Immunization Against Infection Propagation In Heterogeneous Networks

Waseem Abbas, Sajal Bhatia, Yevgeniy Vorobeychik, and Xenofon Koutsoukos
Institute for Software Integrated Systems
Vanderbilt University, Nashville, TN 37212, USA
{waseem.abbas,sajal.bhatia,yevgeniy.vorobeychik,xenofon.koutsoukos}@vanderbilt.edu

Abstract—Modeling spreading processes for infections has been a widely researched area owing to its application in variety of domains especially epidemic spread and worm propagation. Until recently, infection propagation models usually inspired by epidemic spreading, solely relied upon the underlying network properties without taking into account the variation in node specific properties, such as its ability to spread infection or recover from an infection. Owing to this fact, these models have been agnostic to the effects such node heterogeneity might have in the overall infection (or immunization) process. In this paper, we incorporate node properties in a well-known Susceptible Infected Recovered Susceptible (SIRS) model for infection propagation, and propose new heuristics to curb the spread of infection in heterogeneous networks. The proposed heuristics are validated against various network topologies, including a real-world example of an email exchange network.

Keywords-Spreading process, epidemic models, worm propagation, heterogeneous networks, immunization

I. INTRODUCTION

Infection propagation models have constantly evolved over the years and have found their applications in a wide variety of fields ranging from worm propagation in computer networks to information diffusion in social networks (e.g., [1]). However, until recently, these state-of-the-art propagation models, generally inspired by classical epidemic models, took only the underlying network properties into account without paying any attention to the participating node properties. In doing so, these models unknowingly become skeptic to the effects that these node properties might have in determining the overall spread of infection. Moreover, existing immunization strategies proposed for such network-based models generally rely on selecting a subset of nodes from a network and immunizing them with the goal of minimizing the overall spread of infection. Such immunization strategies, in process, fail to incorporate network *heterogeneity* by equally treating all nodes in terms of their properties that can influence the infection propagation. This *homogeneous* treatment of all nodes generally does not hold true in practical networks such as Cyber Physical Systems (CPS), in which nodes can have different capabilities in handling and spreading infections.

In this article, the aforementioned issues are addressed by not only taking forward the inclusion of node heterogeneity in the existing infection propagation models, as suggested recently by Smilkove et al. [2], but also incorporating this notion of heterogeneity while proposing immunization strate-gies to substantially curtail the infection spread. In doing so, the article makes the following contributions:

- A heterogeneous version of the existing Susceptible Infected Recovered Susceptible (SIRS) is presented which incorporates node properties.
- The notion of including node properties while designing immunization strategies is introduced, and a new effective degree based immunization heuristic is proposed to curb infection propagation in heterogeneous networks.

Heuristics proposed in this paper can thus be applied to many real-world scenarios that involve a spreading process. For instance, deployment of anti-virus programs at 'strategic' locations within a CPS network under a scenario in which limited instances of anti-virus programs are available. The following section gives a brief overview of the work done in infection propagation models and immunization strategies.

II. BACKGROUND AND RELATED WORK

The reviewed literature suggests that two models – Susceptible Infected Susceptible (SIS) and Susceptible Infected Recovered (SIR) – have been widely accepted and are the basis of several other variants [1]. A more generic combination of these two models is the Susceptible Infected Recovered Susceptible (SIRS) model (e.g. [3]), which maintains three distinct states, i.e., susceptible (S), infected (I), and recovered (R), as in the case of SIR model, but in addition closes the loop by allowing a transformation from state R to S, as shown in Figure 1. Both SIS and SIR models can be extracted from the SIRS model.

More than a couple of decades ago, Kephart et al. [4] were amongst the first researchers to propose epidemiology-based models (SIS) to analyze virus propagation in computer networks. The incident of the Code Red worm in July 2001, however, instigated the emergence of a plethora of worm propagation models. Moore et al. [5] comprehensively investigated this incident to analyze and model the worm propagation. Kim et al. [3] proposed a modified SIR model by granting nodes a temporary immunity. Chakrabarti et al. [6] developed a discrete time Non-Linear Dynamical System (NLDS) model based on the classical SIS model and assumed universal infection and death rate for each participating node.

