution of the probability p_k is called degree distribution of the network. The directed networks will have in-degree distribution and out-degree distribution corresponding to the number of links ending at and beginning from a node.

2.2.4   Clustering Coefficient

Clustering coefficient reflects the level of clustering in the network and is defined as the average probability that two neighbors of a node are also the neighbours of each other. Local clustering coefficient for a single vertex i having degree k_i is defined as

$$C_i = \frac{\text{number of pairs of neighbors of } i \text{ that are connected}}{\text{number of possible pairs of neighbors of } i} \tag{2.12}$$

where total possible number of pairs of neighbors is $\frac{k_i(k_i-1)}{2}$. The average local clustering coefficient of the network is the mean of local clustering coefficients of all the nodes in the network.

Another way to define clustering on the scale of the whole network is Global clustering coefficients [36].

$$C = \frac{(\text{number of triangles}) \times 3}{(\text{number of connected triplets})} \tag{2.13}$$

Here, a "connected triplet" means three vertices abc with edges (a, b) and (b, c). The edge (a, c) may or may not be present. The factor of three in the numerator arises because each triangle gets counted three times when we count the connected triplets in the network.

2.2.5   Network Generation Models

Network science always aims to build models that can reproduce the properties of real networks. Here, we are going to describe two fundamental models of network generation which helped a lot to understand why a network possess a particular structure and characteristics. Although the networks generated by these models are not matching the real networks exactly, they help to understand the generation mechanism of most of the networks.

2.2.5.1   Random Network

At first inspection, the large network appears to be random. Paul Erdős and Alfréd Rényi [37] used this apparent randomness to generate networks that are truly random. In this model, an edge is created between every possible pair of nodes with an independent linkage probability p. In a network of N nodes, there can be a maximum of $\binom{N}{2}$

distinct pairs of links. Hence, the expected number of edges in the network is $\binom{N}{2}p$, and the average degree of the network is $(N-1)p$. Consequently, the degree distribution of a random network follows the binomial distribution [38]

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}. \tag{2.14}$$

Most real networks are sparse. It means that for them $k << n$. In this limit, the degree distribution given in eq. 2.14 is well approximated by the Poisson distribution

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{2.15}$$

which is often called the degree distribution of a random network. Here, $\langle k \rangle$ is average degree of the network.

2.2.5.2 Scale-free Network

Initially, large networks were assumed to be random and have Poisson degree distribution. With the first map of WWW generated by Hawoong Jeong at the University of Notre Dame, it became clear that all networks do not have Poisson degree distribution instead many of them follow power-law degree distribution [39] given as

$$p_k \propto k^{-\alpha} \tag{2.16}$$

where k is the degree and α is the power constant. Using techniques to generate random graphs with any given degree distribution, it was possible to create a network following power-law. But, these models were not able to explain what is the reason for the network to have a power law degree distribution. This investigation led to generative network models which focus on the mechanism by which networks are created. If the characteristics of the obtained network match with the real network, it can be concluded that the latter might be generated by a similar mechanism.

In recent times, the Barabási-Albert model is the best known generative model [40]. According to this model, two essential features responsible for power-law degree distribution are growth and preferential attachment. Here, nodes are being added to the existing network one by one, and each node connects to a fixed number of pre-existing nodes. Instead of connecting to any node randomly, here node makes a connection to another node with probability proportional to their degrees. In this way, nodes with a higher degree will have more chance to get a new connection. This phenomenon is referred to as a preferential attachment. If a node connects to m of the existing nodes, we will have a network with $2m$ average degree.

In this work, the network generated by both of these mechanisms- Random network and Scale-free network are referred to as model networks.

2.2.5.3 Real Networks

There are many extensions of both of these fundamental models, and all of them try to generate the network as close to real networks as possible, but no one is able to capture the properties of real networks completely. That's why apart from using random and preferential attachment model, it is also necessary to validate our results on real networks. In this thesis, we have considered Hamster network, Email network, and Jazz network. Hamster network contains friendships and family connections between users of the website hamsterster.com. Email network is the email communication network at the University of Rovira i Virgili in Spain. Jazz network is the collaboration network between Jazz musicians, where a node symbolizes a particular musician, and a link between two nodes represents that the two musicians have played together in a band. All these data have been taken from KONECT (The Koblenz Network Collection) [41].

2.2.6   Epidemic Spread over Networks

In a network, different nodes have different degrees. Nodes with a higher (or lower) degree will be able to spread the disease to a more (or less) number of people. Therefore, we need to consider the dynamics of a node based on their degrees.

We are using degree block approximation assuming statistical equivalence among nodes with the same degree. It means that nodes with the same degree will behave in precisely the same manner. We denote the fraction of susceptible and infective nodes with degree k by s_k and i_k respectively. Overall fraction of a particular type of nodes in the network is given by

$$s = \sum_k p_k s_k; \ i = \sum_k p_k i_k$$

where p_k is the degree distribution of the network, explained in the Sec. 2.2.3.

In the homogeneous approach, β was the rate of infection spread by an infectious node in the entire population. The rate was assumed to be equal for each of the nodes. In the network setting, every individual has different reach depending on its degree. Therefore, this assumption of homogeneous model does not look appropriate in the heterogeneous scenario. Instead of going for overall infection rate in the entire population, it is appropriate to consider the rate at which an infectious node spreads the infection to one susceptible individual. We have denoted this rate by β_n and assumed it to be same for nodes of varying degree. For a susceptible node having degree k, number of infectious node in its neighborhood will be ki where i is the fraction of infectious nodes. Hence, the overall rate at which a susceptible node having degree k will receive

the infection will be $\beta_n k i$. System model will be

$$
\begin{aligned}
\frac{\mathrm{d}s_k}{\mathrm{d}t} &= -\beta_n k s_k i + \gamma i_k, \\
\frac{\mathrm{d}i_k}{\mathrm{d}t} &= \beta_n k s_k i - \gamma i_k.
\end{aligned}
\tag{2.17}
$$

In these equations, we have assumed that the fraction of infectives around a node is independent of k (degree of the node) and is equal to the overall fraction of infectives in the population, i.e., i. But, in a real network, this is not the case. As mentioned in [42], Θ_{k_i} is the density function which gives the probability of infectives around a node of degree k and is given by

$$
\Theta_{k_i}(t) = \sum_{k'} p(k'/k) i_{k'}(t)
\tag{2.18}
$$

where $p(k'/k)$ is the probability that a node of degree k will point to a node with degree k' and $i_{k'}(t)$ is the probability that a node with degree k' is in the infectious state. For correlated networks, $p(k'/k)$ is a function of k. For an uncorrelated network, this probability is independent of k and is given by

$$
p(k'/k) = \frac{k' p_{k'}}{\langle k \rangle}.
\tag{2.19}
$$

Some authors use $(k'-1)$ instead of k' in the above equation considering the fact that any susceptible node must have got the infection from one of its infected neighbors and hence it can spread it only to remaining $(k'-1)$ nodes. Both of the formulations have the same order of dependency on nodal degree k, and for the sake of simplicity, we have considered k'. Density function for uncorrelated function thus takes the form

$$
\Theta_{k_i} = \frac{\sum_{k'} k' p_{k'} i_{k'}}{\langle k \rangle} = \Theta_i.
\tag{2.20}
$$

Including density function in the system model, eq. (2.17) modifies to

$$
\begin{aligned}
\frac{\mathrm{d}s_k}{\mathrm{d}t} &= -\beta_n k s_k \Theta_i + \gamma i_k, \\
\frac{\mathrm{d}i_k}{\mathrm{d}t} &= \beta_n k s_k \Theta_i - \gamma i_k.
\end{aligned}
\tag{2.21}
$$

Solving these equations, we can obtain the threshold condition for an epidemic to spread. To know the course of evolution of the epidemic, we carry out simulation which tells the current status (susceptible/infective) of each of the node in the network at a given time. Status of a node keeps on changing with time, but after a steady state is reached, overall fraction of each class remains constant. We can then compare the results obtained by the homogeneous and heterogeneous approach. After discussing the fundamental concepts, we move to our first work that deals with the dynamics of rumor spreading in society. By this work, we want to make the basic foundation of the comparative analysis of homogeneous and heterogeneous approaches.

Chapter 3

Propagation of Rumor and Hoax

3.1 Introduction

We have discussed in the previous chapter that both homogeneous and heterogeneous modelings give crucial insights about the process under study, but the relation between these two methods are mostly unexplored. In this chapter, we try to bridge these two models to understand the effectiveness of both approaches while analyzing the propagation of rumors in the real world.

Spreading of rumors, gossips and hoax in our practical life and social media are quite frequent. Depending on its maliciousness, the spread may sometimes create dangerous situations and should be handled carefully. On March 19, 1935, a rumor was spread in New York city that an innocent Puerto Rican boy was beaten to death by a white man. The rumor led to The Harlem race riot of 1935 which resulted in damage of roughly 2 million dollars. Adjusted for inflation, these figures work out to be 200 trillion dollars worth of property [23]. Similarly, on October 3, 2008, a hoax spread that Apple CEO Steve Jobs had a heart attack. Within an hour the stock lost 10% of its value equivalent to 4.8 billion dollars [43]. In the year 2018, a fake message related to child lifting was

circulated on WhatsApp. It propelled 61 mob lynching incidents leading to 24 deaths in different parts of India [24]. There are numerous examples where a large-scale spread of wrong information has created chaos in the society. As it can be a threatening situation if the rumor is related to any sensitive topic, analysis of rumor dynamics is vital to understand and resist the spread.

Rumor propagates in a society by the individuals who actively take part in the information diffusion [44]. Numerous attempts have been made to model this diffusion to understand the dynamics of the process for effective control and prevention of the rumor propagation. One of the early models in rumor propagation was proposed by Rapoport and Rebhun exploiting the fact that spreading of rumor has an epidemic-like behavior [45]. Later, researchers used the same concept to model diffusion of rumor in society using a set of differential equations and applied homogeneous approaches to analyze the behaviors of the spread in transient and steady-state conditions [46]. Daley and Kendall [47] proposed a diffusion model dividing the entire population into three sub-classes: *Unaware, Believer* and *Inert*. Later, researchers modeled the interaction between these subpopulations to analyze the spreading of rumor. In [48], the author modeled the diffusion using an ordinary differential equation (ODE) with variable rumor strength.

Apart from ODE modeling, researchers have recently started to observe the propagation of epidemic-like phenomena in networks assuming each person as a node and connection between two persons as a link. In [28], authors proposed a shortest path algorithm to spread the rumor in a connected network. Zhao *et al.* [49] modeled the diffusion of rumor for a variable forgetting rate in a small world network. Haeupler proposed a fast rumor spreading model assuming the structure of the network is not known [50]. In [51], authors modeled the spread over the online social network and proposed

effective countermeasure to prevent the spreading. Tripathi *et al.* [52] had proposed novel metrics to model rumor dynamics and introduced an anti-rumor process based on trust. Instead of modeling rumor spread using three sub-populations as mentioned above, Zhou *et al.* [53] divided the population into four sub-classes and modeled the spread over the social network. The significant contribution of Zhou's model [53] is the introduction of a hibernating sub-population to model the forgetting mechanism. In [54], authors proposed a cost-efficient strategy to control rumor in the mobile sensor network.

Though several attempts have been made to analyze rumor propagation using homogeneous and heterogeneous modeling independently, it is not well-explored what are the similarities and dissimilarities between these two approaches. It is required to understand the relationship as in many complex systems, it is not possible to solve the heterogeneous model precisely, and the homogeneous approach is needed to analyze the system [50]. Thus, it is necessary to quantify the differences between homogeneous and heterogeneous approaches explicitly.

In this chapter, a simple model of rumor propagation has been analyzed using homogeneous as well as heterogeneous approach, and the difference in steady-state results have been quantified. In Section 3.2, we discuss the model structure and explore the equilibrium state of the homogeneous system, with discussions about stability and reproduction number of the dynamics. Section 3.3 contains the analysis of the same model for random and scale-free networks, by exploring the degree distributions. Simulation results of homogeneous and heterogeneous approach have been discussed in the subsections 3.2.3 and 3.3.3 respectively. We finally conclude the chapter in Section 3.4, with a discussion on the contrasting results.
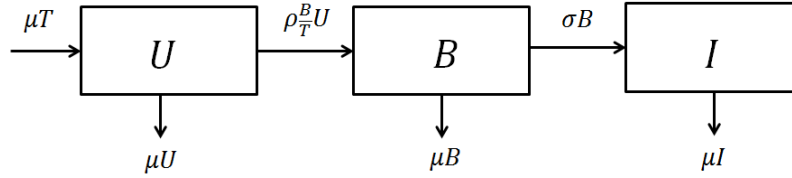
Figure 3.1: Block diagram of the rumor propagation model showing all possible transitions from one class to other.

3.2   Homogeneous Modeling

We consider that total population (T) is categorized into three compartments: Unaware (U), Believer (B) and Inert (I). The fraction of different classes will be $\frac{U}{T}, \frac{B}{T}$ and $\frac{I}{T}$ which is denoted by u, b and i respectively. The nodes in unaware class (U) are not informed about a particular gossip or rumor, whereas nodes in class B believe the rumor and spread it to the nodes in unaware class. Whenever a believer comes in contact with an unaware, believer tries to spread the rumor. The rate of rumor transmission is ρ, which is a combination of the contact rate between a believer and an unaware, and the probability of successful rumor transmission. A class of the population has also been considered which after listening to the rumor, ignores it and do not take part in further spreading. These people are kept in inert category I. This transition is inspired from practical scenarios, where we may lose interest in a particular gossip or a rumor after a certain time. The parameter σ signifies the rate of conversion from believer (B) to inert (I). All three compartments of the model along with possible transitions from one class to another have been shown in Fig. 3.1.

As any real-world population is dynamic, people come and leave a community at a certain rate. To incorporate this feature into the model, birth and death have been

included. The parameter μ is natural death as well as birth rate in the present model. Both rates are kept equal to maintain a constant population. It is assumed that at the time of birth, every individual will be in unaware class. Considering all these assumptions together with homogeneous mixing among the population; the rumor propagation model is defined by the following set of coupled differential equations:

$$
\begin{aligned}
U' &= \mu T - \rho(\frac{B}{T})U - \mu U, \\
B' &= \rho(\frac{B}{T})U - \sigma B - \mu B, \\
I' &= \sigma B - \mu I \cdot
\end{aligned}
$$

(3.1)

Dividing these equations by T gives the rate of change of fractional population of all three classes.

$$
\begin{aligned}
u' &= \mu - \rho bu - \mu u, \\
b' &= \rho bu - \sigma b - \mu b, \\
i' &= \sigma b - \mu i \cdot
\end{aligned}
$$

(3.2)

Since we are analyzing the number or fraction of people in a particular class, all the variables and parameters are assumed to be non-negative.

3.2.1   Equilibrium Analysis

At equilibrium, there is no change in fraction of population corresponding to different classes, hence u', b', and i' will be zero. By equating second equation of the eq. set 3.2 to zero, we get

$$
b(\rho u - \sigma - \mu) = 0 \cdot
$$

(3.3)

The above condition holds if either of the multiplicative terms is zero. One of the solution is $b_0^\star = 0$. We are denoting the equilibrium point value of b by $b^\star$. We obtain

$u_0^\star = 1$ and $i_0^\star = 0$ by substituting $b = 0$ in expression of u' and i' given in eq. 3.2. Hence, one possible equilibrium state is $E_0(1, 0, 0)^1$, when there is not even a single person who believes in the rumor. This steady state is known as rumor-free equilibrium.

Equating the second multiplicative term on the left-hand side of eq. 3.3 to zero, we get

$$u_1^\star = \frac{\sigma + \mu}{\rho}.$$

As all three parameters σ, μ and ρ are positive, value of $u_1^\star$ will be always positive. Substituting the value of $u_1^\star$ in expression of u' and i' given in eq. 3.2, we get the other components, b^* and i^* of second equilibrium point $E_1(u_1^\star, b_1^\star, i_1^\star)$, as follows:

$$b_1^\star = \frac{\mu}{\sigma + \mu}(1 - u^*); \quad i_1^\star = \frac{\sigma}{\sigma + \mu}(1 - u^*).$$

It is to be noted that the fraction of a population can never be greater than 1. So, the endemic equilibrium is practically feasible only if $u_1^\star < 1$ i.e., $\frac{\sigma + \mu}{\rho} < 1$ or $\frac{\rho}{\sigma + \mu} > 1$.

3.2.2   Reproduction Number

As mentioned in Sec. 2.1.2, the average number of new believers generated by a single believer during its lifetime is called the reproduction number. In this model, μ is the rate at which believer dies, and σ is the rate at which it shifts into state I. It means $(\sigma + \mu)$ is the overall rate at which a person departs from believer class. Hence, $\frac{1}{\sigma + \mu}$ is the average time spent by a person in believer class. A believer creates a new believer by spreading the rumor to unaware class at rate ρ. Thus, $\frac{\rho}{\sigma + \mu}$ represents the average number of new believers created by a single believer during its lifetime, which is the reproduction number for this particular dynamics and is denoted by $\mathcal{R}$.

---

[1]E(u,b,i) gives the fraction of unaware, believer and inert population at equilibrium point.

The value of $\mathcal{R}$ depends on three model parameters ρ, σ, and μ. We calculate sensitivity to measure the variation in $\mathcal{R}$, caused by variation in these parameters. Sensitivity analysis tells which parameter affects the quantity of interest ($\mathcal{R}$ in this case) the most. A highly sensitive parameter is crucial to control the diffusion process. The sensitivity of $\mathcal{R}$ for any parameter x is defined as

$$\Gamma_x^{\mathcal{R}} = \frac{x}{\mathcal{R}} \cdot \frac{\partial \mathcal{R}}{\partial x} \tag{3.4}$$

The sensitivities of $\mathcal{R}$ for all three parameters are as follows:

$$\begin{aligned}
\Gamma_\rho^{\mathcal{R}} &= 1, \\
\Gamma_\sigma^{\mathcal{R}} &= -\frac{\sigma}{\sigma + \mu}, \\
\Gamma_\mu^{\mathcal{R}} &= -\frac{\mu}{\sigma + \mu}.
\end{aligned} \tag{3.5}$$

Negative sign indicates that increase in value of the corresponding parameter will decrease the value of $\mathcal{R}$. It can be observed from eq. 3.5 that magnitude of sensitivity with respect to parameters σ and μ are less than 1. Hence, we can say that out of all three parameters, ρ affects $\mathcal{R}$, the most.

In the next section, we have discussed the numerical results of the simulations carried out to verify the analytical findings of homogeneous model obtained so far. All these simulations have been carried out over MATLAB.

3.2.3   Numerical Results

Solving the set of differential equations numerically with a given initial condition (values of u, b and i), we get time evolution of u, b and i. An example with the initial condition $(u, b, i) \equiv (0.95, 0.05, 0)$ has been shown in Fig. 3.2 (a). We can see that as

time progresses, all three values become constant bringing the system to an endemic equilibrium state.

Numerical time-dependent solutions for u, b, and i for a given initial condition can be used to draw the phase plane for the system of equations. We have plotted the $(u - b)$ phase plane for two different set of parameter values in Figs. 3.2 (b) and (c). Arrow along the curves indicate the direction in which the solution moves as time increases. Fig. 3.2 (b), corresponds to parameter set ($\rho = 0.06$, $\sigma = 0.02$, $\mu = 0.05$) which implies the reproduction number $\mathcal{R} = \frac{\rho}{\sigma+\mu} = 0.855 < 1$. We can observe that starting from any initial point, the system eventually reaches to rumor free equilibrium where no believer exists in the entire population. It confirms our analytical result that for $R < 1$, the rumor free equilibrium $E_0(1, 0, 0)$ is the stable state. Similarly, Fig. 3.2 (c) corresponds to parameter set ($\rho = 0.50$, $\sigma = 0.02$, $\mu = 0.05$) which implies $\mathcal{R} = 7.143 > 1$. In this case, the system reaches to an equilibrium state $E_1(0.14, 0.6143, 0.2457)$, which is clearly an endemic state as a finite fraction of population is in believer class and is still active in spreading the rumor.

From all this analysis, we conclude that the steady-state fraction of believers ($b^\star$) can either be zero or non-zero depending on reproduction number $\mathcal{R}$. Thus, we plot $b^\star$ with respect to $\mathcal{R}$ by varying the most sensitive parameter, ρ, in Fig. 3.2 (d). The figure again confirms that for $\mathcal{R} < 1$, $b^\star = 0$ and hence only rumor-free equilibrium E_0 exists whereas for $\mathcal{R} > 1$, along with E_0, the endemic equilibrium E_1 also exists. Different colors in the plot signify whether the equilibrium point is stable or not.
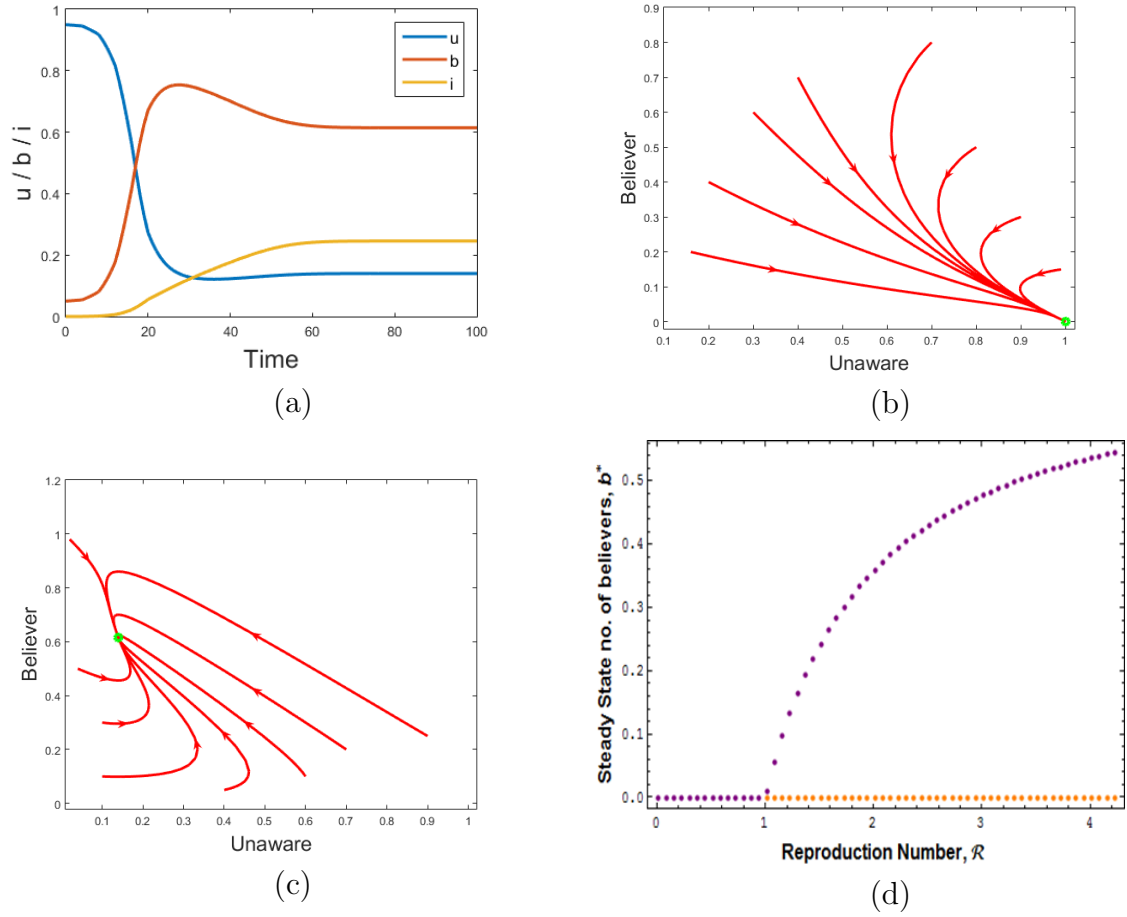
Figure 3.2: (a) Temporal variation of u, b and i for $\mu = 0.05$, $\rho = 0.5$, and $\sigma = 0.02$; (b) Flow diagram of rumor-free steady state starting from various initial conditions for $\mu = 0.05$, $\rho = 0.06$, and $\sigma = 0.02$; (c) Flow diagram of endemic steady state starting from various initial conditions for $\mu = 0.05$, $\rho = 0.5$, and $\sigma = 0.02$; (d) Bifurcation diagram showing fraction of believer with respect to $\mathcal{R}$ for $\mu = 0.05$ and $\sigma = 0.02$.

3.2.3.1 Stability Analysis

To analyze the stability of an equilibrium point, the Jacobian method discussed in Sec. 2.1.3 can be employed. In this case, the components of the rate vector $\mathbf{f}$ necessary to compute Jacobian are

$$
\begin{aligned}
f_1 &= \mu - \rho u b - \mu u, \\
f_2 &= \rho u b - \sigma b - \mu b, \\
f_3 &= \sigma b - \mu i \cdot
\end{aligned}
\tag{3.6}
$$

We obtain $J^\star$ for both the equilibrium points E_0 and E_1 and analyze the eigenvalues. The analysis reveals that for $\mathcal{R} < 1$, the solution (1,0,0) is stable. However, forward transcritical bifurcation occurs at $\mathcal{R} = 1$, as the rumor-free equilibrium loses its stability and becomes unstable. At $\mathcal{R} = 1$, which is also called the bifurcation point, the endemic state appears and remains stable for any $\mathcal{R} > 1$. In Fig. 3.2 (d), we show the stable solution with purple dots, and the unstable solution with orange color.

3.3 Heterogeneous Modeling

We have already established the need for heterogeneous modeling in Chapter 1 and Sec. 2.2. Though the homogeneous model predicts the steady state of the system, that is found to be true in most of the networks; it does not give any information about the dynamics of a particular node. The structure of the network will decide the way rumor will spread through the entire population. Thus, we need to analyze the spread over a network incorporating its structure in the differential equation model.

In this chapter, we have considered two different types of network structure: (a)
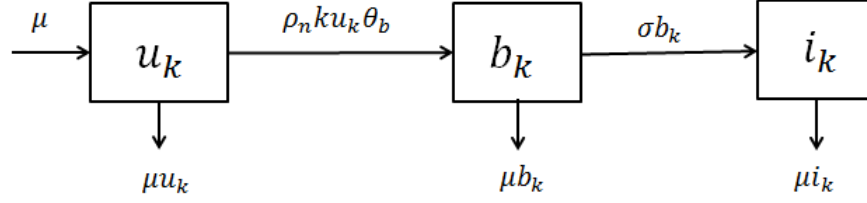
Figure 3.3: Block diagram of the rumor propagation model showing all possible transitions from one class to other for a degree k node.

Random network and (b) Scale-free network. We are going to investigate the following two questions for both of these.

(1) If the equilibrium of the fractions of the population, exists in the heterogeneous model, is it same as calculated in the homogeneous model or is it different? If different, how much is the difference?

(2) Whether or not parametric conditions for equilibrium are same in the homogeneous model and the heterogeneous models.

3.3.1   Degree Block Approximation

Similar to the SIS model of Chapter 2, degree block approximation has been used assuming statistical equivalence among nodes of similar degree. The fraction of unaware, believer and inert nodes having degree k are denoted by u_k, b_k and i_k respectively. All possible transitions with their respective rates have been shown in Fig. 3.3. Differential equations for the rate of change of these degree-wise fractions will be

$$
\begin{aligned}
u'_k &= \mu - \rho_n k u_k \Theta_b - \mu u_k, \\
b'_k &= \rho_n k u_k \Theta_b - \sigma b_k - \mu b_k, \\
i'_k &= \sigma b_k - \mu i_k.
\end{aligned}
\tag{3.7}
$$

Here, Θ_b is the density of believers around a node. As discussed in eq. 2.20 of the previous chapter, the fraction of believers around a node of degree k will be given by density function Θ_{k_b}. For an uncorrelated network, the density function will be independent of k and will take the form

$$
\Theta_{k_b} = \frac{\sum_{k'} k' p_{k'} b_{k'}}{\langle k \rangle} = \Theta_b.
\tag{3.8}
$$

Rumor propagation parameter ρ, used in homogeneous model, discussed in Sec. 3.2, has been replaced by ρ_n as done in Sec. 2.2.6. It is important to note that this modification is needed at all those places where the transition from one class to another depends on the interaction between neighboring nodes of different classes. It is so because the number of such interactions depends on the number of nodes in the neighborhood, i.e., the degree of the node. Whenever transition depends solely on the status of a single node, no such modification will be required e.g. σ in σb_k is left as it is because it represents the probability that a believer will lose interest in rumor and will become inert. This probability is independent of interaction between two nodes and depends on the choice of an individual itself. From the mathematical point of view, rate parameters associated with the non-linear terms in the differential equation (terms containing the product of two variables) will be modified whereas parameters associated with linear terms will remain as it is.

In the last paragraph, we assumed that the rate of rumor spread by a believer should be proportional to its degree and that's why we introduced a new term ρ_n. Now, the next question is what should be the value of ρ_n? Our primary aim is to compare the

results of the homogeneous and the heterogeneous modeling. Value of ρ_n should be such that average rumor spread in the heterogeneous setting is equal to the overall rumor spread by a believer in the homogeneous setting. Mathematically, the condition will be

$$\langle \rho_n k \rangle = \rho \qquad \Rightarrow \qquad \rho_n \langle k \rangle = \rho \qquad \Rightarrow \rho_n = \frac{\rho}{\langle k \rangle}. \tag{3.9}$$

Here, $\langle k \rangle$ represents the average degree of the network which in turn depends on the structure of the network. This condition will let us compare the results of both the approaches.

3.3.2 Early Stage Analysis

In the early stage of the rumor, most of the people are in unaware class. So, u_k can be approximated by 1. Incorporating this approximation, the second equation of the eq. set 3.7 will be simplified to

$$b'_k \approx \rho_n k \Theta_b - \sigma b_k - \mu b_k \cdot \tag{3.10}$$

Multiplying this equation by $\frac{kp_k}{\langle k \rangle}$ and summing over k gives

$$\Theta'_b = \left[\rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} - (\sigma + \mu) \right] \Theta_b. \tag{3.11}$$

This is a linear differential equation having the solution

$$\Theta_b = C e^{\frac{t}{\tau}} \tag{3.12}$$

where

$$C = b_0; \qquad \tau = \frac{\langle k \rangle}{\rho_n \langle k^2 \rangle - (\sigma + \mu) \langle k \rangle} \tag{3.13}$$

Here, C is equal to initial believer density. Assuming uniform distribution of different classes in the initial stage, the density of a particular class around a node can be

assumed to be equal to the initial fraction of that class in the entire population. We have considered $\Theta_b|_{t=0} = b|_{t=0} = b_0$. The value of the time constant τ must be positive so that the density of believers, Θ_b, can grow with time. It gives the condition for the epidemic outbreak as

$$\frac{\rho_n}{\sigma + \mu} > \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{3.14}$$

Replacing ρ_n by $\frac{\rho}{\langle k \rangle}$ as mentioned in eq. 3.13, we get

$$\frac{\rho}{\sigma + \mu} > \frac{\langle k \rangle^2}{\langle k^2 \rangle} \tag{3.15}$$

This equation relates the parameters of homogeneous modeling with the parameters of network structure. LHS of the inequality represents reproduction number $\mathcal{R}$ of the homogeneous rumor dynamics as calculated in Sec. 3.2.2. RHS depends on the first and second moment of degree distribution of the network which will change with the structure of the network. For a random network with Poisson degree distribution, the expectation of square of the nodal degree is

$$\langle k^2 \rangle = \langle k \rangle \left(\langle k \rangle + 1 \right) \cdot \tag{3.16}$$

Putting this value in the inequality eq. 3.15, we get the condition $\mathcal{R} > \frac{\langle k \rangle}{\langle k \rangle + 1}$ which can be approximated by $\mathcal{R} > 1$. This is exactly the same condition what we had in the homogeneous scenario. So, we can say that random network structure behaves like the homogeneously mixed population concerning threshold condition for the epidemic outbreak.

