Analysis of Search Heuristics for Improving the Resistance of Complex Networks to Epidemic Processes

Fernando Concatto[1], **Alex Luciano Roesler Rese**[1], **Rafael de Santiago**[2]

[1]Applied Intelligence Laboratory - LIA
University Vale of Itajaí (UNIVALI)
Mailbox 360 – CEP 88302-202 – Itajaí – SC – Brasil

[2]Department of Informatics and Statistics - INE
Federal University of Santa Catarina (UFSC)
CEP 88040-900 – Florianópolis – SC – Brazil

`fernandoconcatto@gmail.com, alexrese@univali.br, r.santiago@ufsc.br`

Abstract. *The uninhibited propagation of an infectious agent through a network can cause extensive damage of various types on the economy and public health for instance. Therefore, this study sought to develop a strategy to make complex networks more resistant to these agents through modifications in the topology of the network. For this purpose, we employed the epidemic threshold theory, which dictates that the epidemic resistance of a network is inversely proportional to the spectral radius of its adjacency matrix. Thus, we developed a local search to minimize the spectral radius through the removal of links, applying it to a set of real networks, seeking to identify the increase in resistance as a function of the number of links removed.*

Resumo. *A propagação desinibida de um agente infeccioso por uma rede pode causar danos exorbitantes de diversos tipos, como à economia e à saúde. Portanto, este trabalho buscou desenvolver uma estratégia para tornar redes complexas mais resistentes a tais agentes através da realização de mudanças na topologia da rede. Para tanto, utilizou-se a teoria do limiar epidêmico, que dita que a resistência da rede a processos epidêmicos é inversamente proporcional ao raio espectral de sua matriz de adjacências. Assim, desenvolveu-se uma busca local para minimização do raio espectral através da remoção de conexões, e aplicou-se a mesma sobre um conjunto de redes reais, visando identificar o aumento da resistência em função do número de conexões removidas.*

1. Introduction

The human being has been a victim of illness since the beginning of his existence. In this context, a worrying factor for the current society, is the population increase, resulting in a favorable spread of epidemics at an international level. Over the years the scientific community has developed several strategies and models for the prediction of contagion, considering both maximization factors and minimizing factors of global impact epidemics. In the general scenario, the epidemic models are based on the characteristics of real diseases, considering their stages. The application of the models allows representing many kinds of propagation through complex network simulations, being these portrayals of a behavior of the real world.

Complex networks are composed of several structures that are related to each other [Barabasi 2013]. Such systems can be represented and implemented computationally through the graph theory G = (V, E). The vertices represent the elements of the network and the edges represent the connections between pairs of elements. The study related to complex networks intensified at the beginning of the 21st century, being applied to several areas such as Biology and Social Networks [Newman 2003]. In complex systems the propagation of an information can characterize an epidemic, which may have a positive character as an advertisement or news and a negative one as an illness or rumor.

In the study of complex networks, it is important to identify the characteristics of the network structure that resist the dissemination of contents or even diseases. The knowledge about these peculiarities allows us to understand the threshold between a possible local problem and a problem of global proportions. One way to analyze the relationship between the structural properties of a graph is through the spectral theory approach.

Spectral theory studies the properties of networks by using matrix representations and their eigenvalues. This concept arose in Organic Chemistry and has aroused strong interest in mathematicians and scholars.

The study was segmented to define the fundamental concepts and present the strategy addressed to analyze patterns in infection spreadings in complex networks. Thus the work is divided into four sections. The first one is the Fundamental Concepts which describes the concepts of epidemic models, spectral theory and heuristic search. Then the Problem Model is presented. After that, results and analysis are presented and discussed. Finally, conclusions close the text and further works are suggested.

2. Fundamental Concepts

The analysis and control of epidemic spreading consists of a vast interdisciplinary field, involving the areas of public health and other sciences. It is necessary to base some structures, models, and strategies that involve their representations computationally to clarify its behavior. In this section, we present some of the most used methods to represent the epidemiological models; we also describe a strategy used to optimize problems to make them solvable computationally.

2.1. Epidemic Models

The first mathematical modeling of a disease spreading was presented by Daniel Bernoulli in 1760, who found an increase in the life expectancy of people due to inoculation of the smallpox virus [Pastor-Satorras et al. 2015].

Mathematical models have become the main resource for the analysis and proof of theories since the analysis of real behavior is impractical [Kermack and McKendrick 1927]. The evaluation of these models allows to measure uncertainties and to elaborate strategies for the intervention of dissemination. The study on the infection propagation allows to segment infectious diseases into different stages. They allow simulating and predicting behaviors, predicting the course of an outbreak, and evaluating strategies to control such outbreaks.