The infection propagation models seen in the literature have been applied mainly to *homogeneous* networks, in which all the participating nodes are treated equally, especially in terms
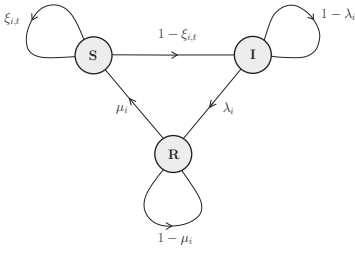
Figure 1. State-transition diagram of a single node in the heterogeneous SIRS model.

Table I
NOTATIONS FOR HETEROGENEOUS SIRS MODEL

$\mathcal{N}_i$	neighborhood of node i
β_i	probability that an infected node i infects its neighbors
λ_i	probability of an infected node i to move into the recovery state
μ_i	probability of a recovered node i to become susceptible.
$\xi_{i,t}$	probability of a susceptible node i at time $t-1$ to remain susceptible at t
$p_{i,t}$	probability of node i being infected at time t
$s_{i,t}$	probability of node i being susceptible at time t
$r_{i,t}$	probability of node i being in the recovery state at time t

of their infection and curing rates. However, recently, a *heterogeneous* version of the SIS model has been proposed [7, 8], in which all the participating nodes are treated differently, i.e., with their own distinct infection and curing rates. Furthermore, recently Smilkove et al. [2] have reported that failure to include heterogeneity, susceptibility of individuals against infections in this case, can lead to over/under estimation of fundamental epidemic parameters, resulting in less accurate descriptions of epidemic spreading processes. Thus, the reviewed literature suggests and highlights the importance of including the notion of *heterogeneity* for a more accurate modeling of the infection propagation process.

From immunization perspective, the goal is to distribute protection resources among nodes in order to minimize the overall effect of infection propagation. Heuristic based on network centrality measures are commonly used to distribute protection resources within the network. These approaches mainly rely on nodes' locations in the overall network to determine their relative importance. Some of the popular strategies include acquaintance immunization [9], targeted immunization (e.g., [10]), page rank centrality based immunization [11] and eigenvalue centrality based immunization [12]. More recently, the problem of optimal allocation of protection resources was formulated and solved using a geometric programming approach by Preciado et al. [8].

A key observation made in all these immunization strategies (excluding [8]) is the non-inclusion of node properties, which can substantially impact the infection propagation. This paper addresses this shortcoming by first incorporating the concept of node properties into existing immunization strategies and subsequently proposing a new degree based heuristics to curb infection propagation in heterogeneous networks.

III. INFECTION PROPAGATION MODEL

A heterogeneous version of Susceptible Infected Susceptible (SIS) model has been proposed recently in [7, 8]. This section presents a discrete-time, heterogeneous SIRS model that incorporates network topology as well as varying node properties to model the spreading process.

A. Heterogeneous SIRS (hSIRS) Model

In hSIRS model, during a time interval Δt, where Δt can be arbitrarily small[1], an infected node i infects its neighbors

[1]We are assuming that Δt is infinitesimally small such that the probability of having multiple events during Δt is negligible.

with probability β_i. At the same time, it can recover and move into the recovery state with probability λ_i. Once a node is recovered, it may stay recovered or immune with probability $(1-\mu_i)$, where μ_i is the probability of a node i in the recovery state to become susceptible again. The state transition diagram for a single node is shown in Figure 1, and the notations used in the SIRS model are described in Table I.

It is to be noted that the state of node i at the next time step is dependent only on its current state. Therefore, this process can be modeled as a Markov chain, and thus the Markov theory can be used to compute the probabilities of a node being in one of the infected, recovered, or susceptible states at a given time.

