SCIENTIFIC REPORTS

OPEN

Received: 02 March 2016
Accepted: 01 June 2016
Published: 22 June 2016

Impact of Information based Classification on Network Epidemics

Bimal Kumar Mishra[1], Kaushik Haldar[2] & Durgesh Nandini Sinha[3]

Formulating mathematical models for accurate approximation of malicious propagation in a network is a difficult process because of our inherent lack of understanding of several underlying physical processes that intrinsically characterize the broader picture. The aim of this paper is to understand the impact of available information in the control of malicious network epidemics. A 1-n-n-1 type differential epidemic model is proposed, where the differentiality allows a symptom based classification. This is the first such attempt to add such a classification into the existing epidemic framework. The model is incorporated into a five class system called the DifEpGoss architecture. Analysis reveals an epidemic threshold, based on which the long-term behavior of the system is analyzed. In this work three real network datasets with 22002, 22469 and 22607 undirected edges respectively, are used. The datasets show that classification based prevention given in the model can have a good role in containing network epidemics. Further simulation based experiments are used with a three category classification of attack and defense strengths, which allows us to consider 27 different possibilities. These experiments further corroborate the utility of the proposed model. The paper concludes with several interesting results.

Scientific efforts to model and accurately approximate the spread of malicious content over the Internet have received significant attention from researchers ever since the appearance of the Morris worm in 1988. The aims and methods employed by attackers, as well as the level of damage inflicted, have all changed significantly, over the years. The aim initially was to infect as many computers as possible. Then the infected computers were used to spread the infection in an automated manner with an exponential rate of spread[1]. This characterized the so-called *fast spreading worms*, which included the highly popular Code Red[2] and Nimda worms[3]. Internet security threats have evolved consistently and considerably over the last one and a half decade as the malware authors have shown constant innovation in their methodologies. Recently the number of attacks which focus primarily on a finite, often small, set of specific IP addresses has started to gain significant popularity. Such attacks are called *targeted attacks*. Symantec reports reveal an ever increasing trend in the global average of reported cases of targeted attacks since 2010. The reported number was 77 in 2010, 82 in 2011[4], and then 116 in 2012[5]. These attacks are characterized by malicious intentions like cyber espionage (Ghostnet attack 2009), cyber sabotage of critical physical resources (Stuxnet attack of 2010), and industrial espionage (Nitro attack of 2011). The traditional threats are now becoming more critical as they are expanding into newer forums like social media and mobile devices. The proportion of mobile malware has also shown a steadily increasing trend over the last few years. Symantec reported a 58% increase in the number of mobile malware families in 2012[5].

The constant use of specialized techniques for intrusion and also customized tools makes it very difficult to defend against such attacks. Stealth techniques incorporating patience and persistence are being used to reduce the detection risk. In the light of such attacking methods, the traditional manual patching approaches to defense are clearly not efficient. The need is to develop detection and response systems which are intelligent enough to identify malicious attacks before they are able to inflict serious damage. The decision making can be improved by including a cooperative strategy where appropriate additional information about the status of infection is readily available. Such information may include data on the infection rate, or the response rates of different nodes in the network. This collective approach can help in a meaningful use of the available evidence on the severity and certainty of an attack, which is rarely used.

[1]Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi, 835215 India. [2]Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi, 835215 India. [3]Adjunct Assistant Professor, Department of Mathematics, Temple University, Philadelphia, USA. Correspondence and requests for materials should be addressed to B.K.M. (email: drbimalmishra@gmail.com)

There are several approaches that have been suggested over the years for a proper understanding of malware and their spread on networks. The initial theory based models proposed by the founding fathers of this domain are responsible for giving birth to what is now called theoretical computer virology[6,7]. Their methods were based on the intriguing similarities that existed between viruses which are computer based and those that are biological. A very novel suggestion was made by Murray when he highlighted that the methods existing for the study of epidemic spread of biological infections could be useful in understanding the propagation of computer viruses[8]. A popular biological epidemic model called the SIS (Susceptible - Infectious - Susceptible) model was then applied as the first such application to study the manner in which computer viruses spread on different kinds of networks[9]. The approaches may roughly be classified into two broad categories. In the first category, we can put those approaches which are based on purely epidemic homogeneous contact models[10–12]. Such models are devoid of the complexities arising from topological considerations. They are also robust enough in providing strong analytical insights about various dynamical properties of the system like epidemic thresholds, equilibrium points of the system, stability of the equilibria, and periodic behavior of solutions, among others. In the second category are included approaches that rely on the topology of networks. Such approaches have provided useful results on the existence of epidemic thresholds for simple models including the SI, SIS and SIR models[13,14]. There is however a difficulty in proving theoretical results like stability of equilibrium points, owing to the large number of different kinds of possible topologies of large scale networks. Instead of theoretical proofs, often simulation and experimentation based proofs have been provided. One of the most important findings of this category of approaches is the lack of the universal epidemic threshold for infinite-size scale-free networks[15]. Another important contribution was the N-intertwined mean-field approximation based model and its fully heterogeneous extension[16,17]. These models provided useful dimensions outlining the relation between network topology and the spreading process on the network. In the epidemic approaches used so far, there has still not been an effort to include the effect of anticipating such attacks before they actually occur. Instead of a *wait-and-watch* approach, anticipation of an epidemic path can be useful in identifying the course of action to pursue.

This paper basically addresses the following research scopes:

- How can we model the spread of an attack in a network with respect to time? How does a network attack start from one or two nodes and propagates to infect often millions of nodes?
- What is the long term behavior of the network with respect to time? A network may recover completely in most scenarios, but is there a possibility that a number of nodes remain infected? If so, what is the stable value of such a fraction of infection that persists?
- Is there a threshold condition that determines the long term behavior of the system with respect to the infection persisting or perishing? Such a threshold exists in epidemic literature and is called the *basic reproduction number*. Based on related ideas, we try to obtain a threshold condition for our system as well.
- Can the symptoms exhibited by nodes infected by different types of attacking agents be used to improve the intelligence level of the underlying detection and response system? What is the impact of such a behavioral classification on the spread of infection? Based on the above classification, can the network be made to react in a more efficient manner?
- Is there a possibility that the nodes use the additional information available with them and also disseminate it, so that they can act in a collaborative manner?

The remaining portion of the paper is structured as follows. The proposed architecture is presented in the next section. Then the various aspects of the model and its mathematical formulation are detailed. The next section analyzes the model to establish the long term behavior of the epidemic system. Experiments and the corresponding results are then discussed. Finally, the paper is concluded with an elaboration on the major findings of the present work.

Proposed Architecture

The proposed architecture for a differential symptom based epidemic classification and defense is shown in Fig. 1. The architecture involves five different components which are as follows:

Data Processing Unit. This component receives raw data from the different hosts and extracts two kinds of information necessary for the working of the other components. It uses the raw data to get meaningful information that is subsequently used for a behavioral classification. This information is sent as an output to the classification unit. It also extracts the epidemic data from the raw data and sends it as an output to the epidemic unit.

Classification Unit. This component uses the behavior based data received to perform a behavioral classification. A number of classifiers exist for an automated malware classification and analysis. The work of Bailey *et al.* can in particular be mentioned[18]. They first examined the effectiveness of existing host based antivirus products in providing semantically meaningful information concerning the malicious software (or malware) and tools used by attackers. Using a large collection of malware that spanned a variety of attack vectors, it was shown that different antivirus products characterize malware in different ways. This characterization is inconsistent across antivirus products, incomplete across malware, and they fail to be concise in their semantics.

