



**HAL**  
open science

# Laplacian Spectra of Graphs and Cyber-Insurance Protection

Thierry Cohignac, Nabil Kazi-Tani

► **To cite this version:**

Thierry Cohignac, Nabil Kazi-Tani. Laplacian Spectra of Graphs and Cyber-Insurance Protection. 2020. hal-02911859

**HAL Id: hal-02911859**

**<https://hal.science/hal-02911859>**

Preprint submitted on 4 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Laplacian Spectra of Graphs and Cyber-Insurance Protection

Thierry COHIGNAC\*      Nabil KAZI-TANI†

August 4, 2020

## Abstract

In this paper, we consider the situation where a given graph has to be protected against communication interruption, through insurance or prevention measures. The goal of the protection buyer is to maintain good connectivity properties of the graph after a malicious attack, giving rise to a virus spread on the network. We model the epidemic spread using the standard Susceptible-Infected-Susceptible (SIS) Markov process. The connectivity of the graph is measured by a function of the average Laplacian spectrum: the second smallest eigenvalue, known as the algebraic connectivity. Using standard results on eigenvalues optimization, we recast the algebraic connectivity maximization as a semidefinite optimization problem, for which a solution exists and can be efficiently numerically computed. Our results allow to hierarchize the edges of a graph, giving more importance to some edges for which the protection demand is high, hence making optimal insurance demand directly depend on the underlying network topology.

**Key words:** Algebraic connectivity, Large graphs, Susceptible-Infected-Susceptible, Cyber Insurance

**AMS 2020 subject classifications:** 91G05, 90C35, 90C22, 05C50.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Summary of the Approach . . . . .	2
1.2	Related Literature . . . . .	3
<b>2</b>	<b>Elements of Graph Theory</b>	<b>4</b>
2.1	Generalities and the Laplacian . . . . .	4
2.2	Spectral Theory . . . . .	6
<b>3</b>	<b>Model and Assumptions</b>	<b>8</b>
3.1	SIS Epidemic Model . . . . .	8
3.2	Demand Model . . . . .	10
3.3	A Solution . . . . .	12
<b>4</b>	<b>Numerical Analysis</b>	<b>16</b>

---

\*CCR, 157 Boulevard Haussmann, 75008 Paris, France, tcohignac@ccr.fr

†Univ Lyon, Université de Lyon 1, Laboratoire de Sciences Actuarielle et Financière, ISFA, 50 Avenue Tony Garnier, F-69007 Lyon, France. nabil.kazi-tani@univ-lyon1.fr.

# 1 Introduction

Cyber risk analysis is a multidisciplinary subject, that has triggered an intensive research activity over the last 15-20 years. In particular, cyber insurance and reinsurance is a promising tool of cyber risk management. However, despite the extensive research activity around cyber insurance, its actuarial analysis is still at an early stage and many practical and theoretical problems remain open.

One of the many difficulties surrounding cyber insurance is that the underlying risk consists of many interconnected entities. Networks are one natural way to model this. In this paper, we focus on the analysis of the network topology and its influence on cyber insurance demand. We use tools from linear algebraic graph theory to formulate an optimal insurance demand problem. Connections in a network are fully characterized by the adjacency matrix or the Laplacian matrix (see Definition 2.1 below). It is thus natural to expect that the spectral analysis of these matrices will provide us with interesting information about the graph itself. More precisely, we focus in this paper on the Laplacian matrix, which can be intuitively understood as a discretization of the continuous Laplacian functional operator, given by the sum of second order derivatives. These techniques are well known in combinatorics, computer science, chemistry etc. [10]. We argue that these tools can be very useful when assessing cyber-insurance risk.

Throughout the paper, we assume that a given graph has to be protected against communication interruption, via prevention measures, or via insurance. Various situations fit in this description: for instance, the graph can be a company's computer network, susceptible to exterior attacks. Due to the increasing digitization of companies' operating models, and due to the related heavy data collection processes, businesses become highly exposed to data breaches and its consequences both in legal matters or in terms of image. Since modern data is stored and represented using graphs, it is crucial to analyze graphs protection, and in particular the impact of the topology of the network on the chosen protection.