To achieve this, the rate at which a node i susceptible at time $t-1$ remains susceptible at time t is computed, and is denoted by $\xi_{i,t}$. In other words, the probability that a susceptible node i does not receive any infection from its neighbors at time t needs to be computed. For this purpose, we proceed as follows: Let $j \in \mathcal{N}_i$, then i cannot get infected from j at time t whenever j is not infected at $t-1$, or is infected at $t-1$, but fails to spread the infection at t. This is true for all $j \in \mathcal{N}_i$. Moreover, we assume that the infection of i from one of its neighbors is *independent* of the other. Using this independence assumption, we get

$$\xi_{i,t} = \prod_{j \in \mathcal{N}_i} [\underbrace{(1 - p_{j,t-1})}_{\substack{\text{Pr. of node } j \text{ being} \\ \text{not infected at } t-1}} + \underbrace{(1 - \beta_j)p_{j,t-1}}_{\substack{\text{Pr. of } j \text{ being infected at } t-1, \\ \text{but failing to infect } i \text{ at } t}}]$$

$$= \prod_{j \in \mathcal{N}_i} (1 - \beta_j p_{j,t-1}) \tag{1}$$

Now, using the state transition diagram in Figure 1, and (1), the equations for the probabilities of a node being infected $(p_{i,t})$, recovered $(r_{i,t})$, or susceptible $(s_{i,t})$ at time t can be written as

$$
\begin{aligned}
p_{i,t} &= (1 - \lambda_i)p_{i,t-1} + (1 - \xi_{i,t})s_{i,t-1} \\
s_{i,t} &= \xi_{i,t}s_{i,t-1} + \mu_i r_{i,t-1} \\
r_{i,t} &= (1 - \mu_i)r_{i,t-1} + \lambda_i p_{i,t-1}
\end{aligned}
\tag{2}
$$

Note that a node i is infected at time t whenever i was infected at time $t-1$ and remained in the infected state at t, or was susceptible at $t-1$ and got infected by its neighbors at t. Similarly, a node is susceptible at time t if it was in the susceptible state at $t-1$ and remained so at t, or was in the recovered state at $t-1$ and moved to the susceptible state
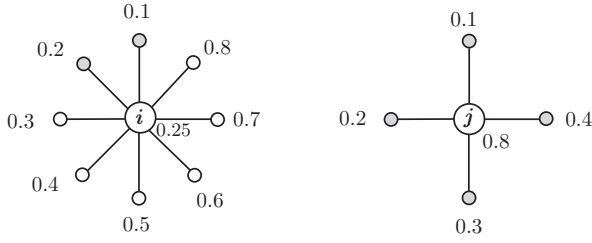
Figure 2. f_i of nodes are shown. Node i has a degree of 8, but $\theta_i = 2$. The degree and θ_j of node j are both equal to 4.

at t. Similarly, $r_{i,t}$ can be obtained. Above equations can be solved numerically to obtain the time evolution of the infected, susceptible, and recovered nodes' populations.

IV. IMMUNIZATION HEURISTICS

If there is a limited budget, under which only a subset of nodes can be immunized (protected against infections), which nodes are the best choice for immunization? Typically, nodes are ranked based on a certain 'centrality measure,' such as degree centrality, eigenvalue centrality etc. Nodes with the highest priority (determined by their centrality scores) are then immunized. For instance, in a degree centrality based approach (such as targeted immunization), nodes with the highest degree or connections with other nodes are selected. Similarly, in an eigenvalue centrality based approach, nodes inducing maximum drop in the largest eigenvalue of the adjacency matrix of the network are selected. These centrality measures rely merely on the underlying network topology.

However, besides connectivity, a node's infection and recovery rates also have a significant impact in determining node's participation in the spreading process. A node with a higher β_i is more likely to infect its neighbors, and a node with a smaller λ_i (or δ_i) is less likely to recover from the infected state. As an example, consider two nodes u and v with degrees 10 and 5 respectively within some network. Let $\beta_u = 0.1$, $\beta_v = 0.8$, and $\delta_u = \delta_v$. Using purely a degree centrality measure, node u has a higher priority for getting immunization. However, once infected, node u infects only 1 of its neighbors on average, whereas node v infects four of its neighbors on average after being infected. Thus, the consequences (in terms of the infection spread) of node v getting infected are more severe than that of node u, even though v has a smaller degree than u. Therefore, ranking of nodes for immunization should be a function of both the network topology based centrality measure and node's characteristics influencing the spreading process (e.g., β_i, δ_i, λ_i). The following section describes one such immunization strategy.