They proposed a new classification method that described malware behavior in terms of number of system state changes like files written, processes created, etc. and not in sequences or patterns of system calls. Also to address the large volume of malware and the diversity of their behavior, a method was provided to automatically categorize these profiles of malware into groups representing similar classes of behaviors. They also demonstrated how behavior based clustering provides a more direct and efficient way of classifying and analyzing Internet malware. In the present paper, we do not attempt to go into the details of the classification unit and the relative

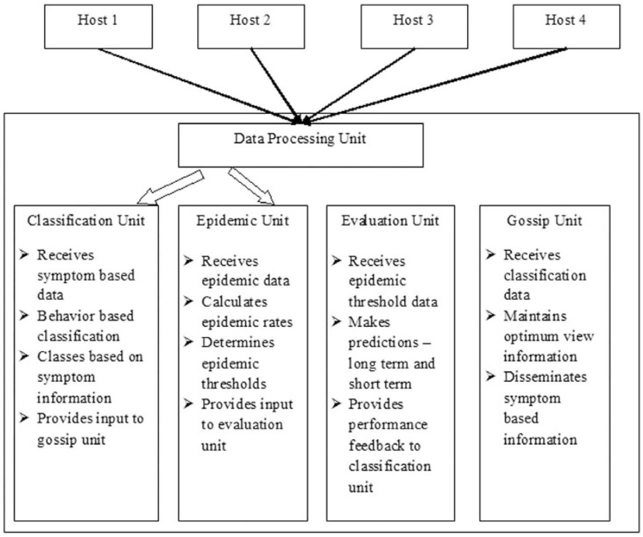
Host 1
Host 2
Host 3
Host 4
Data Processing Unit
Classification Unit
➢ Receives symptom based data
➢ Behavior based classification
➢ Classes based on symptom information
➢ Provides input to gossip unit
Epidemic Unit
➢ Receives epidemic data
➢ Calculates epidemic rates
➢ Determines epidemic thresholds
➢ Provides input to evaluation unit
Evaluation Unit
➢ Receives epidemic threshold data
➢ Makes predictions – long term and short term
➢ Provides performance feedback to classification unit
Gossip Unit
➢ Receives classification data
➢ Maintains optimum view information
➢ Disseminates symptom based information

Figure 1. DifEpGos Architecture.

efficiency of the classification algorithms that can be used, but it can be taken up as a separate work. The information regarding the number of classes, and the symptoms associated for the classification, and the corresponding optimal defense mechanism is sent as an output to the gossip unit.

Epidemic Unit. This component receives the epidemic data from the data processing unit and then uses it to find a number of values, which can be used to effectively describe the epidemic state of the whole network. These values will be in the form of a number of rates and also an epidemic threshold. Subsequent sections of this paper focus on finding these values, and establishing their relevance. Its output is sent to the evaluation unit.

Evaluation Unit. This component uses the data received by it to perform short and long term predictions regarding the epidemic state of the system. This basically allows evaluating the performance of the system, and particularly the classifier involved. A negative feedback may be used as a suggestion to fine tune the performance of the classification unit.

Gossip Unit. This component plays the essential role of disseminating the classification information to the hosts. It also needs to optimize the view that it chooses to use. An efficient working of this unit is important mainly because it is responsible for providing the backup support needed for an intelligent anticipation by the overall system.

The epidemic information needs to be analyzed, for which a suitable epidemic model is necessary. A model incorporating the difference in symptoms is proposed in the subsequent sections. We call our architecture as *DifEpGoss* architecture as it is based on such an *EPidemic* model which uses a *DIF*ference in symptoms, along with a *GOss*ip based information dissemination.

The D-SEIR Model and its mathematical formulation

In this paper, we use the *Susceptible-Exposed-Infectious-Recovered* (SEIR) model[12,19] as the basic underlying framework. Figure 2 provides a schematic representation of this model. We attempt to provide greater significance to the model, by making use of the fact that once a system gets infected and it starts to show specific symptoms, then the role of the defense mechanism can be more targeted and based on intelligent anticipation. If there is a specific response to the stage where identified symptoms are just beginning to appear, then there is a greater chance that even a strong attack can be thwarted before it becomes significant. We do not attempt to make a classification of specific symptoms or the specific defense to be adopted, but use a more abstract approach and consider n different groups or sub-classes based on the symptoms exhibited. The proposed model considers a difference between nodes based on the characteristic features or symptoms exhibited, and hence it is referred hereafter as the *differential – SEIR* or D-SEIR model. The assumptions that lead to a formulation of the model can be enlisted as follows:

Initial susceptibility. All nodes in the network are initially taken in the susceptible (S) class. This accounts for the fact that the modeling process starts at time zero for an attack. All nodes are thus non-infected at that point but have a chance of being infected, as time progresses.

Differential Infection probability. It is assumed that the probability of the susceptible nodes getting infected into the i^th^ exposed sub-class (E_i) is $p_i (i = 1, 2, \ldots, n)$, such that $\sum_{i=1}^{n} p_i = 1$. This assumption allows mathematical tractability but a point that arises out of it is whether it is essential for all nodes to become exposed, before getting recovered. This point can be included by introducing the concept of direct immunity. In the present

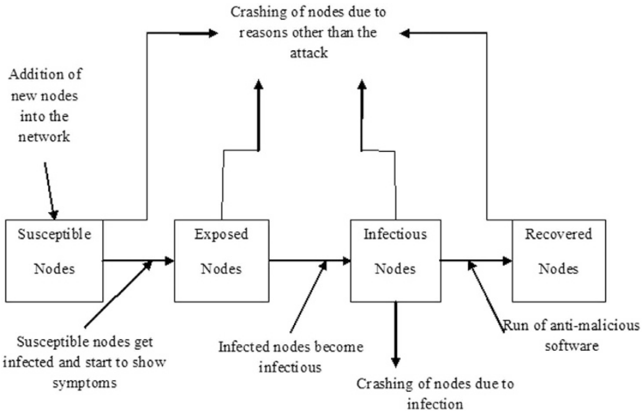
Crashing of nodes due to reasons other than the attack
Addition of new nodes into the network
Susceptible Nodes
Exposed Nodes
Infectious Nodes
Recovered Nodes
Susceptible nodes get infected and start to show symptoms
Infected nodes become infectious
Crashing of nodes due to infection
Run of anti-malicious software

Figure 2. **The SEIR (Susceptible-Exposed-Infectious-Recovered) framework.**

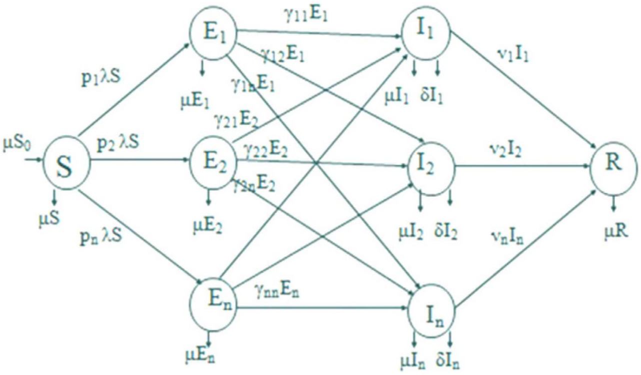

Figure 3. **Schematic Representation of the D-SEIR Model.**

paper, this factor has been ignored for the sake of simplicity, but it can very well be included for a more concrete analysis in some related future work.

Node removal from network. The removal of nodes from the network is assumed to be because of two reasons. Firstly when nodes succumb to the infection (at a rate δ), and secondly due to node failures for other reasons (at a rate μ). Other reasons may include hardware failure, physical damage, or power discharge (in case of sensor and ad hoc networks). Such kind of removal can occur from each of the four classes. Removal because of infection however takes place only from the infectious (I_j) sub-classes.

Post latent infection. A latently infected node in sub-class E_i becomes infectious and moves into sub-class I_j with a rate γ_{ij}. One fact that needs to be accounted for is that it is not at all necessary that the transitions occur only between the corresponding sub-classes (i.e. E_i to I_i only). There is a possibility that a symptom is misclassified or a node shows symptoms of more than one class. Under such a scenario, the probabilities for E_i to I_j transitions will have non-zero value. The situation where all probabilities except the corresponding ones are zero (i.e. only E_i to I_i probabilities exist) will be possible only when we have an *ideal classifier* with no classification error. In the present model, therefore small non-zero values have been assumed for E_i to I_j transitions ($i \neq j$) and for transitions between the corresponding sub-classes (when correct classification is made) higher values have been assumed.

Recovery. The infectious nodes get disinfected on use of anti-malicious measures. Upon recovery the nodes from each of the infectious sub-classes (I_j) move into the recovered class (R). The immunity is considered to be permanent based on assumptions already specified earlier (in case of the SEIR model).

Contact distribution. The average number of contacts per node is assumed as a function c(N) of the population size, i.e. $c(N) = c_0N$, where N is the total population size and c_0 is a constant of proportionality. This fits well into our homogeneity assumption. Here the nodes constituting the network are assumed to have an ability to interact and spread their infection to every other node, which is the most general form of interaction possible. The constant c_0 is the factor by which the number of contacts scales as the population of the network varies. It basically provides a best case situation for the malware to spread, and hence a worst case for the analysis. The distribution may be modified to fit in to other specific topologies.

Based on these assumptions, the model can be schematically represented as in Fig. 3 below.

Nomenclature			
S(t)	Number of nodes in the susceptible class.	n	The number of exposed and infectious sub-classes.
$E_i(t)$	Number of nodes in the i^{th} exposed sub-class.	S_0	Initial number of nodes in the network.
$I_j(t)$	Number of nodes in the j^{th} infectious sub-class.	λ	Rate of infection of susceptible nodes.
R(t)	Number of nodes in the recovered class.	γ_{ij}	Rate at which exposed nodes in the i^{th} subclass become infectious into the j^{th} subclass.
N(t)	Total number of nodes in the network.	δ	The per capita death rate due to infection.
μ	The per system death rate due to reasons other than the infection	ν_j	Rate of recovery of infectious nodes in the j^{th} sub-class.
β_j	Infectivity of nodes in j^{th} infectious sub-class.	$c(N)$	Average number of contacts per node.