In case of a scale-free network, degree distribution has the form

$$p(k) = Ak^{-\gamma} \tag{3.17}$$

and average degree and average of degree square is given by

$$\langle k \rangle = \sum_{k=m}^{\infty} k p_k = A \frac{1}{\gamma} m^{2-\gamma}. \tag{3.18}$$

$$\langle k^2 \rangle = \sum_{k=m}^{\infty} k^2 p_k \approx A \int_{m}^{\infty} k^{2-\gamma} dk. \tag{3.19}$$

When $\gamma \in (2,3]$; $(2-\gamma)$ is in range of $[-1,0)$; $\langle k^2 \rangle$ diverges and time constant obtained in the eq. 3.13 approaches to zero. It indicates that rumor will spread very fast on such a network as the value of $\frac{t}{\tau}$ will approach ∞. In other words, we can say that rumor threshold is absent. But, practically for a finite network $\langle k^2 \rangle$ is never infinite, and hence, we can calculate the threshold, although it will be extremely small as compared to the random network of the same size and same average degree.

For a scale-free network with a degree exponent $\gamma > 3$, $\langle k^2 \rangle$ is finite. Hence, we will observe the same behavior as observed in the random network although with a different value of τ.

Till now, we have carried out the mathematical formulation of degree block approximation model over the network. In the next section, we will simulate the rumor dynamics over a random and a scale-free network and will see whether the numerical results are matching the results of the homogeneous model or not. We will also discuss the additional information that we can gather from the statistical analysis of heterogeneous modeling.

3.3.3   Numerical Results

Simulations have been carried using MATLAB over Erdös-Rényi random network following binomial degree distribution which converges to Poisson distribution for a

very large number of nodes (infinite network) and Barabási-Albert scale-free network
[55] following power-law degree distribution having power exponent 3. In the following
sections, we will discuss the results for both of these networks, one by one.

3.3.3.1   Random Network

Simulations have been carried out on a random network of 1000 nodes with the
probability of linkage[2] 0.02, resulting in a network of average degree 20.

Initially, every node is assigned a particular class. Most of the nodes are unaware
and only a few are believer. Class of an individual node keeps on changing with time
depending upon the rates of transition from one class to another. All transition rates of
the homogeneous model are incorporated in the network simulation in terms of proba-
bility. Due to the involvement of probability, the results of the same simulation carried
out multiple times are not exactly the same. In fact, the steady-state value obtained
in a single simulation has a considerable amount of perturbation. We have therefore
carried out the simulations multiple times and averaged to get the average statistical
result. Time evolution of u, b, and i for the considered random network with initial
value (0.95, 0.05, 0) is shown in Fig. 3.4 (a).

To measure the deviation of the equilibrium point in random network from the
homogeneous values, let us define a measure ϵ as

$$\epsilon = \left(|u_h^* - u_r^*| + |b_h^* - b_r^*| + |i_h^* - i_r^*| \right) \times 100\% \tag{3.20}$$

where (u_h^*, b_h^*, i_h^*) is the steady-state values of different classes derived from the homo-

---

[2]Probability of linkage means the likelihood with which a link is created between two nodes. More
details can be found in the Sec. 2.2.5.1

geneous modeling and (u_r^*, b_r^*, i_r^*) is the experimental steady-state values of different classes in the random network. We observe that for the considered network at $\langle k \rangle = 10$, ϵ is 3.88%.

To understand the effect of the structure of the network on this error, we have simulated many different random networks. Networks have been created by varying the average degree of the network keeping the number of nodes same. In Fig. 3.4(b), we have plotted the total error for networks having 1000 nodes and different average degree. It can be observed that total error decreases with an increase in the average degree of the network. The result can be attributed to the fact that by increasing the average degree, the network is moving closer to the homogeneous setting, where it is assumed that everyone can interact with everyone else in the population.

Steady state analysis gives us the overall fraction of u, b and i in the entire population but it does not tell about what happens to nodes of different degrees. Looking at time evolution of a set of nodes having a particular degree k, we can comment on whether the steady state fraction u_k, b_k, and i_k is independent of degree k or not? In Fig. 3.4(c), we have plotted the steady-state value of these fractions as a function of degree k. It is clear from the figure that higher degree nodes have high believer fraction as compared to lower degree nodes.

To understand the reason behind this observation, we have plotted the fraction of u, b and i in the neighborhood of a node of different degrees in Fig. 3.4(d). It is clear from the figure that statistically the fraction of believers around a node is independent of the degree of the node. But, it does not mean that the number of believers around every node is also equal. A node with a large number of neighbors will have more believers around them. Chances of receiving the rumor will obviously be more for higher degree
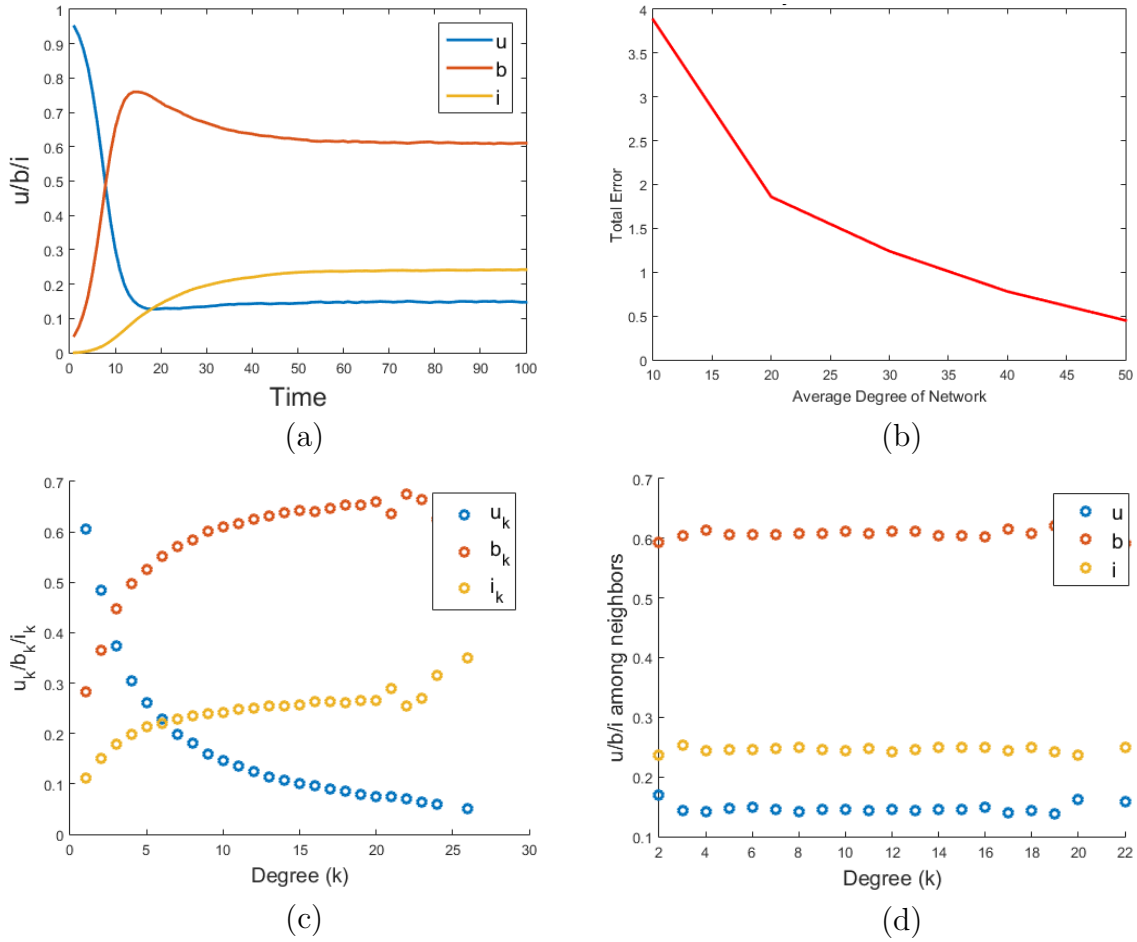
Figure 3.4: (a) Temporal variation of u, b, and i for a random network with average degree $\langle k \rangle = 20$ for rate parameters $\mu = 0.05$, $\rho = 0.02$, and $\sigma = 0.02$; (b) Error ϵ between steady-state values of homogeneous and heterogeneous models with respect to average degree $\langle k \rangle$ in random network; (c) u_k, b_k, and i_k with respect to degree k at steady-state; (d) Fraction of u, b, and i in the neighborhood of a node with degree k.

nodes as more believers are trying to spread the rumor to such nodes. That's why the probability of being in believer state increases monotonically with the degree as shown in Fig. 3.4(c).

3.3.3.2   Scale-free Network

We have used same network size and parameter values as used in random network analysis. Fig. 3.5 (a) shows the temporal evolution of all three classes. We observe that the steady-state reaches the value obtained by homogeneous modeling with a relatively large difference as compared to the random network. We have simulated the scale-free network for average degree varying from 10 to 30. Value of the deviation measure ϵ at $\langle k \rangle = 10$ is 8.98% which has been found to be decreasing with an increase in the average degree of the network in our considered range as shown in Fig. 3.5(b). It can also be noted that the deviation measure does not go below 7% even after increasing the average degree of the network to 30.

A relatively large value of the error is attributed to the structure of the scale-free network. As we have discussed in the Sec. 2.2.5.2, scale-free network follows power-law degree distribution instead of Poisson degree distribution of the random network. In Poisson distribution, most of the nodes have degree closer to the average degree of the network whereas in case of the power-law network, a large number of nodes have smaller degree and a few nodes (hubs) have huge degree. This heterogeneous structure of the scale-free network is far from the homogeneous mixing and is responsible for the comparatively large error.

As done in the case of the random network, to study the effect of degree on node dynamics, we have plotted the fraction of different class of population for varying degrees
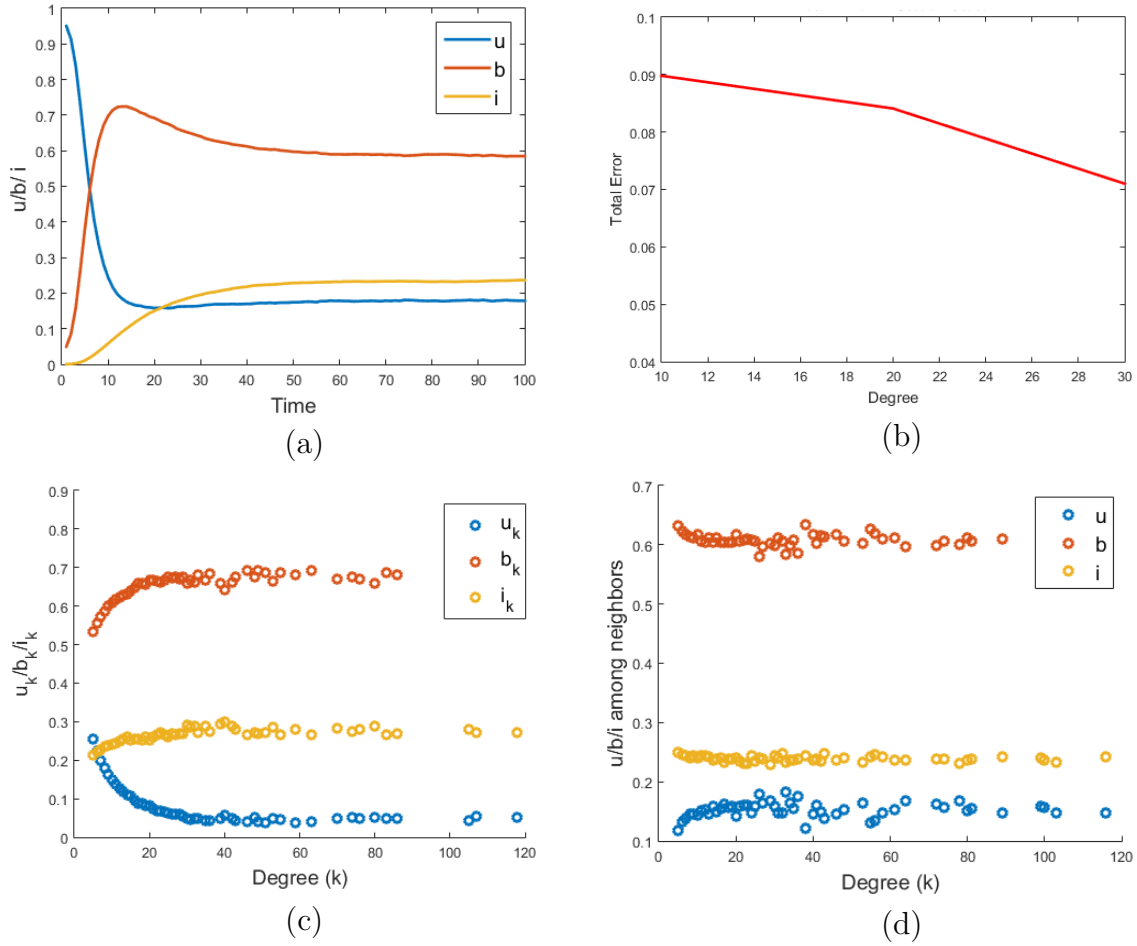
Figure 3.5: (a) Temporal variation of u, b, and i in scale-free network with average degree $\langle k \rangle = 10$ for rate parameters $\mu = 0.05$, $\rho = 0.02$, and $\sigma = 0.02$; (b) Error ϵ between steady-state values of homogeneous and heterogeneous models with respect to average degree $\langle k \rangle$ in scale-free network; (c) u_k, b_k, and i_k with respect to degree k at steady-state; (d) Fraction of u, b, and i in the neighborhood of a node with degree k.

in Fig. 3.5(c). We can observe that similar to random network scenario; smaller degree nodes have fewer chances of receiving rumor as compared to the larger degree nodes. We have also plotted the fraction of different classes in the neighborhood of a node with degree k. Unlike random network, the fraction of believers around nodes of different degree is not identical. It is again due to the heterogeneous structure of the network. Most of the smaller degree nodes have hubs as their neighbors, and these hubs have a high probability to be in believer class. That is why the fraction of believers in the neighborhood of small degree nodes is relatively more as compared to higher degree nodes.

3.4 Summary

In this chapter, we attempted to explore the similarities and dissimilarities of observations between homogeneous and heterogeneous approaches, related to a rumor propagation dynamics. We have observed that though homogeneous ODE approach does not reveal any information about the interconnections and interactions in a complex network, it is extremely useful in predicting the steady-state solutions for the large systems, like Barabasi-Albert network, with a marginal error. These estimations, if computed from network simulations, would have increased the computational complexity and simulation time extensively. The homogeneous approach is also effective to analyze the effect of the parameters and the stability of the solutions at steady states. Network analysis, on the other hand, is not the correct tool to estimate the values of steady-state, parametric contribution or stability with accurate precision. However, because of its considerations of the inherent dynamics and the underlying system structure, the network-based heterogeneous theory can be applied to predict the behavior

of a particular class of nodes with the same degree in both transient state and steady state.

In this chapter, we have considered the well-known SIR model of epidemiology with no backward transition from inert to believer or inert to unaware class. But, if we observe human behaviors, often we find that people change their practices due to internal and external factors. Some of these transitions are indeed induced by social contacts. To understand the behavior of the models with feedback loops, we formulate the problem of referral marketing that has various forward and backward transitions due to complex interactions among classes.

Chapter 4

Spreading of Viral Marketing (VM) Campaigns

4.1 Introduction

According to World Internet Users Statistics, 51% of the world population have found their way online, and this number is rising every day [56, 57]. In this age of the Internet, the importance of social networks is undeniable. People habitually use online social networks for conveying information as well as opinion due to the convenience, competence, and substantial dissemination power. E-commerce together with social networking websites and online life brought notable growth in digital marketing. Introduction of new marketing technologies through the Internet has revolutionized the world of marketing and advertisement.

It is true that creating a viral ad campaign is a cost-effective and fast way to spread the word, but in today's vigorously active social media, there is a huge chance of an ad campaign becoming incredibly short-lived. Similarly, extensive email marketing might help to catch the attention of distracted customers, but sometimes these emails also

cause unnecessary annoyance and irritation to the customers leading to undesirable impact. Therefore it becomes necessary to devise strategies to exploit existing social networks to make a campaign fast spreading as well as sustainable and effective.

One important observation is that in the present time when the fabricated and false promises are prevailing around, people tend to believe the recommendations of their relatives or friends instead of promotional emails or advertisements put up by the companies. Now the question is how a company can make use of its consumers as their promoting agents by encouraging them to share and spread a marketing message through their social contacts? This is what viral marketing (VM) is all about.

Viral marketing (VM) is based on exploiting pre-existing social networks (or other online platforms, like, web forums, blogs, emails, etc.) to accomplish precise marketing goals [58], by considering the existing customers as brand advocates. VM is also known as Internet Word-of-mouth marketing, as it encourages people to share product information (specifications, improvements, campaigns, etc.) with their friends through email or other social media [59]. This prompting is sometimes done by the introduction of some benefits (like credit points, e-cash, extra discounts, cashback, promo codes, etc.) to the existing customers, as a reward for sharing information in their peer network. VM campaigns have several benefits over traditional mass media campaigns, an important one being its ability to reach particular customer groups, as, in many cases, friendship networks arise from common interests [60]. These communications also have more impact and acceptability than third-party advertising among the potential customers, as it comes with an endorsement and recommendation of a friend.

VM is being adopted as a recent marketing strategy and a way of communication with customers, which can potentially reach a large audience very fast [61, 62, 63].

Prominent companies like Amazon, Google, and Hotmail have succeeded with virtually no marketing, based solely on consumer-driven communications [64]. In similar fashion, established organizations such as Procter and Gamble, Microsoft, BMW, and Samsung have successfully used VM, through which the intact marketing message spreads across the market rapidly, imitating an epidemic [25]. While there is no shortage of news publications fighting for the attention of millennials, in 2017, the San Francisco-based daily email newsletter *The Hustle* gained 300,000 subscribers in just a few months, with great copywriting and an aptly planned Milestone Referral program. In 2013, Harry's, a New York-based shaving equipment manufacturer used a credible referral to launch their grooming brand and gathered 100,000 customers in one week before their launch. They have used a strategically planned Milestone Referral campaign where the minimum required referral numbers were kept as low as 5, and 77 % of their initial customers were collected via referrals, where 20,000 people referred about 65,000 friends. Dropbox's referral program is possibly one of the most famous cases of viral marketing executed with exceptional success. Dropbox's metric history shows a 3900% growth just within 15 months (September 2008: 100,000 registered users to December 2009: 4M registered users), with a marketing idea of providing 500MB for 1 referral, coupled with an easy invitation process and clear view of the benefits. In another extraordinarily successful strategy with a well-timed and well-implemented referral marketing program, Airbnb offered a $25 discount for accommodation booking to both sides, which became popular as the 'altruistic referral'.

Since December 1996, when the term VM was coined by a Harvard business school professor, Jeffrey Rayport, in an article *The Virus of Marketing*, the topic has evolved into an interesting area of research. In 2003, Mohammed *et al.* [65] pointed out that users are less likely to trust promotional communications by the company themselves

compared to the peer recommendations about a product or service. In 2012 Nielsen's
Global Trust Advertising survey [66] reported, *"people don't trust advertising, at least
not as much as they trust recommendations from friends and consumer opinions ex-
pressed online."* With data for more than 28,000 Internet respondents in 56 countries,
it has been seen that 92% of consumers trust recommendations from friends and family
above all other forms of advertising. This is one of the major reasons why products
that use a VM campaign tend to succeed very quickly. Moreover, it is rather clear
that VM would cost considerably less than the traditional marketing and promotional
techniques [67]. Starting from gaining new patrons to creating brand intimacy in ex-
isting customers [68], this strategy of Internet word-of-mouth creates several positive
outcomes for the firms. Customers proceed with brand recommendations for diverse
reasons: while someone joining in a referral program may endorse a brand to earn a
monetary incentive [69], another may do the same to communicate identification with
the brand [70, 71]. Going through the studies carried out in this field we can say
that the majority of them propose a conceptual framework, completely ignoring the
mathematical treatment. Studies that attempt to balance qualitative perspectives with
quantitative methods are lacking.

In the last few years, several studies started to conceptualize viral marketing (VM)
as the close derivatives of disease infection models from mathematical epidemiology.
While the usual epidemiology studies aim to contain the epidemic, in the context of
VM, the purpose of the study will be to maximize the spread. The spread of marketing
messages in social networks raises various theoretical and practical questions: How can
an advertisement reach the maximum audience? Are there any factors that affect this
dynamics beyond the design and content? What actions can the companies take to
speed up the diffusion rate?

In marketing, Bass (1969) used the underlying epidemic model as a foundation for his new product diffusion model. But, in the context of online social networks, Sohn *et al.* [72] first demonstrated in 2013, the VM diffusion as SIR and SEIAR processes of epidemiology, although they have developed only the conceptual framework without going into quantitative analysis. A major step forward was the quantitative treatment of the VM dynamics developed by Rodrigues *et al.* [25] in 2016, capturing the epidemiological aspects and consumer behavior into diffusion dynamics study. However, this model was based on the basic SIR model and could not include some important and evident interactions, which are crucial for dealing with real-world scenarios.

It is evident that a clear insight of customer behavior is indispensable for a firm to ensure the relevance and survival of a viral campaign. To understand customer motivation and actual dynamics of VM campaigns, we conducted an extensive questionnaire-based-survey. 331 participants from different age group shared their opinions through both polar as well as qualitative answers in that survey. Considering the inputs from real-world customers, we built a model resting on the key features observed from the survey outcomes.

4.2   Proposed Model of VM

In the homogeneous approach, we compartmentalize the total population (T) and associate each individual to one of three mutually-exclusive *subpopulations*: Unaware (U), Broadcaster (B) and Inert (I). This approach has been adopted before for epidemiological modeling [25] as well as network-level treatment [73] of email-based advertisement campaigns. The behavioral transitions between the mutually-exclusive compartments are driven by several guiding factors, which set up the rules for the model. Results of
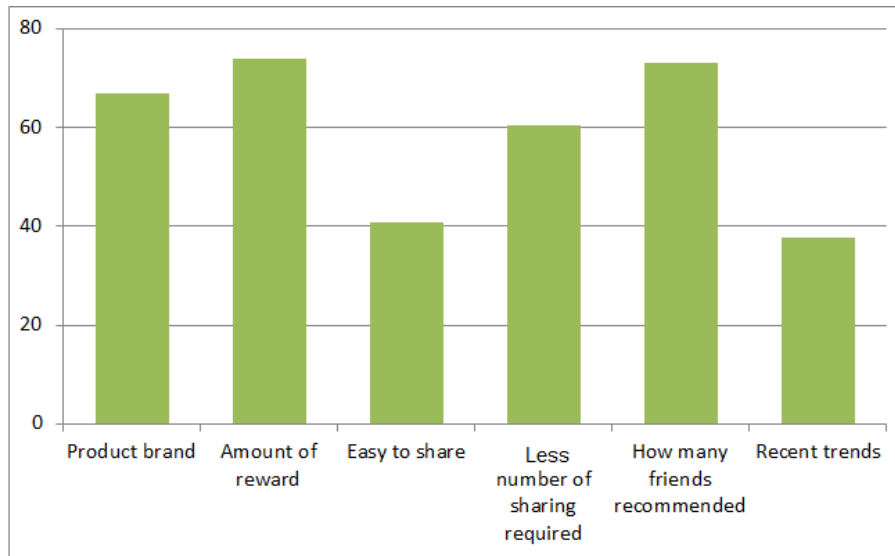
Figure 4.1: Reasons behind sharing an online advertisement. Y-axis indicates the percentage of people motivated by the reasons mentioned along X-axis. Participants could choose more than one option according to their choice.

the survey also indicate that there are several inter-dependent and independent factors that play significant roles in customer's approach towards a VM campaign.

Key observations in the survey to understand the transitions in customer behavior for a particular brand product that is offering some reward/benefit for referring it to others are summarized below. We see that when a person encounters a viral campaign for the first time, few major factors influence the first reaction towards the campaign. As this decision of taking part in a campaign is most of the times, a momentary decision, the fate of the campaign gets decided in a very short time, for that particular customer. These factors are:

1. Reward: The incentive or reward, provided by the advertiser to the advocating customer, is found to greatly influence their motivation to participate in a referral program. In our survey, 55.8% of the total participants admitted that they shared an online advertisement with a friend (or shared the contact detail of the friend)
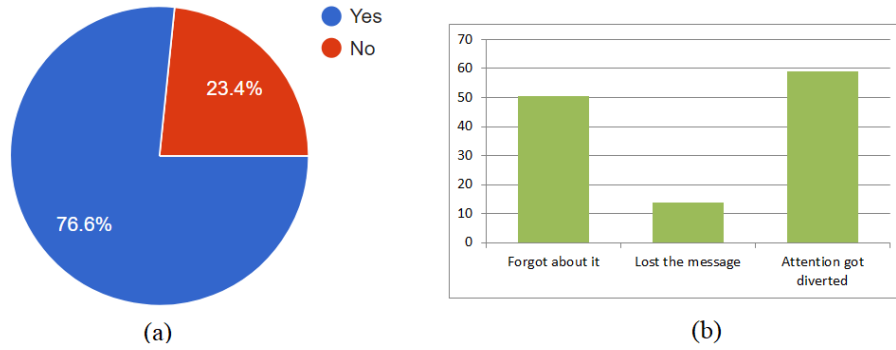
Figure 4.2: Results about missing a viral offer, despite of having initial interest. (a) 'Yes' ('No') signifies the person missed (never missed) such an offer. (b) Reasons for missing an offer. Y-axis indicates the percentage of people motivated by the reasons mentioned along X-axis. While forgetting and diversion of attention were the reasons for majority, some people also claimed that they lost the message. Participants could choose more than one option according to their choice.

to avail an offer/reward/discount. Among these, for 74%, one of the main reasons for sharing was how lucrative the offer was, for themselves.

2. Ease of share: Another driving force for sharing motivation was ease of the sharing procedure and flexibility of the number of referrals. A considerable percentage of customers claimed that they only proceed for sharing if the process is not too complex (40.9% of total population)

These two factors were considered to be the major motivation for a person participating in the viral campaign and moving from Unaware to Broadcasters class. Detailed survey result pertaining to reasons behind sharing has been plotted in the Fig. 4.1. The broadcasters also gradually move to the Inert class as time passes on. The key factors that influence this switch are:

1. Diversion of attention and forgetting: A key finding of our survey was related to the inherent forgetting associated with the customers. In answer to our question, 76.6% participants said they had missed at least one viral offer, which they first
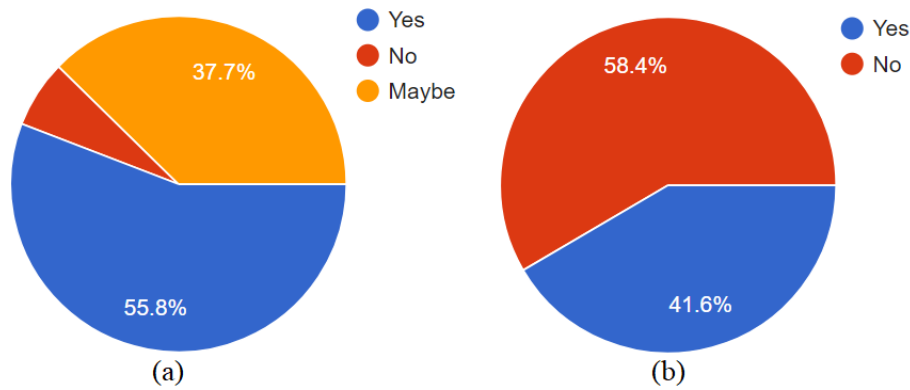
Figure 4.3: Regaining interest in a viral offer which was ignored or missed when (a) a friend wins a good reward from it and (b) when a friend reminds you about it. 'Yes'('No') signifies people who (do not) think the reason will cause them regain of interest in an ongoing offer. "May be" signifies they might regain their interest about it.

decided to avail. Focusing on the reasons that made them miss the offer, diversion of attention and forgetting were detected as the two major contributors (59.1%, 50.4%). Results can be seen in Fig. 4.2.

2. Getting bored or doubtful: We also found that 62% people who used to broadcast the viral message can lose interest and come to inert class due to factors like, getting annoyed (due to low profit-to-effort ratio), bored or suddenly becoming doubtful (about security).

We combine these two factors in a parameter to depict the switching from B to I class. Interestingly, we also found that there are always chances that the inert may regain his interest depending on the influence from another friend who has a common interest (already a broadcaster). Major factors responsible for this switching are:

1. Friendly reminders: The people, who often were interested in the offer, but forgot about it, feel that they *missed* the offer (almost 76%), continue to be submissively interested about the advertisement campaign. A timely reminder by someone or
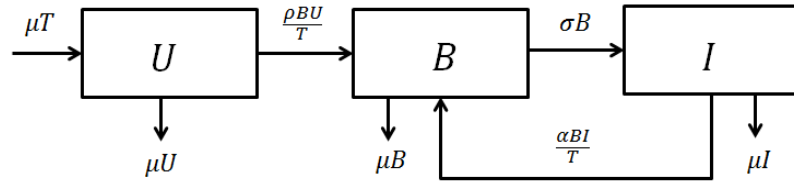
Figure 4.4: Block diagram of the proposed model for viral marketing showing all possible transitions from one state to another.

strategically designed retargeting emails from the company can influence them straightaway to gain their active state back, and they start contributing to the propagation of the campaign again.

2. Lucrative offers: As shown in Fig. 4.3, the people who had left the B class by getting bored or annoyed, more than 92% of them agreed that authentic information or genuine news related to a considerable gain from the same campaign might make them motivated to return to active participation.

On the basis of above discussion, the scheme of the model can be formulated as shown in Fig. 4.4.