In general, the epidemic models determine that the target population of the contamination can be divided into different models related to the stage in which the current disease is found [Pastor-Satorras et al. 2015].

2.1.1. SIR

The Susceptible-Infected-Recovered (SIR) epidemic model can be applied to each network structure. The individuals can be stated on the following stages: (S) susceptible; (I) infected; and (R) recovered. It portrays diseases which an individual can be infected once in its life cycle, representing in its last stage an immunity or its death [Kermack and McKendrick 1927].

In this model there are three stages, in the first an uninfected individual assumes a stage susceptible to the contagion. When infected, there is the possibility of infecting your neighbors and after the infection period, it will be removed from the group and will no longer be infected can be observed in Figure 1.

2.1.2. SIS

Susceptible-Infected-Susceptible (SIS) is a model that lacks the recovery stage. After the infection period, the individual is not withdrawn from the group and is susceptible to infection again, and can be infected multiple times, as is shown in Figure 1. To represent epidemics where individuals can be infected multiple times, the SIS model is used. In this model, individuals can assume the following stages: (S) susceptible; and (I) infected. There is no stage of recovery, after the period of infection individuals become susceptible to contagion again. Tuberculosis and gonorrhea are examples of diseases studied through this model, along with the spread of computer viruses into networks.

2.1.3. Min-SEIS-Cluster

The Susceptible-Exposed-Infected-Susceptible (SEIS) model is designed to take into account a period of incubation of the disease, where the individual is infected but not yet capable of transmitting the condition. A graphical representation of this model is displayed in Figure 1. In this model an infection does not leave any immunity, so the individuals who recover are again susceptible again by moving back into this compartment.

Reducing connectivity in a network has a direct impact on the minimization of spread of epidemics. Approaches involving metaheuristics have been shown to be effective in this context [Concatto et al. 2017].

2.2. Epidemic Threshold and the Spectral Radius

A condition must be able to be propagated to at least one other individual to be considered an epidemic process. Otherwise, the infection process will eventually cease, reaching a state where no individual is infected; in this *absorbing state*, the epidemic is said to be extinguished. Alternatively, if the individuals become immune to the disease after being infected (i.e., the SIR model is being used), the absorbing state is also reached when there are no vulnerable (non-immune) individuals left; in this case, the epidemic reaches a peak and gradually subsides. The average number of infections that an individual transmits is called the *basic reproduction number* R_0, and thus whenever $R_0 > 1$ the condition becomes an epidemic [Pastor-Satorras et al. 2015].

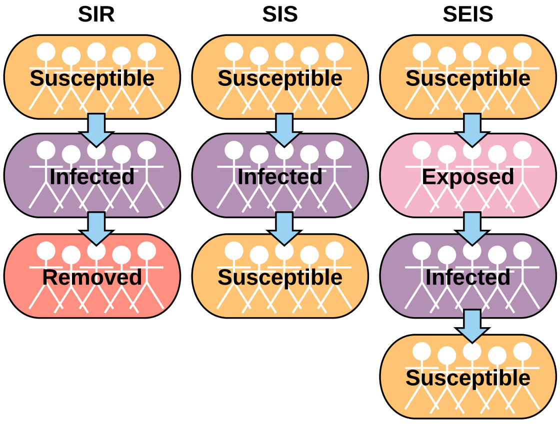

Figure 1. Contagion models SIR, SIS and SEIS

However, in stochastic and heterogeneous models, there are multiple parameters that control the dynamics of the infective agent (typically transmission and recovery rates) as well as a varying number of links by individual, and thus R_0 by itself cannot be used to determine if the condition will become an epidemic or not [Anderson and May 1992]. Instead, the *effective spreading rate* $\tau = \frac{\beta}{\delta}$ can be used to characterize the general potency of the infective agent, where β represents the transmission rate and δ indicates the recovery rate [Wang et al. 2003, Van Mieghem et al. 2009, Pastor-Satorras et al. 2015].