Table 1. Nomenclature used in the Model.

The nomenclature of basic terminology used in the model is summarized in Table 1.

The transformations shown in Fig. 3 can be used to obtain the following system of ordinary differential equations, which gives the mathematical representation of the model.

$$\begin{aligned}
\frac{dS}{dt} &= \mu(S_0 - S) - \lambda S \\
\frac{dE_i}{dt} &= p_i\lambda S - \left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)E_i;\ i = 1, 2, \ldots, n \\
\frac{dI_j}{dt} &= \sum_{i=1}^{n}\gamma_{ij}E_i - (\mu + \delta + \upsilon_j)I_j;\ j = 1, 2, \ldots, n \\
\frac{dR}{dt} &= \sum_{j=1}^{n}\nu_j I_j - \mu R
\end{aligned} \tag{1}$$

where the total population size is

$$N = S + \sum_{i=1}^{n}E_i + \sum_{j=1}^{n}I_j + R$$

considering the average number of contacts per node c(N) as a function of the population size and β_j as the infectivity of nodes in j^{th} infectious class, the rate of infection λ for the nodes in the susceptible class can be given as

$$\lambda = c(N)\sum_{j=1}^{n}\beta_j\frac{I_j}{N} = \frac{c(N)}{N}\sum_{j=1}^{n}\beta_j I_j \tag{2}$$

here $\frac{I_j}{N}$ represents the probability that a contact with a node of infectious sub-class j results in an infection. Assuming $c(N)$ as being directly proportional to the population size N, we have $c(N) = c_0N$, and so the rate of infection reduces to a bilinear form, given as

$$\lambda = c_0\sum_{j=1}^{n}\beta_j I_j \tag{3}$$

The rate of infection is thus dependent on the total sum of infectivity of nodes, where we consider the infectivity of the corresponding infectious sub-class. In the next section, the various analytical aspects of the model are discussed.

Stability Analysis

In this section our focus is on examining the long term behavior of the network with respect to time. The model is analyzed to find conditions under which the network will recover completely or if there is a possibility that a number of nodes will remain infected. In such a case, the stable value of the persisting infectious fraction will also be found. Firstly, we establish an epidemic threshold. It will determine the conditions for long term behavior of the system and would enable us to know if the infection persists or dies out.

Epidemic threshold.

The epidemic threshold will be a value R_0 called the *basic reproduction number* (borrowing terminology from biological epidemics) which may be defined as follows.

Definition 1 (Basic Reproduction Number). The basic reproduction number (R_0) may be defined as the expected number of secondary infections produced by a single node during its entire infectious period, in a population of all susceptible nodes[20].

The value of R_0 will be used to obtain an epidemic threshold (say τ_0) which is a value such that

- the infection dies out over time if $R_0 < \tau_0$
- the infection persists and becomes an endemic if $R_0 < \tau_0$

We first obtain a value of R_0 for our model and then use it to find conditions involving the epidemic threshold.

Theorem 1. The value of the basic reproduction number for the D-SEIR model is given as

$$R_0 = c(S_0)\sum_{k=1}^{n} p_k \sum_{i=1}^{n}\left(\frac{\beta_i \gamma_{ki}}{(\mu + \delta + \nu_i)\left(\mu + \sum_{j=1}^{n}\gamma_{kj}\right)}\right) \tag{4}$$

Proof. The derivation follows along a method called the *next generation matrix* method[20–22].

A quantity that would be useful in the derivation is the partial derivative of the infection rate λ at the infection free equilibrium (IFE), which is given as

$$\left[\frac{\partial\lambda}{\partial I_j}\right]_{IFE} = \frac{\partial}{\partial I_j}\left(\frac{c(N)}{N}\sum_{j=1}^{n}\beta_j I_j\right) = \frac{c(S_0)}{S_0}\beta_j;\ j = 1, 2, \ldots, n \tag{5}$$

The Jacobian at the infection free equilibrium is given as

$$J = \begin{bmatrix} -\mu & 0 & D_{13} & 0 \\ 0 & D_{22} & D_{23} & 0 \\ 0 & D_{32} & D_{33} & 0 \\ 0 & 0 & D_{43} & -\mu \end{bmatrix} \tag{6}$$

where the elements of the block matrix J are given as follows

$$D_{13} = [-c(S_0)\beta_1 - c(S_0)\beta_2 \ldots - c(S_0)\beta_n]$$

$$D_{22} = \text{diag}\left[-\left(\mu + \sum_{j=1}^{n}\gamma_{1j}\right), -\left(\mu + \sum_{j=1}^{n}\gamma_{2j}\right), \ldots, -\left(\mu + \sum_{j=1}^{n}\gamma_{nj}\right)\right] \tag{7}$$

$$D_{23} = \begin{bmatrix} c(S_0)p_1\beta_1 & c(S_0)p_1\beta_2 & \cdots & c(S_0)p_1\beta_n \\ c(S_0)p_2\beta_1 & c(S_0)p_2\beta_2 & \cdots & c(S_0)p_2\beta_n \\ \cdots & \cdots & \cdots & \cdots \\ c(S_0)p_n\beta_1 & c(S_0)p_n\beta_2 & \cdots & c(S_0)p_n\beta_n \end{bmatrix}$$

$$D_{32} = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{n1} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{1n} & \gamma_{2n} & \cdots & \gamma_{nn} \end{bmatrix}$$

$$\begin{aligned} D_{33} &= \text{diag}[-(\mu + \delta + \nu_1), -(\mu + \delta + \nu_2), \ldots -(\mu + \delta + \nu_n)] \\ D_{43} &= [\nu_1 \nu_2 \ldots \nu_n] \end{aligned}$$

We consider the system of equations with the infected classes represented first, and from it we obtain the matrices representing the rate of appearance of new infections ($\hat{F}_{ij}$) and the matrices representing the difference between outward and inward flow of nodes into a compartment ($\hat{V}_{ij}$) as follows

$$\hat{F}_{ij} = \begin{bmatrix} p_i\lambda S \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \hat{V}_{ij} = \begin{bmatrix} \left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)E_i \\ -\sum_{i=1}^{n}\gamma_{ij}E_i + (\mu + \delta + \nu_j)I_j \\ -\mu(S_0 - S) + \lambda S \\ -\sum_{j=1}^{n}\nu_j I_j + \mu R \end{bmatrix}$$

Next, taking the partial derivatives with respect to the infectious classes, we get

$$F_{ij} = \begin{bmatrix} 0 & c(S_0)p_i\beta_j \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V_{ij} = \begin{bmatrix} \left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right) & 0 \\ -\gamma_{ij} & \mu + \delta + \nu_j \end{bmatrix}$$

and then generalizing we have,

$$F = \begin{bmatrix} 0 & D_{23} \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} -D_{22} & 0 \\ -D_{32} & -D_{33} \end{bmatrix}$$

where the sub-matrices are as defined in (7), and the two matrices are observed to be non-negative and non-singular respectively. Now the basic reproduction number is given as the spectral radius (ρ) of the next generation operator FV^{-1} and so we have

$$R_0 = \rho(FV^{-1})$$

where, the inverse of the block matrix V is given as $V^{-1} = \begin{bmatrix} -D_{22}^{-1} & 0 \\ D_{33}^{-1}D_{32}D_{22}^{-1} & -D_{33}^{-1} \end{bmatrix}$

So, the value of R_0 becomes

$$R_0 = \text{trace}\left(\begin{bmatrix} 0 & D_{23} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -D_{22}^{-1} & 0 \\ D_{33}^{-1}D_{32}D_{22}^{-1} & -D_{33}^{-1} \end{bmatrix}\right) = \text{trace}(D_{23}D_{33}^{-1}D_{32}D_{22}^{-1})$$

whose diagonal elements are $c(S_0)\sum_{i=1}^{n} p_1\beta_i\gamma_{1i}(\mu + \delta + \nu_i)^{-1}\left(\mu + \sum_{j=1}^{n}\gamma_{1j}\right)^{-1}$,

$$c(S_0)\sum_{i=1}^{n} p_2\beta_i\gamma_{2i}(\mu + \delta + \nu_i)^{-1}\left(\mu + \sum_{j=1}^{n}\gamma_{2j}\right)^{-1}, \ldots \text{ and}$$

$$c(S_0)\sum_{i=1}^{n} p_n\beta_i\gamma_{ni}(\mu + \delta + \nu_i)^{-1}\left(\mu + \sum_{j=1}^{n}\gamma_{nj}\right)^{-1}$$

and hence we have

$$R_0 = c(S_0)\sum_{k=1}^{n} p_k \sum_{i=1}^{n}\left(\frac{\beta_i\gamma_{ki}}{(\mu + \delta + \nu_i)\left(\mu + \sum_{j=1}^{n}\gamma_{kj}\right)}\right) \tag{8}$$

□

Equilibrium Points. For a system of differential equations, an equilibrium point (also called critical point or equilibrium solution) may be defined as follows:

Definition 2 (Equilibrium Point). For a system of differential equations

$$\frac{dx}{dt} = Ax$$

a substitution of zero in the right hand side gives points that correspond to constant solutions (that do not change with time) and are called equilibrium points[23].

