Epidemics on networks

Muse Aden Ahmed (muse.ahmed@aims.ac.rw)
African Institute for Mathematical Sciences (AIMS) Rwanda

Supervised by: Dr. Philip Knight
Strathclyde University, UK

May 2019

Submitted in partial fulfilment of the requirements of a Master of Science in Mathematical Sciences at AIMS Rwanda

Acknowledgements

First of all, i would like to thank my supervisor Philip Knight for his valuable and useful guidance to fulfil this project. I also thank AIMS founder staff members and all lectures for their initiatives and passion for development African eduction through mathematical science. I extend to thank to AIMS tutors especially Alex for her supporting idea, guidance and comment in writing this project.

All il all, i would like to thank my family and fellow their heartedness and support while doing my master studies.

Abstract

The spread of epidemics on networks has been extensively studied in the last ten years. Epidemic transmission has a variety of important behaviours, independent on the characteristics of the disease and the network its spreads on. In this project, a briefly introduction to the study of epidemics on networks and SIS (susceptible-infection-susceptible) epidemic model will be provided.

We will also observe the difference in epidemic behaviour when comparing real and random networks, which leads to the suggestion that it is insufficient to analyse such behaviour using only random networks.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Muse Aden Ahmed, May 2019

Contents

1. Introduction

In recent years, scientists have began to account for network structure within epidemic modelling by extending the traditional methods. for example, if a disease spreads by human contact, social networks to which individuals belong will clearly play a role in their susceptibility to the disease. Insight into the propagation of diseases can lead to innovative vaccination programmes. The same methods can be used to analyse similar processes, such as the spread of computer viruses.

One method of analysis represents these social networks as special cases of graphs and determines at which threshold (number of susceptible individuals) the disease becomes epidemic (Chakrabarti et al., 2008). This epidemic threshold model has been tested extensively on real and synthesized graphs (providing reliable statistical confidence to algorithmic analysis) and has resulted in threshold condition valid for arbitrary graphs. The model has great predictive power, it can help make policies to limit the spread of disease by evaluating their effectiveness.

Pastor-Satorras and Vespignani (2001) analyse information from malicious program infections, including the typical time period and also the persistence of infectious agent strains on the web. They outline a dynamic model for the spread of infections on scale-free networks, once that they note the absence of an outbreak threshold and its important behaviour. Epidemiological framework streamlines information on PC viruses and will facilitate understanding of alternative propagation phenomena on communication networks and social networks.

Another example of applying networks to traditional epidemic modelling can be found in (Newman, 2002). He considers a simple case of sexual disease transmission amongst a population divided into men and women.

In Chai and Pavlou (2017), They model such a path-based epidemic spreading taking under consideration. The degree and distribution of the Infective agent. Although their model methodology is a general nature, they focus their on communication networks and take into account information packets because of the Infective agent.

Another method of applying networks to traditional of epidemic modelling can be found in (Keeling and Eames, 2005). They review the idea of medical specialist theory(form on random-mixing models) and network theory (form on work from the graph theory, and social sciences). They then describe a range of methodology that enables the blending network. Finally, they look to the future to suggest how the two fields of network theory and epidemiological modelling can deliver an improved understanding of disease dynamics and better public health through effective disease control.

In Meyers et al. (2003), his model expressly capture the patterns of connections with patients also caregivers in an establishment with many wards. Study of this interaction using networks predicts that although the comparatively small occurrence of true bacteria respiratory illness start with caregiver the patterns of caregiver movement, and therefore, the level to that he's protected from disease is also important to the management and prevention of true bacteria outbreaks.

In this project, we will investigate probabilistic models of disease spread which allow for simple simulations, and also for a statistical analysis of the speed and strength of an epidemic on a variety of network.

Many diseases spread in the human population through contact between infective individuals (those carrying the disease) and susceptible individuals (those who don't have the disease but can become

infected). We use three frequently used dynamic epidemic models, SI, SIS and SIR models, that help us to understand the basic building blocks of epidemic models.

1.1 Susceptible-Infected (SI) Model

SI model is the simplest form of disease models. We consider the disease spread in the population containing N persons. Denoted by $S(t)$ the number of persons who are susceptible (those who are not diseased but it can catch it) at time t and $I(t)$ the number individuals that are infected at time t. At $t = 0$ everyone is susceptible $(S(0) = N)$ and no one is infected $(I(0) = 0)$.

Figure 1.1: SI model.

We suppose that a classic person has K links and the disease will be transmitted from infected to susceptible person at a rate of β (per unit time).

In the hypothesis of homogeneous mixing, the probability that an infected person meets a susceptible person is $S(t)/N$. Thus an the infected person comes into contact with $(K)S(t)/N$ susceptible people in a unit of time. Since $I(t)$ infected persons are transmitting the pathogen, in each

β rate, the new average rate of infection $dI(t)$ during the period is

$$\frac{dI(t)}{dt} = \beta(k)\frac{S(t)I(t)}{N} \tag{1.1.1}$$

where

$$s(t) = S(t)/N, \quad i(t) = I(t)/N \tag{1.1.2}$$

once we get $c = i(0)/(1 - i(0))$, we get the fraction of infected individuals increasing at time t as,

$$i(t) = \frac{i(0)e^{\beta(k)t}}{1 - i(0) + i(0)e^{\beta(k)t}}. \tag{1.1.3}$$

1.2 SIS Model

The majority of pathogens are defeated by the immune system. To model this information, we must let infected people recover, hence stopping them spreading the disease. The SI may be extended into the SIS model, where a recovered individual can become susceptible again.