A. Proposed Immunization Strategy

Let $\mathcal{S}$ be the scoring function that ranks nodes for immunization such that the nodes with the highest scores are immunized. Furthermore, let c be the network topology based centrality measure, and f be the measure of node properties influencing spreading process. It is proposed that

$\mathcal{S}_i$ be a function of both c_i and f_i, i.e., $\mathcal{S}_i(c_i, f_i)$ instead of $\mathcal{S}_i(c_i)$.

To achieve this, a measure based on node properties needs to be defined that can capture its influence on the infection propagation in networks. In this regard, it is observed that β_i determines the capability of node i to infect others, once it is infected. Moreover, the ability of node i to remain in the infected state can be determined by its recovery rate λ_i (or δ_i). In fact, for homogeneous networks, in which all nodes are assumed to have the same infection and recovery rates, the ratio of infection rate to the recovery rate has been previously used to determine the epidemic threshold condition.[2] However, this ratio has not yet been used for designing immunization heuristics for heterogeneous networks. This article proposes the use of this ratio – infection rate to recovery rate – to capture the influence of node properties on the infection process, and also for designing the immunization strategy.

$$f_i = \frac{\text{infection rate of node } i}{\text{recovery rate of node } i} \quad (3)$$

For hSIRS model, $f_i = \beta_i/\lambda_i$, and for hSIS model, $f_i = \beta_i/\delta_i$ has been chosen. Simulations done for this paper show good results with these definitions of f_i. However, it is also acknowledged that there might be other ways to capture the influence of node properties in infection propagation.

As mentioned earlier, several different network-centric measures exist. This paper considers two widely used centrality measures – degree based and eigen value based, whose effectiveness has been proven and established in the literature (e.g., [10]). The degree based and eigen value based measures are described below.

$$\begin{aligned} \text{DegCen:} \quad & c_{i,d} = \text{degree of a node } i \\ \text{EegCen:} \quad & c_{i,e} = e_{\max} - e_{\max,i} \end{aligned} \quad (4)$$

Here, $e_{\max}$ is the largest eigenvalue of the adjacency matrix of the network graph and $e_{\max,i}$ is the largest eigenvalue of the adjacency matrix of the network graph after removing node i.

A scoring functions $\mathcal{S}_i$, which incorporate both network effects and node properties can be written as,

$$\begin{aligned} \text{hDegCen:} \quad & S_{i,d} = (c_{i,d})^a \ (f_i)^b \\ \text{hEigCen:} \quad & S_{i,e} = (c_{i,e})^a \ (f_i)^b \end{aligned} \quad (5)$$

Here, a and b can be any real numbers. For this paper, $a \in \{1, 2\}$ and $b \in \{1/2, 1\}$ has been used for experiments and results presented in Section V show that hDegCen and hEigCen scoring functions perform better than the DegCen and EigCen to restrict the number of infected nodes. It is remarked here that $\mathcal{S}$ can be any appropriate function that incorporates network topology, as well as the effects of nodes' properties, i.e., $\mathcal{S}(\mathcal{G}, \mathcal{V})$, where $\mathcal{G}$ corresponds to the function of any network-centric measure, such as (4), and $\mathcal{V}$ corresponds to the function of nodes' properties, such as (3).