Therefore, it is possible to identify a critical value of τ, represented by τ_c, whereby higher values characterize a condition that will always become an epidemic (save for rare occasions where the stochastic nature of the process prevails over the expected behavior) and lower values indicate that it will be quickly extinguished. This value is formally defined as the *epidemic threshold*, and [Wang et al. 2003] proved that for homogeneous, star and power-law networks, the epidemic threshold τ_c is given by:

$$\tau_c = \frac{1}{\lambda_1(A)} \tag{1}$$

where $\lambda_1(A)$ represents the largest eigenvalue – the spectral radius – of the adjacency matrix $\mathbf{A}$. The authors also state that the $\tau < \tau_c$ condition is necessarily true for all circumstances where the epidemic is eventually extinguished in models where individuals become susceptible after the disease is cured (i.e. SIS). This result can be interpreted as a statement that diseases tend to be more efficiently disseminated in groups of individuals that have many connections with each other since the spectral radius is proportional to the number of connections in the network as well as to how tightly connected it is [Nowzari et al. 2016]. In this work, search methods were used to minimize the spectral radius of real networks, with the goal of improving their resistance against infectious agents.

2.3. Minimizing the Spectral Radius

To reduce the propagation of a disease, the most straightforward strategy involves decreasing its transmission rate δ and increasing the recovery rate $beta$, thus diminishing its effective spreading rate τ. However, controlling these parameters directly is often impractical and difficult to model mathematically; therefore, a more feasible option would be to use the $\tau < \tau_c$ condition to indirectly limit the spreading rate of an epidemic. In networked models, this alternative may be implemented through the minimization of the spectral radius $\lambda_1(A)$, which translates to an increase in T_c and consequently in an elevated resistance against epidemics. Since the spectral radius is a function of the adjacency matrix $\mathbf{A}$, changes in the topology of the network have a direct influence in the epidemic threshold [Nowzari et al. 2016].

In [Van Mieghem et al. 2011], the authors investigate two problems related to the minimization of the spectral radius: *i)* Link Spectral Radius Minimization (LSRM), which aims to find the best set of m links to remove from the network, and *ii)* Nodal Spectral Radius Minimization Problem (NSRM), whose goal is to identify the best m nodes to remove. In both cases, *best* is defined as the set of links or nodes that, when removed from G, produce a modified graph G' with the least spectral radius out of every other possible graph obtainable from G.

However, both problems were proved to belong to the NP-complete class by [Van Mieghem et al. 2011], demonstrating their computational intractability for reasonably sized networks. Therefore, heuristic approaches must be used instead. Examples of heuristics for the LSRM, NSRM and similar problems can be found in [Yan et al. 2016, Van Mieghem et al. 2011, Saha et al. 2015]. This work focused in the investigation of the class of heuristics known as *local search*, detailed in Section 2.4.

2.4. Local Search

Local Search algorithms operate in a single state and move to the neighborhood of this state. In many problems, the way to the final solution is irrelevant. In this way, if the way to the solution does not matter, in these cases it is interesting the application of Local Search algorithms. Operating on a single current state, instead of multiple paths, move only to neighbors of this state, where the path traveled is not saved, only the current state. By not storing the previous solutions they use little memory and allow finding reasonable solutions in large search spaces, being extremely useful in solving optimization problems.

3. Methodology

This section describes the methods undertook to accomplish the objectives outlined in Section 1. The study was divided into two different aspects: *i)* epidemic model, which contemplated the implementation of the stochastic network-based Susceptible-Infected-Susceptible (SIS) model, and *ii)* heuristic optimization, which involved the development of the local search aimed at reducing the spectral radius of the network.

3.1. Epidemic model

This work focused mainly on the Susceptible-Infected-Susceptible (SIS) model, since it presents a clearer distinction between simulations where the condition becomes an epidemic and those where it is quickly extinguished.The SIS model implemented in this work followed the specifications described by [Nowzari et al. 2016], where the population is modeled as an undirected and unweighted graph $G = (V, E)$, with each vertex $v \in V$ representing an individual p of population P and each edge in E representing a connection (and, consequentially, the possibility of transmitting the condition) between two individuals. The neighborhood $\mathcal{N}(v)$ is defined as the set of all vertices in V that have an edge towards v. With respect to the epidemic model, the function $\phi_k : P \rightarrow \{Susceptible, Infected\}$ is used to represent the current state of an individual $p \in P$ in time step k. Transitions between the two possible states are given by:

$$\phi_k(p) : Susceptible \rightarrow Infected \ = \ \beta \ |I_k(P) \cap \mathcal{N}(p)| \tag{2}$$

$$\phi_k(p) : Infected \rightarrow Susceptible \ = \ \delta \tag{3}$$

where β represents the infection rate of the condition and δ represents the rate of recovery; both are considered to be inputs to the model. These transition probabilities are tested in every time step. Additionally, the model requires two additional inputs: *i)* N, indicating the total amount of possible time steps executed before the simulation is terminated; and *ii)* M, representing the number of individuals that will begin the simulation in the $Infected$ state, which will be chosen randomly from the entire population.