We wish to relax the hypothesis of stable immunity after recovery to model diseases that can infect us frequently, such as flu. The difference is that the infected individual person recovers at a fixed

rate of μ, to become again susceptible. The equation describing this dynamic model is

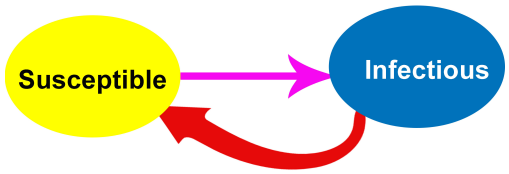


Figure 1.2: SIS model.

$$\frac{di}{dt} = \beta(k)i(1-i) - \gamma i \tag{1.2.1}$$

Where $\beta(k)$ is called the transmission rate and μi is the recovery rate of the disease population.The solution of (1.2.1) gives us the infection of individuals as a function of time,

$$i(t) = \left(1 - \frac{\gamma}{\beta(k)}\right) \frac{ce^{(\beta(k)-\mu)t}}{1 + ce^{(\beta(k)-\gamma)t}} \tag{1.2.2}$$

$i(0)$ at time zero gives us,

$$c = \frac{i(0)}{1 - i(0) - \frac{\gamma}{\beta(k)}}. \tag{1.2.3}$$

1.3 SIR Model

Kermack and McKendrick, in their first article, assume that all members of the population are initially susceptible to disease and that full immunity is conferred after infection.

Figure 1.3: SIR model.

In addition, they divided the population into three distinct classes: susceptible individuals, S in good health, capable of contracting the disease; the infected, I, those who have the disease and can transmit it; and those who are withdrawn, R, who have had the disease and who are now immunized against the infection (or who are removed from the spread of the disease by some other means). Such models are often called SIR models and are explained by the following differential equations

$$\frac{ds}{dt} = -\beta(k)i(I - r - i) \tag{1.3.1}$$

$$\frac{di}{dt} = -\gamma i + \beta(k) i(I - r - i) \tag{1.3.2}$$

$$\frac{dr}{dt} = \gamma i \tag{1.3.3}$$

- The term $\frac{ds}{dt}, \frac{di}{dt}, \frac{dr}{dt}$ in the differential equation, they indicate the rate change of the susceptible infected and recovery population, respectively.
- $\beta =$ the transmission rate (effect contact rate).
- $\gamma =$ the recovery or removed rate.
- High value γ means a person will be infected and high value of β means that the epidemic will spread quickly.

2. Network Epidemics

Epidemiology is the study of how often disease occur in different group of people. Network epidemiology has turn into a fundamental structure for studying the part of individual contact models in the spread of communicable diseases. In epidemiology, the network represents the contact arrangement of network nodes (individuals) related by links (sometimes in the form of a time network, the links are not until now active).

The epidemic patterns discussed in the previous chapter do not include arrangements of an individual's contacts within a network which facilitates the increase in the spread of a pathogen. In its place, they suppose that any one person contact with another person involves a comparable number of contacts, k (homogeneous mixing hypothesis). This assumption is over-simplified, since individuals can only transmit to individuals with whom they contact, the pathogen spreads over a complex network of contacts. In this section we will discuss networks epidemic models.

2.1 Susceptible-Infected (SI) Model on a Network

If the pathogen is spread over a network, the most linked persons are additionally possible to be in contact with an infected person (Barabási et al., 2016). As a result, the mathematical formulation has to consider the degree of every node as an understood variable. That the number of links starting or ending at each node. The graph theory definition or degree or a node and more will be given in chapter three.

The fraction of infected nodes of degree k, i_k in a network is computed by dividing the total number of infected nodes of degree k, I_k by the total number of nodes of degree k, N_k. Thus:

$$i_k = \frac{I_k}{N_k} \tag{2.1.1}$$

The total infected nodes is the total of every infected k nodes and where the P_k is a properly normalized probability distribution,

$$i = \sum_k P_k i_k. \tag{2.1.2}$$

known the different node, we write SI model used for all degree k:

$$\frac{di_k}{dt} = \beta(1 - i_k)k\theta_k \tag{2.1.3}$$

- θ_k is the fraction of infected neighbours of a susceptible node k
- β is the infection rate
- $(1 - i_k)$ is the fraction of degree k nodes with the purpose of are not infected.

From a network lacking correlations the θ_k function is independent of k, cited from (Barabási et al., 2016) and becomes

$$\frac{di_k}{dt} \approx \beta k i(0) \frac{\langle k \rangle - 1}{\langle k \rangle} e^{\frac{t}{\tau^{SI}}} \tag{2.1.4}$$

where τ^{SI} is the characteristics time for the spread of the pathogen

$$\tau^{SI} = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)} \tag{2.1.5}$$

The total infection nodes grows with time is,

$$i = \int_0^k i_k P_k dk \qquad \text{recall equation 2.1.2} \tag{2.1.6}$$

$$= i(0)(1 + \frac{\langle k \rangle^2 - \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle})(e^{t/\tau^{SI}} - 1). \tag{2.1.7}$$

We will now look at some specific types of networks and how SI model is applied to them.

Random network

during the last ten years, importance in social network has been unstable increase. This is addition to work performed even earlier in the area of Physics, especially with random networks. In such networks every new node is linked at random to one other previous node, and its development was studied. The work of Barabasi and others on the physics of Random Network was attractive , it had little to do with 'human made' networks such internet, web, friendship networks etc. This networks follow normal distribution.

The random network follows the Poisson distribution and is almost homogeneous. Most nodes contain approximately the similar number of links. Here $\langle k^2 \rangle = \langle k \rangle (\langle k \rangle + 1)$, obtaining

$$\tau_{ER}^{SI} = \frac{1}{\beta \langle k \rangle} \tag{2.1.8}$$

where

- $\langle k \rangle$ is the average degree of nodes in a network
- $\langle k^2 \rangle$ is the connectivity divergence (sum of squared degree)

Which is same as homogeneous networks.

Homogeneous model assumes that each person contact with every one also in the population and the rate of infection is largely determined by the density of the infected population. Kephart and White adopted

to modify the mode (Khelil et al., 2002). The graph nodes represent individual in the population, from the directed edge from node i to j denotes that i can directly infect j. A rate of infection called the birth rate β is associated with each infected node. A virus death rate γ, is connected with every infected node.