4.3 Homogeneous Modeling

The unaware class, denoted by U, is yet to receive the message; these are susceptible people or the target market, who may receive an advertising message containing marketing offers. The broadcaster class B consists of individuals who came to know about the message and have the potential to forward the message further in the population. If a member in this class decides to participate in the campaigning, (s)he transmits and spreads the message in the entire population by recommending it through (her) his social contacts. We assume that ρ is the rate at which a broadcaster comes in contact

with a member from unaware class, and share the viral message to create new potential broadcasters. The transition from class B to class I has been captured by the parameter σ of the model. Transitions from U to B and B to I were the part of the SIR based model proposed by Rodrigues *et al.* [25]. But according to our survey results, there are always chances that the inert may regain their interest in the campaign. To accommodate this finding, a transition from I to B has been added by allowing feedback from inert class to the broadcaster with a relapse rate α.

In practical scenarios, people enter and leave the population. To include this factor, we have introduced birth and death in our model. Both birth and death rates are kept equal to μ so that fixed population size can be maintained [25, 33, 73, 74, 75, 76]. For a particular VM dynamics, birth and death can be viewed as events when people join or leave the social platform where the campaign is going on. Considering u, b and i to be the fraction of unaware, broadcaster and inert classes normalized by the total population T, the VM dynamics in the population with the mentioned interactions is governed by the following differential equations:

$$
\begin{aligned}
u' &= \mu - \rho b u - \mu u \\
b' &= \rho b u - \sigma b - \mu b + \alpha b i \\
i' &= \sigma b - \alpha b i - \mu i
\end{aligned}
\tag{4.1}
$$

4.3.1 Equilibrium Analysis

At equilibrium, there is no time evolution of the system model defined in eq. 4.1 and the rate of change of u, b and i becomes zero. The system of equations always has a VM-free equilibrium E_0, at which the whole population is unaware. Also, the system exhibits an Endemic equilibrium $E^\star$ with a finite percentage becoming broadcasters.

By setting u', b' and i' of eq. 4.1 to zero, all the components of $E^\star$ can be evaluated. While solving for $E^\star$, the first equation of the eq. set 4.1, gives

$$u^\star = \frac{\mu}{\rho b^\star + \mu}. \tag{4.2}$$

Relevant substitutions from eq. 4.2, replacing $i^\star$ by $(1 - b^\star - u^\star)$, and simple algebra results into $p(b^\star)^2 + qb^\star + r = 0$, where

$$p = \alpha\rho \; ; \qquad q = (\sigma\rho + \mu\rho + \alpha\mu - \alpha\rho) \; ; \qquad r = \mu(\sigma + \mu - \rho) \cdot \tag{4.3}$$

Examining the coefficients, we can conclude that p is always positive; q is positive for small values of α, and r is positive or negative depending on whether $\frac{\rho}{\sigma+\mu} = \mathcal{R}$ is smaller or greater than 1. Two utterly different steady state scenarios can arise:

Case 1: For negative r (i.e., $\mathcal{R} > 1$), the quadratic equation has a unique positive solution $b_+^\star$, as another solution $b_-^\star$ is always negative and so, unphysical, and there exists a unique endemic equilibrium $E^\star$ whenever $\mathcal{R} > 1$.

Case 2: On the other hand, for positive r (i.e., $\mathcal{R} < 1$), the number of physical roots of the equation depends on the sign of q, and therefore, the nonlinear relapse parameter α. Depending on this fact if α is high (or low), multiple (or no) endemic equilibria may exist.

4.3.2   Reproduction Number

Reproduction number is defined as the average number of broadcasters a single broadcaster can create in its lifetime *without* considering its interaction with inert class (which makes α irrelevant for this estimation). If interaction between class B and I is ignored then the model will exactly look like the rumor propagation model discussed in Section

3.2 of the previous chapter. That is why the reproduction number $\mathcal{R}$ of this model is identical to the model of the previous chapter i.e. $\frac{\rho}{\sigma+\mu}$. Sensitivity analysis will also be exactly the same. See section 3.2.2 for details.

4.3.3 Bifurcation

The model exhibits two completely different behaviors depending on the value of the parameter α as described in Sec. 4.3.1. To understand the phenomenon, we observe the steady states of the model for two different α values fixing $\mu = 0.05$, $\sigma = 0.2$ in Fig. 4.5 (a) and (b). Stability of the steady states has been determined by analyzing the eigenvalues of the Jacobian matrix obtained by linear stability analysis described in Sec. 2.1.3. A forward transcritical bifurcation is observed in Fig. 4.5(a), indicating the existence of only the message-free solution before $\mathcal{R} = 1$. On the other hand, a backward bifurcation occurs for high α values as shown in Fig. 4.5(b). In this case, for the parameter regime where $\mathcal{R} \in [\mathcal{R}_c, 1)$, there exists a choice for the system between two distinctly different responses. This regime is known as a region of bistability where both the endemic and the message-free solutions can be achieved by the system depending upon the initial conditions. The initial condition here corresponds to the initial number of people in different classes. This history dependence is commonly known as *hysteresis*, drawing an analogy from a similar phenomenon in ferromagnetic systems. This phenomenon of bistability gives the system sustainability, so that, once a transition occurs from the message-free state to the endemic state, the nonlinearity of the dynamics inherently makes it difficult to any switch-back driven by the immediate fluctuations of the parameters; the whole system works as a very robust switch. Thus, it can be concluded that *high value of nonlinear relapse rate α makes it difficult to eradicate the message from the system;* the message-epidemic will be present for a
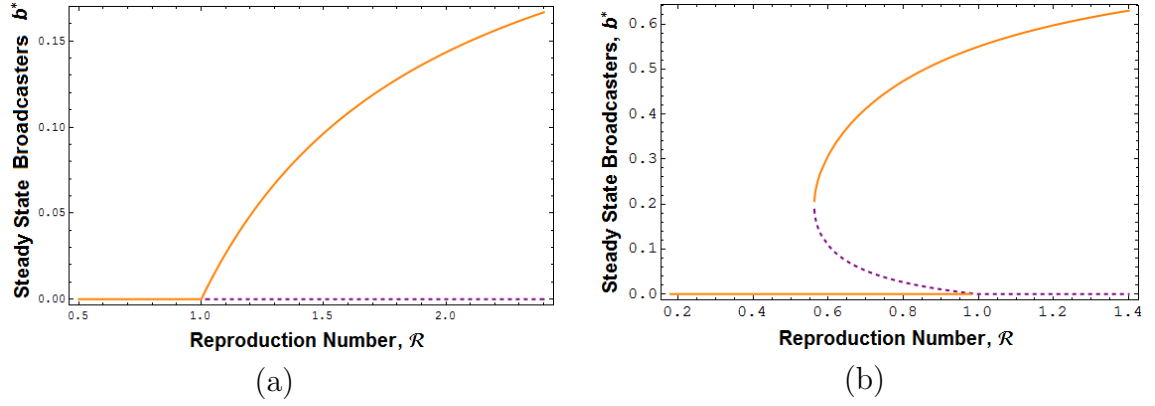
Figure 4.5: Variation in steady state fraction of B with reproduction number $\mathcal{R}$ for (a) $\alpha = 0.1$, when only a single epidemic state persists beyond $\mathcal{R} = 1$ and for (b) $\alpha = 1$, when bistability can be observed in range $\mathcal{R}_c$ to 1. Parameter values are $\sigma = 0.2$ and $\mu = 0.05$. In these figures, orange (and continuous) lines indicate stable solutions and purple (and dashed) lines indicate unstable solutions. For these parameter values, we calculated $\mathcal{R}_c = 0.562$ from eq. 4.6.

broad parameter regime, even when $\mathcal{R} < 1$. The condition for the existence of this bistable region will be discussed in next section.

4.3.4 Conditions for Bistability

To ensure bistability, the necessary conditions are $q < 0$ and $q^2 - 4pr > 0$ where, p, q, and r are given by eq. 4.3. We can figure out the limiting condition for bistability from these relations. Though the nonlinear relapse rate, α, does not appear in the expression of $\mathcal{R}$, it causes a drastic change in the behavior of the system. By equating $q = 0$, we can figure out the minimum threshold for α as:

$$\alpha_{th} = \frac{\rho(\sigma + \mu)}{\rho - \mu}. \tag{4.4}$$

For a given set of parameters, *iff* $\alpha > \alpha_{th}$, then bistable solutions can be expected. Once we satisfy the condition for α, it should be noted that both the endemic states can exist (i.e., have real roots) only if $q^2 - 4pr > 0$. The region of bistability extends for
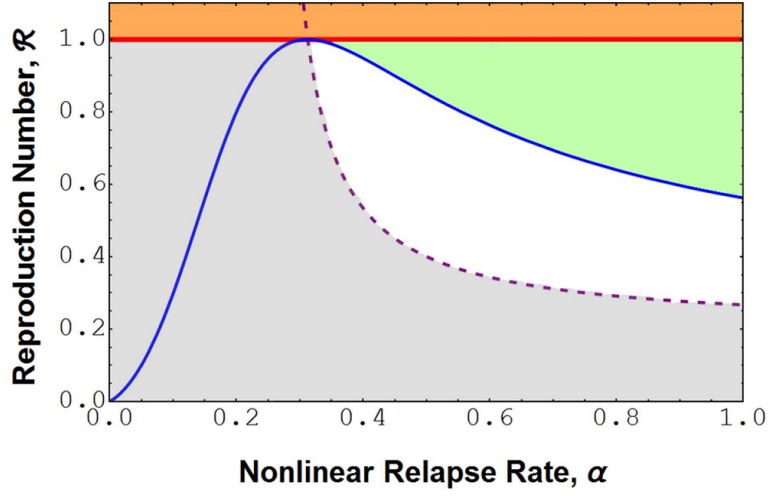
Figure 4.6: Phase diagram of the model in $\alpha - \mathcal{R}$ space for $\sigma = 0.2$ and $\mu = 0.05$. The blue line indicates $\mathcal{R}_c$, purple dashed line indicates α_{th} and red line indicates $\mathcal{R} = 1$. The region filled with orange color always exhibits monostable endemic state as $\mathcal{R} > 1$, the gray region exhibits monostable VM free state as $\alpha < \alpha_{th}$. For both the white region and green region, $\alpha \geq \alpha_{th}$. For the white region, $\mathcal{R} < \mathcal{R}_c$ and the region contains monostable VM free state. However for the green region, $\alpha > \alpha_{th}$, $\mathcal{R} > \mathcal{R}_c$ and $\mathcal{R} < 1$. Thus, this area exhibits bistability, where either VM free state or the endemic state is chosen by the system depending upon the initial state.

a range of $\mathcal{R}$ values, from $\mathcal{R}_c$ to 1, as mentioned in the Sec. 4.3.3. For $\mathcal{R} < \mathcal{R}_c$ neither of the endemic solutions is feasible and message-free state is the only steady state. $\mathcal{R}_c$ or the critical threshold for bistability can be evaluated by equating $q^2 - 4pr$ to zero. With some additional algebraic manipulations, we get

$$\rho = \frac{\alpha\mu}{\alpha + \mu + \sigma - 2\sqrt{\sigma\alpha}}. \tag{4.5}$$

Calling this ρ as ρ_c, and substituting in the expression of $\mathcal{R}$, discussed in Sec. 4.3.2, we get

$$\mathcal{R}_c = \frac{1}{(\sigma + \mu)} \frac{\alpha\mu}{(\alpha + \mu + \sigma - 2\sqrt{\sigma\alpha})}. \tag{4.6}$$

Eq. 4.4 and 4.6 provide us with two limits for ensuring the bistable dynamics of the system. From our previous discussions, it is evident that depending upon the two key

parameters of the model, $\mathcal{R}$ and α, *only* endemic, *only* VM free or *both* solutions can be obtained. To illustrate this idea, we explored the phase diagram of the system in $\alpha - \mathcal{R}$ space in Fig. 4.6. The only region where bistable dynamics can be observed is shaded in green, bounded by the $\mathcal{R} = 1$, α_{th} and $\mathcal{R}_c$. Bifurcation diagram for two different values of α shown in Fig. 4.5 (a) and (b) can be obtained by tracking the system behavior while moving across the $\alpha - \mathcal{R}$ phase space through vertical lines $\alpha = 0.1$ and $\alpha = 1$ respectively. We note that the region of bistability increases gradually as the value of α increases. This shows that high values of the relapse rate ensure the survival of the campaign in the steady state. From the phase diagram, it can also be noted that without the relapse (i.e., $\alpha = 0$), no bistability is possible.

4.4 Heterogeneous Modeling

In this chapter, along with our model networks (Random and Scale-free networks), we have studied the diffusion of the viral campaign on real networks too. In upcoming sections, we will formulate degree block approximation for VM diffusion dynamics. We will also find threshold condition required for the spread of campaign in its initial phase, and its sustenance after steady state is reached.

4.4.1 Degree Block Approximation

In contrast to the homogeneous approach, diffusion in networks will be dependent on degree distribution of the network. We denote with u_k, b_k and i_k, the fraction of unaware, broadcaster and inert nodes with degree k. Instead of rate parameters ρ and α of the homogeneous analysis, we have used ρ_n and α_n respectively, as we had done in section 3.3.1 of the previous chapter. Using Θ_u, Θ_b, and Θ_i as density of various classes

around a given node, system eq. 4.1 modifies to

$$
\begin{aligned}
u'_k &= \mu - \rho_n k u_k \Theta_b - \mu u_k, \\
b'_k &= \rho_n k u_k \Theta_b + \alpha_n k i_k \Theta_b - (\sigma + \mu) b_k, \\
i'_k &= \sigma b_k - \alpha_n k i_k \Theta_b - \mu i_k.
\end{aligned}
\tag{4.7}
$$

Multiplying all these three equations by $\frac{kp_k}{\langle k \rangle}$ and then performing summation over k, we get

$$
\begin{aligned}
\Theta'_u &= \sum_k \frac{kp_k}{\langle k \rangle} \mu - \rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \Theta_b - \mu \sum_k \frac{kp_k}{\langle k \rangle} u_k, \\
\Theta'_b &= \rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \Theta_b + \alpha_n \sum_k \frac{k^2 p_k}{\langle k \rangle} i_k \Theta_b - (\sigma + \mu) \sum_k \frac{kp_k}{\langle k \rangle} b_k, \\
\Theta'_i &= \sigma \sum_k \frac{kp_k}{\langle k \rangle} b_k - \alpha_n \sum_k \frac{k^2 p_k}{\langle k \rangle} i_k \Theta_b - \mu \sum_k \frac{kp_k}{\langle k \rangle} i_k.
\end{aligned}
\tag{4.8}
$$

4.4.2 Early Stage Analysis

In the initial phase of message spreading [77], u_k can be approximated by by 1, and b_k as well as i_k can be considered to be negligible. Using these values in nonlinear terms so that they can be simplified to a linear equation, we get

$$
\begin{aligned}
\Theta'_u &= \mu - \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \Theta_b - \mu \Theta_u, \\
\Theta'_b &= \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \Theta_b - (\sigma + \mu) \Theta_b, \\
\Theta'_i &= \sigma \Theta_b - \mu \Theta_i.
\end{aligned}
\tag{4.9}
$$

Integrating second equation of the above system and using b_0 as initial value of Θ_b, we get $\Theta_b = b_0 e^{\frac{t}{\tau_b}}$ where

$$
\tau_b = \frac{\langle k \rangle}{\rho_n \langle k^2 \rangle - (\sigma + \mu) \langle k \rangle}.
\tag{4.10}
$$

Putting the value of Θ_b in the third equation of the system and using i_0 as an initial value of Θ_i, we get $\Theta_i = C_1 e^{\frac{t}{\tau_b}} + C_2 e^{-\frac{t}{\tau_i}}$, where

$$C_1 = \sigma \tau_b b_0; \quad C_2 = \sigma \tau_b b_0 - i_0; \quad \tau_i = \frac{1}{\mu}. \tag{4.11}$$

Similarly, first equation of the system with u_0 as the initial value of Θ_u gives $\Theta_u = \mu t + C_3 e^{\frac{t}{\tau_b}} + C_4 e^{\frac{-t}{\tau_u}}$, where

$$C_3 = -\rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \tau_b b_0; \quad C_4 = u_0 + \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \tau_b b_0; \quad \tau_u = \frac{1}{\mu}. \tag{4.12}$$

For an epidemic to spread, τ_b must be positive. This condition gives a relation between various rate constants of the model and network parameters to ensure epidemic, i.e.,

$$\frac{\rho_n}{\sigma + \mu} > \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{4.13}$$

As per eq. 3.13, we have again chosen ρ_n to be $\frac{\rho}{\langle k \rangle}$ so that the overall infection by a node in homogeneous as well as heterogeneous approaches remain same and it will be possible to compare the results. Substituting this value, the inequality 4.13 modifies to

$$\frac{\rho}{\sigma + \mu} > \frac{\langle k \rangle^2}{\langle k^2 \rangle}. \tag{4.14}$$

We have the same inequality which we had in eq. 3.15 in the previous chapter. Further analysis of the condition for random and scale-free networks will also be exactly the same. We are skipping those details here. See Sec. 3.3.1 for further details.

4.4.3   Steady State Analysis

In large time limit, the system will reach a steady state. The rate of change of fractions u, b and i will be zero. In case of degree based compartment scheme, u_k, b_k and i_k will not change. Equating the three system evolution equations of eq. 4.7 to

zero, we have

$$b_k = \frac{\rho_n k \Theta_b (\mu + \alpha_n k \Theta_b)}{(\mu + \rho_n k \Theta_b)(\alpha_n k \Theta_b + \sigma + \mu)}. \tag{4.15}$$

Multiplying b_k by $\frac{k p_k}{\langle k \rangle}$ and performing summation over k we get

$$\Theta_b = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k^2 \rho_n \Theta_b (\mu + \alpha_n k \Theta_b)}{(\mu + \rho_n k \Theta_b)(\sigma + \mu + \alpha_n k \Theta_b)}. \tag{4.16}$$

This is a self-consistency equation where $\Theta_b = f(\Theta_b)$. At $\Theta_b = 0$, $f(\Theta_b)$ is also zero.

Hence, $\Theta_b = 0$ is a solution of the equation. Value of the function at $\Theta_b = 1$ is

$$f(1) = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k^2 \rho_n (\mu + \alpha_n k)}{(\mu + \rho_n k)(\sigma + \mu + \alpha_n k)} \tag{4.17}$$

which after a slight arrangement can be written as

$$f(1) = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k}{(1 + \frac{\mu}{\rho_n k})(1 + \frac{\sigma}{\mu + \alpha_n k})}. \tag{4.18}$$

It is clear from the above expression that $f(1) < 1$. A sample diagram for the scenario has been given in Fig. 4.7. It can be observed from the figure that in order to have a non-zero solution in the interval 0 to 1, the slope of the function at $\Theta_b = 0$ must be greater than 1.

$$\left. \frac{\mathrm{d} f(\Theta_b)}{\mathrm{d} \Theta_b} \right|_{(\Theta_b = 0)} = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k^2 \rho_n}{(\sigma + \mu)} = \frac{\rho_n}{(\sigma + \mu)} \frac{\langle k^2 \rangle}{\langle k \rangle} \geq 1$$

which is the same condition we had from linear approximation in the initial phase of the epidemic. We conclude that satisfying this single condition is sufficient for an initial spread of the campaign to reach out to a large fraction of the population as well as to achieve a steady state without dying out sooner.

4.5   Numerical Results

Simulations for homogeneous as well as heterogeneous analysis have been carried out on MATLAB. To understand the effect of population heterogeneity and network
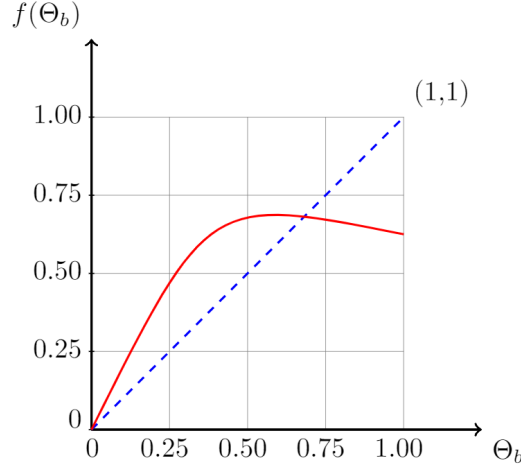
Figure 4.7: Condition for existence of a non-zero steady state value of Θ_b in the range 0 to 1. Curve $f(\Theta_b)$, represented by a red solid line, is starting from $(0,0)$ and at $\Theta_b = 1$, its value is less than 1. If initial slope of the curve $f(\Theta_b)$ will be less than 1, then it will never intersect with the line $f(\Theta_b) = \Theta_b$, represented by a dotted blue straight line, and there will be no solution of the equation $f(\Theta_b) = \Theta_b$, other than zero.

topology, we have considered random network, scale-free network, and a few real social networks in our simulations.

4.5.1 Simulation of Homogeneous Model

As discussed in Sec. 4.3.4, the system may lead to any one of three possible steady-state situations: when only message-free state exists; when only an endemic state exists; when an endemic, as well as a message-free state, can exist depending on the initial population of different classes. We have shown two cases where at least one of the steady state is endemic, in Fig. 4.8(a)-(b) with their respective parameter values.
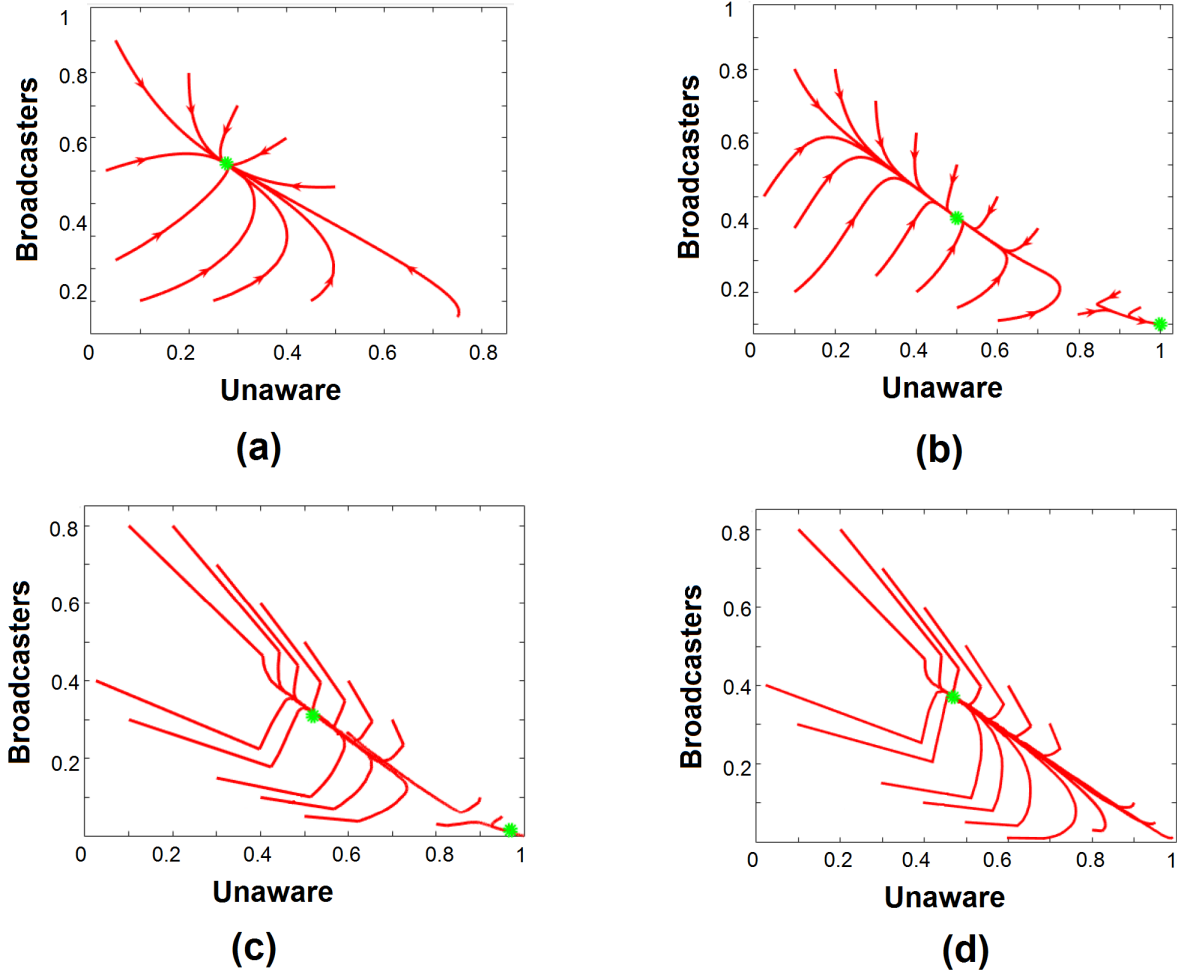
Figure 4.8: Numerical simulation of convergence to the steady state for different initial conditions, for (a) Homogeneous system with single endemic steady state for $\mu = 0.05$, $\sigma = 0.1$, $\rho = 0.25$ and $\alpha = 0.4$; (b) Homogeneous system with bistable dynamics for $\mu = 0.05$, $\sigma = 0.15$, $\rho = 0.15$ and $\alpha = 0.75$; Temporal variation of u and b with different initial conditions for equivalent parameter regime as (b) in (c) random network and (d) scale-free network. To ensure the equivalence with the homogeneous analysis, the infection rate and the relapse rate for network dynamics are taken as $\rho/\langle k \rangle$ and $\alpha/\langle k \rangle$ respectively. In all of these figures, X and Y coordinates of the initial point of any flow represents the initial fractional population of unaware and broadcaster class of population.

4.5.2   Simulation over Model Networks

The same set of parameters have been used to compare the results of the homogeneous model with the heterogeneous model. We have carried out our simulation over random and scale-free network having 1024 nodes and average degree 20.

In the case of a random network, we obtained similar results as predicted by the homogeneous model. Using the values of the bistable steady-state configuration of Fig. 4.8(b), we similarly observed two stable states in Fig. 4.8(c), one endemic and one message-free. Steady state value obtained by the simulation is also in agreement with the homogeneous approach.

In the case of a scale-free network, endemic steady state values are not exactly the same. For the considered set of parameters, there is a maximum of 4% error in the steady-state fraction of different classes. This error is due to non-homogeneity present in the scale-free network in the form of hubs. The second point of difference is that message-free state never appears in a scale-free network. The fact is in alignment with our analytic result regarding the absence of an epidemic threshold in a scale-free network. It can be observed in Fig. 4.8(d) where the temporal variation of u and b leads to an endemic steady state in every set of initial conditions, in contrast with the random network scenario of Fig. 4.8(c), where we can see few flows terminating at message-free steady state.

To understand the dynamics of nodes of different degree, we have plotted steady state value of u, b, and i in their neighborhood. As shown in Fig. 4.9(a), for a random network, these fractions are independent of the degree of nodes. Hence, the number of broadcasters around a higher degree node is large as compared to a lower degree node which makes them more prone to receive the message. It is evident from Fig. 4.9(b) where the fraction of broadcasters is shown to be monotonically increasing with the
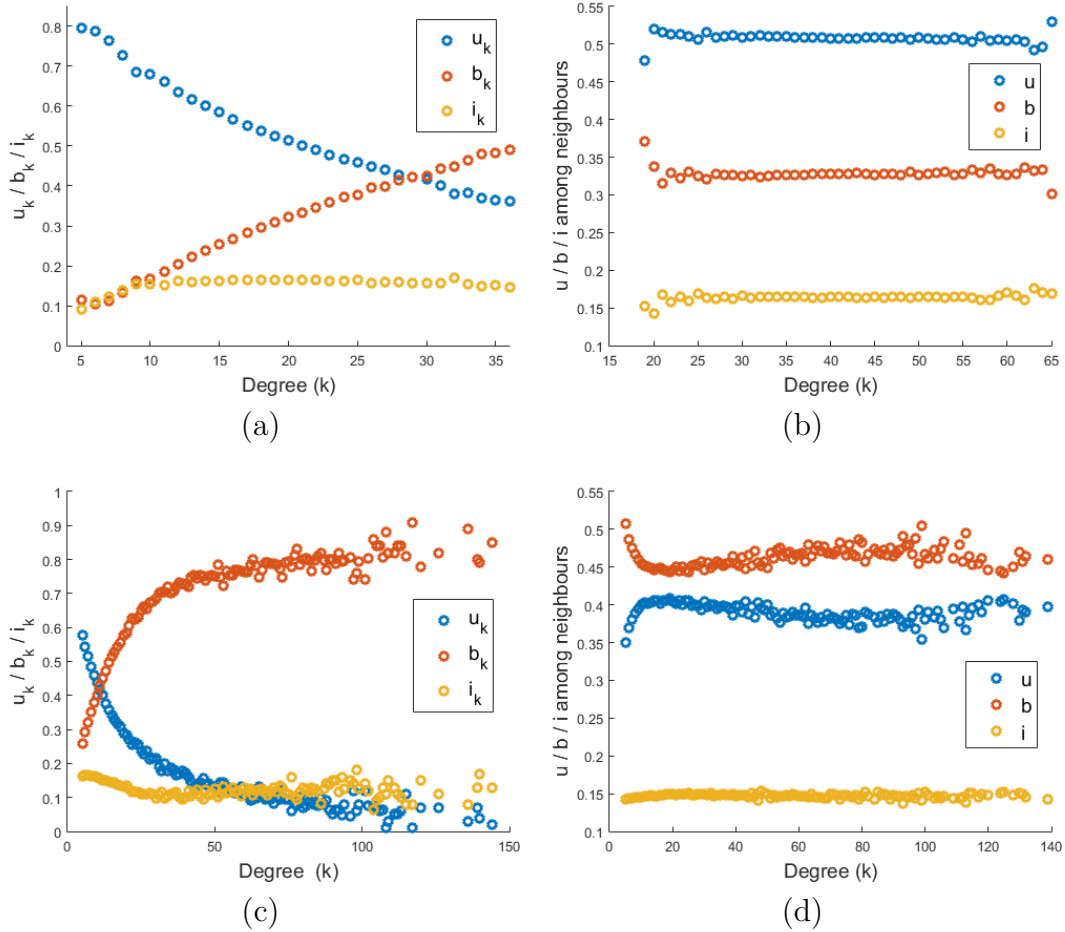
Figure 4.9: (a) Degree-wise fraction of u, b and i with respect to degree k at steady-state in random network; (b) Fraction of u, b and i in the neighborhood of a node with degree k in random network; (b) u_k, b_k and i_k with respect to k at steady-state in random network; (c) Degree-wise fraction of u, b and i with respect to degree k at steady-state in scale-free network; (d) Fraction of u, b and i in the neighborhood of a node with degree k in scale-free network; u_k, b_k and i_k with respect to k at steady-state in scale-free network.

degree. Even in a scale-free network, b_k increases with degree k as shown in Fig. 4.9(d). But, in Fig. 4.9(c) fraction of broadcasters around any node is more than the fraction of unawares which is entirely opposite to the random network case shown in Fig. 4.9(a). This outcome is a result of these two features of the scale-free network: (i) chances of higher degree nodes (hubs) getting infected is very high as shown in Fig. 4.9(c) and (ii) same hubs are present in the neighborhood of multiple nodes while counting broadcasters around a node.