The D-SEIR model has two equilibrium points. The first of these has a zero value for all I_j (and in fact for all E_i as well) and hence is referred to as the *infection free equilibrium* (IFE) point. The second one, on the other hand, has a positive component of infection and hence is called an *endemic equilibrium* point.

Infection free equilibrium: The infection free equilibrium point for the D-SEIR model is given as $(S = \frac{\mu}{\mu + \lambda}S_0,\ E_i = 0,\ I_j = 0;\ i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n,\ R = 0)$

Endemic equilibrium: The endemic equilibrium point for the D-SEIR model is given as

$$
\begin{aligned}
S^* &= \frac{\mu S_0}{\mu + \lambda} \\
E_i{}^* &= \frac{p_i \lambda \mu S_0}{(\mu + \lambda)\left(\mu + \sum_{j=1}^{n} \gamma_{ij}\right)} \\
I_j{}^* &= \frac{\lambda \mu S_0}{(\mu + \delta + \nu_j)(\mu + \lambda)} \sum_{i=1}^{n} \frac{p_i \gamma_{ij}}{\mu + \sum_{k=1}^{n} \gamma_{ik}} \\
R^* &= \sum_{j=1}^{n} \frac{\lambda \nu_j S_0}{(\mu + \delta + \nu_j)(\mu + \lambda)} \sum_{i=1}^{n} \frac{p_i \gamma_{ij}}{\mu + \sum_{k=1}^{n} \gamma_{ik}}
\end{aligned}
$$

where

$$
\lambda = c_0 \sum_{j=1}^{n} \beta_j I_j = \frac{c_0 \lambda \mu S_0}{\mu + \lambda} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\beta_j p_i \gamma_{ij}}{(\mu + \delta + \nu_j)(\mu + \sum_{k=1}^{n} \gamma_{ik})} \tag{9}
$$

The relevance of the endemic equilibrium point is that it gives a quantitative measure for the infected population, when the infection survives. This allows us to have an estimate for the number of nodes that are expected to be infected in the long run.

Value of the threshold τ_0. The definition of the basic reproduction number as the expected number of secondary infections induced by a single infected host, leads us to an intuitive idea the threshold has to be one. This is because when an infected node infects at least one other node, then only we can expect the infection to spread. In the following we give a mathematical reasoning for this intuitive idea.

We consider a function f to be defined as

$$
f(\lambda) = c_0 \mu S_0 \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\beta_j p_i \gamma_{ij}}{(\mu + \delta + \nu_j)(\mu + \sum_{k=1}^{n} \gamma_{ik})(\mu + \lambda)} - 1
$$

where λ is given by (9). Then the endemic equilibrium exists if and only if there exists a positive solution of $f(\lambda) = 0$.

Now its derivative with respect to λ is $f'(\lambda) = -c_0 \mu S_0 \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\beta_j p_i \gamma_{ij}}{(\mu + \delta + \nu_j)\left(\mu + \sum_{k=1}^{n} \gamma_{ik}\right)(\mu + \lambda)^2} < 0$

Also $f(\lambda) \to -1$ *as* $\lambda \to \infty$ and further $f(0) = c_0 S_0 \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\beta_j p_i \gamma_{ij}}{(\mu + \delta + \nu_j)\left(\mu + \sum_{k=1}^{n} \gamma_{ik}\right)} - 1 = R_0 - 1$

Therefore, a solution exists for $f(\lambda) = 0$, and hence an endemic equilibrium exists if and only if $f(0) > 0$ or $R_0 > 1$.

Stability of the Infection Free Equilibrium. The infection free equilibrium, as explained above, corresponds to a state where the infection disappears in the long run from the network. The epidemic threshold condition for this scenario was already seen to be $R_0 < \tau_0$ and we also established the value of τ_0 to be one. Next we consider the impact of a small or a large perturbation on the stability of the infection free equilibrium point. Stability on the face of a large perturbation will give us a guarantee that the infection will continue to disappear, even if the attack uses a large number of nodes initially. We thus consider the impact of a minor attack (corresponding to a small perturbation – referred to as local stability) and that of a major attack (corresponding to a large perturbation – referred to as global stability) on the stability of the infection free equilibrium. A mathematical proof is provided for the more stronger case of global stability.

Theorem 2. The infection free equilibrium is globally asymptotically stable when $R_0 < 1$.

Proof. Here we prove the global stability of the infection free equilibrium (IFE) using Lyapunov's method. For the total collection of nodes, we have

$$
\frac{dN}{dt} = \mu(S_0 - N) - \delta \sum_{j=1}^{n} I_j
$$

and so, $N \in [0, S_0]$. Then the domain $\Gamma = \{(S, E, I, R) \mid 0 \le N \le S_0\}$ where, $E = (E_1, E_2, \ldots, E_n)^T$ and $I = (I_1, I_2, \ldots, I_n)^T$, is positive time-invariant set for the system (1). A real-valued function L defined on Γ is selected, which is analogous to the potential function of classical dynamics, which is popularly referred to as the *Lyapunov function*. The function needs to have a non-negative value at all points in tis domain, and for stability at an equilibrium point it needs to have a zero value there and its time derivative at nearby points needs to be negative. This corresponds to the energy of a system which will dissipate as it approaches an equilibrium point. The choice of L considers the transitions for the infectious classes, and in general the i^{th} exposed sub-class and the j^{th} infectious sub-class has been taken into consideration. We consider the function

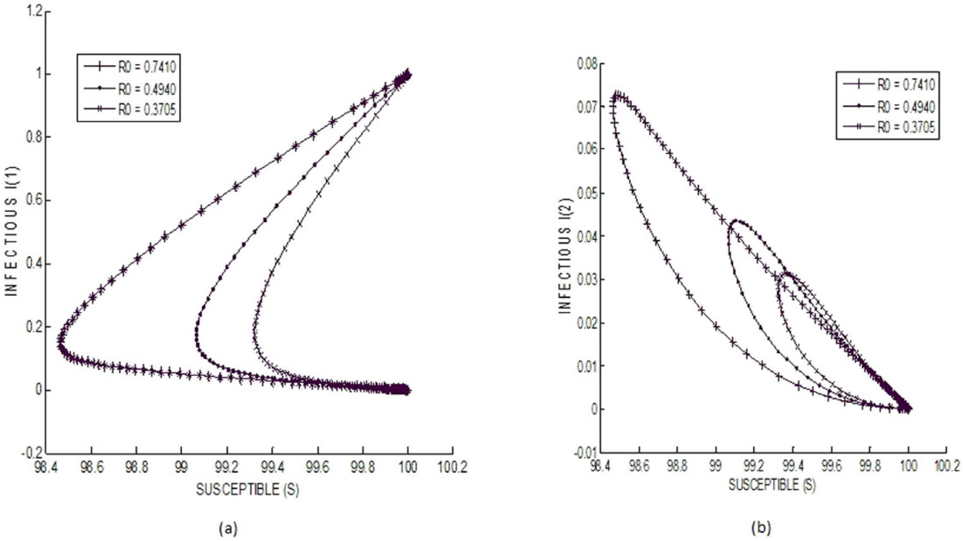

Figure 4. Local stability of infection free equilibrium.

$$L = \sum_{i=1}^{n}\gamma_{ij}E_i + \left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)I_j \tag{10}$$

Then at the infection free equilibrium $L(IFE) = 0$ and otherwise $L(x \neq IFE) > 0$, for all $x \in \Gamma$. Also, the time derivative of L is given as

$$\frac{dL}{dt} = \sum_{i=1}^{n}\gamma_{ij}\frac{dE_i}{dt} + \left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)\frac{dI_j}{dt}$$

which on substitution of values of the derivatives from (1) becomes

$$\frac{dL}{dt} = \sum_{i=1}^{n}\gamma_{ij}\left(p_i\lambda S - \left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)E_i\right) + \left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)(\gamma_{ij}E_i - (\mu + \delta + \nu_j)I_j)$$