If we indicate the population infection at time t as η_t, a deterministic time evolution of η_t in the Kephart-White model can be represented as,

$$\frac{d\eta_t}{dt} = \beta\langle k\rangle\eta_t(1 - \frac{\eta_t}{N}) - \gamma\eta_t \tag{2.1.9}$$

The steady state of equation (2.1.9) is,

$$\eta_t = N\left(1 - \frac{\gamma}{\beta\langle k\rangle}\right) \tag{2.1.10}$$

where $\langle k\rangle$ is the average of connectivity and N is the number of nodes.

Scale-free network

A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. This is when the number of nodes with low degree is high and this number decreases as you increase the degree of the vertex. Therefore, the fraction $P(k)$ of nodes in the network having k links to other nodes for a large value of k becomes,

$$P(k) \sim k^{-\gamma} \tag{2.1.11}$$

where, $2 < \gamma \leq 3$. This implies every node has a statistically significant probability of having an extremely large number of links to comparing the other standard connectivity of the network (Bornholdt and Schuster, 2006).

Scale-free network with $\gamma \geq 3$

If the contract network is scale-free with degree exponent $\gamma \geq 3$, $\langle k^2\rangle$ and $\langle k\rangle$ are both finite. τ^{SI} is also finite and the spreading dynamics is same as random network.

Scale-free network with $\gamma \leq 3$

For $\gamma \leq 3$ in the $N \to \infty$ limit $\langle k^2\rangle \to \infty$.Therefore $\tau^{SI} = \frac{\langle k\rangle}{\beta(\langle k^2\rangle - \langle k\rangle)}$ and we predict that $\tau^{SI} \to 0$, the spread of pathogen on scale-free is instantaneous.

On the other hand, Pastor-Satorras developed an analytic model (referred to as SV model) for the Barabasi-Albert (BA) power topology $(\gamma = 3)$. There is steady state prediction is,

$$\eta_t = 2e^{\frac{-\gamma}{m\beta}} \tag{2.1.12}$$

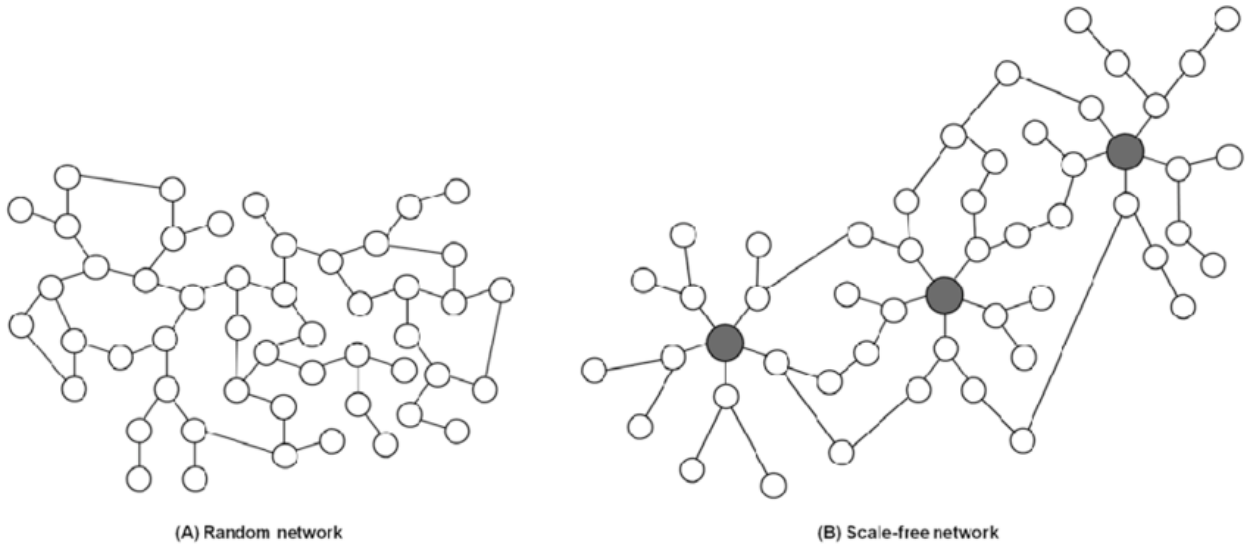


Figure 2.1: Random network and Scale-free network.

where m is the minimum connectivity in the network.

Figure 2.1: shows Random network and Scale-free networks, cited form (Seo et al., 2013). (A) is a Random network which follow normal distribution and it is nearly homogeneous. (B) is a Scale-free network follow power-law and is inhomogeneous. Most nodes have one or two links but a small number of nodes have a large numbers of connectivity, so it is called *hub*.

2.2 SIS Model and the Vanishing Epidemic Threshold

We adapt the equation describing the dynamics of the SIS model on the simple transfer network as well its extended SI model in equation 2.1.3

$$\frac{di_k}{dt} = \beta(1 - i_k)k\theta_k - \gamma i_k \tag{2.2.1}$$

The difference between (2.2.1) and (2.1.3) is γi_k.This changes the characteristic time of the epidemic, which now is

$$\tau^{SIS} = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle) - \gamma\langle k \rangle} \tag{2.2.2}$$

For large values of μ, the characteristic time is negative, so the i_k decay exponentially. The decay condition is not only a function of the recovery rate and $\langle k \rangle$, but also the heterogeneity of the network up to $\langle k^2 \rangle$.

The rate of spread (λ) of a pathogen,i.e. the ratio between the probability of transmission β and the recovery rate μ, is given as

$$\lambda = \frac{\beta}{\gamma} \tag{2.2.3}$$

The greater λ is, the more likely it is that the diseases spreads. The number infected individual does not increase gradually with λ; rather the pathogen will become epidemic spread if the spreading rate reaches an epidemic threshold λ_c.

Random Network

For a random network the epidemic threshold is,

$$\lambda_c = \frac{1}{k+1} \tag{2.2.4}$$

- If $\lambda > \lambda_c$, the pathogen will spread until it becomes an epidemic state.
- If $\lambda < \lambda_c$ the pathogen dies out.