The second feature can be observed clearly in the portion of a network shown in Fig. 4.10. There are 10 nodes in the network and 2 of them having higher degree (hubs) are broadcasters. Overall fraction of broadcasters in the network is thus 0.2. Both of these nodes have 5 neighbors and only one of the neighbors is the broadcaster. The fraction of broadcaster in the neighborhood of these 2 nodes is thus 0.2. Remaining 8 nodes are having degree 2 and one of their neighbors (one of the hubs) is broadcaster and second one is unaware. The fraction of broadcasters in their neighborhood is hence 0.5. Statistically, the average fraction of broadcasters in the neighborhood of a node in the network is $\frac{2 \times 0.2 + 8 \times 0.5}{10} = 0.44$, which is quite larger than the overall fraction of believers in the population. Here we can see that same hub is repeatedly counted as a broadcaster neighbor while calculating the fraction of broadcasters in the neighborhood. This redundancy leads to increment in the local fraction of broadcasters in neighborhood of a node.

4.5.3 Simulation over Real Networks

Though Figs. 4.8 and 4.9 indicate that the information diffusion in the proposed model follows the dynamics as discussed in Sec. 4.2 and 4.4, it is important to analyze
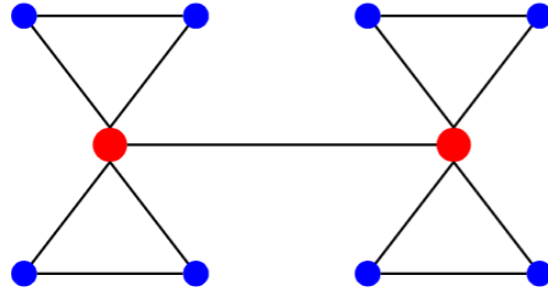
Figure 4.10: A sample network structure where individuals are denoted by nodes and interaction between them by links. Red and blue color nodes represent broadcaster and unaware individuals respectively.
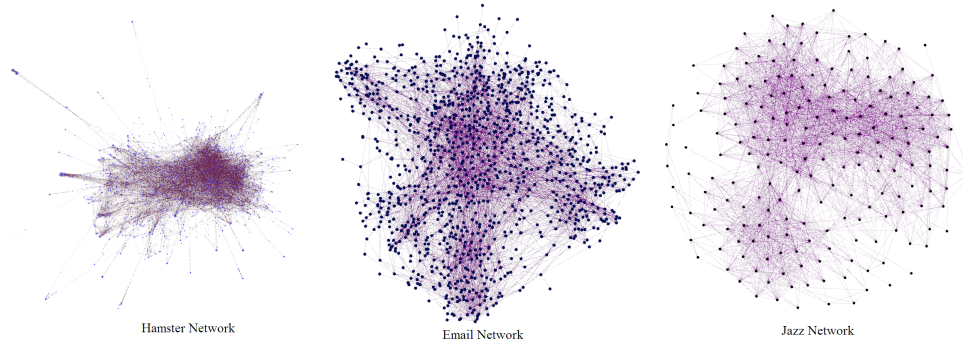
Figure 4.11: Physical topologies of real networks (Hamster, Email, and Jazz) used in our work

the model over real-world networks as most of these networks do not follow the typical characteristics of any particular model network. Thus, we have studied the proposed viral marketing model over some real networks collected from KONECT database [41], to understand its behavior in real social interactions scenarios.

The networks that we have used for testing our VM models are referred to as Hamster network, Email network and Jazz network in rest of the chapter. The first real network that we considered is the friendship network of website www.hamsterster.com that has 2,426 users (nodes) with 16,631 friendships edges. The second one, the Arena email network has been collected from University Rovira i Virgili of Spain. The network

Table 4.1: Important characteristics of different network

Network characteristics	Hamster Network	Email Network	Jazz Network
Number of nodes	2426	1133	198
Number of edges	16631	5451	2742
Average degree	13.71	9.624	27.7
Maximum degree	273	71	100
Power law Exponent	2.46	6.77	5.27

has 1,133 users (nodes) with 5,451 connections (edges). The third one, referred to as the Jazz network, is the collaboration network between jazz musician that can be visualized as a network of people with common interest or skill. The jazz network has 198 musicians (nodes) with 2,742 collaborations (edges). In Fig. 4.11 we show the topologies of the real networks used in this work. The network parameters for the real networks that are considered in this chapter are summarized in Table 4.1.

We studied the flow of a viral message in all the three real networks mentioned above. As email network has almost the same number of nodes as the model networks considered in our simulations, we show the time evolution of email network for different parameters and initializations in Fig. 4.12 as the message diffuses. In Fig. 4.12(a), we set the parameters equivalent to Fig. 4.5(a) along with $\mathcal{R} = 0.64$, which gives only VM free state in homogeneous analysis. For all different initializations, we get a complete broadcaster-free state in the email network as well. To analyze bistability, we set the parameters equivalent to Fig. 4.5(c) with $\mathcal{R} = 0.64$, which belongs to a bistable region in homogeneous case. To locate the lower branch, we generated five different realizations of simulations for 10^4 time units, and results were obtained where 2% of the nodes were broadcasters initially. If the infected fraction went to zero in any of the five runs, the message-free state was considered stable [78]. Simulation results show that the email network exhibits VM free state when the number of broadcasters
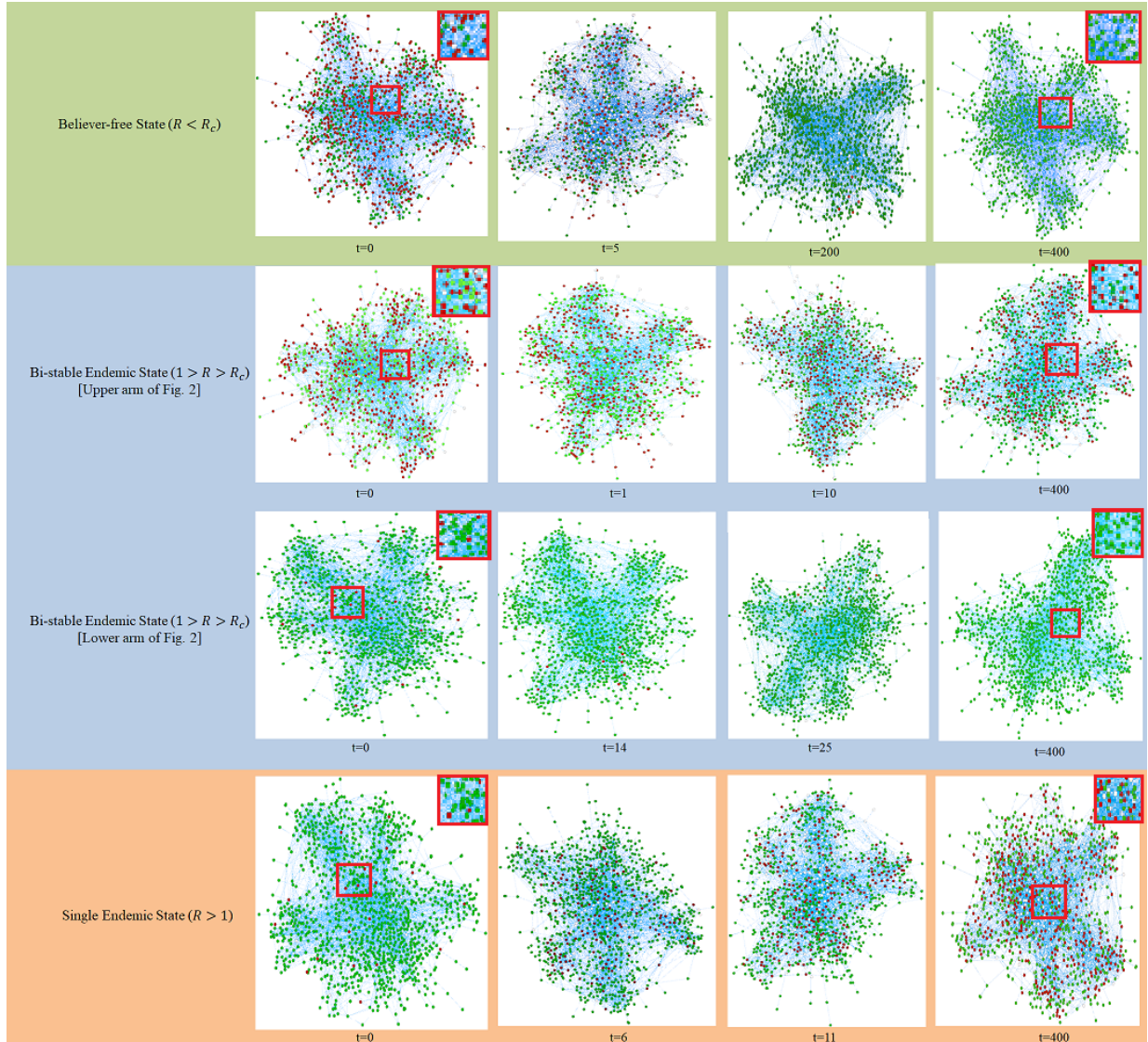
Figure 4.12: Time evolution of email network for $\mathcal{R} = 0.64$: (a) when the network parameters are equivalent to Fig. 4.5(a); steady state is completely free of broadcasters. (b) When the network parameters are equivalent to Fig. 4.5(b), which satisfies $\mathcal{R}_c < \mathcal{R} < 1$ with 70% broadcasters initially; steady state is endemic, having 30% broadcasters. (c) When the network parameters are equivalent to Fig. 4.5(b) but with 2% broadcasters initially; steady state is completely free of broadcasters. (d) When the network parameters are equivalent to Fig. 4.5(b) but for $\mathcal{R} = 1.4$, that satisfies $\mathcal{R} > 1$ with 2% broadcasters initially; steady state is endemic. green, red and white colors represent unaware, broadcaster and inert nodes respectively. Some portions (marked in red squares) of the network at the initial state and the final state are enlarged and shown in corner. Please refer to the online version of the chapter at maximum zoom to fully appreciate the results.

Table 4.2: Comparison of steady states in different networks for $\mathcal{R} = 0.64$

Steady state fraction	Homogeneous Setting	Random Network	Scale-free Network	Hamster Network	Email Network	Jazz Network
u^*	0.5	0.505	0.49	0.59	0.53	0.50
b^*	0.33	0.325	0.37	0.31	0.35	0.37
i^*	0.17	0.17	0.14	0.1	0.12	0.13

is low initially. To locate the upper branch, the system was run to a steady state where initially 70% of the total population were acting as broadcaster and none were in the inert state in each network. For each run, we studied the system for 10^4 time units and then averaged over 200 samples. Like the homogeneous model, with a high number of broadcasters initially, the email network also shows the endemic state. We show the steady states of the email network for both the initializations in Figs. 4.12(b) and 4.12(c) respectively; different final states for different initial conditions demonstrates the existence of hysteresis. It is also noted that for all different networks, systems' propensity for the endemic state starts to dominate as $\mathcal{R}$ goes beyond $\mathcal{R}_c$, for a specific parameter set. As $\mathcal{R}_c$ is always less than 1, even in real networks we acquire a state of endemic for $\mathcal{R} < 1$, where the message is being spread throughout the population. The final steady-state conditions for the real networks are compared in Table 4.2 for $\mathcal{R} = 0.64$ with 70% broadcasters initially and parameters equivalent to Fig. 4.5(b). It shows that 30-35% of the population belongs to the broadcaster class, ensuring the survival of the advertisement campaign in the steady state, even when $\mathcal{R} < 1$.

4.5.4   Behavior for Unequal Birth and Death Rate

It is to be noted that for a more realistic modeling of the dynamics, the birth and death rates could be considered different, i.e., we can assume that people enter and leave the population at different rates. Considering that in a social media platform,

new people migrate in at a much faster rate than the rate at which people leave the platform, it might be assumed that the birth rate of unaware people is μ while the death rate for them is μ_1 ($< \mu$). The death rate for broadcasters, as they are more active in the media platform, could be taken as much smaller than μ (we can call it μ_2), while the rate at which inerts leave the population could be a bit more than that of μ, which could be taken as μ_3. A model with a variable population instead of a fixed one was considered. We tested our model with a typical set of parameters chosen according to the above logical relation, with $\mu = 0.05, \mu_1 = 0.03, \mu_2 = 0.005$ and $\mu_3 = 0.07$. With the other parameters unchanged, the equilibrium analysis shows no qualitatively different results; bistable (for $\alpha = 1$) as well as monostable (for $\alpha = 0.1$) dynamics were observed for higher and lower values of α respectively, but the mathematical handling becomes complicated.

4.6   Extended Viral Marketing Model

A major finding of the conducted survey was the possibility of people returning to broadcaster class from inert one. It was accommodated in the model by adding a transition from I to B. In the extended version of this model, we are including a few more real-life observations.

Till now, we have assumed that reason for I to B transition is the campaign information shared to inert people from their friends in the broadcaster class. Transition rate was, therefore, αbi, proportional to the fractional population of B as well as I. But, there is a fraction of people from inert class who goes back to class B by their own choice without persuasion of anyone else. The transition is generally noticed when an
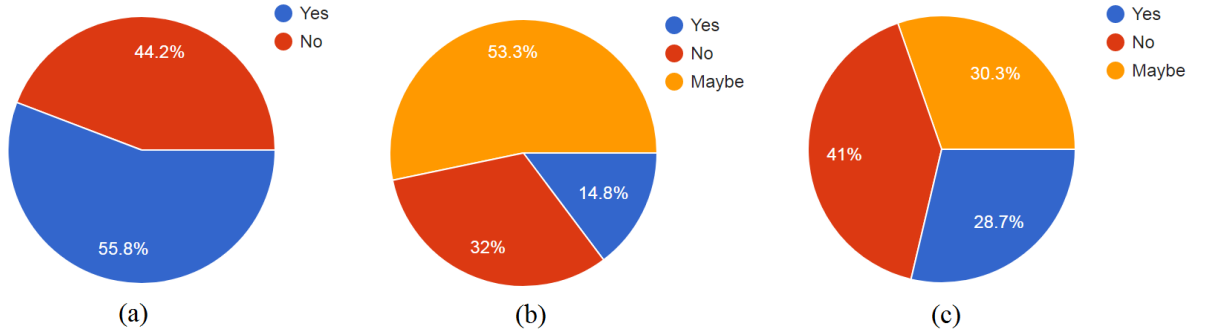
Figure 4.13: Rigidly inert people in a population : (a) 'Yes' ('No') signifies people who (never) contributed in VM campaign; (b) Possibility of regaining interest of inert class in a viral offer when a positive feedback or positive review is circulated by a company. (c) Possibility of regaining interest of inert class in a viral offer when friends request to avail, or discuss about in social platform. We mark the people as rigidly inert who do not contribute in VM campaign, and never gain interest in such activity even in presence of positive review or friend's requests.

inert individual remembers the campaign s(he) had once forgotten or sees their friends getting benefits from any particular campaign. This behavior has been included in the model by adding a linear transition from inert to the broadcaster with rate λi.

The second observation, shown in Fig. 4.13, is that almost 23% of the total population was rigidly inert, and they show no interest to be a part of a viral campaign ever. Even a good amount of reward or an exceptionally good review of the product does not entice them to share these messages with their peers. The reason behind that was noted as follows:

1. Spamming: While preparing the questionnaire, the factor of spamming was not explicitly mentioned by us; hence, it was not among the given options. Nevertheless, the words related to spamming (e.g., spamming, spam, spammer, bulk, etc.) found 18 appearances in our final excel sheet, as participants mentioned them while explaining their reasons to be against viral campaigns. One partic-

ipant clearly stated that *" . . . If I send this kind of messages, though it's a spam, it will find its way into the main Inbox. Important emails getting lost in the junk of Spam, I hate that."* Another respondent explained, *" . . . Whats the point? Nine out of ten people would ignore it as its spam. dont wanna be spammer."* Moreover, spam messages can never be trusted, as those can contain viruses, malware that can end up disrupting the computerized environment of the user. Many respondents also mentioned that they assumed these emails being associated with a virus *(" Not felt secured, link has some virus")* or phishing *(" Asking about personal information"* or *" . . . stealing personal data")*.

2. Brands: We also found that the idea of brands closely affects this. While reputed brands are connected with trust and comfort, 39% respondents thought that recognized brands would not be associated with spurious offers. Interestingly, two consecutive surveys in 2012 and 2015 pointed out, that customers identify a brand as unprofessional and choose to email opt-out (or unsubscribe) in case of excessive frequency of promotional emails.

These observations indicate that whenever a broadcaster sends a referral message to an unaware individual, apart from entering broadcaster class, unaware may directly move to inert class as well. These two factors were compiled into a parameter that gives a room for people to move from U to I class directly.

In our previous viral marketing model, an unaware always has to become a broadcaster before entering the inert class, but in extended model, we have relaxed this assumption and allowed people to move directly from U to I class. Now, whenever a broadcaster sends a referral message to an unaware individual, unaware moves to broadcaster class with probability p and to inert class with probability $(1 - p)$.
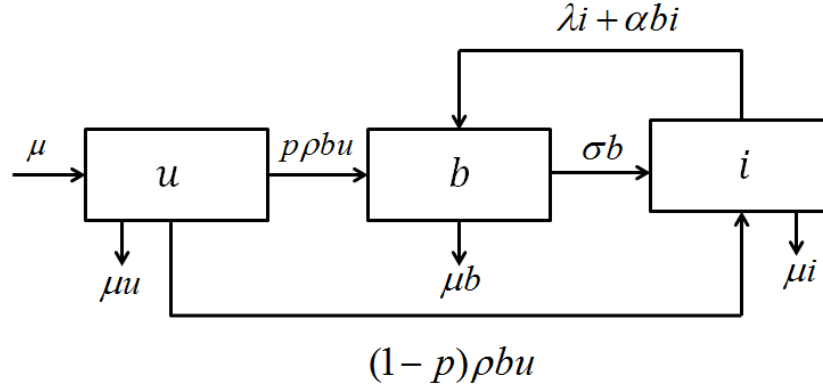
Figure 4.14: Block diagram of the extended viral marketing model showing all possible transitions from one class to another.

Adding these two observations to the previous model, we get an extended version of the viral marketing model as shown in Fig. 4.14.

4.7   Homogeneous Analysis

Transition between three sub-populations Unaware (U), Broadcaster (B), and Inert (I) is now modeled by the following set of coupled ordinary differential equations:

$$
\begin{aligned}
u' &= \mu - \rho bu - \mu u, \\
b' &= p\rho bu + \lambda i + \alpha bi - \sigma b - \mu b, \\
i' &= \sigma b + (1 - p)\rho bu - \lambda i - \alpha bi - \mu i \cdot
\end{aligned}
\tag{4.19}
$$

Similar to the previous model, here also the birth and death rate of the individual has been kept the same (μ) to maintain a fix population size. The parameters ρ, σ, and α have the same meaning. Two new parameters λ and p have been added as per the discussion in the previous section. All the steps carried out in the previous model

will be repeated to analyze the extended model. We will briefly discuss all the steps emphasizing upon the new findings.

4.7.1 Equilibrium Analysis

Similar to our previous model discussed in Sec.4.3, this system model also has two steady states E_0 and $E^\star$. At E_0, the VM marketing campaign fails to sustain, and the whole population eventually becomes unaware about it, whereas at $E^\star$ the system exhibits an endemic equilibrium where a certain percentage of people acts as broadcasters and helps to sustain the campaign. While solving for $E^\star$, the first equation of the eq. set 4.19 gives

$$u^\star = \frac{\mu}{\rho p b^\star + \mu}. \tag{4.20}$$

Substituting this expression of $u^\star$ in second equation of the eq. set 4.19 and replacing $i^\star$ by $(1 - b^\star - u^\star)$, simple algebraic manipulation results into $l(b^\star)^2 + mb^\star + n = 0$, where

$$l = \alpha\rho; \; m = (\sigma\rho + \mu\rho + \lambda\rho + \alpha\mu - \alpha\rho); \; n = \mu(\sigma + \mu + \lambda) - \rho(\lambda + \mu p). \tag{4.21}$$

Examining the coefficients, we conclude that l is always positive; m is positive for small values of α, and n is positive or negative depending on whether $\frac{\rho(\lambda+\mu p)}{\mu(\lambda+\mu+\sigma)} = \mathcal{R}$ is smaller or greater than 1. Two utterly different steady state scenarios can arise:

Case 1: For negative n (i.e., $\mathcal{R} > 1$), the quadratic equation has a unique positive solution $b^\star_+$, as another solution $b^\star_-$ is always negative and so, unphysical, and there exists a unique endemic equilibrium $E^\star$ whenever $\mathcal{R} > 1$.

Case 2: On the other hand, for positive n (i.e., $\mathcal{R} < 1$), the number of physical roots of the equation depends on the sign of m, and therefore, the nonlinear relapse parameter

α. Depending on this fact if α is high (or low), multiple (or no) endemic equilibria may exist.

We can see that the description of the equilibrium analysis is equivalent to the description of the previous VM model discussed in the Sec. 4.3.3.

4.7.2　Reproduction Number

As found in the Section 4.7.1, reproduction number of the model is

$$\mathcal{R} = \frac{\rho(\lambda + \mu p)}{\mu(\lambda + \mu + \sigma)}. \tag{4.22}$$

It is evident that replacing p by 1 and λ by 0, the extended system model of eq. 4.19 will reduce to the previous model of the eq. 4.1. Reproduction number $\mathcal{R}$ mentioned in eq. 4.22 will also reduce to $\frac{\rho}{\mu+\sigma}$ which is the reproduction number of our previous model as discussed in Sec. 4.3.2. From eq. 4.22, we can observe that, for smaller value of p, a larger value of ρ is required to satisfy the basic condition for an epidemic to spread, i.e., $\mathcal{R} > 1$. In a practical sense, it means that if more people are switching directly to inert class from unaware class, more marketing effort will be required to attain an endemic steady state.

The sensitivities of $\mathcal{R}$ for various parameters are as follows:

$$
\begin{aligned}
\Gamma_{\rho}^{\mathcal{R}} &= 1, \\
\Gamma_{p}^{\mathcal{R}} &= \mu p, \\
\Gamma_{\sigma}^{\mathcal{R}} &= -\frac{\sigma}{(\lambda + \mu + \sigma)}, \\
\Gamma_{\lambda}^{\mathcal{R}} &= \frac{\lambda}{(\lambda + \mu p)}\frac{\sigma + \mu(1 - p)}{(\lambda + \mu + \sigma)}, \\
\Gamma_{\mu}^{\mathcal{R}} &= -\left(\frac{\lambda}{\lambda + \mu p} + \frac{\mu}{\lambda + \mu + \sigma}\right).
\end{aligned}
\tag{4.23}
$$
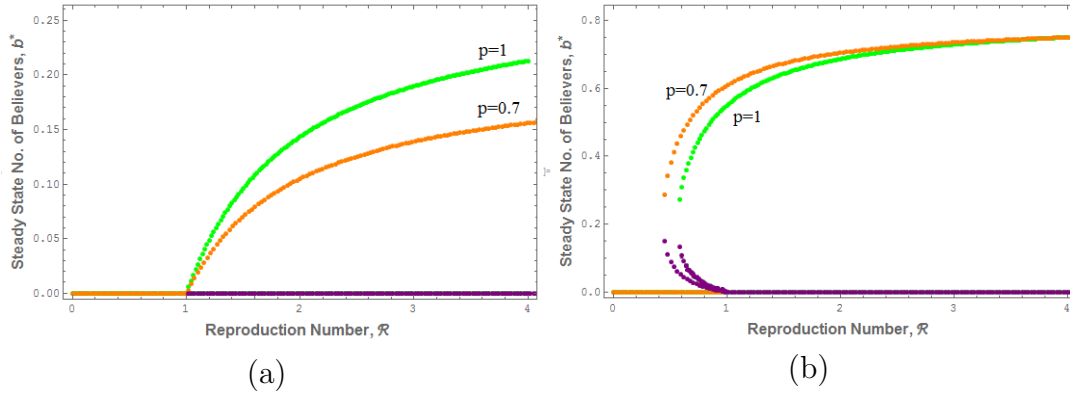
Figure 4.15: Variation in steady state fraction of b with reproduction number $\mathcal{R}$ for (a) $\alpha = 0.1$, when only a single epidemic state persists beyond $\mathcal{R} = 1$ and for (b) $\alpha = 1$, when bistability can be observed in range $\mathcal{R}_c$ to 1. Parameter values are $\sigma = 0.2$, $\lambda = 0.0002$, and $\mu = 0.05$. In these figures, orange and green lines indicate stable solutions and purple lines indicate unstable solutions. For these parameter values, we calculated $\mathcal{R}_c = 0.562$ from eq. 4.26.

4.7.3 Bifurcation

As discussed in Sec. 4.7.1, the extended model also exhibits different behavior for smaller and larger value of α. In Fig. 4.15, we plot the steady-state fraction of class B for two different values of α. To highlight the impact of the parameter p, results have been shown for $p = 1$ and $p = 0.7$.

Results for a small value of α, shown in Fig. 4.15 (a) depicts that the endemic steady-state fraction of broadcasters, $b^\star$, reduces with a decrease in value of p. For a smaller value of p, the probability of an unaware individual to move to broadcaster class will be less, and more people will be moving from unaware class to inert class directly. As the relapse rate, α is also small, switching from inert to broadcaster will take place at a smaller rate. Reduction in $b^\star$ can be attributed to the combined effect of these two phenomena.

In Fig. 4.15 (b), we can see that for a higher value of α, there is no significant change in $b^\star$ for different values of p, in the region $\mathcal{R} > 1$. Although a fraction of unaware

population is coming directly to inert class, the large value of α brings them back to the broadcaster class. In this case, the significant change is observed in the region $\mathcal{R} < 1$, where we can observe the expansion in the bistable region as $\mathcal{R}_c$ decreases. We will derive the expression of $\mathcal{R}_c$ in the next section.

Though for the sake of completeness we have discussed all the parameter space of the system model given by eq. 4.19 which also includes a high value of α with a low value of p (close to 0), but this region may not be observed in real cases. A small value of p signifies that a large fraction of population goes to inert class as soon as they listen about the campaign for the first time, whereas a large value of α means broadcasters can bring a large fraction of inert people back to the broadcaster class. If the first impression of the campaign for most of the unaware class is not good, then it is very much unlikely that they can be easily brought back to the broadcaster class later. That is why in practical scenarios, a low value of p and the high value of α may not exist simultaneously.

4.7.4   Conditions for Bistability

We have found in Sec. 4.7.1 that multiple solutions (bistable region) may exist when $\mathcal{R} < 1$, for large value of α. Equating the coefficient m in the quadratic eq. 4.21 to 0, we get threshold value of α

$$\alpha_{th} = \frac{\rho(\sigma + \mu + \lambda)}{(\rho - \mu)}. \tag{4.24}$$

Interestingly, α_{th} does not depend on the value of p. Comparing with the α_{th} of the previous model (Sec. 4.3.4), the value is larger due to the addition of λ in the numerator.

The critical value of ρ is calculated by equating $m^2 - 4ln$ to 0 and is given by

$$\rho_c = \frac{\alpha\mu}{(\alpha + \lambda + \mu + \sigma) - 2\sqrt{\alpha(\mu - p\mu + \sigma)}}. \tag{4.25}$$

Substituting ρ_c in place of ρ in the expression of reproduction number $\mathcal{R}$ (eq. 4.22) gives the critical value of reproduction number $\mathcal{R}_c$ as

$$\mathcal{R}_c = \frac{\rho_c(\lambda + \mu p)}{\mu(\lambda + \mu + \sigma)} = \frac{\alpha\mu}{(\alpha + \lambda + \mu + \sigma) - 2\sqrt{\alpha(\mu - p\mu + \sigma)}} \frac{(\lambda + \mu p)}{\mu(\lambda + \mu + \sigma)}. \tag{4.26}$$

At $p = 1$ and $\lambda = 0$, $\mathcal{R}_c$ is equal to the value of the threshold in our previous model (Sec. 4.6). To see the limiting case, equating $p = 0$ we observe that value of $\mathcal{R}_c$ will be smaller than the value obtained in the previous model. It means that region of bistability $\mathcal{R} \in [\mathcal{R}_c, 1)$ will expand.

4.8 Heterogeneous Analysis

Similar to the heterogeneous analysis of the previous VM model, we will formulate the differential equation using degree block approximation in Sec. 4.8.1. We will then derive the threshold condition for a viral marketing message to spread in the population and its persistence in the population at the steady state.

4.8.1 Degree Block Approximation

Differential equations representing the evolution of degree based compartments of different classes will be

$$
\begin{aligned}
u_k' &= \mu - \rho_n k u_k \Theta_b - \mu u_k, \\
b_k' &= p\rho_n k u_k \Theta_b + \lambda i_k + \alpha_n k i_k \Theta_b - (\sigma + \mu)b_k, \\
i_k' &= \sigma b_k + (1-p)\rho_n k u_k \Theta_b - \lambda i_k - \alpha_n k i_k \Theta_b - \mu i_k.
\end{aligned}
\tag{4.27}
$$

All the parameters used in the above model have the same meaning what they had in the previous model of eq. 4.1, apart from λ and p which have been introduced in the extended model. ρ_n and α_n are the network counterparts of ρ and α used in the homogeneous setting. The density function of broadcasters in the neighborhood of a node is Θ_b.

Multiplying all these three equations by $\frac{kp_k}{\langle k \rangle}$ and then performing summation over k, we get

$$
\begin{aligned}
\Theta'_u &= \sum_k \frac{kp_k}{\langle k \rangle}\mu - \rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \Theta_b - \mu \sum_k \frac{kp_k}{\langle k \rangle} u_k, \\
\Theta'_b &= p\rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \Theta_b + \lambda \sum_k \frac{kp_k}{\langle k \rangle} i_k + \alpha_n \sum_k \frac{k^2 p_k}{\langle k \rangle} i_k \Theta_b - (\sigma + \mu) \sum_k \frac{kp_k}{\langle k \rangle} b_k, \\
\Theta'_i &= \sigma \sum_k \frac{kp_k}{\langle k \rangle} b_k + (1-p)\rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \Theta_b - \lambda \sum_k \frac{kp_k}{\langle k \rangle} i_k - \alpha_n \sum_k \frac{k^2 p_k}{\langle k \rangle} i_k \Theta_b - \mu \sum_k \frac{kp_k}{\langle k \rangle} i_k.
\end{aligned}
\tag{4.28}
$$

4.8.2 Early Stage Analysis

As done in the Sec. 4.4.2 of the previous model, in the initial phase of message spreading u_k can be approximated by 1, and b_k as well as i_k can be considered to be negligible. Using these values in nonlinear terms so that they can be simplified to a linear equation, we get

$$
\begin{aligned}
\Theta'_u &= \mu - \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle}\Theta_b - \mu\Theta_u, \\
\Theta'_b &= p\rho_n \frac{\langle k^2 \rangle}{\langle k \rangle}\Theta_b + \lambda\Theta_i - (\sigma + \mu)\Theta_b, \\
\Theta'_i &= \sigma\Theta_b + (1-p)\rho_n \frac{\langle k^2 \rangle}{\langle k \rangle}\Theta_b - (\lambda + \mu)\Theta_i.
\end{aligned}
\tag{4.29}
$$

Last two eqs. of the system 4.29 form a system of simultaneous linear differential equations with constant coefficients.

$$\Theta_b' = C_1\Theta_b + C_2\Theta_i \tag{4.30}$$

$$\Theta_i' = C_3\Theta_i + C_4\Theta_b \tag{4.31}$$

Solving the above simultaneous equations, we get

$$\Theta_b'' - (C_1 + C_3)\Theta_b' + (C_1C_3 - C_2C_4)\Theta_b = 0. \tag{4.32}$$

It is clear from the form of the equation that Θ_b will be the summation of two exponential, the exponents of which depend on the roots of the auxiliary equation of the differential eq. 4.32. For a viral message to spread in the population, Θ_b needs to be an increasing function in time. If $C_1C_3 - C_2C_4 > 0$ holds, it can be shown that C_1 will surely be negative and in turn $C_1 + C_3$ will be negative resulting into negative roots of the auxiliary equation. Hence Θ_b will be an exponentially decaying function. So, necessary condition for initial growth in class B is $C_1C_3 < C_2C_4$. Substituting the expression of all these constant terms, the condition modifies to

$$\frac{\rho_n}{\mu}\frac{(\lambda + p\mu)}{(\sigma + \lambda + \mu)} > \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{4.33}$$

Replacing ρ_n by $\frac{\rho}{\langle k \rangle}$ the condition modifies to

$$\frac{\rho}{\mu}\frac{(\lambda + p\mu)}{(\sigma + \lambda + \mu)} = \mathcal{R} > \frac{\langle k \rangle^2}{\langle k^2 \rangle}. \tag{4.34}$$

In the extended model also, we have been able to find the analytical expression of the epidemic threshold. While finding threshold condition in every heterogeneous model, we observe that left-hand side of the inequality is the reproduction number $\mathcal{R}$ of the homogeneous model and right-hand side is $\frac{\langle k \rangle^2}{\langle k^2 \rangle}$ which depends on average degree and

average of square of individual degrees of the nodes in the network. The exact value of the expression will depend on the type of network. Values for the random and scale-free network has been discussed in the Sec. 3.3.1.