---

[2]If the epidemic threshold condition is satisfied, the infection dies out from the network over time, i.e., epidemic dies out whenever β_i/λ_i is lesser than some specific value [6, 13].
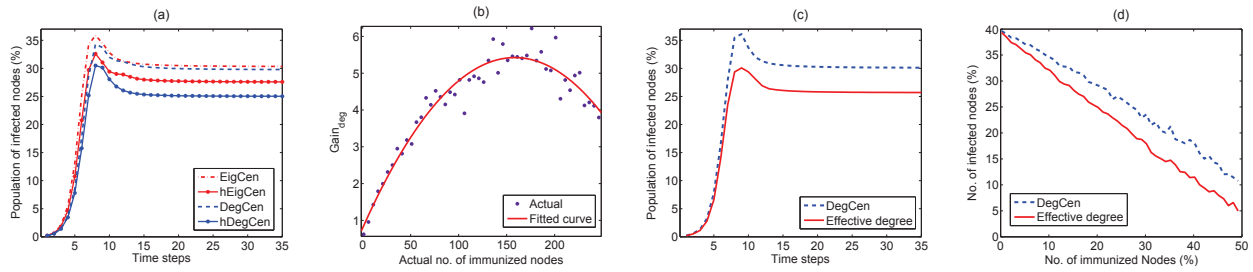
Figure 3. hSIRS model and ER graphs. (a) A comparison of hDegCen (hEigCen) with DegCen (EigCen) based immunization with 20% of the nodes immunized. (b) Gain$_{deg}$ in (6) is plotted against different number of immunized nodes. (c) A comparison of infected nodes' population with DegCen and effective degree based immunizations when 20% nodes are immunized. (d) A comparison of DegCen and effective degree over a range of number of immunized nodes.

Inspired by the degree based approach, a simple measure called *effective degree*, which incorporates node degree (as a measure of node connectivity in the network) as well as f_i (as a measure of node's properties influencing infection process) to determine its score for immunization, has been proposed in this paper. It is to be noted that f_i is a measure of the extent to which a node i, once infected, spreads the infection to its neighbors. Thus, the effective degree measure is defined as

Definition 4.1: (Effective degree, θ_i) If $\mathcal{N}_i$ is the set of nodes adjacent to node i, then θ_i is the number of nodes in the set $\{v_j \in \mathcal{N}_i : f_j < f_i\}$.

θ_i is simply the number of nodes in the neighborhood of i with a smaller infection rate to recover rate ratio 3 than i. An example is shown in Figure 2.

Since f_i measures node's ability to spread infection, θ_i ranks node i's contribution to the infection propagation. A higher θ_i implies that in the neighborhood of node i, there are more nodes with smaller infection rate to recovery rate ratios as compared to f_i, i.e., i is one of the more harmful nodes (from infection spreading view-point) in its neighborhood. That is the reason $\theta_i < \theta_j$ in Figure 2, as majority of the neighbors of i have a higher capability to spread infection, even though the degree of i is more than that of j.

V. EXPERIMENTAL RESULTS AND EVALUATION

In this section, the proposed immunization strategies, which combine node properties as well as node connectivity, are evaluated in heterogeneous networks for the following question:

Compared to purely network-centric immunization heuristics such as DegCen and EigCen, how useful are the proposed immunization strategies in terms of the reduction in the size of the infected population?

The above question is explored by simulating various widely used network topologies. It is observed that the proposed approach outperforms existing immunization heuristics.

A. hSIRS Model

The results in this sub section are obtained with the proposed hSIRS model in Section III-A. Each plot is averaged over 30 individual runs, unless specified otherwise. For experimental purposes, the infection rate β_i, recovery rate λ_i, and re-susceptibility rate μ_i, are chosen randomly from the

standard uniform distribution. In each iteration, a random node is infected with a virus that eventually spreads the infection as per (2). Simulation proceeds in steps of one time unit. A certain number of nodes K, as determined by some immunization strategy, are immunized, i.e., they will not get infected from the virus, and therefore, will not spread infection. The overall percentage of infected nodes in the steady state is the parameter of interest. The proposed immunization strategies are now validated for three graph types – random graphs, Modular graphs and a real-world email exchange graph.