3. Basics of Network Theory

Network theory is the learning of graphs the same as illustration of either symmetric associations linking discrete objects. However, network theory is piece of graph theory (study of graphs). Network can be defined us graph has a nodes or edges contain attributes.

The first time in the theory of networks was Euler's solution of the seven Bridges of Konigsberg problem (Estrada and Knight, 2015). Since then graph theory has been applied to problems in money different areas of maths and to real world problems, such as designing computer chips and modelling disease spread, as we show in this essay.

3.1 Degree of a vertices

The number of edges incident on a vertex it is called the degree of the vertices. It is denoted $d(u_i)$.

Since each edge contributes two degree so the sum of the degree of all vertices in a graph is twice the number of edges of that graph.

Hence,

$$\sum_{i-1}^{n} d(u_i) = 2e \tag{3.1.1}$$

where

- e is the number of edges
- n is the number of vertex.

3.2 Adjacency matrix

3.2.1 Definition. Adjacency matrices: $A(A_G)$ is a nxn zero-one matrix with 1 as $(A_{ij})^{th}$ entry when i and j are adjacency and 0 otherwise.

This is a simple undirected graph of G.

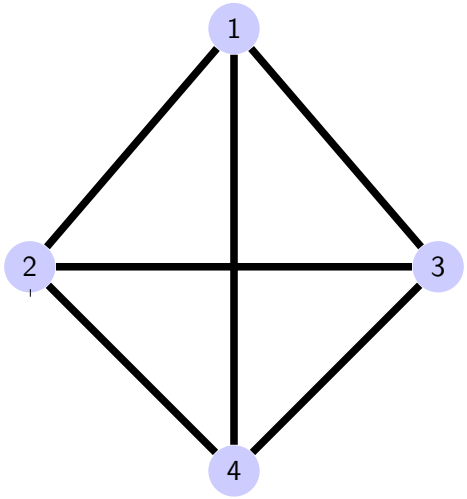


Figure 3.1: simple undirected graph

therefore,

- the number of edges $E(G)$= {(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)}
- the number of vertex $V(G)$={1,2,3,4}

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad \text{where} \quad A_{i,j} = \begin{cases} 1 & \text{if } i,j \in E \\ 0 & \text{otherwise .} \end{cases}$$

However, rows and columns represent vertices.

3.3 Walks, trails and paths

3.3.1 Definition. Walk: let 'G' be graph and u,v be any two vertices is 'G'. A walk from u and v is an alternating sequence of vertices and edges. The walk always starts from a vertex and ends to a vertex.

A walk from u to v can be written

$$w = v_1 e_1 v_2 e_2, ... v_{n-1} e_{n-1} v_n \tag{3.3.1}$$

where $v_1, ..., v_n \in V(G)$ and $e_1, ..., \in e_n E(G)$ and we let $u = v$ and $v = V_n$.

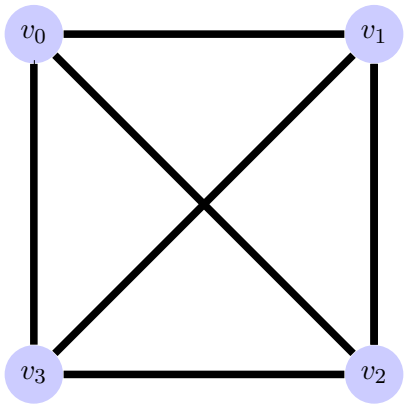


Figure 3.2: example of walk

When G is simple, we say write the walk by indicating the vertices only, eg

$$w = v_1v_2v_0v_2v_3. \tag{3.3.2}$$

3.3.2 Definition. **A trail** is a walk such that all of the edges are distinct, eg

$$t = v_1v_2v_1v_3. \tag{3.3.3}$$

3.3.3 Definition. **A path** is a walk such that all of the vertices and edges are distinct, eg

$$p = v_1v_2v_0v_3. \tag{3.3.4}$$

3.4 Node Centrality

Centrality is one of the foremost necessary properties of the network analysis. If a node contains a central position it can have an important role in the spread data or will be one of the foremost dangerous points that we must always affect with care. Those nodes as "well-connected" with alternatives, and that they have contact with several other necessary nodes (Burgess, 2004).

3.5 Degree Centrality

Degree centrality exposure to the network; opportunity to directly influence. We see in here, the degree of node i in a simple network G is defined by using adjacency matrix A, as,

$$D_i = \sum_{j=1}^{n} a_{ij} = (e^tA)i = (Ae)i \tag{3.5.1}$$

Where

- A is an adjacency matrix
- e^t is a vector of centrality

Therefore, the degree centrality, i is further centrality than j if $D_i > D_j$.

3.6 Closeness Centrality

The closeness centrality of a node i is the average of the shortest distance from i to every other node in the graph. Closeness can be interpreted as the approximation time to hear information; indirect influence; position of fast diffusion.

The closeness of node i in an indirect network G is defined as

$$CC(i) = \frac{n-1}{S(i)} \tag{3.6.1}$$

where the distance total is calculated the shortest path distance $d(i,j)$ as

$$S(i) = \sum_{j \in V(G)} d(ij). \tag{3.6.2}$$

3.7 Betweenness Centrality

The betweenness centrality extend to which a particular node lies on the shortest path between other nodes. The betweenness of the node i an undirected network G is defined as

$$BC(i) = \sum_i \sum_j \frac{\rho(j,i,k)}{\rho(j,k)} \quad \text{where} \quad i \neq j \neq k \tag{3.7.1}$$

where

- $\rho(j,k)$ is the number of connecting node j to the node k
- $\rho(j,i,k)$ is the number of shortest path through node i in the network.

3.8 Eigenvector Centrality

Eigenvector centrality assigns importance proportional to the importance of the nodes neighbours. Let us we denote the centrality of node i by x_i then we can allow for the effect by making x_i to be the proportional average of the centrality of i's network neighbours.