Without loss of generality, we assume that the considered networks will be connected. Indeed, if a network has two connected components, which are completely independent, it is natural to analyze these networks separately, for pricing or risk management issues.

## 1.1 Summary of the Approach

The approach we take in this paper is to determine optimal insurance purchase parameters on a graph, by maximizing the algebraic connectivity. Our focus is on an endemic situation: the agent buys protection to maintain the connectivity of its network as high as possible in a situation of cascading virus infection.

**Epidemic Model.** We model a malicious attack diffusion in the agent's network using the Susceptible-Infected-Susceptible (SIS) epidemic model. It is difficult to analyze the exact SIS model in its full generality, since the needed number of equations to describe it grows exponentially with the dimension (see Section 3.1). That is why, on one hand, we adopt the standard mean-field approximation developed in [38], which upper bounds the infection probabilities. On the other hand, we consider the long time steady state associated to this approximation, which is accurate when infection probabilities are high, corresponding to cascading infections. Note that even in this situation, the obtained graph with optimal insurance parameters remains connected (see Remark 3.4). The SIS model incorporates a vector of curing rates, describing the speed at which an infected edge gets cured. We make the assumption that the level of insurance purchased directly influences the curing rates (Assumption

3.2).

Indeed, as reported in [14] by a pool of European insurance companies, most cyber insurance contracts include data restoration coverages, and the services of IT experts for network interruption management. These measures, partly covered by insurance contracts, clearly help restore infected edges back into service. Measures taken by a given company to increase its network resilience are intricately related to available insurance coverages. We include this particular feature of cyber insurance markets in our model by assuming that the level of insurance protection has an impact on curing edges.

**The Optimization Problem.** Since the seminal work of Fiedler [18], we know that the lowest strictly positive eigenvalue of the Laplacian matrix describes the connectedness of a graph. This particular eigenvalue is known as the *algebraic connectivity*. Maintaining network resilience after a highly spreading malicious attack can be done through algebraic connectivity maximization. The insurance demand model that we introduce in this paper is in line with this idea: the protection buyer optimally chooses the level of purchased insurance, which amounts in our case to determining optimal curing rates, in order to maximize connectivity, under a budget constraint. This problem thus naturally falls into the class of eigenvalue optimization problems. The main result of this paper is the reformulation of the optimal insurance demand problem under budget constraint as a semidefinite programming problem (SDP), for which there is an established theory and powerful numerical algorithms, that can be applied to large graphs.

## 1.2 Related Literature

This paper connects several topics: cyber insurance risk, which is already a multidisciplinary field, epidemic models and linear algebraic graph techniques. There is an extensive literature for all these subjects, we do not aim here at being exhaustive, but we rather indicate strongly related papers and monographs.

**Cyber Insurance.** The early 2000's have seen the development of the field known as economics of information security, in which authors have tried, in particular, to determine the optimal amount to invest to protect a firm's information system or data. The model introduced in [20] has been widely discussed in the literature (see e.g. [21, 3] and the references therein). By comparison, the results presented here allow to determine *where on a network* it is judicious to invest a *fixed* amount dedicated to security. It can be indeed interesting, from a risk management point of view, to concentrate the effort on a subpopulation, rather than having the whole group engage into immunization measures [24]. Several other papers adopt a game theoretic approach to analyze incentives for cybersecurity investments, in particular through insurance [27], and the effect of network topology on the resulting equilibria [26, 2]. The survey paper [30] details some of these approaches.

[16] develops a mathematical model of insured losses generated by cyber threats on a network, where the authors provide a methodology for standard reinsurance contracts pricing. The recent contribution [40] also provide a pricing methodology based on a simulation scheme. These papers both assume a SIS epidemic model on the underlying network. [23] introduces a model of cyber aggregate loss distribution, adapted to local network topologies. To make use of these theoretical models, statistical estimation of cyber insurance claim parameters is needed: some recent contributions employ regression trees [17], or Hawkes processes [1].