4.8.3   Steady State Analysis

In large time limit, system will reach steady state and rate of change of fractions u_k, b_k and i_k will become zero. Equating first and third equation of eq. set 4.27 to zero, we have

$$u_k = \frac{\mu}{\mu + \rho_n k \Theta_b} \qquad and \qquad i_k = \frac{\sigma b_k + (1-p)\rho_n k u_k \Theta_b}{\lambda + \mu + \alpha_n k \Theta_b}. \tag{4.35}$$

Putting these values in second equation of the same set will give

$$b_k = \frac{\rho_n k \Theta_b (\mu p + \lambda + \alpha_n k \Theta_b)}{(\mu + \rho_n k \Theta_b)(\lambda + \mu + \sigma + \alpha_n k \Theta_b)}. \tag{4.36}$$

Multiplying b_k by $\frac{k p_k}{\langle k \rangle}$ and performing summation over k, we get

$$\Theta_b = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k^2 \rho_n \Theta_b (\mu p + \lambda + \alpha_n k \Theta_b)}{(\mu + \rho_n k \Theta_b)(\lambda + \mu + \sigma + \alpha_n k \Theta_b)}. \tag{4.37}$$

This is a self consistency equation where $\Theta_b = f(\Theta_b)$. At $\Theta_b = 0$; $f(\Theta_b)$ is also zero. Hence $\Theta_b = 0$ is a solution of the equation. Value of the function at $\Theta_b = 1$ is

$$f(1) = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k^2 \rho_n (\mu p + \lambda + \alpha_n k)}{(\mu + \rho_n k)(\lambda + \mu + \sigma + \alpha_n k)}. \tag{4.38}$$

After slight rearrangement of the terms, we get

$$f(1) = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k}{(1 + \frac{\mu}{\rho_n k})(1 + \frac{\sigma + (1-p)\mu}{\lambda + \mu p + \alpha_n k})}. \tag{4.39}$$

It is clear from the above expression that $f(1) < 1$. As shown in Fig. 4.7, to have another solution in the interval 0 to 1, slope of the function at $\Theta_b = 0$ must be greater

than 1.

$$\frac{\mathrm{d}f(\Theta_b)}{\mathrm{d}\Theta_b}\bigg|_{(\Theta_b=0)} = \frac{1}{\langle k \rangle} \sum_k \frac{p_k k^2 \rho_n (\mu p + \lambda)}{\mu(\lambda + \mu + \sigma)} = \frac{\rho_n (\mu p + \lambda)}{\mu(\lambda + \mu + \sigma)} \frac{\langle k^2 \rangle}{\langle k \rangle} > 1$$

After replacing ρ_n by $\frac{\rho}{\langle k \rangle}$, we will get the same condition what we had from linear approximation in early stage analysis discussed in Sec. 4.8.2.

4.9 Numerical Results

Like previous model, we have again carried out the simulations for homogeneous as well as heterogeneous approach. Along with random and scale free networks, few real network structures have also been considered. Size of the network and simulation are same as used for previous model discussed in Sec. 4.5.

Depending upon the parameter values, system may lead to message-free state or endemic state. Two different cases for homogeneous setting have been shown in Fig 4.16(a)-(b). As discussed in Sec. 4.7.3, bi-stability can be observed in the system for value of α greater than α_{th}. It is observed in Fig. 4.16 (b), where depending on initial fraction of different classes, system reaches to endemic or message-free equilibrium.

Parameter values for bistable case have been used to plot the results for both of the model networks, random as well as scale-free network in Fig. 4.16 (c) and (d) respectively. Results of random network almost matches the findings of homogeneous model. Along with similar endemic steady-state values, bistability can also be observed in random network scenario of Fig. 4.16 (c). In case of scale-free network, endemic steady state values are not exactly same and maximum error in endemic steady-state fraction of a particular class is 5%. Reasons for the difference in steady state values are the same what have been discussed in Sec. 3.3.3.2. Under bistable parameter set,
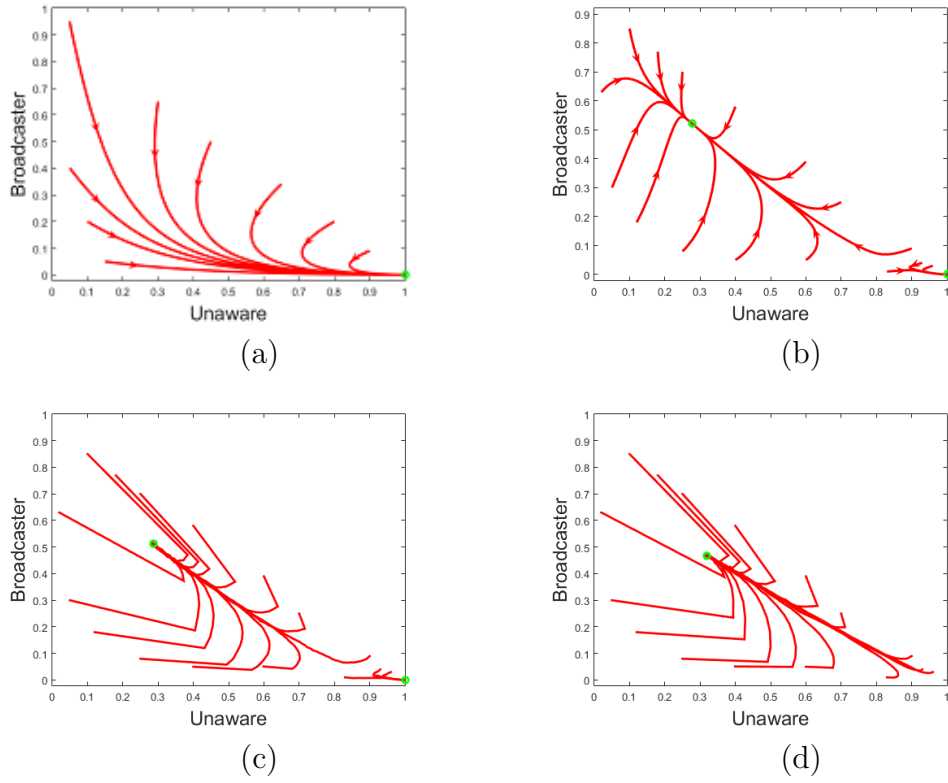
Figure 4.16: Numerical simulation of convergence to the steady state for different initial conditions with parameter values $\mu= 0.05$, $\rho= 0.25$, $\sigma= 0.2$, $\lambda= 0.0002$, $p=0.7$, and (a) $\alpha=0.1$ for a homogeneous system with a single campaign free steady state; (b) $\alpha=1$ for a homogeneous system with bistable steady states ; Temporal variation of u and b with different initial conditions for equivalent parameter regime as for (b) in (c) random network and (d) scale-free network. In all of these figures, X and Y coordinates of the initial point of any flow represents the initial fractional population of unaware and broadcaster class of population.

message-free steady state never appears and system leads to endemic steady state for every set of initial conditions. It can be observed in Fig. 4.16(d) where every flow terminates at endemic steady state. This observation is in alignment with our analytic result regarding absence of epidemic threshold in scale-free network as mentioned in Sec. 4.8.2.

Degree-wise steady state fraction u_k, b_k, and i_k has been plotted for random and scale free networks in Fig. 4.17 (a) and (c) respectively. Fraction of u, b and i in the neighborhood of a node of different degrees has also been plotted for both the networks in Fig. 4.17 (b) and (d) respectively. Figures are showing the same trend what we have observed in Fig. 4.9 of our previous model. A node with higher degree has higher probability to be in broadcaster class, in random as well as scale free network. For random network, the fraction of believers around a node is independent of its nodal degree, but for scale free network this fraction is not identical. It is again due to the heterogeneous structure of the network and presence of hubs in the network.

We have analyzed the model over a few real world networks as most of these networks do not follow the typical characteristics of any particular model network. Analysis helps us to understand the behavior in real social interaction scenarios. Same set of real networks as collected earlier from KONECT database [41], and have been used in the previous model discussed in Sec. 4.5.3, have been considered here.

To see the node dynamics in the real network scenario, we have simulated all three real networks obtained from KONECT database. Steady state values for various networks are mentioned in Table 4.3. Degree-wise steady state fraction and fraction of different classes in neighborhood of a node of a particular degree for email network has also been plotted in Fig. 4.18 (a) and (b) respectively. Observations are close to the results obtained for scale free network shown in Fig. 4.17 (c) and (d).
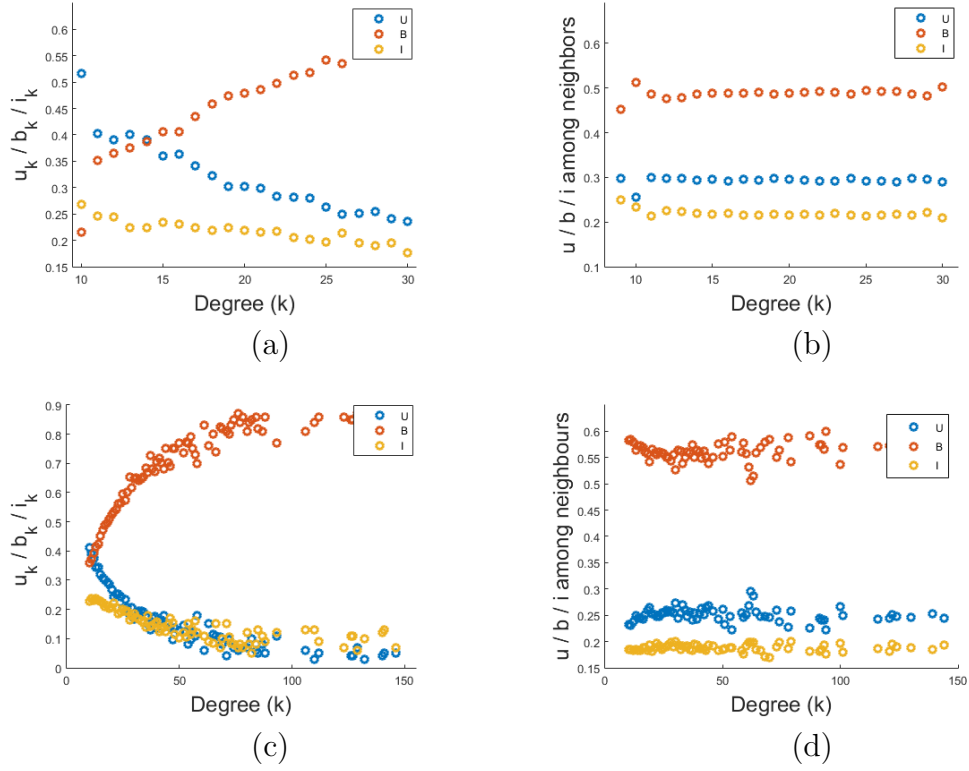
Figure 4.17: (a) u_k, b_k and i_k with respect to k at steady-state in random network; (b)Fraction of u, b and a in the neighborhood of a node with degree k in random network; (c) u_k, b_k and i_k with respect to k at steady-state in scale-free network; (d) Fraction of u, b and i in the neighborhood of a node with degree k in scale-free network.

Table 4.3: Comparison of steady states in extended VM for $\mathcal{R} = 0.64$

Steady state fraction	Homogeneous Setting	Random Network	Scale-free Network	Hamster Network	Email Network	Jazz Network
u^*	0.277	0.288	0.319	0.512	0.394	0.348
b^*	0.521	0.510	0.467	0.342	0.417	0.469
i^*	0.202	0.202	0.214	0.146	0.189	0.183
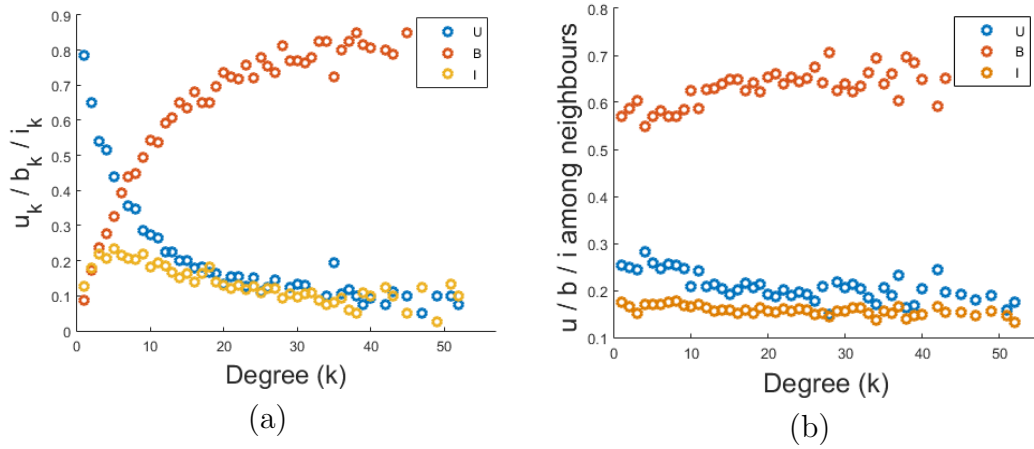
Figure 4.18: (a) u_k, b_k and i_k with respect to k at steady-state in email network; (b) Fraction of u, b and a in the neighborhood of a node with degree k in email network.

4.10   Summary

Marketing is always considered as one of the key components, not an auxiliary arrangement, for a successful business [79]. Surely, a viral marketing campaign works as a less expensive and unexpected way to reach the customers; but nowadays, when almost 85 million videos and photos get uploaded every day in a popular social networking website like Instagram [80], the main challenge is making that advertisement execute long lasting iterations in the population, so that it can reach a bigger audience. If we consider just the case of Instagram, with almost 500,000 advertisers using this popular website as campaigning platform [81], most of the uploads get tossed like a needle in a haystack. The model we propose in this chapter establishes a principle of sustainability for online advertisement campaigns relying on rigorous consumer psychology survey data [82]. Extensive analysis with differential equation based homogeneous approach as well as network simulation based heterogeneous approach shows that the region of bistability grows as the value of α, the nonlinear relapse rate increases. Bistability gives the system a chance to retain its viral state for adverse parametric conditions as well.

While we have observed that the steady states of diffusion in networks are closely related to homogeneous system dynamics, we also appreciated the importance of network structures in this issue. Not only on model systems, but in real social networks, with consideration to all heterogeneity that exists in the population, it has been shown that sustainability of a viral campaign is actually dependent on drawing the attention of those who are not participating in spite of being aware of the campaign. If a certain percentage of this inert population start broadcasting in favor of the campaign, it retains its endemic state in the entire population. We have also shown that over-usage of advertising mails and too much lucrative offers are many a times treated as spam and create negative brand value instead of a positive impact. For a new firm having no brand history, these points must be included in their referral policy, otherwise an initial adverse acceptability of the brand might lead to the failure of the entire campaign process.

The model presented in this chapter is the first to include the relapse rate while analyzing the epidemic spread and sustainability of viral marketing messages. The remarkable effect that this relapse rate has on the sustainability of the campaign has deeper marketer-level implications. Through last couple of years, advertisers have slowly understood the importance of capturing the attention of lost customers. We are already familiar with Facebook retargeting for products, where by adding a code snippet (often called a *pixel*), the online websites retarget attention of the customers, whom they lost from their website due to unknown reasons [83]. Firms are becoming quite inclined to get the services of companies like Adroll, Retargeter, Perfect Audience, etc. or going directly to the exchanges like Google, Facebook, Twitter [84] for running their own retargeting campaigns to re-engage anonymous users. But recent studies show that continuous retargeting leads to a definite privacy concern and skepticism among customers, which results into a lower purchase intention [85]. Our findings in this chapter

point out that the relapse has a major effect, especially if a social-circle-level remarketing technique can be devised where factors like authenticity and security are ensured. The campaigns should adopt clear privacy policies about protection of consumer data, as well as consider adding a social context to encourage spontaneous reminders among the population. As friends and peers have a substantial influence, close proximity and often share similar interest, it is both more plausible and effective, if they assure the lost customers about the genuineness and usefulness of a campaign.

In this chapter, we demonstrate a typical case where the endemic solution is desired for the sustainability of a viral campaign in the population. We discussed the critical parameters in this chapter that helps to sustain the endemic state even in unfavorable conditions by exploiting the properties of bistability. However, in several epidemiological spreading models, often we want to inhibit the diffusion in the population by identifying the contributions of the parameters. In the next chapter, we model the diabolical habit of piracy as an epidemic spreading through a population and discuss the effectiveness of one-to-one awareness policy and mass awareness policy by analyzing various parameters of the flow.

Chapter 5

Contagious Habit of Online Media Piracy

5.1 Introduction

Digital contents like movies, songs, games, software, etc., are protected with copyright and distribution laws so that production companies can maintain the supply chains. Any illegal usage, preservation or distribution of such contents are commonly known as digital piracy. Internet, being the most convenient way to share data across the world, magnifies digital piracy by order of magnitudes [86]. Internet-based piracy, also known as online piracy, is one of the biggest concerns of digital content manufacturers and has become a burning problem for them since last few decades.

Though internet is the prime platform for digital piracy nowadays, the history of illegal distributions of digital contents can be traced back to even early '80s [87]. Back then, movies and music cassette tapes were used to be illegally replicated and sold as pirated copies. However, with the advancement of internet, digital piracy not only becomes fast and easy, but it also becomes extremely difficult to control. In the last decade, the scenario has deteriorated even more rapidly as the connectivity and the number

of internet users have grown rapidly [88]. Though the effects of online piracy on the different industries were not well-understood initially [89, 90, 91], several researchers later discussed the adverse impact of online piracy [92, 93, 94], and pointed out online piracy as a major threat to various industries. In [92], authors found out that countries with higher internet access suffer more due to online piracy. Zenter [95] calculated that online piracy can reduce the profit of music industry by 30%, whereas Ma *et al.* [96] measured that a pirated copy of early released movie can cause 19.1% reduction in the revenue. A more alarming situation was figured out by Danaher and Smith [97] that showed that the problem is not limited to any particular country. Sudler [87] reported that countries like Venezuela, Indonesia, China and Thailand lost more than 70% of their software markets only because of piracy. According to a recent report [98], India is one of the top-3 countries that suffer most due to online piracy, and the cumulative loss was around $3 billion in 2017. According to a survey conducted in the same report [98], 49% out of 10,343 people responded that they have some pirated software in their computers. Even new laws that are imposed to reduce this trend is making little effect in restricting online movie piracy in India [99]. Because of its substantial severity, online piracy is often associated with an epidemic for the digital industry [100, 101, 102].

Several extended studies have been conducted to understand the psychological, social and behavioral aspects of online 'pirates'. Surprisingly, it is observed in several studies that most of the people who are engaged in online piracy know that it is illegal to distribute copyrighted materials [103]. But several demographic (age, gender), economic (price, availability, income), social (recent trends, education, culture) and technical (internet speed, quality of the available product, malware) aspects control the behavior and participation of an individual in online piracy [104, 105]. A recent report [105] shows that online piracy is spreading alarmingly for mobile devices also, and users in the age group 18-24 contribute the most (∼60%) in online piracy in middle east Asian and

African countries. In an another independent study [106], 56.3% participants among 116 university students told that they are well aware of the fact that internet piracy is a serious crime. However, 61.6% of them started downloading pirated media contents online below the age of 12. The survey also revealed that there is a very strong peer influence that attracts people towards online piracy. In the study, 62.5% of students reported that they were introduced to online piracy by their close friends and parents. The main motivations behind engaging in piracy are the availability of media contents without any price (58.5%), and the convenience to avail recent products (31.6%). In many studies, it is found that online piracy among young people is an acquired habit even in the presence of their sense of moral obligation [107, 108, 109]. In [107], authors emphasized that a person acquires the habit of piracy depending on the risk of punishment that he estimates by direct or indirect encounters. Ramayah *et al.* [106], found out that people have a strong tendency to recommend or share a friend any illegal download information. It also depicts that social contacts influence the habit of online piracy. Lee *et al.* [110] had shown that people with prominent presence in social media sites have a strong tendency to share music at online sites.

Even when major file sharing sites are going down regularly, the habit of online piracy and adaptation to new ways of sharing illegal contents are not diminishing [97, 111]. Though several studies have been conducted to understand the economic and behavioral aspects of the online piracy, very little is investigated about the spreading of this habit in a population. To the best of our knowledge, no one has investigated the spread of the habit of online piracy in the light of social contacts.

In this work, we argue that someone who does not know about online piracy or someone who does not know about a particular way to download pirated files is introduced to specific piracy techniques by their friends and social contacts. However, because of the substantial adverse effects on industries and individuals, there are always a group

of aware people who want to reduce the habit of piracy using word-of-mouth in their respective social circles.

5.2   Proposed Model with Word-of-Mouth Awareness

In our proposed model, T is the total population which is categorized in three compartments: U, B, and A. Class U is called unaware as they are unaware of the digital piracy techniques. Class B is called bootleggers[1] or illegal downloaders. They are the people who themselves are habituated to piracy and also encourage others to get involved in such activities. The effective rate at which bootleggers succeed in transferring this habit to an unaware person is denoted by ρ. Class A is the population who are well-aware of the fact that piracy is an anti-social activity and they spread this awareness among their contacts. The parameter β signifies the rate at which an aware node is able to change the mentality of bootleggers and bring them to aware class. There are scenarios when people from aware class themselves cannot resist the temptation to download certain pirated contents. This transformation from aware to bootleggers have been included in the model by adding a linear transition from class A to B at the rate λ. As mentioned in earlier chapters, μ is the birth and death rate of the individual node. Both rates have been kept same to maintain a fix population size. All these transitions are shown in Fig. 5.1 with appropriate rate parameters.

---

[1]Bootlegger is someone who makes, copies, distributes or sells something illegally. The term was first coined in Omaha, Nebraska in 1889 when liquor was illegally supplied by concealing bottles in high boots of the bootlegger. More recently, the term has broadened to include a whole range of other illegal or pirated goods.
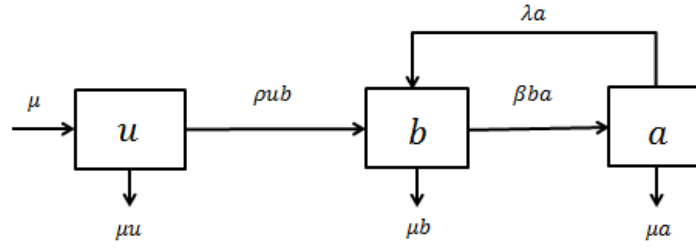
Figure 5.1: Block diagram of the proposed model for the propagation of habit of on-line piracy when only word-of-mouth awareness prevails in society.

5.3   Homogeneous Modeling

In Fig. 5.1, the fractions of unaware, bootlegger and aware classes are denoted by u, b, and a respectively. The rate of change of u, b, and a are given by the set of ordinary coupled differential equation mentioned below:

$$
\begin{aligned}
u' &= \mu - \rho u b - \mu u, \\
b' &= \rho u b - \beta b a + \lambda a - \mu b, \\
a' &= \beta b a - \lambda a - \mu a.
\end{aligned}
\tag{5.1}
$$

5.3.1   Equilibrium Analysis

At steady state, rate of change of u, b and a will be 0. Equating third equation of the system model described in eq. 5.1 to zero, it can be observed that $a^\star = 0$, is an obvious solution. Substituting $a^\star = 0$ in remaining two equations of the eq. set 5.1, we get $E_0(1, 0, 0)$ as one equilibrium point where triplets of E_0 are in order (u, b, a). Equating first equation of the eq. set 5.1 to zero, we get

$$
u^\star = \frac{\mu}{\mu + \rho b^\star}.
\tag{5.2}
$$

Substituting u from eq. 5.2 and using $a^\star = (1 - u^\star - b^\star)$ in second equation of the eq. set 5.1, we get a quadratic equation in $b^\star$ i.e., $p(b^\star)^2 + qb^\star + r = 0$, where

$$p = \beta\rho \; ; \quad q = -\beta(\rho - \mu) + \rho(\mu + \lambda) \; ; \quad r = (\mu + \lambda)(\rho - \mu) \cdot \tag{5.3}$$

Roots of the quadratic equation are $b_1^\star = \frac{\rho - \mu}{\rho}$ and $b_2^\star = \frac{\mu + \lambda}{\beta}$. By relevant substitutions, we get

$$u_1^\star = \frac{\mu}{\rho}; \; a_1^\star = 0 \quad \text{and} \quad u_2^\star = \frac{\mu\beta}{\mu\beta + \rho(\mu + \lambda)}; \; a_2^\star = \frac{(\mu + \lambda)((\rho - \mu)\beta - (\mu + \lambda)\rho)}{\beta(\mu\beta + \rho(\mu + \lambda))}.$$

Thus, we have two more possible steady states: $E_1^\star(u_1^\star, b_1^\star, a_1^\star)$ and $E_2^\star(u_2^\star, b_2^\star, a_2^\star)$. As fraction of bootleggers are non-zero in both of these equilibrium states, they are referred to as endemic equilibria. A contrasting difference between these endemic equilibria is the steady state value of aware class. In $E_2^\star$, there remains a fraction of aware people in the steady state, but in $E_1^\star$, aware class depletes completely from the population.

Examining the coefficients of the quadratic eq. 5.3, we can infer that system of equation can have two different set of feasible steady states depending on sign of r, which in turn depends on whether $\frac{\rho}{\mu}(= \mathcal{R})$ is smaller or greater than 1. $\mathcal{R}$ is reproduction number of the model.

Case 1 : When $\mathcal{R} < 1$, steady state values $b_1^\star$ and $a_2^\star$ are negative. Therefore, neither of the endemic steady states $E_1^\star$ and $E_2^\star$ are physically realisable in this region. Irrespective of the initial condition i.e., initial fractional population of all the three compartments, system eventually reaches to only possible steady state $E_0(1, 0, 0)$.

Case 2: When $\mathcal{R} > 1$, both $b_1^\star$ and $b_2^\star$ are positive. Moreover, $b_1^\star$ will surely be less than 1 as numerator $(\rho - \mu)$ is less than the denominator ρ. Hence, $E_1^\star$ is always a possible steady state if $\mathcal{R} > 1$ but, feasibility of $E_2^\star$, will depend on whether $a_2^\star$ is positive or not. Value of $a_2^\star$ will be positive only if $(\rho - \mu)\beta - (\mu + \lambda)\rho > 0$ i.e., $\beta > \frac{(\mu + \lambda)\rho}{(\rho - \mu)}$.

We are denoting right hand side of the inequality by β_{th}.

Subcase 2.1: For $\beta < \beta_{th}$, $a_2^\star$ will be negative and only $E_1^\star$ will be the feasible steady state.

Subcase 2.2: For $\beta > \beta_{th}$, $a_2^\star$ will be positive. Rearranging the condition of $\beta > \beta_{th}$, we get $\frac{\rho}{\mu} > \frac{\beta}{\beta - \lambda - \mu}$. Denoting right hand side of the inequality by $\mathcal{R}_c$, the condition is $\mathcal{R} > \mathcal{R}_c$. We can conclude that for $\beta > \beta_{th}$, $E_2^\star$ will also be a feasible steady state, along with $E_1^\star$.

5.3.2 Bifurcation

Bifurcation diagrams have been plotted in Fig. 5.2 for two different values of β by varying the parameter ρ. It can be observed from the figures that in the range $\mathcal{R} < 1$ system leads to piracy-free steady state in both the cases. As discussed in case 2 of Sec. 5.3.1, for $\beta < \beta_{th}$, a single endemic steady state can be observed in the region $\mathcal{R} > 1$ in Fig. 5.2 (a). Similarly, we can observe two endemic steady states in Fig. 5.2 (b) for $\beta > \beta_{th}$ in the region $\mathcal{R} > \mathcal{R}_c$.

Stability of all these equilibrium points have been obtained by linear stability analysis as discussed in Sec. 2.1.3. Forward transcritical bifurcation can be observed at $\mathcal{R} = 1$ (in both the cases) and $\mathcal{R} = \mathcal{R}_c$ (in second case). At $\mathcal{R} = 1$, endemic free equilibrium E_0 looses its stability and an endemic equilibrium $E_1^\star$ appears. Similarly, at $\mathcal{R} = \mathcal{R}_c$, $E_1^\star$ becomes unstable and $E_2^\star$ emerges.