A mathematically we can write as:

$$x_i = \frac{1}{\lambda} \sum_{j}^{N} A_{i,j} x_j \tag{3.8.1}$$

where

- N is the number of nodes in the network
- λ an eigenvalue of the adjacency $matrix(A)$
- $A_{i,j}$ an elements in the adjacency matrix A.

3.9 Immunization

The immunization of the disease has involved the awareness of both experimental and theoretical

researchers. As a result, vast majority of existing achievements are limited to the simple assumption of single layer networked population. This seems obviously incompatible with the recent development of the theory of complex networks, each of which may have several roles in different topology connections.

We consider the SIS model and then test the effect of immunization strategies on disease propagation in complex network. With regards networks, we select ER random network and Barasi-Elbert scale-free network (Estrada and Knight, 2015).

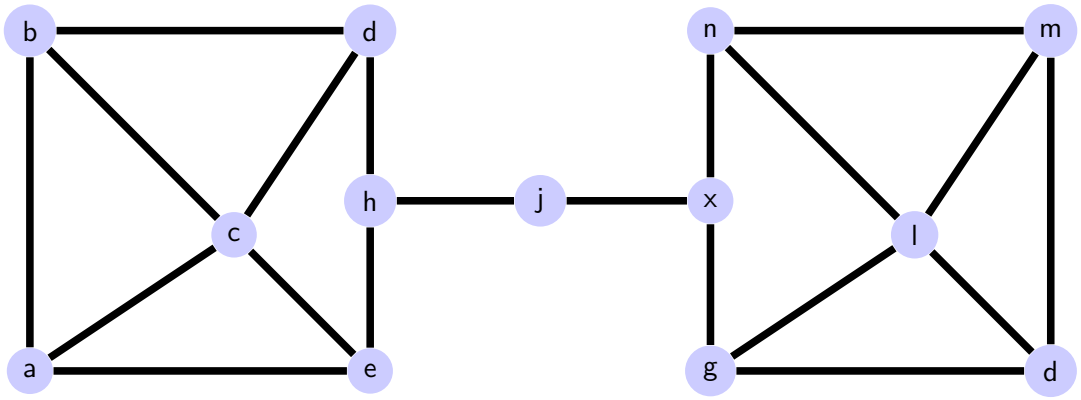


Figure 3.3: Two graphs of the same size connected with a bridge

If we have k nodes which nodes should be immunized? The immunization target is the nodes with the highest degree centrality.

Consider graphs (3.3) consisting of two equal sizes connected by j. After that, we assume that $k = 1$, so that we are allowed to immunize a single node. The best immunization in this graph is j because it is connected on both sides. On the other hand, immunization strategy of node h and x would be equal

effective, but immunizing node c or i both have highly degree centrality, so these are also important in the immunization strategy. All in all we can say immunization strategy is dependent on the important nodes in this case c, j and i.

4. PROPOSED MODEL

In this chapter, we explain a model that does not suppose homogeneous connectivity or topology. We expect a linked undirected network $G = (N, E)$, where N is the number of nodes in the network and E is the set of edges. The same as traditional models, we suppose a universal infection rate β for each edge connected to an infected node, and rate γ for each infected node.

Symbol	Description
β	Virus birth rate on a link from an infected neighbour
γ	Virus death rate on an infected node
t	timestamp
$P_{i,t}$	Probability that node i is infected at time t
$\zeta_{i,t}$	Probability that a k -linked node will not receive infections from its neighbours at time t
G	An undirected connected graph
N	Number of nodes in G
E	Number of edges in G
η_t	Number of infected nodes at time t

Table 4.1: Table of Symbols

4.1 Model

Our model works with small discrete time-steps $\triangle t$, with $\triangle t \to 0$. At some stage in every occasion interval, an infected node i seeks to effect a neighbour with probability β. At the same time, i may be cured with probability γ.

This procedure we will be modelled as Markov chain. Every state of Markov chain corresponds to one particular system (Susceptible or infected). We denote the chance that a node i is infected t as $P_{i,t}$ and also we can define $\zeta_{i,t}$ the chance that k-linked node will acquire infection from its neighbour at time t,

$$\zeta_{i,t} = \prod_{\text{j: neighbor of i}} (P_{j,t-1}(1-\beta) + (1 - P_{j,t-1})) \tag{4.1.1}$$

$$= \prod_{\text{j: neighbor of i}} (1 - \beta * P_{j,t-1}) \tag{4.1.2}$$

In the above model, a node i is healthy at time t if,

- i was healthy before t and not infected at t
- i was infected at $t-1$, cured at t and not infected at t

Figure 4.1 The SIS model as seen in the figure: Every node in time step t is either Susceptible (which have not got the disease but can catch it) or Infective. A susceptible node i is presently cured, but it may be infected $(1 - P_{j,t})$ by its neighbour node i. An infected node may be cured with probability γ and return again to susceptible.

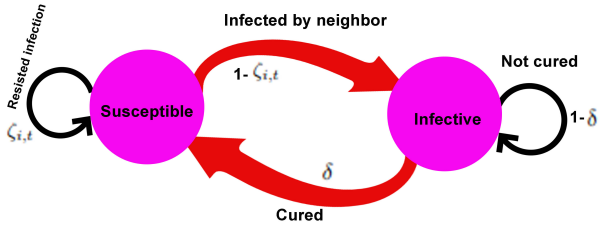


Figure 4.1: SIS model.

This is an independent assumption: we expect probabilities $(1 - P_{j,t-1})$ to be independent in each other (Chakrabarti et al., 2008).