1) Random (Erdős-Rényi (ER)) Graphs: Erdős-Rényi (ER) graphs $G(N, p)$, in which $N = 500$ and $p = 0.016$, are used in simulations in Figure 3. In Figure 3(a), populations of infected nodes during the infection processes are shown when 20% of the nodes are immunized. The number of immunized nodes in a network is denoted by K. It can be seen that the application of hDegCen and hEigCen immunization strategies result in a smaller number of infected nodes in the steady state as compared to the exisiting DegCen and EigCen respectively. Gain$_{deg}$ given in (6) compares hDegCen and DegCen over a range of number of immunization nodes is shown in Figure 3(b). Again hDegCen turns out to be a better approach than DegCen for a wide range of K since Gain$_{deg}$ is always positive. Here,

$$\text{Gain}_{deg} = \begin{bmatrix} \text{Infected nodes'} \\ \text{population in steady} \\ \text{state with DegCen} \end{bmatrix} - \begin{bmatrix} \text{Infected nodes'} \\ \text{population in steady} \\ \text{state with hDEgCen} \end{bmatrix} \tag{6}$$

In plots (c) and (d) of Figure 3, a comparison of DegCen and effective degree is demonstrated, in which effective degree based immunization strategy outperforms DegCen.

2) Modular Graph: Modular graphs consist of clusters, in which nodes within a cluster are densely connected, whereas the inter-cluster connectivity is sparse as observed in various CPS, network sciences, and energy control systems. A modular graph with 500 nodes arranged in 10 clusters, where each cluster is itself an ER graph with $N = 50$ and $p = 0.35$, is used for simulation. Inter-cluster interactions are also modeled as an ER graph, in which each cluster is a node, and interacts with four other clusters on average. The results are illustrated in Figure 4 showing improvements in the number of infected
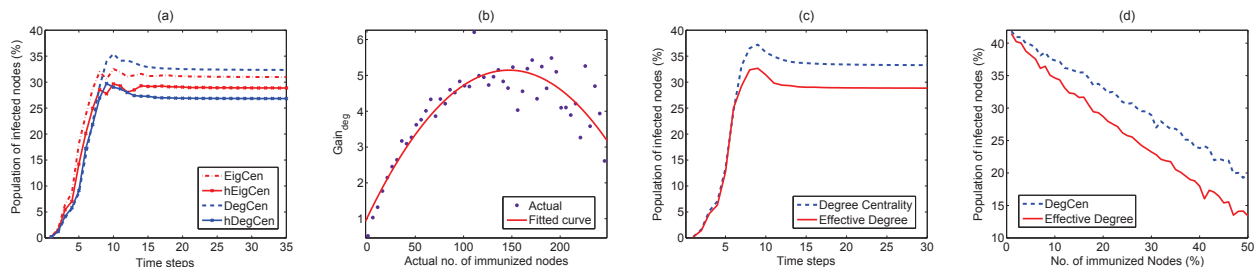
Figure 4. hSIRS model and Modular graphs. (a) Infected nodes' population when 20% of the nodes are immunized based on different immunization strategies. (b) hDegCen and DegCen are compared in terms of Gain$_{deg}$ over a range of number of immunization nodes. (c) A comparison of effective degree with DegCen with 20% of the nodes immunized. (d) Infected nodes' population in the steady state with DegCen and effective degree based selection of immunized nodes.
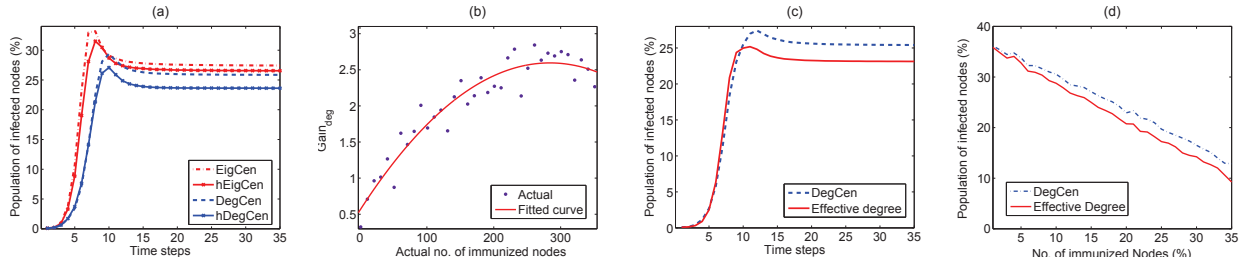