We get a a straight horizontal line beyond $\mathcal{R} = \mathcal{R}_c$ in Fig 5.2 (b), because the value of $b_2^\star$ is $\frac{\mu + \lambda}{\beta}$, which is independent of ρ, and our bifurcation diagram has been plotted by varying the parameter ρ.
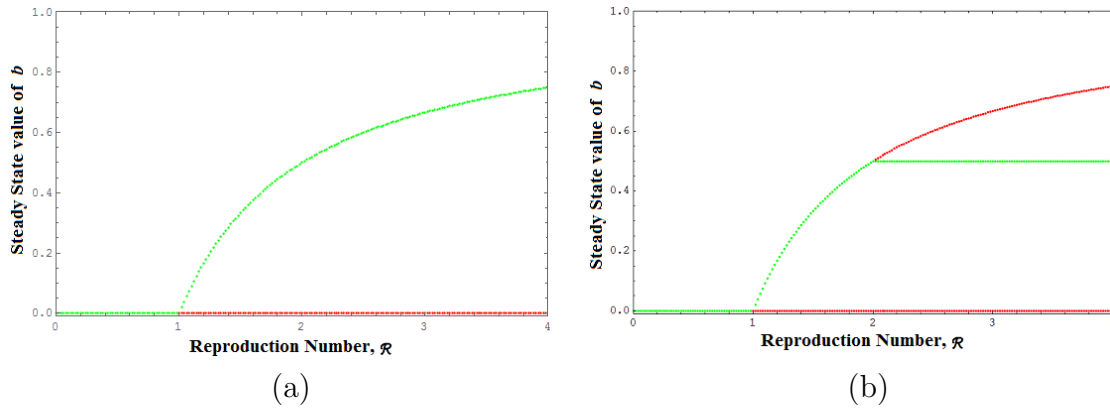
Figure 5.2: Variation in steady state fraction of b with reproduction number $\mathcal{R}$ for (a) $\mu = 0.05$, $\lambda = 0.2$, and $\beta = 0.15$, where only one type of endemic steady state persists beyond $\mathcal{R} = 1$;(b)$\mu = 0.05$, $\lambda = 0.2$, and $\beta = 0.5$, where two types of endemic steady state persists − one in the range $1 < \mathcal{R} < \mathcal{R}_c$ and another in $\mathcal{R} > \mathcal{R}_c$. In these figures, green lines indicate stable solutions and red lines indicate unstable solutions.

5.3.3   Effect of Word-of-Mouth

The impact of word-of-mouth awareness has been modeled by transition of people from bootlegger to aware class at the rate β. We have seen in Sec. 5.3.1 that system shows two different behaviors depending on whether β is less or more than $\beta_{th} = (\mu + \lambda)\frac{\rho}{\rho - \mu}$. In most of the real applications birth-death rate, u, is very small as compared to other rates. Considering this fact we can assume that $\frac{\rho}{\rho - \mu}$ will have value very close to 1, and β_{th} will be approximately $(\mu + \lambda)$.

Relating this mathematical findings to the physical phenomenon, λ and μ are the rates at which people are departing from aware class. The only way to generate aware class is by their interaction with bootleggers and making them aware. If this rate of conversion from B to A is slower than the overall rate of departure from class A, i.e., $(\beta < \lambda + \mu)$, then eventually the population of aware will decrease to zero. That is why the system attains the equilibrium point $E_1^\star$, where the fraction of aware is zero. If the overall awareness is significant i.e, value of β is large, then system can be brought to a steady

state where aware class will persist in the society.

5.4   Heterogeneous Modeling

To understand the diffusion of piracy habit in nonhomogeneous social structure, heterogeneous modeling has been investigated. In next sections, we will discuss the degree block approximation model of the diffusion process and epidemic threshold in heterogeneous case.

5.4.1   Degree Block Approximation

As used in earlier chapters, degree block approximation has been used to model the process over heterogeneous network structure. In this approach, nodes having equal degree are considered to be statistically equivalent assuming that nodes with similar degree are able to impact their neighbours to the same extent. An unaware node having degree k is in class U_k and fraction of such nodes is denoted by u_k. Similar notations like B_k, b_k, and A_k, a_k are used for bootleggers and aware nodes respectively.

In this approach, impact rate of a node of a particular class on another is assumed to be proportional to degree of the node. Nonlinear rates of homogeneous approach, ρ and β modifies to $\rho_n k$ and $\beta_n k$ respectively. As discussed in Sec. 3.3.1, ρ_n and β_n are chosen to be $\frac{\rho}{\langle k \rangle}$ and $\frac{\beta}{\langle k \rangle}$ respectively so that overall infection or awareness spread by a single person is same in homogeneous and heterogeneous approaches. Fraction of bootlegger or aware class around a node is given by the density function Θ_b and Θ_a respectively.

As discussed in Sec. 4.4, the density functions have the expression as follows:

$$
\begin{aligned}
\Theta_u &= \frac{\sum_{k'} k' p(k') u_{k'}}{\langle k \rangle}, \\
\Theta_b &= \frac{\sum_{k'} k' p(k') b_{k'}}{\langle k \rangle}, \\
\Theta_a &= \frac{\sum_{k'} k' p(k') a_{k'}}{\langle k \rangle}.
\end{aligned}
\tag{5.4}
$$

Including all these factors in eq. set 5.1, the system model is modified to

$$
\begin{aligned}
u'_k &= \mu - \rho_n k u_k \Theta_b - \mu u_k, \\
b'_k &= \rho_n k u_k \Theta_b - \beta_n k b_k \Theta_a + \lambda a_k - \mu b_k, \\
a'_k &= \beta_n k b_k \Theta_a - \lambda a_k - \mu a_k.
\end{aligned}
\tag{5.5}
$$

Multiplying by $\frac{k p_k}{\langle k \rangle}$ and summing over k, we get the expression for rate of change of density functions of all three classes as follows

$$
\begin{aligned}
\Theta'_u &= \mu - \rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \Theta_b - \mu \Theta_u, \\
\Theta'_b &= \rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \Theta_b - \beta_n \sum_k \frac{k^2 p_k}{\langle k \rangle} b_k \Theta_a + \lambda \Theta_a - \mu \Theta_b, \\
\Theta'_a &= \beta_n \sum_k \frac{k^2 p_k}{\langle k \rangle} b_k \Theta_a - \lambda \Theta_a - \mu \Theta_a.
\end{aligned}
\tag{5.6}
$$

5.4.2 Early Stage Analysis

Analysing the system from early stage when very few people are having the habit of piracy or very few people are aware of a particular piracy strategy, we can approximate u_k by 1, and b_k and a_k will be negligible. Using this approximation in nonlinear terms,

eq. set 5.6 simplifies to

$$
\begin{aligned}
\Theta'_u &= \mu - \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \Theta_b - \mu \Theta_u \\
\Theta'_b &= \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \Theta_b + \lambda \Theta_a - \mu \Theta_b \\
\Theta'_a &= -\lambda \Theta_a - \mu \Theta_a.
\end{aligned}
\tag{5.7}
$$

Initial values of Θ_u, Θ_b, and Θ_a are assumed to be u_0, b_0, and a_0 respectively. Solving third equation of the eq. set 5.7 with these initial conditions, we get

$$
\Theta_a = a_0 e^{\frac{t}{\tau_a}}
\tag{5.8}
$$

where $\tau_a = -\frac{1}{\lambda + \mu}$. Substituting Θ_a from eq. 5.8 in second equation of the eq. set 5.7, we get

$$
\Theta_b = \lambda a_0 \tau_a e^{\frac{t}{\tau_a}} + C_1 e^{\frac{t}{\tau_b}}
\tag{5.9}
$$

where

$$
C_1 = b_0 - \lambda a_0 \tau_a = b_0 + \frac{\lambda a_0}{\lambda + \mu}; \quad \text{and} \quad \tau_b = \frac{\langle k \rangle}{\rho_n \langle k^2 \rangle - \mu \langle k \rangle}.
$$

Substituting Θ_a and Θ_b in first equation of the eq. set 5.7, we get

$$
\Theta_u = \mu t - \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \left(C_1 \tau_b e^{\frac{t}{\tau_b}} + a_0 \lambda \tau_a^2 e^{\frac{t}{\tau_a}} \right) + C_2 e^{\frac{t}{\tau_u}}
\tag{5.10}
$$

where

$$
C_2 = u_0 + \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} (c_1 \tau_b + \lambda \tau_a^2 a_0) \quad \text{and} \quad \tau_u = -\frac{1}{\mu}.
$$

For the habit of piracy to spread, rate of change of Θ_b should be positive. Differentiating the expression of Θ_b (eq. 5.9), we get

$$
\Theta'_b = \lambda a_0 e^{\frac{t}{\tau_a}} + C_1 \frac{1}{\tau_b} e^{\frac{t}{\tau_b}}.
$$

Substituting values of C_1, τ_b and τ_a we get

$$\Theta_b' = \lambda a_0 e^{\frac{t}{\tau_a}} + (b_0 + \frac{\lambda a_0}{\lambda + \mu})(\frac{\rho_n\langle k^2\rangle - \mu\langle k\rangle}{\langle k\rangle})e^{\frac{t}{\tau_b}}. \tag{5.11}$$

If $\rho_n\langle k^2\rangle > \mu\langle k\rangle$ holds, every term in equation 5.11 will be positive which will ensure the increment of Θ_b in early stage. Replacing ρ_n by $\frac{\rho}{\langle k\rangle}$, we get the necessary condition for the habit of piracy to spread, i.e.,

$$\mathcal{R} = \frac{\rho}{\mu} > \frac{\langle k\rangle^2}{\langle k^2\rangle} \tag{5.12}$$

This inequality relates parameters of the model to parameters of network structure and gives epidemic threshold in terms of first and second moment of degree distribution.

5.4.3   Steady State Analysis

In steady state (after a large time) u_k, b_k and a_k will not vary with time. Thus equating eq. set 5.5 to zero, we have

$$
\begin{aligned}
\mu - \rho_n k u_k \Theta_b - \mu u_k &= 0, \\
\rho_n k u_k \Theta_b - \beta_n k b_k \Theta_a + \lambda a_k - \mu b_k &= 0, \\
\beta_n k b_k \Theta_a - \lambda a_k - \mu a_k &= 0\cdot
\end{aligned}
\tag{5.13}
$$

Solving these equations, we get

$$u_k = \frac{\mu}{\mu + \rho_n k \Theta_b}; \; a_k = \frac{\beta_n k b_k \Theta_a}{\lambda + \mu}. \tag{5.14}$$

Substituting these expression of u_k and a_k in $(u_k + b_k + a_k = 1)$, we get

$$b_k = (\frac{\rho_n k \Theta_b}{\rho_n k \Theta_b + \mu})(\frac{\lambda + \mu}{\lambda + \mu + \beta_n k \Theta_a}). \tag{5.15}$$

Multiplying the above equation by $\frac{kp_k}{\langle k \rangle}$ and summing over k, we have

$$\Theta_b = \sum_k \frac{kp_k(\lambda + \mu)(\rho_n k \Theta_b)}{\langle k \rangle(\lambda + \mu + \beta_n k \Theta_a)(\rho_n k \Theta_b + \mu)}. \tag{5.16}$$

This is a self consistency equation of the form $\Theta_b = f(\Theta_b, \Theta_a)$ having 0 as an obvious solution. At $\Theta_b = 1$

$$f(1, \Theta_a) = \sum_k \frac{kp_k(\lambda + \mu)(\rho_n k)}{\langle k \rangle(\lambda + \mu + \beta_n k \Theta_a)(\rho_n k + \mu)} \tag{5.17}$$

which is less than 1. As discussed in Sec. 4.4.3 and shown in Fig. 4.7, to have a solution of eq. 5.16 in the interval $\Theta_b = (0,1)$, slope of the function f must be greater than 1 at the point $(\Theta_b = 0, \Theta_a = 0)$.

$$\frac{\partial f(\Theta_a, \Theta_b)}{\partial \Theta_b} = \frac{(\lambda + \mu)\rho_n}{\langle k \rangle(\lambda + \mu + \beta_n k \Theta_a)} \sum_k \frac{\mu k^2 p_k}{(\mu + \rho_n k \Theta_b)^2}. \tag{5.18}$$

At the point $(\Theta_b = 0, \Theta_a = 0)$ slope is $\frac{\rho_n \langle k^2 \rangle}{\mu \langle k \rangle}$. Hence, the condition to have a non zero solution is

$$\frac{\rho_n}{\mu} > \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{5.19}$$

which can be further simplified to

$$\frac{\rho}{\mu} > \frac{\langle k \rangle^2}{\langle k^2 \rangle}. \tag{5.20}$$

This is the same condition, what we had in early stage analysis.

5.5   Numerical Results

All the analytical results claimed till now have been verified in this section by numerical results of the simulations. Results have been obtained for homogeneous as well as heterogeneous approach.

(a)

(b)
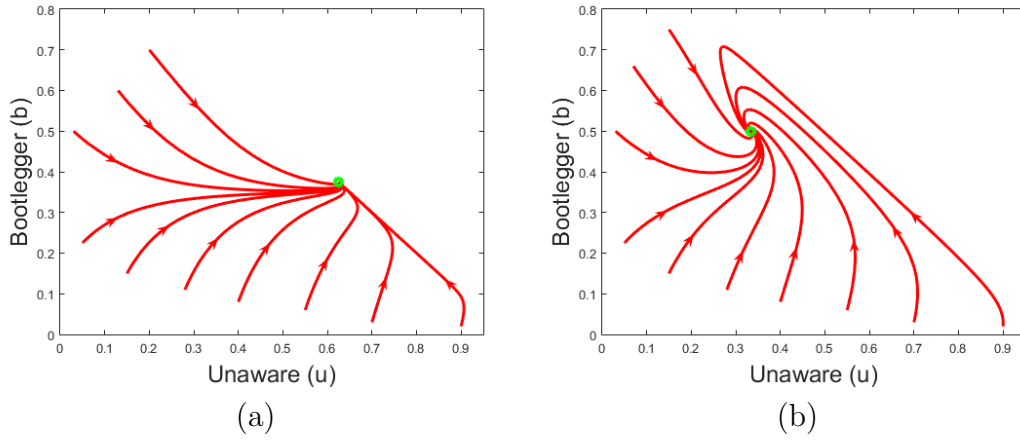
Figure 5.3: Temporal variation of u and b with different initial conditions, for homogeneous settings with $\mu = 0.05$, $\lambda = 0.2$, $\beta = 0.5$ with two cases $-$ (a) $\rho = 0.08$ satisfying the condition $1 < \mathcal{R} < \mathcal{R}_c$ and resulting into a steady state value $(0.625, 0.375, 0)$, having no aware individual in the population; (b) $\rho = 0.2$ satisfying the condition $\mathcal{R} > \mathcal{R}_c$ and resulting into a steady state value $(0.333, 0.5, 0.167)$, having a nonzero fraction of aware in the steady state. Value of $\mathcal{R}_c$ for both cases is 2. Values of $\mathcal{R}$ for first and second figures are 1.6 and 4 respectively.

5.5.1 Simulation of Homogeneous Model

As discussed in the Sec. 5.3.1, system may lead to piracy free steady state (or endemic steady state) if reproduction number $\mathcal{R}$ is less (or more) than 1. For endemic scenario, the nature of equilibrium point depends on value of β, which signifies the impact of word-of-mouth awareness. For $\beta < \beta_{th}$, aware class vanishes from the population, whereas for $\beta > \beta_{th}$, aware class will persist if $\mathcal{R}$ is greater than $\mathcal{R}_c$. In Fig. 5.3, we have considered the endemic scenario for $\beta > \beta_{th}$ and have shown the temporal variation of fraction of unaware (u), and fraction of bootlegger (b) for both the cases $-$ $\mathcal{R} < \mathcal{R}_c$ and $\mathcal{R} > \mathcal{R}_c$.
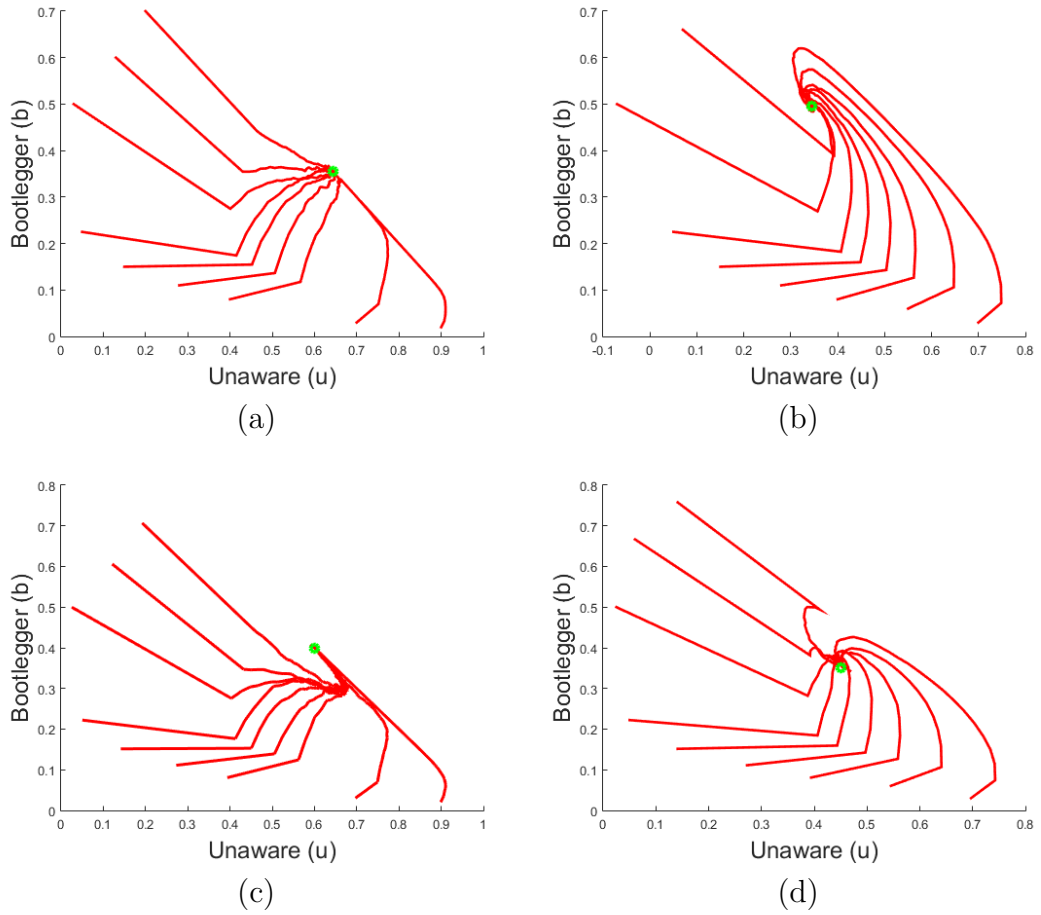
Figure 5.4: Temporal variation of u and b with different initial conditions, for random network with $\mu = 0.05$, $\lambda = 0.2$, $\beta = 0.5$, with two cases - (a) $\rho = 0.08$ satisfying the condition $1 < \mathcal{R} < \mathcal{R}_c$, and (b) $\rho = 0.2$ satisfying the condition $\mathcal{R} > \mathcal{R}_c$. Value of $\mathcal{R}_c$ is 2. Values of $\mathcal{R}$ for first and second case are 1.6 and 4 respectively. Both these results for Jazz network are plotted in (c) and (d) respectively.

5.5.2   Simulation over Model Networks

To understand the propagation of the piracy habit in a heterogeneous society, simulations have been carried over a random and a scale-free network. To compare the result with the homogeneous approach, we have plotted the temporal variation in the $u - b$ plane using the parameter set of Sec. 5.5.1, in Fig. 5.4(a)-(b). Similar to our findings in the previous chapter, the steady-state value in case of the random network is almost the same as in the homogeneous setting. Error in the steady-state fraction of different classes is bounded by 3 % for the considered parameter set in case of random network.

To analyze the impact of diffusion on nodes with different degrees in random network, we have plotted degree wise steady-state fraction of different classes in Fig. 5.5 (a). It can be observed from the figure that the probability of a node to be in unaware class is decreasing with the nodal degree. To understand the reason behind this observation, we have plotted the fraction of different classes in the neighborhood of a node in Fig. 5.5 (b). It is independent of the degree in case of a random network. It indicates that higher degree nodes are surrounded by more number of bootleggers. That is why they are more prone to infection. Similar pattern can be observed in case of scale-free networks, in Fig 5.5 (c) and (d). We can also observe that in case of the random network, the fraction of aware class in the neighborhood of a node with degree k is zero, which is matching with the results of homogeneous analysis observed in Fig. 5.3 (a). But, in case of the scale-free network, this fraction is not completely zero, especially for nodes with a higher nodal degree. It reminds us about the absence of epidemic threshold in case of a scale-free network, we have discussed in the Sec. 3.3.2. It is very difficult to completely remove any particular class from the population if the social contacts are following scale-free degree distribution.
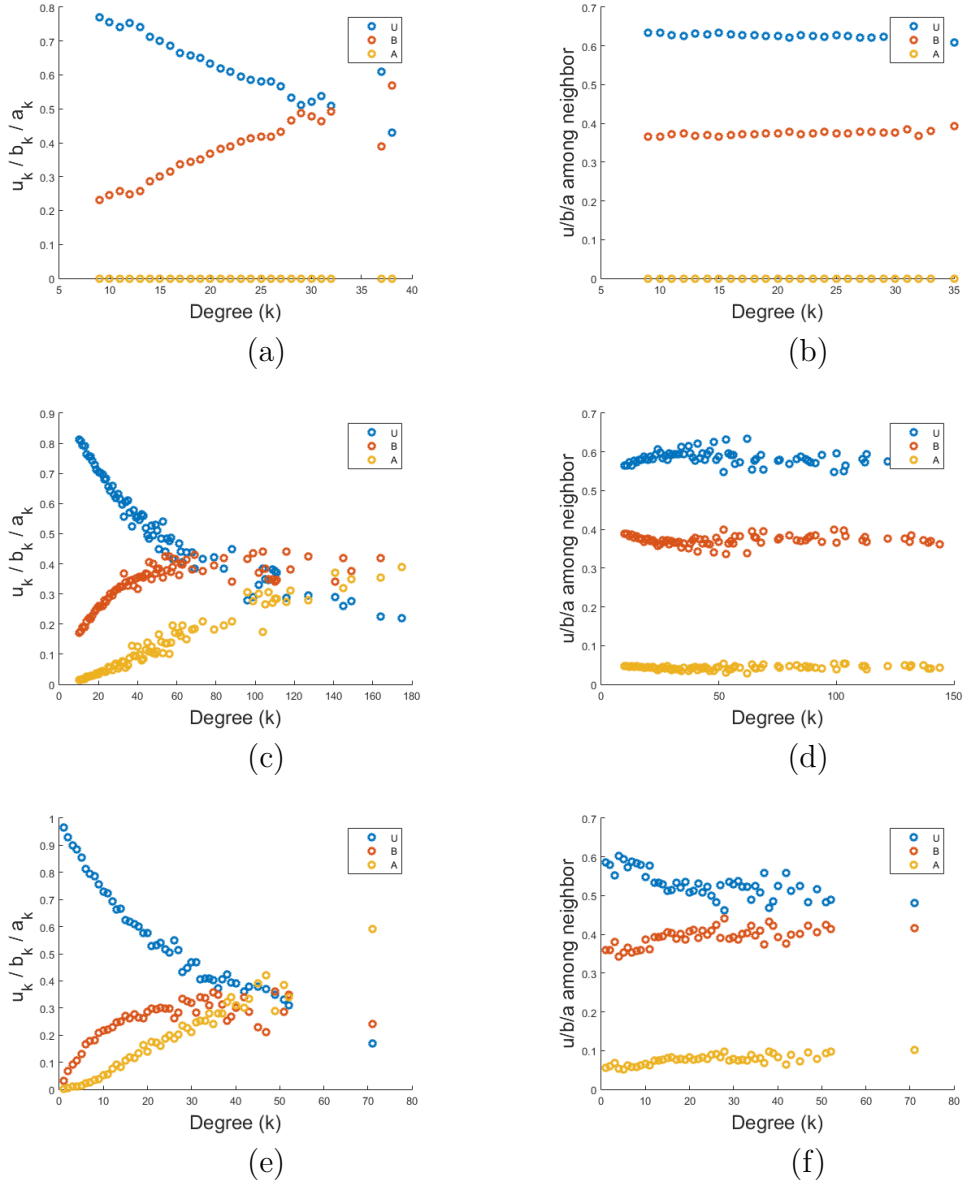
Figure 5.5: Degree-wise fraction of u, b and a with respect to degree k at steady-state in case of (a) Random network; (c) Scale-free network; and (e) Jazz network. Fraction of u, b and a in the neighborhood of a node with degree k in (b) Random network; (d) Scale-free network; and (f) Jazz network. Parameter values for all these cases are $\mu = 0.05$, $\lambda = 0.2$, $\beta = 0.5$, and $\rho = 0.08$.

Table 5.1: Comparison of steady state values of different classes for different networks in presence of word-of-mouth awareness when $\mathcal{R} < \mathcal{R}_c$.

Steady state fraction	Homogeneous Setting	Random Network	Scale-free Network	Hamster Network	Jazz Network	Email Network
$u^\star$	0.625	0.640	0.740	0.619	0.733	0.749
$b^\star$	0.375	0.360	0.210	0.381	0.228	0.216
$a^\star$	0	0	0.050	0	0.039	0.035

5.5.3   Simulation over Real Networks

In this section, we have shown the results of the diffusion dynamics on real networks. Same parameter set, used in the analysis of model networks, has been used to plot the temporal variation in $u - b$ plane in Fig. 5.4 (c)-(d), for Jazz network. Although the system finally reaches to a steady state closer to the prediction of homogeneous model, considerable difference can be observed in the trajectory of temporal evolution in real network and homogeneous setting shown in Figs. 5.4 (c)-(d) and 5.3 (a)-(b) respectively. Steady state fraction of all three classes in case of different networks has been shown in Tables 5.1 and 5.2, for $\mathcal{R} < \mathcal{R}_c$ and $\mathcal{R} > \mathcal{R}_c$ respectively. Degree-wise steady state fraction of different classes and fraction of different class in neighborhood of a degree k node, in case of Jazz network has been shown in Figs. 5.5 (e) and (f) respectively. These degree specific results again show that in any real network, impact of the diffusion process is more on higher degree nodes. These nodes have very high probability to undergo a transition from one class to another. Once, they undergo a behaviorial change, they can propagate this transition to a large number of nodes connected to them. Hence, they are crucial in the diffusion process.

As we have seen in the previous chapters, we also observe here that the nature of the steady-state values in homogeneous and heterogeneous populations closely resemble each other. Moreover, the rate parameters have equivalent effects on the diffusion in

Table 5.2: Comparison of steady state values of different classes for different networks in presence of word-of-mouth awareness when $\mathcal{R} > \mathcal{R}_c$.

Steady state fraction	Homogeneous Setting	Random Network	Scale-free Network	Hamster Network	Jazz Network	Email Network
$u^{\star}$	0.333	0.345	0.480	0.426	0.565	0.525
$b^{\star}$	0.500	0.485	0.340	0.441	0.271	0.319
$a^{\star}$	0.167	0.170	0.180	0.133	0.164	0.156

both the cases. The epidemic thresholds that control the final steady-state values of the endemic classes are also analogous in homogeneous and heterogeneous mixing. Even though word-of-mouth awareness model succeeds to control the outbreak of piracy in a population, it is still difficult to eradicate the habit only using one-to-one awareness. In practice, we also see many mass campaign programs to generate social awareness to reduce piracy. This social enlightenment cannot be modeled using conventional one-to-one awareness model. Rather, it requires a more complex model of the diffusion process with more number of nonlinearities.

5.6 Proposed Model with Mass Media Awareness

In the last section, we analyzed the effect of word-of-mouth awareness in spreading of online piracy habit. Though this approach has some effect on the diffusion dynamics, it is certainly not very effective to minimize piracy in adverse conditions. In reality, other means like law enforcement, advertisement, media campaign etc. can be also used to reduce the severity of piracy. It can be observed that spreading awareness through various means can significantly reduce the habit of piracy in a population. Several researchers [108, 109, 112] have shown that the fear of punishment is one of the prime factors that controls the habit of piracy.

In a recent survey [113] of more than 25000 people, 48% people informed that they would

stop or watch less number of pirated contents after knowing the damages incurred by piracy, and the probable punishments related to it. In [103], authors mentioned that though a person associated with piracy can perceive momentary enjoyment and sense of benefit, ethical efficacy and fear dominates one's tendency to engage in piracy. According to Moore and Chang [114], this habit is guided by moral psychology which in turn is controlled by externally received information. Also in [115], authors pointed out legal and ethical concerns of an individual as one of the leading reasons that control the habit of piracy. Both the aspects can be evoked using external influences like mass campaigns or recent cases of active law enforcement [97]. The study of Cronan and Sulaiman[116] also aligns with this finding. The study of Gupta *et al.* [117] had shown that mere knowledge of the fact that piracy may lead to penalty does not often control the habit of piracy. Thus, a word-of-mouth awareness might not be the optimal strategy to control the habit of piracy. Rather several external factors control the attitude of an individuals habit [103], and the diffusion of the habit of piracy should be analyzed in the presence of active efforts from different sources to reduce piracy, along with the word-of-mouth awareness process.

In the next section, we consider the external efforts other than word-of-mouth that are made to reduce the severity of online piracy. Instead of distinguishing different parameters like media campaigns, advertisements, punishments, law enforcement, drives to boost ethical and moral values etc., we combine them in a single time varying parameter, and for sake of simplicity, we will call it as 'effect of media'.
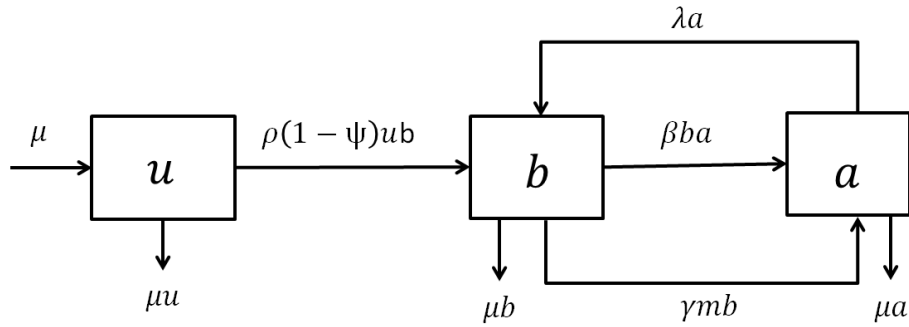
Figure 5.6: Block diagram of the proposed model for the propagation of habit of on-line piracy in presence of mass media awareness.

5.7 Homogeneous Analysis

Schematic diagram containing all transitions between different classes have been shown in the Fig. 5.6. Cumulative density of awareness programs driven by the media is denoted by m. Rate of change of media is given by the following differential equation

$$m' = \phi b - \phi_0 (m - m_0).$$
(5.21)

As clear from this equation, extent of media program implementation is assumed to be increasing linearly with proportion of bootleggers in the population at rate ϕ. Depletion rate of these programs due to ineffectiveness of social barriers in the population is ϕ_0. Awareness level of the society before implementation of any awareness program is denoted by m_0. It shall be noted that m is always higher than m_0. When $m = m_0$, then $m' = \phi b$.