We denote the chance of a node i being infected at time t by $P_{i,t}$;

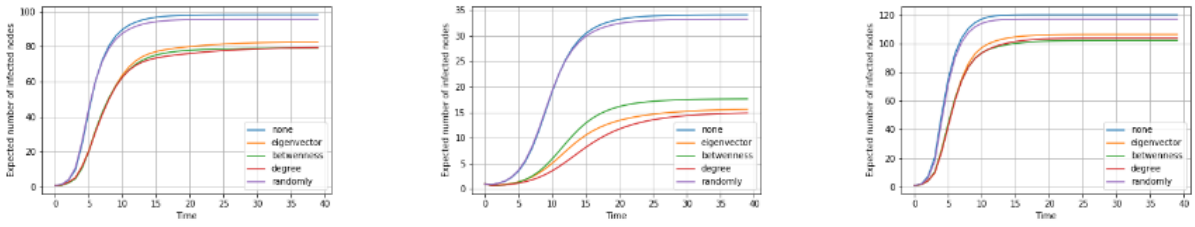
$$1 - P_{i,t} = (1 - P_{i,t-1})\zeta_{i,t} + \delta P_{i,t-1}\zeta_{i,t} \quad \text{where} \quad i = 1, ..., N \tag{4.1.3}$$

This equation represents NLDS (Non-Linear Dynamical System), Figure 4.1 indicates the transition diagram. Note for the closing term at the right-hand side of the equation (4.1.3), we expect the chance cured an event at node i takes place before infection from the neighbour is approximately 50%.

Given a particular network topology and values of β and δ, we resolve equation 4.1.3 numerically and procure the time evaluation of the infected population length $\zeta_{i,t}$ therefore,

$$\zeta_{i,t} = \sum_{i=1}^{N} P_{i,t}. \tag{4.1.4}$$

In the simulation graphs below, each individual performs on average 40 times runs. Different value of β and were tested with γ and we are looking at how population vaccination is differs.



$\gamma = 0.65$ and $\beta = 0.45$ $\gamma = 0.65$ and $\beta = 0.20$ $\gamma = 0.65$ and $\beta = 0.60$

Figure 4.2: Experiments on software digital graph.

The graph above shows the temporal evaluation of vaccination in the software digital graph. Three simulations were performed with a fixed γ but varying β. We start each simulation with a set of random initial infectivity. Each graph shows the number of nodes removed which are random nodes or nodes with highest betweenness centrality, eigenvector centrality or degree centrality. However, we can see how vaccination strategy changes when we fixed γ and vary β.

4.2 Epidemic Threshold and Eigenvalue

As shown in the previous chapter, we denote the infection population at time as η_t, a deterministic time evolution of η_t in the Kephart-White model. We will now use this model to derive results of steady state,

$$\frac{d\eta_t}{dt} = \beta\langle k\rangle\eta_t(1 - \frac{\eta_t}{N}) - \gamma\eta_t \tag{4.2.1}$$

Let $\frac{d\eta_t}{dt}$=0 to obtain

$$\beta\langle k\rangle\eta_t(1 - \frac{\eta_t}{N}) - \gamma\eta_t = 0 \tag{4.2.2}$$

We factorize to get

$$\eta_t \left(\beta\langle k\rangle(1 - \frac{\eta_t}{N}) - \gamma\right) = 0 \tag{4.2.3}$$

$$\left(1 - \frac{\eta_t}{N}\right) = \frac{\gamma}{\beta\langle k\rangle} \tag{4.2.4}$$

Therefore, the steady state solution is,

$$\eta_t = N\left(1 - \frac{\gamma}{\beta\langle k\rangle}\right). \tag{4.2.5}$$

An epidemic threshold is the critical state beyond which the infection become epidemic. Therefore, predicting the epidemic threshold on important part of the epidemiological model.

In this chapter, we describe a general theory of epidemic threshold on an arbitrary graph. We observe that epidemic threshold links the conditions of birth and death rates to the adjacency matrix, such that the infection become epidemic if this conditions holds,

$$\frac{\beta}{\gamma} < \tau \implies \text{infection dies out over time} \tag{4.2.6}$$

$$\frac{\beta}{\gamma} > \tau \implies \text{infection survives and becomes an epidemic} \tag{4.2.7}$$

The epidemic threshold is only depend on a single parameter the largest eigenvalue of the adjacency matrix. The system is stable if the first eigenvalue of the appropriate of the matrix is small, otherwise unstable.

A stable NLDS implies that a small perturbation (a few initial infected nodes) will eventually return all nodes to the stable healths while the unstable states move away.

Our estimate for the epidemic threshold τ is

$$\tau = \frac{1}{\lambda_{1,A}} \tag{4.2.8}$$

Where $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix A of the network. A proof is shown in (Wang et al., 2003).

5. Experiments

In this chapter, we present a set of simulation results. These simulations test vaccination strategies of different proportions of the population. We then compare the Erdös–Rényi (ER) random network and the Real-word network.

The data sets we used were:

- Soft ware Digital: Collaboration networks with six different open-source soft ware systems provides more detail of the construction or origin of these network are provided Myers(Myers, 2003). These graph contains 150 nodes and 198 edges.
- Small world citation: A graph containing 233 nodes and 994 edges. Citation network from cite S.milgram's 1967 Psychology Today paper (Batagelj and Mrvar, 2006).
- Electronic: Electronic sequential logic circuit parsed from the ISCAS89 benchmark set, where the nodes represent logic gates and Flip-Flop (Milo et al., 2002). The graph contains 122 nodes and 189 edges.
- Centrality literature: The graph contains 118 nodes and 613 edges. Citation network of paper of published in the field of network centrality (Hummon et al., 1990).
- Scotch Broom: The contains 154 nodes and 370 edges. More details you can get (Memmott et al., 2000).

For each dataset, all the nodes were initially infected randomly, we then remove 5 nodes, 10 nodes, 20 nodes and 50 nodes with highest degree centrality, highest betweenness centrality, highest eigenvector centrality. We also remove 5 nodes, 10 nodes, 20 nodes and 50 nodes randomly, so that we are testing vaccination strategies of different proportions of the population.

5.1 Testing vaccination strategies of difference proportions of the population

Here we test vaccination strategies of different proportions of the population. Each table and graph shows the number of nodes removed which are random nodes or nodes with highest betweenness centrality, eigenvector centrality or degree centrality.