On simplification and cancellation of common terms and also using the value of the rate of infection λ, the equation reduces to the following form

$$\frac{dL}{dt} = (\mu + \nu_j)\left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)\left(c(S_0)\sum_{i=1}^{n}p_i\sum_{j=1}^{n}\frac{\beta_j\gamma_{ij}}{(\mu + \nu_j)(\mu + \sum_{k=1}^{n}\gamma_{ik})} - 1\right)I_j$$

where the dummy index in the denominator has been changed to avoid repetition. Replacing the indices i, j and k inside the parenthesis with k, i and j respectively, the expression becomes

$$\frac{dL}{dt} = (\mu + \nu_j)\left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)\left(c(S_0)\sum_{k=1}^{n}p_k\sum_{i=1}^{n}\frac{\beta_i\gamma_{ki}}{(\mu + \nu_i)\left(\mu + \sum_{j=1}^{n}\gamma_{kj}\right)} - 1\right)I_j$$

Using the value of R_0 from (8), gives

$$\frac{dL}{dt} = (\mu + \nu_j)\left(\mu + \sum_{j=1}^{n}\gamma_{ij}\right)(R_0 - 1)I_j \tag{11}$$

hence $dL/dt < 0$ if $R_0 < 1$ and $dL/dt = 0$ if and only if $R_0 = 1$ or $Ij = 0$, i.e. at the infection free equilibrium. As such the function L is positive definite over Γ and its time derivative is negative definite. Moreover when $R_0 > 1$, then $dL/dt > 0$ if $Ij > 0$, which shows that the infection free equilibrium is unstable when $R_0 > 1$. This validates the claim that the infection free equilibrium is globally asymptotically stable if $R_0 \leq 1$. □

Next we use numerical simulations on the D-SEIR model to illustrate the fact that the infection free equilibrium is both locally and globally stable (asymptotic stability follows because the trajectories approach the equilibrium point for large values of time).

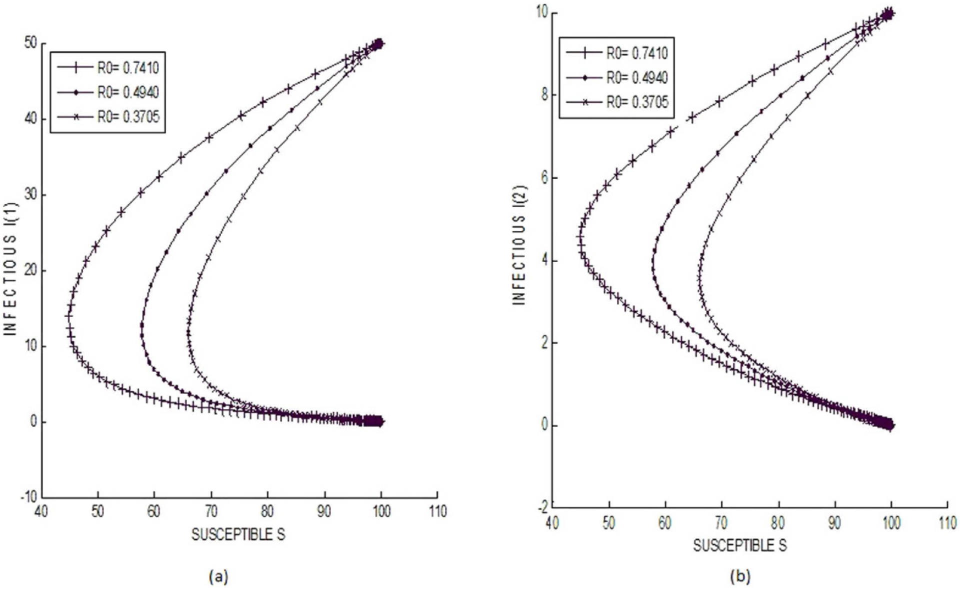

Figure 5. **Global stability of infection free equilibrium.**

Starting with an initial population of 100 susceptible nodes and 1 I_1 node (for small perturbation – minor attack), the system is seen to stabilize to the infection free equilibrium (Fig. 4). Three different values of R_0 have been used (0.7410, 0.4940 and 0.3705), all of which do not violate the threshold condition of $R_0 \leq 1$. In all the simulated cases, a total of three sub-classes have been considered for both the exposed (E_1, E_2, E_3) as well as infectious (I_1, I_2, I_3) classes. In Fig. 4, the asymptotic behaviors of the sub-classes I_1 and I_2 are shown. A difference is observed in the two sets of curves, owing only to the fact that they have a different initial infectious value (initially I_1 is 1 and I_2 is zero). Behaviorally they are similar behavior because in both cases the final value of I (i.e. I_1 and I_2) are both zero.

Next we consider 50 I_1 nodes initially and 10 nodes each of I_2 and I_3 subclasses (large perturbation). In Fig. 5, both I_1 and I_2 populations are seen to converge to the infection free equilibrium point. This shows the global stability of the infection free equilibrium for $R_0 \leq 1$.

Stability of Endemic Equilibrium.

Theorem 3. The endemic equilibrium is globally stable when $R_0 > 1$.

Proof. We use the geometric approach suggested by Li and Muldowney[24] to prove the global stability condition for the endemic equilibrium point. Based on this approach, it is known that if the mapping f: $D \subset R^n \rightarrow R^n$, where D is an open set, be such that each solution x(t) of the differential equation $x' = f(x)$ is uniquely determined by its initial value $x(0) = x_0$, and an equilibrium point $\overline{x} \in D$ satisfies the following assumptions

(A1) D is simply connected
(A2) There is a compact absorbing set $K \subset D$
(A3) $\overline{x}$ is the only equilibrium point in D,

then the global stability of $\overline{x}$ in D is given by the additional *Bendixson criteria*

$$\overline{q}_2 = \lim_{t\to\infty} \sup \sup_{x_0 \in K} \frac{1}{t} \int_0^t \mu(B(x(s, x_0)))ds < 0$$

In this criteria, $x(t, x_0)$ denotes the solution $x(t)$ determined by the initial point x_0, and B is given as

$$B = A_f A^{-1} + A\frac{\partial f^{[2]}}{\partial x} A^{-1}$$

where $J^{[2]} = \frac{\partial f^{[2]}}{\partial x}$represents the second compound Jacobian matrix given as

$$J^{[2]} = \begin{bmatrix} j_{11} + j_{22} & j_{23} & -j_{13} \\ j_{32} & j_{11} + j_{33} & j_{12} \\ -j_{31} & j_{21} & j_{22} + j_{33} \end{bmatrix}$$

and A is a matrix-valued function satisfying

$$\mu(A_f A^{-1} + AJ^{[2]} A^{-1}) \leq -\delta < 0$$

on K and μ denotes the Lozinskii measure, given as

$$\mu(M) = \lim_{h \to 0^+} \frac{|I + hM| - 1}{h}$$

for an $N \times N$ matrix M.

Now the existence of a compact set that is absorbing in the interior of Γ follows from the uniform persistence of the system, where it can be shown that

$$\lim_{t\to\infty} \inf\ S(t) > c,\ \lim_{t\to\infty} \inf\ E_i(t) > c \ \text{ and } \ \lim_{t\to\infty} \inf\ I_j(t) > c \text{ for some } c > 0.$$

The proof for the Bendixson criteria $\bar{q}_2 < 0$, can be enumerated in the form of the following steps:

(1) Jacobian of reduced system: The reduced system obtained by neglecting the recovered class, which is possible because of its non-involvement in the dynamics of the other classes, is given as

$$\begin{aligned} \frac{dS}{dt} &= \mu(S_0 - S) - \lambda S \\ \frac{dE_i}{dt} &= p_i \lambda S - \left(\mu + \sum_{j=1}^{n} \gamma_{ij}\right) E_i;\ i = 1, 2, \ldots, n \\ \frac{dI_j}{dt} &= \sum_{i=1}^{n} \gamma_{ij} E_i - (\mu + \delta + \upsilon_j) I_j;\ j = 1, 2, \ldots, n \end{aligned}$$

where $\lambda = \frac{c(N)}{N} \sum_{j=1}^{n} \beta_j I_j$

Then the Jacobian matrix of the reduced system is given as

$$J = \begin{bmatrix} -\mu - \lambda & 0 & D_{13} \\ D_{21} & D_{22} & D_{23} \\ 0 & D_{32} & D_{33} \end{bmatrix}$$

where

$$D_{13} = \left[-\frac{c(N)}{N}\beta_1 S \ \ -\frac{c(N)}{N}\beta_2 S \ \ \ldots \ \ -\frac{c(N)}{N}\beta_n S \right]$$