Figure 5. hSIRS model and Email exchange network. (a) 15% of the nodes are immunized based on different immunization schemes, and the resulting population of infected nodes is plotted. (b) A comparison of DegCen and hDegCen based on Gain$_{deg}$ over a range of number of immunization nodes. (c) Infected nodes' population when 15% of the nodes are immunized based on DegCen and effective degree. (d) A comparison of DegCen and effective degree over a range of the number of immunized nodes.

nodes as a result of hDegCen, hEigCen, and effective degree based immunization strategies.

3) Email Exchange Graph: A real-world email exchange network [14] is also used to validate the proposed immunization heuristics. The network consists of 1133 nodes and 5451 undirected edges, which indicate e-mail exchanges between nodes. Plots in Figure 5 illustrate the effectiveness of proposed immunization strategies in terms of reduction of infected nodes' population.

VI. CONCLUSION AND FUTURE WORK

Until recently, infection propagation models have been mainly applied to homogeneous networks without taking any node heterogeneity into account, which does exists in real networks. This paper extends the notion of incorporating node heterogeneity when modeling infection propagation and designing immunization strategies. It is observed that heuristics for the selection of immunization nodes incorporating both network centric measures and node properties outperform the ones that merely rely on the network centric measures. As a part of the future work, authors plan to extend this work to quantify the notion of node heterogeneity and its effect on the characteristics of spreading processes in a formal manner.

ACKNOWLEDGMENT

REFERENCES

[1] M. Newman, *Networks: an introduction*. Oxford University Press, 2010.
[2] D. Smilkov, C. A. Hidalgo, and L. Kocarev, "Beyond network structure: How heterogeneous susceptibility modulates the spread of epidemics," *Scientific Reports*, vol. 4, 2014.
[3] J. Kim, S. Radhakrishnan, and S. Dhall, "Measurement and analysis of worm propagation on internet network topology," in *Proc. of 13th Intl. Conf. on Computer Communications and Networks,*, 2004, pp. 495–500.
[4] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *Proc. of IEEE Symposium on Research in Security and Privacy*, 1991, pp. 343–359.
[5] D. Moore, C. Shannon, and J. Brown, "Code-red: a case study on the spread and victims of an internet worm," in *Proc. of the 2nd ACM SIGCOMM Workshop on Internet Measurement*, 2002, pp. 273–284.
[6] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Trans. on Information and System Security*, vol. 10, no. 4.
[7] P. Van Mieghem and J. Omic, "In-homogeneous virus spread in networks," *arXiv preprint arXiv:1306.2588*, 2013.
[8] V. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. Pappas, "Optimal resource allocation for network protection against spreading processes," *IEEE Trans. on Control of Network Systems*, vol. 1, no. 1, pp. 99–108, 2014.
[9] R. Cohen, S. Havlin, and D. Ben-Avraham, "Efficient immunization strategies for computer networks and populations," *Physical Review Letters*, vol. 91, no. 24, 2003.
[10] R. Pastor-Satorras and A. Vespignani, "Immunization of complex networks," *Physical Review E*, vol. 65, no. 3.
[11] F. Chung, P. Horn, and A. Tsiatas, "Distributing antidote using pagerank vectors," *Internet Mathematics*, vol. 6, no. 2, pp. 237–254, 2009.
[12] H. Tong, B. A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, and D. H. Chau, "On the vulnerability of large graphs," in *Proc. of 10th IEEE Intl. Conf. on Data Mining*, 2010, pp. 1091–1096.
[13] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 1–14, 2009.
[14] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," *Physical Review E*, vol. 68, no. 6, p. 065103, 2003.