The graph theory and network analysis indicates the centrality and identifies the most important vertex in the graph. Centrality is the one of the foremost necessary properties of the network analysis as we seen in chapter three. On the other hand, as we seen in chapter two, the random network follows the Poison distribution and it is almost homogeneous. Pal Erdos and Alfred Renyi, have played an important role in understanding of this network. In their honour a random network is called Erdos-Renyi network. To come back our exprement, Figure (5.1) represents the results for software digital graph. The graph shows that a centrality based immunisation strategy is much more effective than a random network.

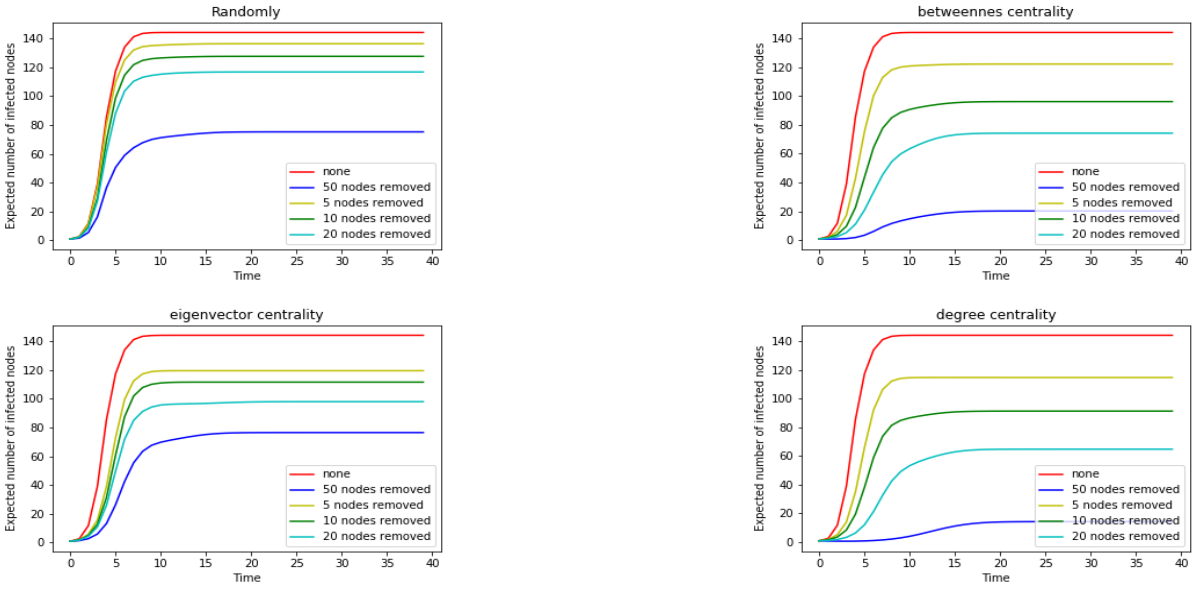


Figure 5.1: Software Digital graph.

Social network analysis measure is important tool for understanding behaviour network and graphs. In this case, we will take a look at some real world network measures for small immunisation programmes (not many individual receive a vaccine) in Figure (5.2) shows that all centrality strategies work equally well in immunisation strategy.

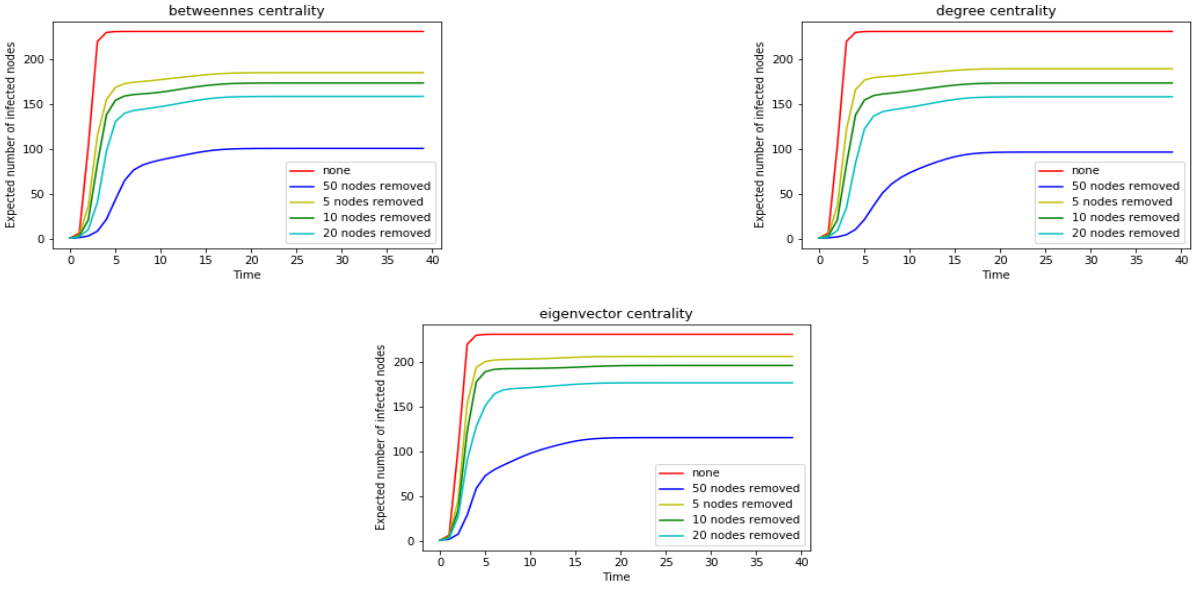


Figure 5.2: Small World Citations graph.

Table(5.1) Shows the results of the software digital dataset, when we look at the table, the values represent the number of vaccinations corresponding to each method of removal. Therefore, the best is

degree centrality, second is betweenness centrality, third is eigenvector centrality while randomly removed nodes, results in the lowest number of vaccinations.

Number of nodes removed	Randomly	Betweenness centrality	Eigenvector Centrality	Degree centrality
5	7	22	25	30
10	13	48	33	53
20	31	70	47	80
50	67	124	68	130

Table 5.1: Software Digital table

Table(5.2) shows result for the centrality literature dataset. We can see from the table the best vaccination is betweenness centrality and the second one is degree centrality.