3.2. Heuristic optimization

To achieve the reduction in the spectral radius of the adjacency matrix of a network represented by a graph $G = (V, E)$ through the removal of links, an optimization strategy involving a local search was implemented. More specifically, a monotonic exploration method was adopted, which does not admit moving to solutions worse than the current one and thus gets stuck in local minima; given the absence of publications that investigate this category of problems using metaheuristics, the proposed methodology provides a foundational baseline for future research in the topic.

The search described in this section aims at finding a higher quality solution to the LSRM problem considering an initial solution $s_{initial}$. Solutions are represented as a binary vector s of length $|E|$, where E indicates the set of edges of the graph. Each component s_i can either have a value of 1, meaning that the link should be left as it is, or 0, signifying that it should be removed from the network. Solutions are subject to the constraint $\sum_{i=1}^{|E|} s_i = |E| - w$, which demands that the amount of zeroes in the solution be equal to w; also, w should be a positive integer and less than $|E|$. The quality of a solution s (also referred to as the objective function) is given by:

$$\psi(s) = \frac{1}{\lambda_1(A_s)} \tag{4}$$

where A_s indicates the reduced adjacency matrix of the original graph, with its set of edges given by $\{e_i \in E \mid s_i = 1\}$, and $\lambda_1(A)$ represents the spectral radius of matrix A. This equation is equivalent to Equation 1, which describes the epidemic threshold. Since the quality of the solution should always increase, this problem can be interpreted as a maximization of the epidemic threshold, indicating that the network will tend to become resistant to higher values of the effective spreading rate τ.

3.2.1. Neighborhood

In every iteration, the search moves to a neighbor of the current solution, which must have a higher value of ψ; if there are no neighbors that fit this criterion, the search stops. In the context of this study, the neighborhood of a solution s is defined as any solution s' initialized with the same values of s in every position, but with $s'_i = 0$ and $s'_j = 1$ where, in the original solution, $s_i = 1$ and $s_j = 0$. In other words, every position which contains a zero in the original solution can have its value exchanged with another position containing a value of one. Therefore, the number of candidate neighbors is given by the amount of positions containing a value of one multiplied by the amount of positions containing zero; equivalently, it can be defined as the cardinality of the set $N(s) = \{(i, j) \mid s_i = 1 \text{ and } s_j = 0\}$. The worst case occurs when the number of zeros is the same as the number of ones, resulting in a neighborhood of size $\frac{|E|^2}{4}$.

Using a *best improvement* updating policy, every neighbor s' of s has to be evaluated using Equation 4, which implies in the necessity to decompose the adjacency matrix of the graph produced by s' into its spectrum, the execution of the search can become computationally expensive for networks with a large number of connections, both due to the larger amount of neighbors and also due to the increased difficulty of computing the spectral radius, which possesses a time complexity of $O(n^3)$ for an n by n matrix [Golub and Van Loan 2012]. To overcome this problem, a parameter named $\alpha \in [0, 1]$ was introduced, which limits the amount of neighbors of s that have to be evaluated to $\alpha|N(s)|$, which are randomly picked (without duplicates) from the set of all possible neighbors of s. A second option, also analyzed in this work, is to use a *first improvement* policy, where the first neighbor that possesses a quality greater than s is accepted as the current solution in the next iteration.

4. Results

To accomplish the goals of this study, a variety of experiments have been executed; namely, the procedures undertaken were: *i)* execution of the monotone local search over the set of networks listed in Table 1 [Batagelj and Mrvar 2006], with various updating policies and fractions of edges to be removed; and *ii)* simulation of the SIS epidemic model with the same networks, varying the parameters of transmission, recovery and initially infected individuals, both with the original and post-improvement configurations.

Table 1. List of real-world networks used in the experiments.