Effect of media awareness on bootleggers has been incorporated in the model by introducing a new transition from bootlegger to aware at the rate γm where γ is the success rate at which a person moves from Class B to class A. In presence of external awareness programme, conversion rate from unaware to bootleggers also decreases. This

reduction has been incorporated by multiplying ρ by a factor $(1 - \psi)$ where ψ is $\frac{m}{c+m}$. The positive constant c limits the effect of awareness programs on unawares and known as half saturation point for Holling type-II functional response [118]. Other parameters are similar to the word-of-mouth awareness model as discussed in Sec. 5.2. Coupled differential equations for the model are as following:

$$
\begin{aligned}
u' &= \mu - \rho(1 - \psi)ub - \mu u, \\
b' &= \rho(1 - \psi)ub - \beta ba + \lambda a - \gamma mb - \mu b, \\
a' &= \beta ba - \lambda a + \gamma mb - \mu a, \\
m' &= \phi b - \phi_0(m - m_0)\cdot
\end{aligned}
\tag{5.22}
$$

5.7.1 Equilibrium Analysis

After attaining the steady state, fraction of all the classes will remain static and rate of change will be zero. Solving the eq. set 5.22, after equating every equation to zero, we get

$$
m^\star = m_0 + zb^\star; \quad u^\star = \frac{(c + m_0 + zb^\star)\mu}{(c + m_0 + zb^\star)\mu + \rho cb^\star}; \quad a^\star = \frac{(m_0 + zb^\star)\gamma b}{\lambda + \mu - \beta b^\star}
\tag{5.23}
$$

where $z = \frac{\phi}{\phi_0}\cdot$ Substituting the expression of $u^\star$ and $a^\star$ in $(u^\star + b^\star + a^\star = 1)$ we get a cubic equation in $b^\star$ i.e., $p(b^\star)^3 + q(b^\star)^2 + rb^\star = 0$ where

$$
\begin{aligned}
p &= (\mu z + \rho c)(\gamma z - \beta), \\
q &= (\mu z + \rho c)(\lambda + \mu + m_0 \gamma) + \mu(c + m_0)(\gamma z - \beta) + \rho c\beta, \\
r &= \mu(c + m_0)(\lambda + \mu + m_0 \gamma) - \rho c(\lambda + \mu)\cdot
\end{aligned}
\tag{5.24}
$$

It can be observed from the equation that $b^\star = 0$ is always a solution. Corresponding values of $u^\star$ and $a^\star$ are 1 and 0. Hence, $E_0(1, 0, 0)$ is always a steady state solution of

the system. To Investigate other roots, we consider $b^\star \neq 0$ and are left with quadratic equation $p(b^\star)^2 + qb^\star + r = 0$, where on the basis of sign of the coefficient p there are two different situations:

Case 1: $p > 0$ implies $\gamma z > \beta$,which in turn also implies $q > 0$. There will be one (or no) positive solution depending on whether r is negative (or positive). Hence, required condition to have a positive solution of the quadratic equation is

$$\rho c(\lambda + \mu) > \mu(c + m_0)(m_0\gamma + \lambda + \mu) \cdot$$

It gives the expression of reproduction number $\mathcal{R}_m$. Here, m signifies that mass media awareness has been included in the model. A single endemic steady state exists only when

$$\mathcal{R}_m = \frac{\rho c(\lambda + \mu)}{\mu(c + m_0)(\lambda + \mu + m_0\gamma)} > 1 \cdot \tag{5.25}$$

otherwise, only a piracy free steady state E_0 exists.

Case 2: When $p < 0$ i.e, $\gamma z < \beta$. The case can be further divided in two sub-cases depending on sign of r.

Sub-case 2.1: When $r > 0$ ($\mathcal{R}_m < 1$), similar to Case 1, one of the roots is positive and other is negative.

Sub-case 2.2: When $r < 0$ ($\mathcal{R}_m > 1$), q will surely be positive. It can be observed after a bit of rearrangement in expression of q and r given in eq. set 5.24.

$$q = (\mu z + \rho c)(\lambda + \mu + m_0\gamma) + \mu(c + m_0)\gamma z - \beta(\mu c + \mu m_0 - \rho c),$$

$$r = \mu(c + m_0)(m_0\gamma) + (\lambda + \mu)(\mu c + \mu m_0 - \rho c) \cdot$$

Term responsible for negative sign of r i.e., $(\mu c + \mu m_0 - \rho c)$ is appearing in expression of q with a negative sign resulting in positive q. With $p < 0$, $q > 0$ and $r < 0$ it is

ensured that both roots of the quadratic equation will be positive.

Quadratic equation in $b^\star$ tells that there will be one or two positive solutions when $(\mathcal{R}_m < 1)$ or $(\mathcal{R}_m > 1)$, but it does not guarantee that all these solutions will be physical (i.e., $0 \le b^\star \le 1$). To investigate it further, along with the quadratic equation in $b^\star$, we have also analyzed a quadratic equation in $a^\star$. At steady state, the rate of change of all three classes will be zero. Equating third equation of the eq. set 5.22 to zero, we get

$$\beta ba - \lambda a + \gamma mb - \mu a = 0 \cdot$$

At steady state, value of b, a, and m will be $b^\star$, $a^\star$, and $m^\star$ respectively. Replacing $b^\star$ by $(1 - u^\star - a^\star)$ and substituting $m^\star$ and $a^\star$ from eq. 5.23, we get a quadratic equation $p_a(a^\star)^2 + q_a a^\star + r_a = 0$ where

$$
\begin{aligned}
p_a &= (\gamma z - \beta), \\
q_a &= (\beta - 2z\gamma)(1 - u^\star) - \gamma m_0 - (\mu + \lambda), \\
r_a &= \gamma m_0(1 - u^\star) + \gamma z - \gamma z u(2 - u^\star) \cdot
\end{aligned}
\tag{5.26}
$$

As $\frac{dr_a}{du^\star} < 0$ for $0 < u^\star < 1$, hence the constant term r_a is a strictly decreasing function of $u^\star$ in the interval [0,1] with values $\gamma(m_0 + z)$ and 0 at corresponding end points. For a physical solution, $u^\star$ can only vary from 0 to 1 and in this whole range r_a is positive. Opposite signs of p_a and r_a ensure that if roots of the above quadratic equation are real then, only one root will be positive.

On the basis of above discussion, we conclude that there exists only one physical endemic steady state beyond $\mathcal{R}_m > 1$.

For $\mathcal{R}_m < 1$, examining the roots of quadratic equations in $b^\star$ and $a^\star$ (eq. 5.24 and 5.27), it can be observed that one of the roots is positive and another is negative

in both the cases. For any physical solution $u^\star$, $b^\star$, and $a^\star$ must be in the range $[0,1]$ individually. For $0 \leq u^\star < 1$, quadratic equation in $b^\star$ and $a^\star$ gives two roots $(b_1^\star, b_2^\star)$ and $(a_1^\star, a_2^\star)$ respectively which finally form steady state triplets $(u_1^\star, b_1^\star, a_1^\star)$ and $(u_2^\star, b_2^\star, a_2^\star)$. For $\mathcal{R}_m < 1$, if $b_1^\star$ is positive then $b_2^\star$ is negative or vice versa. As discussed earlier, the statement also holds true for $a_1^\star$ and $a_2^\star$ in this range. To check the feasibility of the triplets, let us assume that $(u_1^\star, b_1^\star, a_1^\star)$ is physical, i.e., $u_1^\star$, $b_1^\star$, and $a_1^\star$ are in the range 0 to 1 individually, and $u_1^\star + b_1^\star + a_1^\star = 1$. However, if our assumption is true, then it implies that $b_2^\star$ and $a_2^\star$ will surely be negative. But, for $0 \leq u_2^\star < 1$, along with negative values of $b_2^\star$ and $a_2^\star$, the required condition of $u_2^\star + b_2^\star + a_2^\star = 1$ will never hold. Hence, our initial assumption that both $b_1^\star$ and $a_1^\star$ are positive, is not right. Actually, for positive $b_1^\star$, $a_1^\star$ will be negative and for negative $b_1^\star$, $a_1^\star$ will be positive. Hence, we conclude that for $\mathcal{R}_m < 1$, neither of the endemic state is physical and only piracy free steady state prevails.

We conclude that scenario is same whether $\gamma z < \beta$ or $\gamma z > \beta$. For $\mathcal{R}_m < 1$ there exist only piracy free steady state and for $\mathcal{R}_m > 1$ there exist a unique endemic steady state.

5.7.1.1   Sensitivity

Value of $\mathcal{R}_m$ depends on ρ, c, λ, m_0, μ and γ. As discussed in Sec. 3.2.2, sensitivity of $\mathcal{R}_m$ for any parameter x is defined as

$$\Gamma_x^{\mathcal{R}_m} = \frac{x}{\mathcal{R}_m} \cdot \frac{\partial \mathcal{R}_m}{\partial x} \tag{5.27}$$

Using this definition, sensitivities of $\mathcal{R}_m$ for all six parameters are as follows:

$$
\begin{aligned}
\Gamma_{\rho}^{\mathcal{R}_m} &= 1, \\
\Gamma_{c}^{\mathcal{R}_m} &= \frac{m_0}{c + m_0}, \\
\Gamma_{\lambda}^{\mathcal{R}_m} &= \frac{\lambda}{\lambda + \mu} \frac{m_0 \gamma}{\lambda + \mu + m_0 \gamma}, \\
\Gamma_{m_0}^{\mathcal{R}_m} &= -(\frac{m_0 \gamma}{\lambda + \mu + m_0 \gamma} + \frac{m_0}{c + m_0}), \\
\Gamma_{\mu}^{\mathcal{R}_m} &= -(\frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu + m_o \gamma}), \\
\Gamma_{\gamma}^{\mathcal{R}_m} &= -\frac{m_0 \gamma}{\lambda + \mu + m_0 \gamma}.
\end{aligned}
$$

5.7.2 Bifurcation

Bifurcation diagram for the system has been shown in Fig. 5.7. Forward transcritical bifurcation occurs at $\mathcal{R}_m = 1$, where endemic free equilibrium looses its stability and a new endemic steady state appears. Main difference from the word-of-mouth awareness model discussed in Sec. 5.6, is existence of a single endemic steady state for complete range of reproduction number beyond $\mathcal{R}_m = 1$, instead of two different steady states before and after $\mathcal{R}_c$. The endemic steady state where fraction of aware class was becoming zero (in the range $\mathcal{R} = [1, \mathcal{R}_c]$ in case of word-of-mouth awareness model) does not exist any more. Another important difference is population of class B in the steady state. With considered parameter values, the steady state value of b, after applying media becomes near by 5%, which is very small as compared to the model without media.

Bifurcation diagram has been drawn for two different values of β and it can be observed that steady state fraction of bootleggers dereases with increase in β.
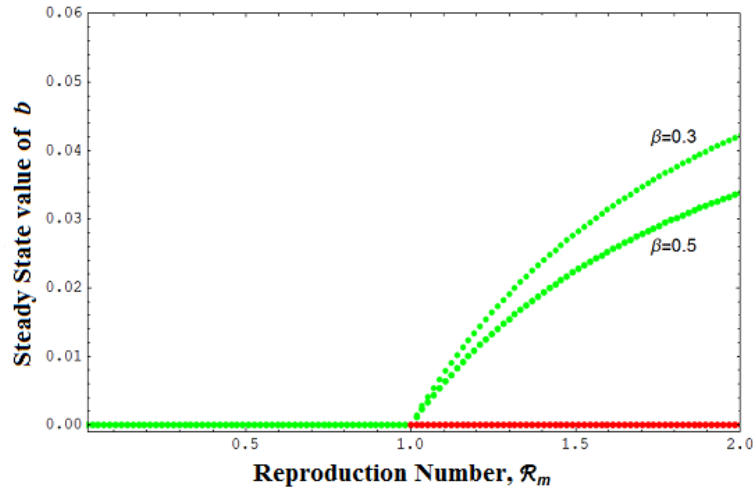
Figure 5.7: Variation in steady state fraction of b with reproduction number $\mathcal{R}_m$ for two different values of β i.e., 0.3 and 0.5. Other parameter values are $\mu = 0.05$, $\lambda = 0.2$, $\gamma = 0.08$, $\phi = 0.05$, $\phi_0 = 0.01$, $c = 5$ and $m_0 = 4$. In the figure, green lines indicate stable solutions and red lines indicate unstable solutions.

5.7.3   Effect of Mass Media

As discussed in the previous section, system behavior depends on value of $\mathcal{R}_m$, which in turn depends on other parameters of the model. To understand the effect of media on the process, we have investigated the individual contribution of parameters c and m_0 which controls the effect of media campaign.

5.7.3.1   Effect of c

The parameter c controls the value of ψ which in turn modifies the rate of conversion from class U to class B in presence of mass media awareness. For piracy habit to extinct from the population, required condition is $\mathcal{R}_m < 1$ i.e.,

$$\mathcal{R}_m = \frac{\rho c(\lambda + \mu)}{\mu(c + m_0)(\lambda + \mu + m_0\gamma)} < 1.$$

(a)   (b)
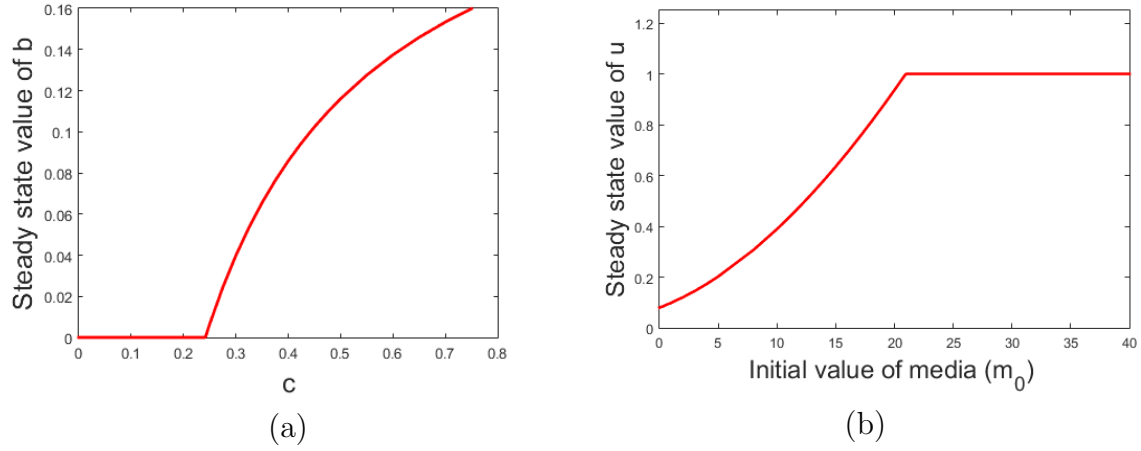
Figure 5.8: Variation in steady state fraction of (a) bootleggers with constant c which in turn changes the value of ψ and decides the rate of conversion from class U to B and (b) unaware class with different initial level of intrinsic social awareness m_0. Parameters having same value in both the cases are $\mu = 0.05$, $\rho = 2$, $\lambda = 0.2$, $\beta = 0.3$, $\gamma = 0.08$, $\phi = 0.5$, and $\phi_0 = 0.1$. In (a)$m_0 = 4$ and in (b) c=5. Threshold c_{th} in (a) is 0.24178 and $m_{0_{th}}$ in (b) is 20.955.

Rearranging the above equation, we get the threshold condition on the value of c as

$$c < \frac{m_o \mu}{(\rho - \mu) - \left(\frac{\rho m_0 \gamma}{\lambda + \mu + m_0 \gamma}\right)}.$$

Right hand side of the inequality is denoted by c_{th}. As shown in Fig. 5.8 (a), for $c < c_{th}$ value of b is 0 , which means that complete population is in endemic free state and the problem of piracy does not exist. For c greater than c_{th}, non-zero value of b can be observed, signifying the presence of people indulging the habit of piracy.

5.7.3.2   Effect of m_0

Analyzing the effect of intrinsic level of social awareness m_0, we get the following quadratic expression from the threshold condition $\mathcal{R}_m < 1$

$$\mu \gamma m_0^2 + \mu(\lambda + \mu + c\gamma)m_0 + c(\lambda + \mu)(\mu - \rho) > 0. \tag{5.28}$$

For $(\mu - \rho) > 0$, both roots of m_0 will be negative. It means for any positive value of m_0, the required condition for extinction of piracy will hold. For $(\mu - \rho) < 0$, one of the roots will be positive and that will be the threshold value of m_0, denoted by $m_{0_{th}}$. If value of intrinsic social awareness is more than this positive root of the quadratic equation, the society will be free from piracy. In Fig. 5.8 (b), we can observe that beyond this threshold value, fraction of unaware becomes 1 indicating the absence of piracy.

5.8   Heterogeneous Analysis

Though the homogeneous model predicts the steady state of the system, it does not give any information about the dynamics of a particular node. In this section, we will analyze the diffusion of piracy habit in heterogeneous population incorporating its network structure in the differential equation model. We will also see how nodes of different degree behave in the diffusion process.

5.8.1   Degree Block Approximation

In terms of density functions, system of equations can be written as

$$
\begin{aligned}
u_k' &= \mu - \rho_n k(1 - \psi)u_k\Theta_b - \mu u_k, \\
b_k' &= \rho_n k(1 - \psi)u_k\Theta_b - \beta_n k b_k\Theta_a + \lambda a_k - \gamma m b_k - \mu b_k, \\
a_k' &= \beta_n k b_k\Theta_a - \lambda a_k + \gamma m b_k - \mu a_k, \\
m' &= \phi b - \phi_0(m - m_0) \cdot
\end{aligned}
\tag{5.29}
$$

In heterogeneous setting, u, b, and a has been modified to u_k, b_k, and a_k. But, the impact of mass media has been considered same for all nodes irrespective of their degrees. That's why in expression of m', we have used b, not b_k. Here, b signifies the fraction of bootlegger in entire population. Multiplying every equation of the eq. set 5.29 by $\frac{kp_k}{\langle k \rangle}$ and summing over k, we get

$$
\begin{aligned}
\Theta'_u &= \mu - \rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \frac{c}{c+m} \Theta_b - \mu \Theta_u, \\
\Theta'_b &= \rho_n \sum_k \frac{k^2 p_k}{\langle k \rangle} u_k \frac{c}{c+m} \Theta_b - \beta_n \sum_k \frac{k^2 p_k}{\langle k \rangle} b_k \Theta_a + \lambda \Theta_a - \gamma m \Theta_b - \mu \Theta_b, \quad (5.30) \\
\Theta'_a &= \beta_n \sum_k \frac{k^2 p_k}{\langle k \rangle} b_k \Theta_a - \lambda \Theta_a + \gamma m \Theta_b - \mu \Theta_a.
\end{aligned}
$$

5.8.2 Early Stage Analysis

Similar to early stage analysis in word-of-mouth awareness model, here also we are approximating u_k by 1 and b_k as well as a_k is considered to be negligible. Using this approximation to linearize the nonlinear terms, eq. set 5.30 is simplified to

$$
\begin{aligned}
\Theta'_u &= \mu - \rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \frac{c}{c+m} \Theta_b - \mu \Theta_u, \\
\Theta'_b &= \left(\rho_n \frac{\langle k^2 \rangle}{\langle k \rangle} \frac{c}{c+m} - (\gamma m + \mu) \right) \Theta_b + \lambda \Theta_a, \quad (5.31) \\
\Theta'_a &= -(\lambda + \mu) \Theta_a + \gamma m \Theta_b.
\end{aligned}
$$

Last two equation of the eq. set 5.31 forms a system of simultaneous linear differential equations with constant coefficients.

$$
\begin{aligned}
\Theta'_b &= C_1 \Theta_b + C_2 \Theta_a \\
\Theta'_a &= C_3 \Theta_a + C_4 \Theta_b. \quad (5.32)
\end{aligned}
$$

The equations are exactly the same, what we have got while analyzing the extended VM model in Sec. 4.8.2. We will use the final result from the VM analysis that necessary condition for initial growth in class B is $C_1 C_3 < C_2 C_4$. Substituting the expression of all these constant terms, the condition modifies to

$$\frac{\rho_n}{\mu}\left(\frac{c}{c+m}\right)\left(\frac{\lambda+\mu}{\lambda+\mu+\gamma m}\right) > \frac{\langle k \rangle}{\langle k^2 \rangle}.$$

Replacing ρ_n by $\frac{\rho}{\langle k \rangle}$, we get

$$\frac{\rho}{\mu}\left(\frac{c}{c+m}\right)\left(\frac{\lambda+\mu}{\lambda+\mu+\gamma m}\right) = \mathcal{R}_m > \frac{\langle k \rangle^2}{\langle k^2 \rangle}.$$

Right hand side of the inequality is *epidemic threshold* in terms of network parameter. First and second moment of degree distribution will depend on the network structure.

5.8.3 Steady State Analysis

In steady state, rate of change of u_k, b_k, a_k and m will be zero. Equating last equation of the eq. set 5.29 to zero we get $m = m_0 + zb$ where $z = \frac{\phi}{\phi_0}$. Using this expression of m in eq. set 5.29, we get

$$u_k = \frac{\mu(c+m_0+bz)}{\mu(c+m_0+bz)+\rho_n kc\Theta_b}; \quad a_k = \frac{\beta_n k\Theta_a + \gamma(m_0+bz)}{\lambda+\mu}b_k.$$

Substituting these expressions of u_k and a_k in $(u_k + b_k + a_k = 1)$, we get

$$b_k = \left(\frac{\lambda+\mu}{\lambda+\mu+\beta_n k\Theta_a+\gamma(m_0+bz)}\right)\left(\frac{\rho_n kc\Theta_b}{\rho_n kc\Theta_b+\mu(c+m_0+bz)}\right).$$

Multiplying above equation by $\frac{kp_k}{\langle k \rangle}$ and summing over k we get

$$\Theta_b = \sum_k \frac{kp_k(\lambda+\mu)\rho_n kc\Theta_b}{\langle k \rangle(\lambda+\mu+\beta_n k\Theta_a+\gamma(m_0+bz))(\rho_n kc\Theta_b+\mu(c+m_0+bz))}. \tag{5.33}$$

This is a self consistency equation of the form $\Theta_b = f(\Theta_b, \Theta_a)$ having 0 as an obvious solution. At $\Theta_b = 1$, which also implies $b = 1$

$$f(1, \Theta_a) = \sum_k \frac{kp_k(\lambda+\mu)(\rho_n kc)}{\langle k \rangle(\lambda+\mu+\beta_n k\Theta_a+\gamma(m_0+z))(\rho_n kc+\mu(c+m_0+z))}. \tag{5.34}$$

Observing the numerator and denominator of the eq. 5.34, we can see that for every term in the numerator, there is a corresponding larger term in the denominator. Hence, $f(1, \Theta_a)$ will surely be less than 1. To have a solution of eq. 5.33 in the interval $\Theta_b = (0, 1)$, slope of the function f must be greater than 1 at the point $(\Theta_b = 0, \Theta_a = 0)$. Slope of the function is

$$\frac{\partial f(\Theta_a, \Theta_b)}{\partial \Theta_b} = \frac{(\lambda + \mu)\rho_n c}{\langle k \rangle (\lambda + \mu + \beta_n k \Theta_a + \gamma(m_0 + bz))} \sum_k \frac{\mu(c + m_0 + bz)k^2 p_k}{(\mu(c + m_0 + bz) + \rho_n kc\Theta_b)^2}.$$

At point $(\Theta_b = 0, \Theta_a = 0)$, value of the slope is

$$\frac{(\lambda + \mu)\rho_n c \langle k^2 \rangle}{(\lambda + \mu + \gamma m_0)\mu(c + m_0)\langle k \rangle}.$$

Hence, the required condition to have a desired solution for the eq. 5.33 is

$$\frac{\rho_n(\lambda + \mu)c}{\mu(\lambda + \mu + \gamma m_0)(c + m_0)} > \frac{\langle k \rangle}{\langle k^2 \rangle}.$$

Substituting ρ_n by $\frac{\rho}{\langle k \rangle}$, the condition is equivalent to

$$R_m > \frac{\langle k \rangle^2}{\langle k^2 \rangle}$$

which is the same condition we have obtained in the Sec. 5.8.2.

5.9   Numerical Results

In this section, we are going to discuss the results of spread of piracy habit after applying mass media awareness in society. We will compare the results of homogeneous and heterogeneous analysis. In heterogeneous part, along with model networks, results over real networks will also be discussed.

5.9.1   Simulation of Homogeneous Model

In Sec. 5.7.2, we found that after mass media awareness, people with habit of piracy will exist only beyond $\mathcal{R}_m = 1$. Comparing the reproduction number of word-of-mouth awareness, $\mathcal{R}$, with reproduction number of mass media awareness model, $\mathcal{R}_m$, it can be observed that both multiplying factors in $\mathcal{R}_m - \frac{\lambda+\mu}{\lambda+\mu+\gamma m_0}$ and $\frac{c}{c+m_0}$ are less than 1. Hence, effective implementation of mass media awareness program makes it difficult to cross the epidemic threshold and enter the endemic region.

Temporal evolution in $u - b$ plane for different initial points have been shown in the Fig. 5.9 (a), for $\mathcal{R}_m > 1$. We can observe that at steady state, fraction of bootleggers in the population is very less (around 5% in present case) as compared to its value in case of word-of-mouth awareness.

5.9.2   Simulation over Model Networks

Just like homogeneous case, we are able to notice the significant decrement in steady state fraction of bootleggers in case of model networks also. Temporal evolution for random and scale-free network has been shown in Figs. 5.9 (a) and (b) respectively. Steady state value in case of random network almost matches with homogeneous scenario. Error in the steady-state fraction of different classes is bounded by 2 % for the considered parameter set in case of random network. For scale-free network the error is comparatively large attributed to the heterogeneity in its structure.

To understand the diffusion process from perspective of nodes with different degrees, we have plotted the degree wise steady state fraction of different classes in Figs. 5.10 (a) and (c) for Random and Scale-free network respectively. Fraction of different classes
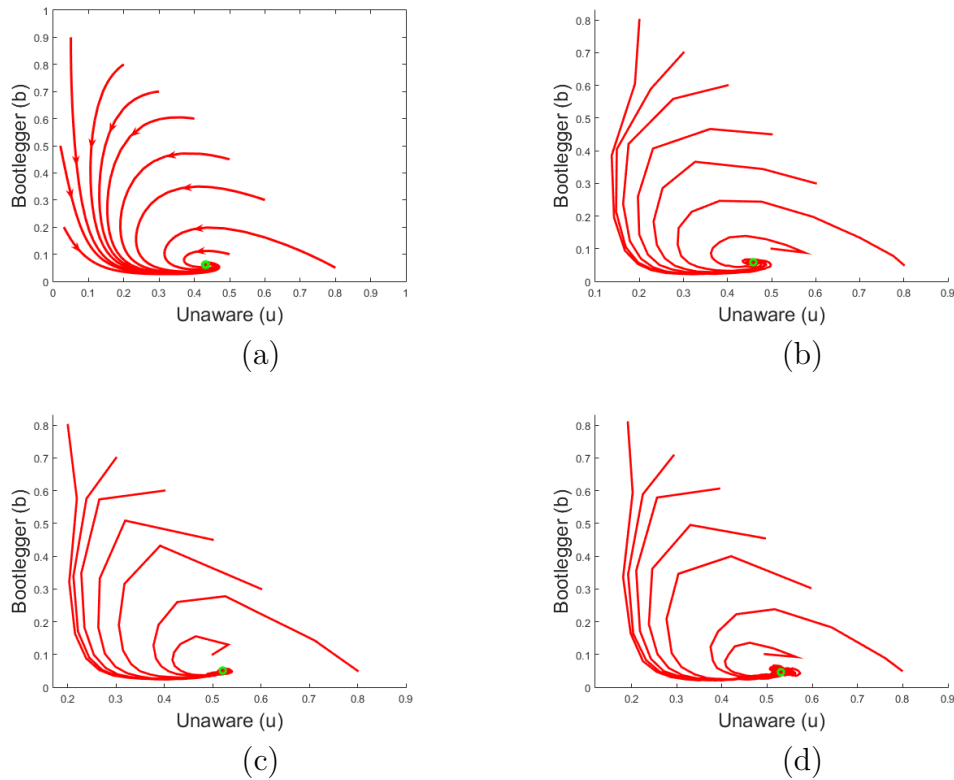
Figure 5.9: Temporal variation of u and b in case of endemic steady state with different initial conditions for parameter set $\mu = 0.05$, $\rho = 2$, $\lambda = 0.01$, $\beta = 0.3$, $\gamma = 0.08$, $m_0 = 4$, $c = 5$, $\phi = 0.5$ and $\phi_0 = 0.1$ in case of (a) homogeneous setting; (b) Random network; (c) Scale-free network; and (d) Jazz network. Value of $\mathcal{R}_m$ for considered parameter set is 3.5, which is greater than 1.
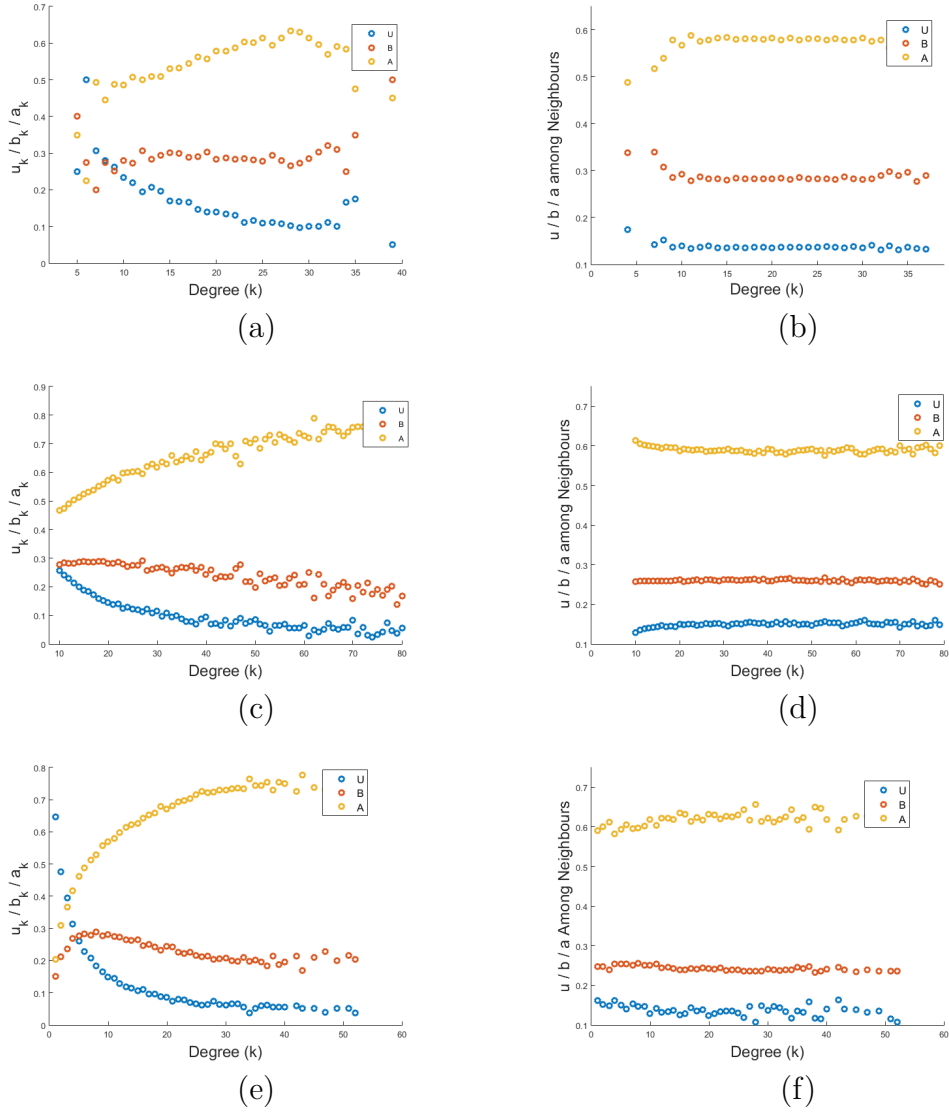
Figure 5.10: Degree wise fraction of u, b and i with respect to degree k at steady-state in case of (a) Random network; (c) Scale-free network; and (e) Jazz network. Fraction of u, b and i in the neighborhood of a node with degree k in (b) Random network; (d) Scale-free network; and (f) Jazz network.