Number of nodes removed	Randomly	Betweenness centrality	Eigenvector Centrality	Degree centrality
5	5	10	6	6
10	30	21	13	17
20	24	38	25	30
50	61	89	75	86

Table 5.2: Centrality literature

Table 5.3: shows result for the Electronic dataset, the best vaccination centrality in this case is degree centrality.

Number of nodes removed	Randomly	Betweenness centrality	Eigenvector Centrality	Degree centrality
5	5	5	5	6
10	12	12	11	11
20	23	22	26	24
50	60	64	53	75

Table 5.3: Electronic

In Table (5.4) one can clearly the difference between degree centrality, betweenness centrality, eigenvector centrality and random nodes when we use the Scotch Broom dataset. However, the best centrality in the table again betweenness centrality, when choosing vaccination strategy.

Number of nodes removed	Randomly	Betweenness centrality	Eigenvector Centrality	Degree centrality
5	6	32	13	29
10	12	73	45	55
20	24	111	64	99
50	68	143	75	128

Table 5.4: Scotch Broom

Table (5.5) represents the results for the small citation world dataset. Here we can see the best vaccination strategy is a degree centrality.

Number of nodes removed	Randomly	Betweenness centrality	Eigenvector Centrality	Degree centrality
5	5	44	24	41
10	10	57	35	57
20	32	72	54	72
50	60	130	115	134

Table 5.5: Small citation world

After analysing the different datasets, we see how the immunization strategies effects each dataset differently. It seems that all the centralities works well but when we use a random strategy is the results are not as good.

To sum up, for large scale immunisation strategies degree centrality seems to be particularly effective and this is easy to do in practice because you only need to ask individuals how many people they know.

5.2 Comparing real world and random networks

Here we look how our simulation of immunisation strategies shows a great difference in the behaviour when we compare real world and random networks.

Table(5.6) shows there exists a big difference between real world and random networks when we remove 10 nodes in all datasets for each vaccination strategy. All in all, we can say the real world networks performs better than random network.

Datasets	ER random network	Real-world
Software Digital	17	53
Centrality literature	10	17
Electronic	12	11
Scotch Broom	11	55
Small citation world	10	57

Table 5.6: ER random and real-world networks

Comparing real world and random networks in the above table shows that for the five datasets, we can say that the results on random networks are different from real-world networks the strategies are much more effective on real-world networks than on random networks with similar numbers of nodes and edges. This suggests that the analysis provided by Barabasi et al (discussed in chapter 2) does not provide a full picture and justifies the need for models such as Chakrabati's that allows for empirical testing.

6. Conclusion

During the project, we investigate epidemic real world and random networks with the aim of testing vaccination strategies in different proportions of a population using a number of datasets. For each dataset, all the nodes were initially infected randomly, we then remove nodes with the most important centrality (degree centrality, eigenvector centrality and betweenness centrality) in addition to removing some nodes randomly. We observed how the method of removal made a difference to the number of vaccinations, before comparing real world and networks as seen in the Table (5.6).

Then, when we use different datasets, we see how the vaccination is different. However, we know that the best vaccination here is a degree of centrality, since the five data sets with three degrees of centrality have become better.

On the other hand, we also compare real world and random network. We see Table(5.6), real network is better performer than random network.

All in all, to understand the behaviour of epidemics on networks it is not sufficient to analyse their behaviour on random networks. For instance, our simulation of immunisation strategies show a great difference in behaviour when we compare real world and random networks.

References

Barabási, A.-L. et al. *Network science*. Cambridge university press, 2016.

Batagelj, V. and Mrvar, A. Pajek datasets. http://vlado.fmf.unilj.si/pub/networks/data/, 2006.

Bornholdt, S. and Schuster, H. G. *Handbook of graphs and networks: from the genome to the internet*. John Wiley & Sons, 2006.

Burgess, M. *Analytical network and system administration: managing human-computer networks*. John Wiley & Sons, 2004.

Chai, W. K. and Pavlou, G. Path-based epidemic spreading in networks. *IEEE/ACM Transactions on Networking*, 25(1):565–578, 2017.

Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., and Faloutsos, C. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, 10(4):1, 2008.

Chen, L.-C. and Carley, K. M. The impact of countermeasure propagation on the prevalence of computer viruses. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(2):823–833, 2004.

Estrada, E. and Knight, P. A. *A first course in network theory*. Oxford University Press, USA, 2015.

Hummon, N. P., Doreian, P., and Freeman, L. C. Analyzing the structure of the centrality-productivity literature created between 1948 and 1979. *Knowledge*, 11(4):459–480, 1990.

Keeling, M. J. and Eames, K. T. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005.

Khelil, A., Becker, C., Tian, J., and Rothermel, K. Directed-graph epidemiological models of computer viruses. In *Proc. 5th ACM Int'l workshop on Modeling analysis and simulation of wireless and mobile systems*, pages 54–60, 2002.

Memmott, J., Martinez, N. D., and Cohen, J. Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web. *Journal of Animal Ecology*, 69(1):1–15, 2000.

Meyers, L. A., Newman, M., Martin, M., and Schrag, S. Applying network theory to epidemics: control measures for mycoplasma pneumoniae outbreaks. *Emerging infectious diseases*, 9(2):204, 2003.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

Myers, C. R. Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs. *Physical Review E*, 68(4):046116, 2003.

Newman, M. E. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

Newman, M. E. A measure of betweenness centrality based on random walks. *Social networks*, 27(1): 39–54, 2005.

Pastor-Satorras, R. and Vespignani, A. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

Seo, H., Kim, W., Lee, J., and Youn, B. Network-based approaches for anticancer therapy. *International journal of oncology*, 43(6):1737–1744, 2013.

Wang, Y. and Wang, C. Modeling the effects of timing parameters on virus propagation. In *WORM*, volume 3, pages 61–66, 2003.

Wang, Y., Chakrabarti, D., Wang, C., and Faloutsos, C. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34. IEEE, 2003.