ID	**Instance**	**|V|**	**|E|**	$\lambda_1(\mathbf{A})$
1	strike	24	38	3.8702
2	sawmill	36	62	4.9718
3	karate	34	78	6.7257
4	dolphins	62	159	7.1936
5	polbooks	105	441	11.9326
6	football	115	613	10.7806

The first experiment aimed to produce a comparison between different values for the α parameter and the *first improvement* policy, both described in Section 3.2.1, as well as the impact of removing a larger fraction of the edges of the network. Additionally, the results were compared to a Monte Carlo search, which consists in an iterative random sampling strategy that stores the solution with the highest quality seen up to the current iteration. This search was configured to halt after executing 1000 iterations without any

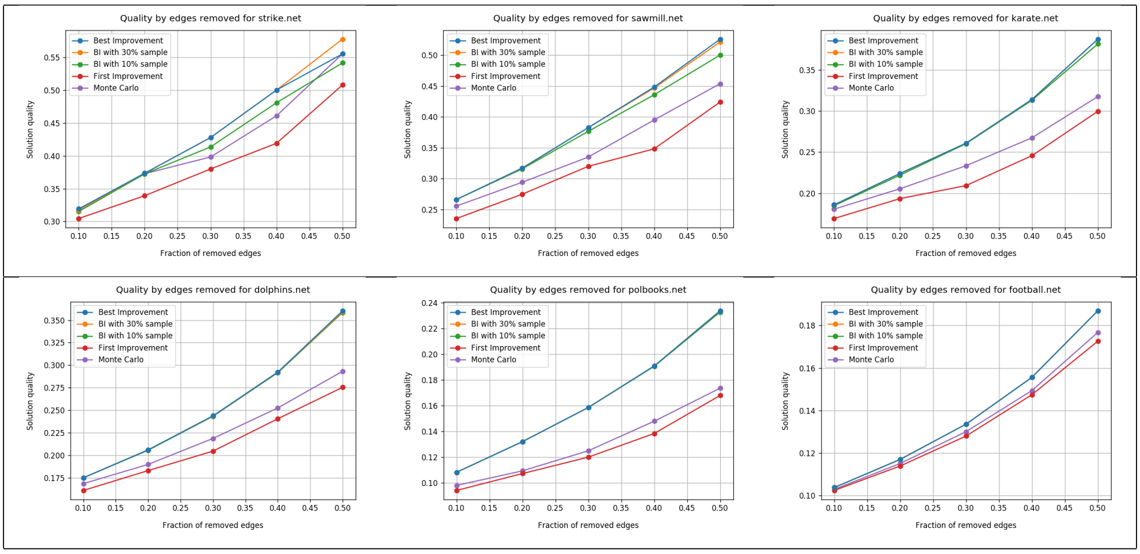

Figure 2. Quality as a function of the fraction of links removed.

change in the best-known solution. The parameter α was set to 0.1, 0.3 and 1.0, the latter representing a *best improvement* policy, and the number of removed edges w was set to 10%, 20%, 30%, 40% and 50% of the total number of edges of the network. The highest values of ψ obtained from 30 replications are presented in Figure 2.

A few trends can be inferred from the results. First, the sampling parameter α shows that the difference between exploring the full neighborhood of a solution and a small portion of it is negligible for the LSRM problem. Therefore, since evaluating solutions is computationally expensive, as discussed in Section 3.2.1, using a limited sample is much more preferred than examining the entire neighborhood. Secondly, the Monte Carlo method consistently produced higher quality solutions than the local search using a first improvement policy, indicating that the search space might be very noisy, with a lot of local minima. However, it displayed worse results than the sampled local search in all but one experiment (*strike* instance with 50% of its edges removed), suggesting that taking steeper steps in the search space might avoid getting stuck in a minimum too early. Finally, a quasi linear relationship can be seen between the number of removed edges and the quality of the best solutions attained, tending towards a low degree polynomial as the number of total edges in the graph grows larger. These observations apply to all six networks tested.

Next, an experiment was developed to validate the solutions obtained by the search methods, verifying the impact of reducing the spectral radius upon the resistance of the network against an epidemic. To this end, the SIS model was simulated with various configurations, first with the original network and then with a subset of its links removed. To quantify the resistance of the network, a measurement named *vulnerability* was defined, given by:

$$\frac{1}{|C|}\sum_{(\beta,\delta,M)\in C}\begin{cases}0 & \text{if } P_{\beta\delta M}=0\\ 1 & \text{otherwise}\end{cases} \tag{5}$$

where C is a set of 3-tuples containing every combination of infection rate β, recovery rate δ and initially infected individuals M available and $P_{\beta\delta M}$ represents the observed prevalence (ratio between the number of infected individuals and the size of the population) of the condition in the final time step of the simulation with the specified configuration, averaged across all replications. Essentially, the vulnerability metric expresses the proportion of simulations where the condition has become an epidemic, i.e. it was still active when the simulation ended.