Table 5.3: Comparison of steady state values of different classes for different networks in presence of mass media awareness when $\mathcal{R} > \mathcal{R}_m$.

Steady state fraction	Homogeneous Setting	Random Network	Scale-free Network	Hamster Network	Jazz Network	Email Network
$u^\star$	0.432	0.458	0.528	0.535	0.597	0.636
$b^\star$	0.061	0.059	0.051	0.050	0.042	0.038
$a^\star$	0.507	0.483	0.402	0.414	0.361	0.325

(a)   (b)
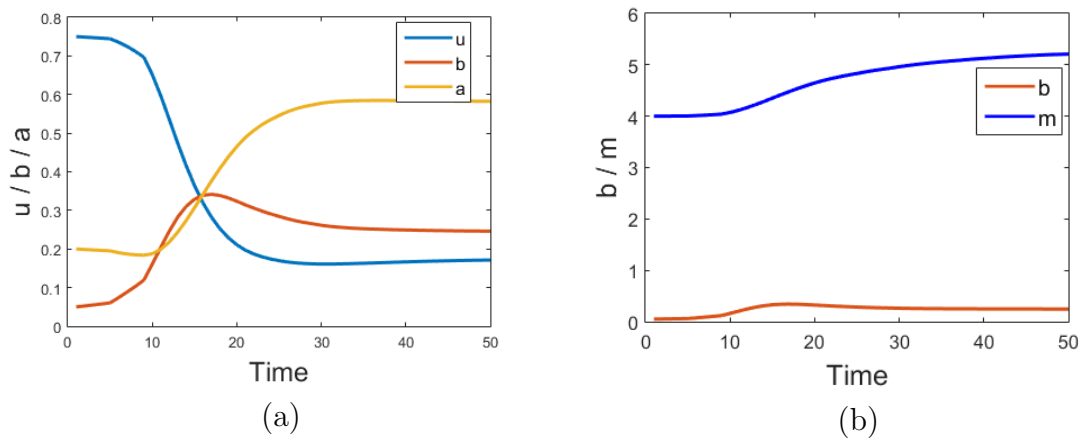
Figure 5.11: (a) Temporal evolution of u, b, and a in presence of mass media awareness with an initial condition (0.75,0.05,0.2) (b) Similar Variation of media level and population of bootleggers with time. Parameter set for both the plots are $\mu = 0.05$, $\rho = 2$, $\lambda = 0.2$, $\beta = 0.3$, $\gamma = 0.08$, $m_0 = 4$, $c = 5$, $\phi = 0.5$, and $\phi_0 = 0.1$.

in neighborhood of a node with a particular degree has been plotted in Fig. 5.10 (b) and (d) for random and scale-free network respectively. Results follow the same pattern what we have observed in case of word-of-mouth awareness, with a difference that in presence of mass media awareness, steady state fraction of aware class is very large and fraction of bootleggers is very less as compared to word-of-mouth awareness.

5.9.3   Simulation over Real Networks

All the results shown for model networks have been obtained for a real network also. Steady state values of different classes under same parameter set, for all three real networks (Hamster, Jazz, and Email) has been listed in Table 5.3. Temporal evolution in $u - b$ plane, for endemic scenario has been shown in Fig. 5.9 (d) for Jazz network. Degree wise steady state fraction of all three classes has been shown in Fig. 5.10 (e). Fraction of different classes in neighborhood of nodes with particular degree has been shown in Fig. 5.10 (f). All the results obtained for Jazz network also confirm the crucial role played by higher degree node in the diffusion process.

We have also shown the time evolution of u, b, and a in case of jazz network in Fig 5.11 (a) for a particular initial condition. In Fig. 5.11 (b), we have shown the variation of media awareness level and the variation of bootleggers population in a single plot to emphasize that for success of media awareness campaign, the level of media must be adjusted proportionally with population of bootleggers in the society.

5.10   Summary

Piracy is a burning problem for any digital industry. Though several strategic counter-measures are evolving, digital piracy is growing day by day. Even legal measures are not very effective to stop the spreading of the habit of online piracy globally. In this chapter, we explore the effect of word-of-mouth awareness programs and mass media awareness campaigns. Though we have explored the effect on the habit of online media piracy, the approach can be adapted to model any adversarial habit or addiction spreading in society. While analyzing word-of-mouth awareness, we observe that this technique can resist the outbreak of the addiction. However, in severe conditions, only

word-of-mouth awareness strategy may not be very useful to restrict the spreading of the habit in the population. In both homogeneous and heterogeneous analyses, we observe similar kind of parameters thresholds that control the existence of endemic states. Comparing both the models, we observe that the similarity between the homogeneous approach and the heterogeneous approach do not break even after introducing more number of nonlinear interactions. Obviously, the differences in the steady-state values of homogeneous and heterogeneous populations change depending on the diffusion dynamics and network architectures. Moreover, we observe that for particular diffusion parameters, the nature of the spreading and the steady state conditions closely resemble in both the approaches. Our study also shows that along with word-of-mouth awareness, the presence of mass media awareness program is much more effective to restrict the spread of the habit. The study also reveals that in a society that initially has a basic morality or awareness and law enforcement, the habit of piracy is relatively easier to eradicate.

Chapter 6

Conclusion and Future Scope

In entire thesis, we have analyzed the process of diffusion observed in different phenomena. Diffusion has been widely studied by the homogeneous approach based on differential equations since eighteenth century. Recent development in network science has helped us to understand the process using simulation based heterogeneous approach. This approach includes the view point of network structure while analyzing the diffusion process. Although a significant amount of work has recently been done by heterogeneous approach also, but a comparison of both of them is still missing. In our work, we have tried to bridge this gap and by highlighting the similarities, dissimilarities and appropriateness of both the approaches. Starting from the spread of rumors in society, we have studied the online propagation of viral marketing campaigns and the propagation of the habit of online piracy.

6.1 Conclusion

We discussed the fundamental concepts required to analyze homogeneous and heterogeneous diffusion models in chapter 2. We also briefly explained the different termi-

nologies associated with the heterogeneous network structures.

In chapter 3, we introduced a diffusion phenomenon using both homogeneous and heterogeneous approach. We also exhibited that for a simple case, the steady state values for diffusion process are quite similar in both the approaches, however, it still remains unclear that what would be the major similarities and dissimilarities among both these approaches.

In chapter 4, we proposed a new model for Viral marketing which also included a backward transition. We concluded that by increasing the relapse parameter beyond a particular threshold, a region of bistability can be observed. The bistable region helps to sustain the online advertisement even in adverse conditions when reproduction number is less than one. Apart from homogeneous settings, we were able to observe the bistability in heterogeneous setting also, over a real network. The model was further enhanced by considering the presence of rigidly inert people who are not interested in taking part in the referral at any cost. We noticed that if considerable fraction of people are rigid, in such a scenario, only way to sustain the campaign is to increase the overall relapse from inert to broadcaster class. The model also suggests that special care of rigidly inert people should be taken while designing the referral campaign.

In chapter 5, habit of online piracy has been modeled as an epidemic. We analyzed the effectiveness of word-of-mouth awareness and media awareness to control the habit of piracy in a society. We observe that though word-of-mouth awareness has effectiveness to control the problem with limited scope, presence of media campaigns that are proportional to the severity of the problem can play a crucial role to eliminate the problem even in adverse conditions. Presence of media campaigns helps to maintain the number of aware people in the population, and directly restricts the spread of the

problem. We observe similar behavior of the population in both homogeneous and heterogeneous analysis.

With a broad goal to compare the homogeneous and heterogeneous approaches in diffusion modeling, we proposed different ways to model behavioral habits of individuals that spread in a society due to their social contacts. In each case, we found the key parameters and the limiting thresholds that control the nature of the diffusion. Comparing the homogeneous and heterogeneous diffusion processes, it can be concluded that both the approaches have close resemblances with enough diversities. Execution of homogeneous diffusion process is computationally fast and completely mathematically tractable. Even a minute difference from two parameter settings can be well-observed in homogeneous models. Heterogeneous models, being a better representation of the actual social structures, depict the diffusion dynamics in more realistic way. However, a straightforward analysis of complex heterogeneous models is not always possible, and we had to rely on certain assumptions to simplify the problem. However, independent simulations for ideal networks and real networks show that the assumptions do not change the conclusion that we had drawn from the heterogeneous analyses drastically. Moreover, we also observed that though the diffusion processes in homogeneous structures and heterogeneous structures are not equivalent with time, the steady state conditions are quite similar in both the cases. Even some critical behaviors, like bistability, bifurcation etc., that can be observed in homogeneous analysis can also be observed in the diffusion process of heterogeneous structures. However, due to presence of structural characteristics, the observability of the phenomena might either be difficult or can have some fluctuations as compared to from their homogeneous counter parts.

6.2   Future Scope

In our entire thesis, we not only attempt to model some physical phenomena in the light of epidemiological diffusion, but also draw a connection between homogeneous analysis and heterogeneous analysis of such diffusion processes. In heterogeneous approach, we consider all our networks to be uncorrelated during our derivations for the sake of simplicity. However, in real scenarios, this assumption may often fail. For example, in social networks, people tend to interact with people having similar degree whereas in biological networks, many a times, high degree nodes tend to attach to low degree nodes. In future, the diffusion process for the correlated network structures can be analyzed for more practical settings.

In this thesis we have overlooked some of the network structures that are observed in the real world e.g., community network, directed network, adaptive network etc. These network structures hold some typical network properties, and any diffusion process over these types of networks might exhibit atypical behaviors. Thus, it will be interesting to observe the diffusion spread over these kinds of networks in future.

All rate parameters in our current work are constant, but these contact rates may change with time. Generally, in the beginning, people are more interested and excited to spread any rumor or campaign, but their enthusiasms eventually decrease with time. This decrement or change in the level of participation of an individual can be represented by time varying rate parameters. In future, it will be more realistic to model these phenomena with time varying rate parameters to gain more insight in the events under study.

Bibliography

[1] Black death: Pandemic, Medieval europe. *https://www.britannica.com/event/Black-Death*, 2018.

[2] Stephen G Wheatcroft. Famine and epidemic crises in Russia, 1918-1922: The case of saratov. In *Annales de démographie historique*, pages 329–352. JSTOR, Editions Belin, Paris, 1983.

[3] West african Ebola virus epidemic. *https://en.wikipedia.org/wiki/West_African_Ebola_virus_epidemic*, 2018.

[4] Global health observatory (GHO) data on HIV/AIDS. *http://www.who.int/gho/hiv/en*, 2018.

[5] Poliomyelitis: World Health Organization. *http://www.who.int/news-room/fact-sheets/detail/poliomyelitis*, 2018.

[6] Klaus Dietz and JAP Heesterbeek. Daniel bernoulli's epidemiological model revisited. *Mathematical biosciences*, 180(1-2):1–21, 2002.

[7] William Heaton Hamer. *The Milroy lectures on epidemic disease in England: The evidence of variability and of persistency of type*. Bedford Press, London, 1906.

[8] Ronald Ross. Some quantitative studies in epidemiology. *Nature*, 87:466–467, 1911.

[9] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics, III- Further studies of the problem of endemicity. *Bulletin of mathematical biology*, 53(1-2):89–118, 1991.

[10] William O Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics, II. —The problem of endemicity. *Proc. R. Soc. Lond. A*, 138(834):55–83, 1932.

[11] Lila R Elveback, John P Fox, Eugene Ackerman, Alice Langworthy, Mary Boyd, and Lael Gatewood. An influmza simulation model for immunization studies. *American journal of epidemiology*, 103(2):152–165, 1976.

[12] Herbert W Hethcote and James A Yorke. *Lecture notes in biomathematics.* Springer, New York, 1984.

[13] Hui-Ming Wei, Xue-Zhi Li, and Maia Martcheva. An epidemic model of a vector-borne disease with direct transmission and time delay. *Journal of Mathematical Analysis and Applications*, 342(2):895–908, 2008.

[14] Herbert W Hethcote and P Van Den Driessche. Two SIS epidemiologic models with delays. *Journal of mathematical biology*, 40(1):3–26, 2000.

[15] James M Hyman and E Ann Stanley. Using mathematical models to understand the AIDS epidemic. *Mathematical Biosciences*, 90(1-2):415–473, 1988.

[16] David JD Earn, Pejman Rohani, and Bryan T Grenfell. Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1390):7–10, 1998.

[17] QJA Khan and David Greenhalgh. Hopf bifurcation in epidemic models with a time delay in vaccination. *Mathematical Medicine and Biology: A Journal of the IMA*, 16(2):113–142, 1999.

[18] Dan Li, Shengqiang Liu, and Jing'an Cui. Threshold dynamics and ergodicity of an SIRS epidemic model with markovian switching. *Journal of Differential Equations*, 263(12):8873–8915, 2017.

[19] Mohammad Reza Parsaei, Reza Javidan, Narges Shayegh Kargar, and Hassan Saberi Nik. On the global stability of an epidemic model of computer viruses. *Theory in Biosciences*, 136(3-4):169–178, 2017.

[20] Rahil Sachak-Patwa, Nabil T Fadai, and Robert A Van Gorder. Understanding viral video dynamics through an epidemic modelling approach. *Physica A: Statistical Mechanics and its Applications*, 502:416–435, 2018.

[21] MohammadSidarth Trisal. Most expensive computer viruses of all time. `https: // cyware. com/ news` , 2016.

[22] Subhankar Ghosh and SS Kumar. Video popularity distribution and propagation in social networks. *Int. J. Emerg. Trends Technol. Comput. Sci.(IJETTCS)*, 6(1):001–005, 2017.

[23] Terry Ann Knopf. *Rumors, race and riots*. Taylor & Francis, New York, 2017.

[24] Alison Saldanha, Pranav Rajput, and Jay Hazare. 24 persons killed in mob attacks in 2018. `https: // www. firstpost. com/ india` , 2018.

[25] Helena Sofia Rodrigues and Manuel José Fonseca. Can information be spread as a virus? Viral marketing as epidemiological model. *Mathematical Methods in the Applied Sciences*, 39(16):4780–4786, 2016.

[26] Roy M Anderson and RM May. Infectious diseases of humans, vol. 1. *Oxford University Press, England*, 176:188, 1991.

[27] Stephen Eubank, VS Anil Kumar, and Madhav Marathe. Epidemiology and wireless communication: Tight analogy or loose metaphor? In *Bio-inspired computing and communication*, pages 91–104. Springer, 2008.

[28] Alan M Frieze and Geoffrey R Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10(1):57–77, 1985.

[29] Leon Danon, Ashley P Ford, Thomas House, Chris P Jewell, Matt J Keeling, Gareth O Roberts, Joshua V Ross, and Matthew C Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.

[30] Matthew J Silk, Darren P Croft, Richard J Delahay, David J Hodgson, Nicola Weber, Mike Boots, and Robbie A McDonald. The application of statistical network models in disease research. *Methods in Ecology and Evolution*, 8(9):1026–1041, 2017.

[31] Ken TD Eames and Jonathan M Read. Networks in epidemiology. In *Bio-inspired computing and communication*, pages 79–90. Springer, 2008.

[32] S Waqar Jaffry and Jan Treur. Agent-based and population-based simulation: A comparative case study for epidemics. In *Proceedings of the 22nd European Conference on Modelling and Simulation*, pages 123–130. Citeseer, 2008.

[33] Shannon Gallagher. Comparing compartment and agent-based models. *Joint Statistical Meeting, Baltimore*, 2017.

[34] Grazziela P Figueredo, Peer-Olaf Siebers, Markus R Owen, Jenna Reps, and Uwe Aickelin. Comparing stochastic differential equations and agent-based modelling and simulation for early-stage cancer. *PloS one*, 9(4):e95150, 2014.

[35] Johan Andre Peter Heesterbeek. A brief history of R_0 and a recipe for its calculation. *Acta biotheoretica*, 50(3):189–204, 2002.

[36] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016.

[37] Paul Erdos and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[38] Albert-László Barabási. Scale-free networks: A decade and beyond. *Science*, 325(5939):412–413, 2009.

[39] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: The topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4):69–77, 2000.

[40] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1-2):173–187, 1999.

[41] Jérôme Kunegis. Konect: The koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350. ACM, 2013.

[42] Albert-László Barabási. Network Science. *Philosophical Transactions of the Royal Society A*, 371(1987):20120375, 2013.

[43] David Dubois and Derek Rucker. How to stop rumors before they ruin your brand. *https://www.forbes.com/sites/onmarketing/2011/09/16/*, 2011.

[44] William Goffman and VA Newill. Generalization of epidemic theory. *Nature*, 204(4955):225–228, 1964.

[45] Anatol Rapoport and Lionel I Rebhun. On the mathematical theory of rumor spread. *The bulletin of mathematical biophysics*, 14(4):375–383, 1952.

[46] K Thompson, R Castro Estrada, D Daugherty, A Cintron-Arias, et al. A deterministic approach to the spread of rumors. *Cornell's digital repository*, 2003.

[47] Daryl J Daley and David G Kendall. Epidemics and rumours. *Nature*, 204(4963):1118, 1964.

[48] Kazuki Kawachi. Deterministic models for rumor transmission. *Nonlinear analysis: Real world applications*, 9(5):1989–2028, 2008.

[49] Laijun Zhao, Wanlin Xie, H Oliver Gao, Xiaoyan Qiu, Xiaoli Wang, and Shuhai Zhang. A rumor spreading model with variable forgetting rate. *Physica A: Statistical Mechanics and its Applications*, 392(23):6146–6154, 2013.

[50] Bernhard Haeupler. Simple, fast and deterministic gossip and rumor spreading. *Journal of the ACM (JACM)*, 62(6):47, 2015.

[51] Zaobo He, Zhipeng Cai, and Xiaoming Wang. Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In *Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference on*, pages 205–214. IEEE, 2015.

[52] Rudra M Tripathy, Amitabha Bagchi, and Sameep Mehta. Towards combating rumors in social networks: Models and metrics. *Intelligent Data Analysis*, 17(1):149–175, 2013.

[53] Laijun Zhao, Jiajia Wang, Yucheng Chen, Qin Wang, Jingjing Cheng, and Hongxin Cui. Sihr rumor spreading model in social networks. *Physica A: Statistical Mechanics and its Applications*, 391(7):2444–2453, 2012.

[54] Zaobo He, Zhipeng Cai, Jiguo Yu, Xiaoming Wang, Yunchuan Sun, and Yingshu Li. Cost-efficient strategies for restraining rumor spreading in mobile social networks. *IEEE Transactions on Vehicular Technology*, 66(3):2789–2800, 2017.

[55] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[56] World internet users and 2018 population stats. *https://www. internetworldstats. com/ stats. html*, 2018.

[57] John O'Connor, Eamonn Galvin, and Martin J Evans. *Electronic marketing: Theory and practice for the twenty-first century*. Pearson Education, 2004.

[58] Danilo Cruz and Chris Fill. Evaluating viral marketing: Isolating the key criteria. *Marketing Intelligence & Planning*, 26(7):743–758, 2008.

[59] Ralph F Wilson. The six simple principles of viral marketing. *Web Marketing Today*, 70(1):232, 2000.

[60] Lada A Adamic and Eytan Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.

[61] Richard Hanna, Andrew Rohm, and Victoria L Crittenden. We're all connected: The power of the social media ecosystem. *Business horizons*, 54(3):265–273, 2011.

[62] Clay Shirky. The political power of social media: Technology, the public sphere, and political change. *Foreign affairs*, pages 28–41, 2011.

[63] Homero Gil de Zúñiga, Nakwon Jung, and Sebastián Valenzuela. Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication*, 17(3):319–336, 2012.

[64] Renee Dye. Buzz on Buzz. *Harvard business review*, page 139, 2000.

[65] Rafi Mohammed, Robert J Fisher, Bernard J Jaworski, and Gordon Paddison. *Internet marketing: Building advantage in a networked economy*. McGraw-Hill, Inc., 2003.

[66] Paul Chaney. word of mouth still most trusted resource, says nielsen, implications for social commerce. `https://digitalwellbeing.org`, 2012.

[67] David Godes and Dina Mayzlin. Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science*, 28(4):721–739, 2009.

[68] Vera Blazevic, Wafa Hammedi, Ina Garnefeld, Roland T Rust, Timothy Keiningham, Tor W Andreassen, Naveen Donthu, and Walter Carl. Beyond traditional word-of-mouth: An expanded model of customer-driven influence. *Journal of Service Management*, 24(3):294–313, 2013.

[69] Gangseog Ryu and Lawrence Feick. A penny for your thoughts: Referral reward programs and referral likelihood. *Journal of Marketing*, 71(1):84–94, 2007.

[70] Jonah Berger and Katherine L Milkman. What makes online content viral? *Journal of marketing research*, 49(2):192–205, 2012.

[71] Andrea C Wojnicki and David Godes. Signaling success: Word of mouth as self-enhancement. *Customer Needs and Solutions*, 4(4):68–82, 2017.

[72] Kyongsei Sohn, John T Gardner, and Jerald L Weaver. Viral marketing: More than a buzzword. *Journal of Applied Business and Economics*, 14(1):21–42, 2013.

[73] Mauro Bampo, Michael T Ewing, Dineli R Mather, David Stewart, and Mark Wallace. The effects of the social structure of digital networks on viral marketing performance. *Information systems research*, 19(3):273–290, 2008.

[74] Ernest S Shtatland and T Shtatland. Early detection of epidemic outbreaks and financial bubbles using autoregressive models with structural changes. *Proceedings of the NESUG*, 21, 2008.

[75] Hazhir Rahmandad and John Sterman. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014, 2008.

[76] David P Fan and R Dennis Cook. A differential equation model for predicting public opinions and behaviors from persuasive information: Application to the index of consumer sentiment. *Journal of Mathematical Sociology*, 27(1):29–51, 2003.

[77] Jingzhou Liu, Yifa Tang, and ZR Yang. The spread of disease with birth and death on networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(08):P08008, 2004.

[78] Ilker Tunc and Leah B Shaw. Effects of community structure on epidemic spread in an adaptive network. *Physical Review E*, 90(2):022801, 2014.

[79] Peter Drucker. The society of organizations. *Harvard business review*, 95104, 1992.

[80] Kit Smith. 41 incredible instagram statistics. *https://www.brandwatch.com/blog/instagram-stats*, 2018.

[81] Instagram Business Team. `https://business.instagram.com/blog/500000-advertisers`, 2016, Accessed on 27.03.2018.

[82] Sayantari Ghosh, Kumar Gaurav, Saumik Bhattacharya, and Yatindra Nath Singh. Going viral: The epidemiological strategy of referral marketing. *arXiv preprint arXiv:1808.03780*, 2018.

[83] Dennis Yu and Alex Houg. Facebook analytics, advertising, and marketing. In *Facebook Nation*, pages 117–138. Springer, 2014.

[84] Todd A Finkle. Adroll: A case study of entrepreneurial growth. *New England Journal of Entrepreneurship*, 16(1):47–50, 2013.

[85] Brahim Zarouali, Koen Ponnet, Michel Walrave, and Karolien Poels. "do you like cookies?" Adolescents' skeptical processing of retargeted facebook-ads and the moderating role of privacy concern and a textual debriefing. *Computers in Human Behavior*, 69:157–165, 2017.

[86] Hong Heather Yu, Deepa Kundur, and Ching-Yung Lin. Spies, thieves, and lies: The battle for multimedia in the digital era. *IEEE multimedia*, 8(3):8–12, 2001.

[87] Hasshi Sudler. Effectiveness of anti-piracy technology: Finding appropriate solutions for evolving online piracy. *Business Horizons*, 56(2):149–157, 2013.

[88] John D Martin III. Piracy, public access, and preservation: An exploration of sustainable accessibility in a public torrent index. In *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology*, page 123. American Society for Information Science, 2016.

[89] Brett Danaher, Michael D. Smith, and Rahul Telang. The truth about piracy. `https://idea.heinz.cmu.edu/2016/02/02/`, 2016.

[90] Felix Oberholzer-Gee and Koleman Strumpf. The effect of file sharing on record sales: An empirical analysis. *Journal of political economy*, 115(1):1–42, 2007.

[91] Birgitte Andersen and Marion Frenz. Don't blame the P2P file-sharers: The impact of free music downloads on the purchase of music CDs in Canada. *Journal of Evolutionary Economics*, 20(5):715–740, 2010.

[92] Alejandro Zentner. File sharing and international sales of copyrighted music: An empirical analysis with a panel of countries. *The BE Journal of Economic Analysis & Policy*, 5(1), 2005.

[93] Norbert J Michel. The impact of digital file sharing on the music industry: An empirical analysis. *Topics in Economic Analysis & Policy*, 6(1), 2006.

[94] Rafael Rob and Joel Waldfogel. Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students. *The Journal of Law and Economics*, 49(1):29–62, 2006.

[95] Alejandro Zentner. Measuring the effect of file sharing on music purchases. *The Journal of Law and Economics*, 49(1):63–90, 2006.

[96] Liye Ma, Alan L Montgomery, Param Vir Singh, and Michael D Smith. An empirical analysis of the impact of pre-release movie piracy on box office revenue. *Information Systems Research*, 25(3):590–603, 2014.

[97] Brett Danaher and Michael D Smith. Gone in 60 seconds: The impact of the megaupload shutdown on movie sales. *International Journal of Industrial Organization*, 33:1–8, 2014.

[98] Kumar Vikram. Use of pirated software is rampant in india. *http://www. newindianexpress.com/nation/2017/sep/25*, 2017.

[99] Arul George Scaria. Online piracy of indian movies: Is the film industry firing at the wrong target. *Mich. St. U. Coll. L. Int'l L. Rev.*, 21:647, 2013.

[100] Majid Yar. The global 'epidemic'of movie 'piracy': Crime-wave or social construction? *Media, Culture & Society*, 27(5):677–696, 2005.

[101] Ian Condry. Cultures of music piracy: An ethnographic comparison of the US and Japan. *International journal of cultural studies*, 7(3):343–363, 2004.

[102] Sudip Bhattacharjee, Ram D Gopal, Kaveepan Lertwachara, and James R Marsden. Impact of legal threats on online music sharing activity: An analysis of music industry legal actions. *The Journal of Law and Economics*, 49(1):91–114, 2006.

[103] Yi-Shun Wang, Ching-Hsuan Yeh, and Yi-Wen Liao. What drives purchase intention in the context of online content services? The moderating role of ethical self-efficacy for online piracy. *International Journal of Information Management*, 33(1):199–208, 2013.

[104] Sudip Bhattacharjee, Ram D Gopal, and G Lawrence Sanders. Digital music and online sharing: Software piracy 2.0? *Communications of the ACM*, 46(7):107–111, 2003.

[105] Katie Walsh. Irdeto survey: Millennials shaping video piracy in the middle east and north africa. *https://irdeto.com/news*, 2017.

[106] T Ramayah, Noor Hazlina Ahmad, Lau Guek Chin, and May-Chiun Lo. Testing a causal model of internet piracy behavior among university students. *European Journal of Scientific Research*, 29(2):206–214, 2009.

[107] Rajiv K Sinha and Naomi Mandel. Preventing digital music piracy: The carrot or the stick? *Journal of Marketing*, 72(1):1–15, 2008.

[108] Ankur Nandedkar and Vishal Midha. It won't happen to me: An assessment of optimism bias in music piracy. *Computers in Human Behavior*, 28(1):41–48, 2012.

[109] Cheolho Yoon. Theory of planned behavior and ethics theory in digital piracy: An integrated model. *Journal of business ethics*, 100(3):405–417, 2011.

[110] Dongwon Lee, Jaimie Yejean Park, Junha Kim, Jaejeung Kim, and Junghoon Moon. Understanding music sharing behaviour on social network services. *Online Information Review*, 35(5):716–733, 2011.

[111] Christian Peukert, Jörg Claussen, and Tobias Kretschmer. Piracy and box office movie revenues: Evidence from megaupload. *International Journal of Industrial Organization*, 52:188–215, 2017.

[112] Yu-Chen Chen, Rong-An Shang, and An-Kai Lin. The intention to download music files in a P2P environment: Consumption value, fashion, and ethical decision perspectives. *Electronic Commerce Research and Applications*, 7(4):411–422, 2008.

[113] Katie Walsh. Nearly half of consumers around the globe are willing to stop or watch less pirated video content. *https://irdeto.com/news/*, 2017.

[114] Trevor T Moores and Jerry Cha-Jan Chang. Ethical decision making in software piracy: Initial development and test of a four-component model. *Mis Quarterly*, pages 167–180, 2006.

[115] James R Coyle, Stephen J Gould, Pola Gupta, and Reetika Gupta. "to buy or to pirate": The matrix of music consumers' acquisition-mode decision-making. *Journal of Business Research*, 62(10):1031–1037, 2009.

[116] Timothy Paul Cronan and Sulaiman Al-Rafee. Factors that influence the intention to pirate software and media. *Journal of business ethics*, 78(4):527–545, 2008.

[117] Pola B Gupta, Stephen J Gould, and Bharath Pola. "to pirate or not to pirate": A comparative study of the ethical versus other influences on the consumer's software acquisition-mode decision. *Journal of Business Ethics*, 55(3):255–274, 2004.

[118] David Greenhalgh, Sourav Rana, Sudip Samanta, Tridip Sardar, Sabyasachi Bhattacharya, and Joydev Chattopadhyay. Awareness programs control infectious disease: Multiple delay induced mathematical model. *Applied Mathematics and Computation*, 251:539–563, 2015.

List of Publications

1. **Kumar Gaurav**, Sateeshkrishna Dhuli, and Y. N. Singh, " Fraction of Connections Among Friends of Friends as a New Metric for Network Analysis," in *National Conference on Communications (NCC)*, 2018.

2. **Kumar Gaurav**, Sayantari Ghosh, Saumik Bhattacharya, and Y. N. Singh, "Equilibria of Rumor Propagation: Deterministic and Network Approaches," in *Region 10 Conference, TENCON 2017, IEEE*, pp. 2029–2034, 2017.

3. Sateeshkrishna Dhuli, **Kumar Gaurav** and Y. N. Singh, "Convergence Analysis for Regular Wireless Consensus Networks," *IEEE Sensors Journal* , no. 8, pp. 4522–4531, 2015.

<u>Submitted:</u>

1. **Kumar Gaurav**, Sayantari Ghosh, Saumik Bhattacharya, and Y. N. Singh, "Going viral: Factor influencing the public attitude towards referral marketing in India," in *Journal of Retailing and Consumer Services*, Dec 2018.