$$D_{21} = [p_1 \lambda \ p_2 \lambda \ldots p_n \lambda]^T$$

$$D_{22} = \text{diag}\left[-\left(\mu + \sum_{j=1}^{n} \gamma_{1j}\right), -\left(\mu + \sum_{j=1}^{n} \gamma_{2j}\right), \ldots., -\left(\mu + \sum_{j=1}^{n} \gamma_{nj}\right) \right]$$

$$D_{23} = \begin{bmatrix} p_1 \frac{c(N)}{N}\beta_1 S & p_1 \frac{c(N)}{N}\beta_2 S & \ldots & p_1 \frac{c(N)}{N}\beta_n S \\ p_2 \frac{c(N)}{N}\beta_1 S & p_2 \frac{c(N)}{N}\beta_2 S & \ldots & p_2 \frac{c(N)}{N}\beta_n S \\ \ldots & \ldots & \ldots & \ldots \\ p_n \frac{c(N)}{N}\beta_1 S & p_n \frac{c(N)}{N}\beta_2 S & \ldots & p_n \frac{c(N)}{N}\beta_n S \end{bmatrix}$$

$$D_{32} = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{n1} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{1n} & \gamma_{2n} & \cdots & \gamma_{nn} \end{bmatrix}$$

$$D_{33} = \text{diag}[-(\mu + \delta + \nu_1), -(\mu + \delta + \nu_2), \ldots -(\mu + \delta + \nu_n)]$$

(2) Second Compound Matrix of the Jacobian: For the Jacobian matrix of the reduced system obtained above, the second additive compound matrix is given as

$$
\begin{aligned}
J^{[2]} &= \begin{bmatrix} j_{11} + j_{22} & j_{23} & -j_{13} \\ j_{32} & j_{11} + j_{33} & j_{12} \\ -j_{31} & j_{21} & j_{22} + j_{33} \end{bmatrix} \\
&= \begin{bmatrix} -\mu - \lambda + D_{22} & D_{23} & -D_{13} \\ D_{32} & -\mu - \lambda + D_{33} & 0 \\ 0 & D_{21} & D_{22} + D_{33} \end{bmatrix}
\end{aligned}
$$

(3) Definition of matrix B in the Bendixson Criteria: We consider a diagonal matrix A defined as

$$A(S, E_i, I_j) = \operatorname{diag}\left(1, \frac{E_i}{I_j}, \frac{E_i}{I_j}\right)$$

considering in general the i^th exposed sub-class and the j^th infectious sub-class. If f denotes the vector field of the system, then

$$A_f A^{-1} = \operatorname{diag}\left(0, \left(\frac{E_i}{I_j}\right)_f \frac{I_j}{E_i}, \left(\frac{E_i}{I_j}\right)_f \frac{I_j}{E_i}\right)$$

and

$$
\begin{aligned}
AJ^{[2]}A^{-1} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{E_i}{I_j} & 0 \\ 0 & 0 & \frac{E_i}{I_j} \end{bmatrix} \begin{bmatrix} -\mu - \lambda + D_{22} & D_{23} & -D_{13} \\ D_{32} & -\mu - \lambda + D_{33} & 0 \\ 0 & D_{21} & D_{22} + D_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{I_j}{E_i} & 0 \\ 0 & 0 & \frac{I_j}{E_i} \end{bmatrix} \\
&= \begin{bmatrix} -\mu - \lambda + D_{22} & D_{23}\frac{I_j}{E_i} & -D_{13}\frac{I_j}{E_i} \\ \frac{E_i}{I_j}D_{32} & -\mu - \lambda + D_{33} & 0 \\ 0 & D_{21} & D_{22} + D_{33} \end{bmatrix}
\end{aligned}
$$

using these two matrices, gives the block matrix

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where

$$B_{11} = [-\mu - \lambda + D_{22}]$$

$$B_{12} = \left[D_{23}\frac{I_j}{E_i} - D_{13}\frac{I_j}{E_i}\right]$$

$$B_{21} = \left[\frac{E_i}{I_j}D_{32}\ 0\right]$$

$$B_{22} = \begin{bmatrix} -\mu - \lambda + D_{33} + \frac{I_j}{E_i}\left(\frac{E_i}{I_j}\right)_f & 0 \\ D_{21} & D_{22} + D_{33} + \frac{I_j}{E_i}\left(\frac{E_i}{I_j}\right)_f \end{bmatrix}$$

(4) Lozinskii Measure of matrix B: The Lozinskii measure for matrix B can be estimated as

$$\mu(B) \leq \sup\{g_1, g_2\}$$

where g_1 and g_2 are defined as

$$\begin{aligned} g_1 &= \mu_1(B_{11}) + |B_{12}| \\ &= -\mu - \lambda - \left(\mu + \sum_{j=1}^{n} \gamma_{ij}\right) + p_i \lambda \frac{S}{E_i} \end{aligned}$$

and,

$$\begin{aligned} g_2 &= \mu_1(B_{22}) + |B_{21}| \\ &= -\mu - \lambda + p_i \lambda - (\mu + \delta + \nu_1) + \frac{I_j}{E_i}\left(\frac{E_i}{I_j}\right)_f + \sum_{i=1}^{n} \frac{E_i}{I_j} \gamma_{ij} \end{aligned}$$

here the Lozinskii measure μ is with respect to the vector norm defined as

$$|(u, v, w)| = \sup\{|u|, |v| + |w|\}$$

while the Lozinskii measure μ_1 is with respect to the l_1 norm and the norms of matrices B_{12} and B_{21} are also obtained with respect to the l_1 vector norm.

$$\text{Now } \frac{I_j}{E_i}\left(\frac{E_i}{I_j}\right)_f = \frac{I_j}{E_i}\left(\frac{I_j E_i' - E_i I_j'}{I_j^2}\right) = \frac{E_i'}{E_i} - \frac{I_j'}{I_j}$$

and so from the reduced set of equations,

$$\frac{E_i'}{E_i} + \left(\mu + \sum_{j=1}^{n} \gamma_{ij}\right) = p_i \lambda \frac{S}{E_i}$$

and

$$\frac{I_j'}{I_j} + (\mu + \delta + \nu_j) = \sum_{i=1}^{n} \gamma_{ij} \frac{E_i}{I_j}$$

using these values in the equations for g_1 and g_2, gives

$$g_1 = -\mu - \lambda + \frac{E_i'}{E_i}$$

and

$$g_2 = -\mu - \lambda + p_i \lambda + \frac{E_i'}{E_i}$$

so, the Lozinskii measure of matrix B becomes

$$\begin{aligned} \mu(B) &\le \sup\{g_1, g_2\} \\ &\le \frac{E_i'}{E_i} - \mu - \lambda + \sup\{0, p_i \lambda\} \end{aligned}$$

which finally gives

$$\frac{1}{t}\int_0^t \mu(B)\,dt \le \log\ E_i(t) - (\mu + \lambda) < -\frac{1}{2}(\mu + \lambda) < 0$$

for all ($S(0)$, $E_i(0)$, $I_j(0)$) in the absorbing set, where the bound on the sizes of the classes are implied by the uniform persistence of the system.

Hence it is shown that the additional criterion $\bar{q}_2 < 0$ is also satisfied and thus the endemic equilibrium is globally stable. This condition also itself proves the local stability of the endemic equilibrium[24]. □

Numerical simulations are once again used to clearly illustrate the situation in a phase plane. In Fig. 6, it is shown that for a value of $R_0 = 1.1116$ (which exceeds the threshold value), there exists a stable endemic equilibrium point at ($S^* = 94.6992$, $E_1^* = 2.0738$, $E_2^* = 1.0371$, $E_3^* = 0.3456$, $I_1^* = 0.2592$, $I_2^* = 0.2304$, $I_3^* = 0.2034$, $R^* = 0.4575$). The figure shows the phase plane formed by the variables S (susceptible class) and I_1 (first infectious sub-class). The trajectories are seen to asymptotically approach the stable endemic equilibrium point. The equilibrium point is unique and globally stable in the entire phase plane, as can be clearly seen.

In Fig. 7, the stability condition is verified using the phase plane formed by the variables S (susceptible class) and E_1 (first exposed sub-class). In this case, the equilibrium point is observed to be ($S^* = 65.6816$, $E_1^* = 13.4289$, $E_2^* = 6.7145$, $E_3^* = 2.2378$, $I_1^* = 1.6786$, $I_2^* = 1.4921$, $I_3^* = 1.3166$, $R^* = 2.9623$) and can be seen to be globally asymptotically stable. Here, each of the trajectories assumes initially 10 infective nodes in the population.