Results of this experiment are presented in Table 2, where BI represents the local search with a best improvement policy and $\alpha = 1$, MC represents the Monte Carlo method and FI represents the local search with a first improvement policy. For β and δ, the values tested ranged between 0.025 and 0.9 with increments of 0.025, while M ranged from 10% to 90% with increments of 10%. Simulations were replicated 10 times, and were forcibly terminated after 1000 time steps.

Table 2. Vulnerability values for each network, before and after removing links

		10% less edges			**30% less edges**			**50% less edges**		
ID	**Orig.**	BI	MC	FI	BI	MC	FI	BI	MC	FI
1	0.411	0.367	**0.366**	0.383	0.300	0.308	**0.299**	**0.151**	0.162	0.172
2	0.487	**0.445**	0.448	0.460	**0.333**	0.371	0.388	**0.198**	0.235	0.264
3	0.555	**0.525**	**0.525**	0.533	**0.449**	0.469	0.482	**0.331**	0.379	0.394
4	0.620	**0.589**	**0.589**	0.597	**0.512**	0.531	0.534	**0.395**	0.446	0.452
5	0.748	**0.720**	0.727	0.727	**0.648**	0.677	0.680	**0.550**	0.601	0.599
6	0.771	**0.752**	0.753	0.754	**0.701**	0.702	0.704	**0.619**	0.628	0.631

The results display a clear relationship between a lower spectral radius and an increased resistance against epidemics, since the best improvement (BI) search, which produced solutions with the highest quality (as seen in Figure 2), was also able to make the networks less vulnerable than the other search methods in the majority of the cases. Furthermore, it can be seen that when the fraction of links removed is small (10%), both BI and MC produce virtually equivalent reductions in vulnerability. The only exception is in the *polbooks* (5) instance, which also possesses the highest original spectral radius (presented in Table 1). Conversely, with half of the links of the network removed, a larger margin can be seen between BI and MC, while it becomes smaller between MC and FI. This result reinforces the benefit of being able to explore the search space more effectively as it becomes larger.

5. Concluding remarks

In this paper, an analysis between four variations of a monotone local search and a Monte Carlo method applied to the problem of identifying the best set of links to be removed in order to minimize the spectral radius of the adjacency matrix of a network has been presented. Experimentally, it was verified that a local search that chooses the neighbor that offers the best improvement for the objective function performs considerably better than the tested alternatives, even if only a small fraction of the full neighborhood of a solution is examined in each iteration. The results obtained indicate that the resistance of a complex network against epidemic processes grows as the number of links to be removed increases in the form of a polynomial with a low degree, resembling a linear relationship for smaller networks. Despite this, the removal of links can be very costly for some contexts, such as transportation; in these cases, a constrained multi-objective approach is recommended.

Suggestions for further research include: *i)* using more sophisticated search algorithms, especially strategies that avoid local minima such as genetic algorithms, as well as comparing execution times; and *ii)* employing the vulnerability metric, discussed in Section 4 and defined in Equation 5, as an objective function to be minimized and verifying if the spectral radius behaves in a manner similar to the one observed in this study.

References

Anderson, R. M. and May, R. M. (1992). *Infectious diseases of humans: dynamics and control*. Oxford university press.

Barabasi, A.-L. (2013). Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.*

Batagelj, V. and Mrvar, A. (2006). Pajek datasets. Technical report.

Concatto, F., Zunino, W., Giancoli, L. A., Santiago, R., and Lamb, L. C. (2017). Genetic algorithm for epidemic mitigation by removing relationships. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '17. ACM.

Golub, G. H. and Van Loan, C. F. (2012). *Matrix Computations*, volume 3. JHU Press.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.

Nowzari, C., Preciado, V. M., and Pappas, G. J. (2016). Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems.*

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979.

Saha, S., Adiga, A., Prakash, B. A., and Vullikanti, A. K. S. (2015). Approximation algorithms for reducing the spectral radius to control epidemic spread. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 568–576. SIAM.

Van Mieghem, P., Omic, J., and Kooij, R. (2009). Virus spread in networks. *IEEE/ACM Transactions on Networking (TON)*, 17(1):1–14.

Van Mieghem, P., Stevanović, D., Kuipers, F., Li, C., van de Bovenkamp, R., Liu, D., and Wang, H. (2011). Decreasing the spectral radius of a graph by link removals. *Phys. Rev. E*, 84:016101.

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

Yan, J., Gong, M., Ma, L., Wang, S., and Shen, B. (2016). Structure optimization based on memetic algorithm for adjusting epidemic threshold on complex networks. *Applied Soft Computing*, 49:224 – 237.