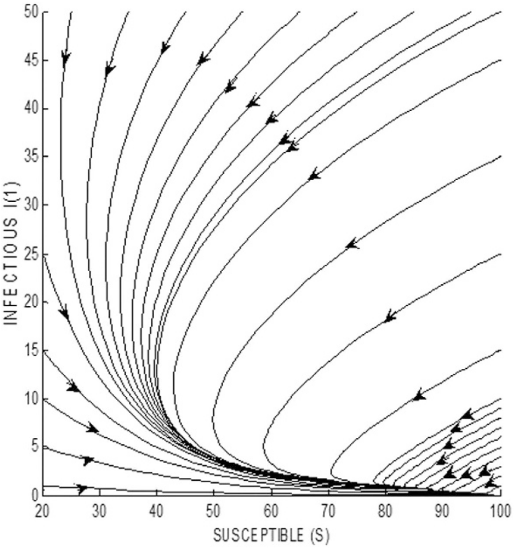

Figure 6. Global stability of endemic equilibrium point when $R_0 > 1$ depicted in $S - I_1$ phase plane.

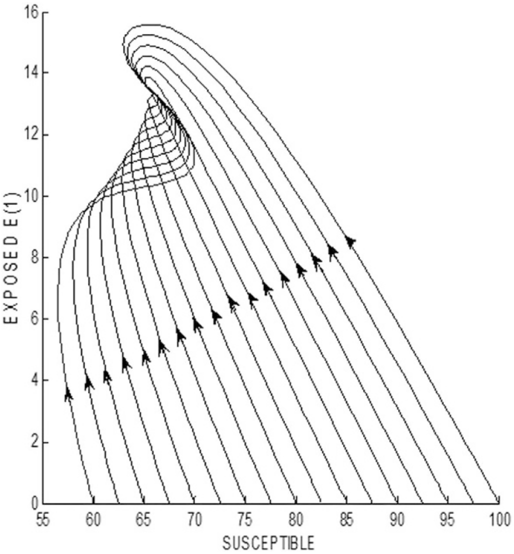

Figure 7. Global stability of endemic equilibrium point when $R_0 > 1$ depicted in $S - E_1$ phase plane.

In the next section, experiments are performed for both real and synthetic data to explore the validity of the proposed model.

Numerical Simulations

In the previous section we obtained a threshold condition defined in terms of the basic reproduction number. Conditions for the infection to disappear with time or to persist into an endemic were also highlighted. Experiments through simulations were already used to verify the results as they were analytically obtained in section 4. In this section we perform some more experiments, mainly to bring forth the impact of the classification into sub-classes that was suggested in the model. We first perform the experiments with a real network dataset. It shows that the results can be generalized and applied to networks, with varying underlying topology.

Real network data. The experiments on real network datasets which includes an AS (autonomous systems) graph instance containing AS-level connectivities inferred from the Oregon route-views. Three datasets are used-peer.oregon.010331, peer.oregon.010414 and peer.oregon.010505, which are available online at http://topology.eecs.umich.edu/data.html. The datasets contain pairs of interconnected ASs according to the Oregon route-views of a given collection date. The total number of undirected edges in the resulting AS graph for the three datasets were 22002, 22469 and 22607 respectively.

In the experiments three situations are explored:

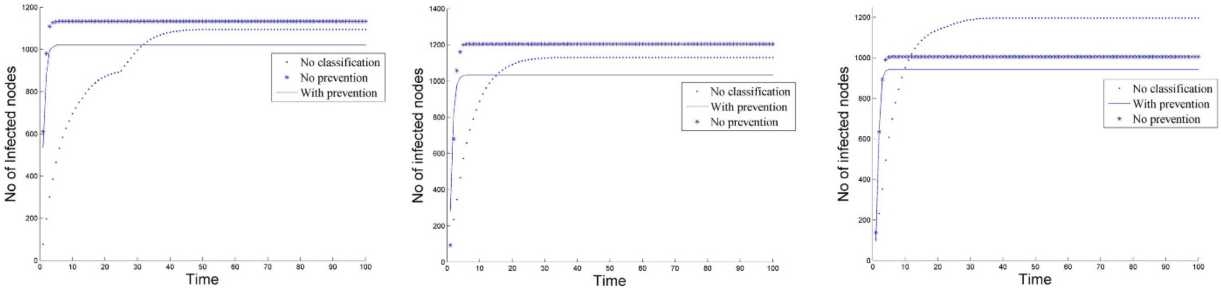

Figure 8. Performance of the D-SEIR model on the Oregon datasets.

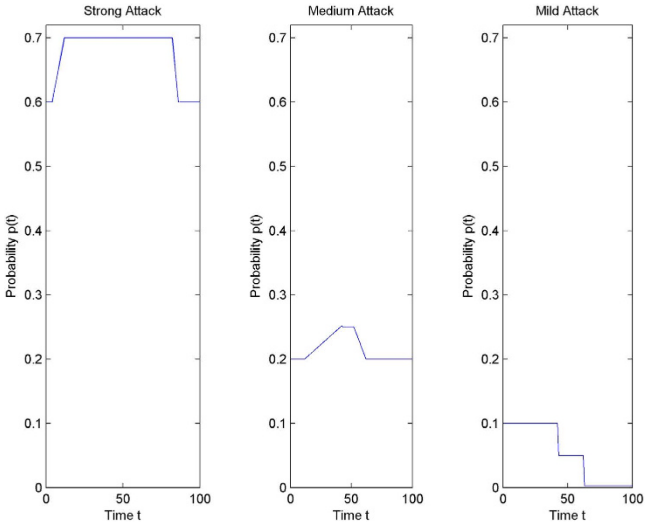

Figure 9. Characterization of Attack Categories.

- **No classification:** Spread of infection on the network without any classification (with an infection of every fifth susceptible node encountered and a recovery of one node after every 10 units of time).
- **No prevention:** Spread of infection on the network with classification into two kinds of attack (one with infection as earlier and the other for every susceptible node encountered; the recovery remains same).
- **D-SEIR model (with prevention):** Spread of infection on the network with classification and prevention of more severe attack (infection same; recovery of critical class is one node for every time unit, and for other class it is every fifth time unit).

The results are shown in Fig. 8. The third case is clearly seen to have the least infection value once the network stabilizes. For all the three datasets, initially 20 nodes were randomly selected to spread the infection. The asymptotic values were plotted in each of the cases.

Simulative experiments.

The experiments performed in this section are based on the following assumptions:

Attack and defence categories. Three abstract categories of attacks have been considered – strong, medium and mild. Analogously the defence mechanism is also considered to be of three types – strong, moderate and weak. Three classes have been assumed in the model for the exposed (E_1, E_2, E_3) and infectious (I_1, I_2, I_3) populations. It is also assumed that the first class in the model is equipped with a strong defence system, the second class with a moderate defence and the third class with a weak defence.

Variable parameters. The parameter values are assumed to be variable. This makes it possible to quantitatively analyze more realistic scenarios with respect to the interaction between the attacking and defence mechanisms. The attacking scenario is represented through a variable probability of infection (p_i), whose characterization is as shown in Fig. 9. From the figure, it can be seen that:

- A strong attack takes effect instantaneously, i.e. it spreads at a very fast rate and infects as many systems, as quickly as possible. It has a large impact for a long duration. It then slowly starts to lose its effect, because of

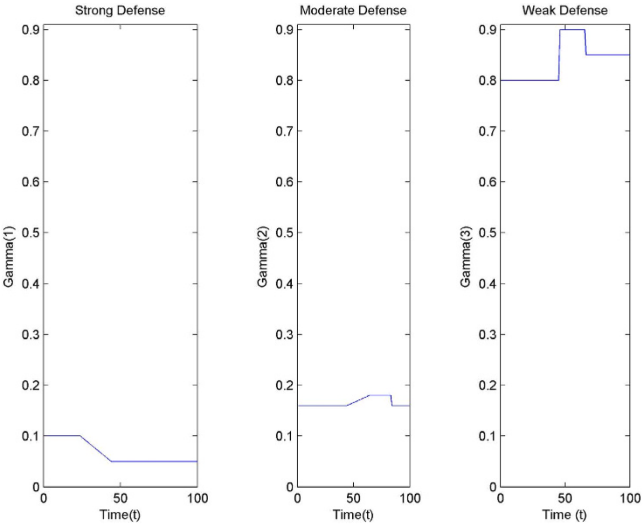

Figure 10. Characterization of Defense Categories .

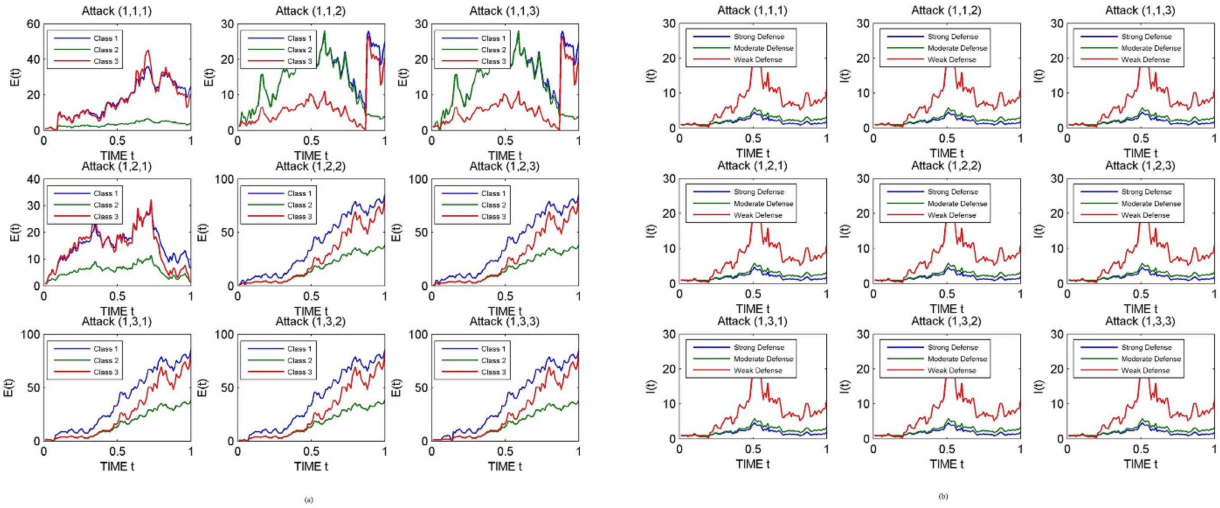

Figure 11. (**a**) Dynamical behavior of system for a strong attack in class 1 (**b**). Impact of Model on Dynamical Behavior of System for a Strong Attack in Class 1

various reasons like network congestion, or inability of the scanning algorithm to further detect more vulnerable systems.

- A medium attack begins with a lesser impact compared to a strong attack. It then spreads but again to a comparatively smaller scale. It also takes greater time to reach its peak stage, and also spends relatively lesser time in this stage. Both the strong and medium attacks have been characterized using trapezoidal functions, but with different slopes, peak values and time spent at the peak.
- Mild attacks have been characterized using a two step decreasing function. This allows us to consider a very less starting impact, which subsequently becomes negligible.

The characterization of the defense scenario is shown in Fig. 10. Here the modeled parameters are γ_i where the value $i = 1$ represents a strong defense, $i = 2$ represents a moderate defense while $i = 3$ represents a weak defense scenario.

Impact of D-SEIR model. Next we consider 27 possibilities arising from our consideration of 3 attack types and 3 defense types for 3 classes. This number will change depending on the actual considerations. In Fig. 11(a), a strong attack has been considered in the first class (shown by a 1 as the first element in the attack triplet). Consequently greater and faster spread of infection is observed in class 1. In Fig. 11(b), it can be seen that even a strong attack is not able to survive when the D-SEIR model uses a strong prevention.

It is thus clear that a distributed defense is able to change the course of even a very strong attack.

Conclusion

Epidemic studies are known to provide important insights on network epidemics. Various kind of information may be obtained including the scale and long-term behavior of an attack. Epidemic models however still do not use available information to improve the model performance. In this paper, the utility of including available information in controlling the spread of a network epidemic was explored. A 1-n-n-1 type differential epidemic model has been proposed and analyzed to see the improvement in quality of an epidemic system. An overall epidemic architecture is also suggested that can be useful in providing a more practical utility to the epidemic models. An epidemic threshold of the system was obtained which clearly demarcated the long-term behavior of a network epidemic into two exhaustive classes, one with persistent infection and the other without any infection. An analysis of real network datasets also revealed a better performance for the model in controlling an epidemic when compared to previous models. Simulation based experiments allowed us to perform generalized scenario based experiments, which again corroborated the analytical findings. In future, the model can be extended to deal with different specific network topologies.

References

1. Chen, Z., Gao, L. & Kwait, K. Modeling the spread of active worms, *INFOCOM 2003, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, California, USA: IEEE*. (doi: 10.1109/INFCOM.2003.1209211) (2003, March 30-April 3).
2. Zou, C. C., Gong, W. & Towsley, D. Code Red Worm Propagation Modeling and Analysis. *9th ACM Conference on Computer and Communications Security, Washington DC, USA: ACM*. (doi: 10.1145/586110.586130) (2002, November 18–22).
3. Machie, A., Roculan, J., Russell, R. & Velzen, M. V. Incident Analysis Report on Nimda Worm Analysis. http://www.di.unisa.it/~ads/corso-security/www/CORSO-0102/NIMDA/link_locali/010921-Analysis-Nimda-v2.pdf. (2001) (Date of access: 10th June, 2015).
4. Wood, P. *et al.* Symantec Corporation Internet Security Threat Report 2011 Trends. http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_2011_21239364.en-us.pdf. (2012) (Date of access: 10th June, 2015).
5. Symantec Corporation Internet Security Threat Report 2012 Trends. http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v18_2012_21291018.en-us.pdf. (2013) (Date of access: 10th June, 2015).
6. Cohen, F. Computer Viruses, PhD thesis, University of Southern California. http://all.net/books/Dissertation.pdf. (1985) (Date of access: 19th June, 2015).
7. Adleman, L. M. An abstract theory of computer viruses. *Advances in Cryptology - CRYPTO'88 - Lecture Notes in Computer Science* **403,** 354–374 (1988).
8. Murray, W. H. The application of epidemiology to computer viruses. *Comp. and Sec.* **7,** 139–145 (1988).
9. Kephart, J. O. & White, S. R. Directed-graph epidemiological models of computer viruses. *IEEE Computer Society Symposium on Research in Security and Privacy, Oakland CA, USA: IEEE*. (doi: 10.1109/RISP.1991.130801) (1991 May 20–21).
10. Piqueira, J. R. C., de Vasconcelos, A. A., Gabriel, C. E. C. J. & Araujo, V. O. Dynamic models for computer viruses. *Comp. and Sec.* **27,** 355–359 (2008).
11. Ren, J., Yang, X., Yang, L.-X., Xu, Y. & Yang, F. A delayed computer virus propagation model and its dynamics. *Chaos, Solitons & Fractals*. **45,** 74–79 (2012).
12. Mishra, B. K. & Saini, D. K. SEIRS epidemic model with delay for transmission of malicious objects in computer network. *Appl. Math. and Comput*. **188,** 1476–1482 (2007).
13. Barth´elemy, M., Barrat, A., Pastor-Satorras, R. & Vespignani, A. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* **92,** 178701, 1–4 (2004).
14. Pastor-Satorras, R. & Vespignani, A. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* **63,** 066117, 1–9 (2001).
15. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86,** 3200–3203 (2001).
16. van Mieghem, P., Omic, J. & Kooij, R. Virus spread in networks. *IEEE/ACM Trans. Netw.* **17(1),** 1–14 (2009).
17. van Mieghem, P. & Omic, J. In-homogeneous virus spread in networks. Delft University of Technology, Report2008081. http://www.nas.ewi.tudelft.nl/people/Piet/TUDelftReports (2008) (Date of access: 22nd March, 2016).
18. Bailey, M. *et al.* Automated Classification and Analysis of Internet Malware. 10th International Symposium on Recent Advances in Intrusion Detection, Gold Coast, Australia. (doi: 10.1007/978-3-540-74320-0_10) (2007 Sep 5–7).
19. Li, M. Y., Smith, H. L. & Wang, L. Global dynamics of an SEIR epidemic model with vertical transmission. *SIAM J. Appl. Math.* **62,** 58–69 (2001).
20. Driessche, P. van den & Watmough, J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180,** 29–48 (2002).
21. Heffernan, J. M., Smith, R. J. & Wahl, L. M. Perspectives on the basic reproductive ratio. *J. R. Soc. Interface* **2,** 281–293 (2005).
22. Jones, J. H. Notes on R_0. www.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf. (2007) (Date of access: 12th May, 2015).
23. Boyce, R. DiPrima. *Elementary Differential Equations and Boundary Value Problems* Ch. 9, 459–468 (Wiley, 2004).
24. Li, M. Y. & Muldowney, J. S. A geometric approach to global stability problems *SIAM J. Math. Anal.* **27,** 1070–1083 (1996).

Author Contributions

B.K.M. planned and formulated the research design and made important contributions both in the model formulation process and its analysis. K.H. analyzed the model and also formulated the different characterizations. The partial simulation part is taken care by D.N.S. All authors have reviewed the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mishra, B. K. *et al.* Impact of Information based Classification on Network Epidemics. *Sci. Rep.* **6**, 28289; doi: 10.1038/srep28289 (2016).