**Epidemic Models.** We consider here an attack spread following the Susceptible-Infected-Susceptible (SIS) model. It is a continuous time Markov chain, known as the *contact process* in the statistical physics literature [29], in which each infected node (or edge in this paper) gets infected at a rate proportional to the number of infected neighbors, and gets cured afterwards at a constant rate. We refer the reader to the survey [34] and the references therein for a review of the main issues and results on epidemic spreading on complex networks, including transition phase phenomena, approximation strategies, and spreading on random networks. In the first order mean field approximation known as NIMFA [38, 37], one assumes that neighbour nodes are statistically uncorrelated, which lowers the number of equations needed to describe the finite dimensional distributions of the SIS model and upper bounds infection probabilities. Other approximation schemes are possible, with higher moment approximations ([12, 16]), not assuming independence (see the Hilbert approximation in [16]) or via a degree-based mean field procedure [35], with the inconvenient of losing the information on local topology of the network.

**Graph Theory.** The optimization criteria we retain to determine the insurance demand function is a graph theoretic one: the algebraic connectivity [18], which measures how well a graph is connected, and for which there is a huge body of literature. The reader is referred to the survey papers [10, 9] and to the monographs [25, 6] for a detailed analysis of graphs spectra and their various fields of applications. The problem of choosing the weights of a graph to maximize its algebraic connectivity has already been analyzed by Fiedler [19], in the particular case of trees: he shows that the optimal algebraic connectivity, under an average constraint on the weights, is given by 1 divided by the variance of the graph, i.e. the average squared distance to the center of gravity of the tree. This maximization problem is also considered in [36], where the authors analyze the so called Fastest Mixing Markov Process on a graph, and give various physical interpretations. Their analysis allows us to give a probabilistic interpretation to our insurance related problem (see Section 3.3). Maximizing functions of the Laplacian spectra of graphs is a particular case of an eigenvalue optimization problem, which have been extensively discussed in the literature [33, 28, 4] and for which efficient numerical procedures are available [39].

## Structure of the Paper

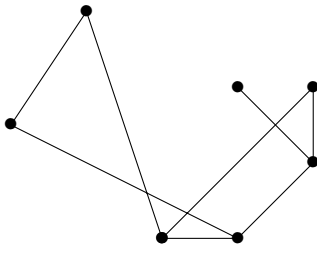
This paper is organized as follows: Section 2 very briefly provides introductory elements on graph theory for non specialists, and in particular some elements and intuitions on algebraic connectivity, since it plays a crucial role in our demand model. In Section 3, we introduce the epidemic transmission model that we use, along with some simplifying assumptions, and we state and solve the optimization problem. In Section 4, relying on a SDP rephrasing of the problem, we numerically solve it on some small and larger graph examples.

## 2 Elements of Graph Theory

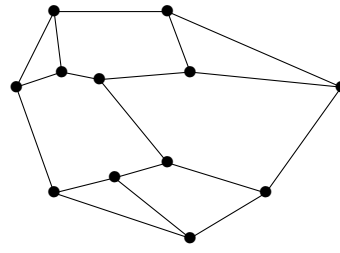
### 2.1 Generalities and the Laplacian

In this section, we introduce some tools from graph theory that will be used in the paper. This presentation is intended for non specialists and we try to insist on intuitions rather than formalism.

Let us first fix some terminology. Depending on the context, a graph can have various meanings, but the common idea is to represent relations between entities. In a cyber-insurance context, the entities can be computers, servers, softwares, routers, data sets, etc. And the relations can be internet



(a) A graph on  $n = 7$  vertices, with 8 edges.



(b) A 3-regular graph on  $n = 12$  vertices.

Figure 1: Graphs examples

connections, distant access to servers, links between softwares etc. Here, we will use the most common definition: a graph is given by a pair  $G = (V, E)$  formed by a countable set  $V$  and a set  $E$  of unordered pairs of elements of  $V$ .  $V$  represents the set of vertices of the graph and  $E$  the set of edges. We will assume throughout the paper that the sets  $V$  and  $E$  are finite, and we will write  $n := |V|$  and  $m := |E|$ , to denote respectively the number of vertices and the number of edges of  $G$ . If two vertices  $a$  and  $b$  are adjacent, we denote the associated edge by the unordered pair  $(a, b)$ . Occasionally, we will also write the edges with a single index running from 1 to  $m$ .

The degree of a vertex  $a \in V$ , denoted  $d(a)$ , is the number of edges attached to it. The graph is called  $k$ -regular when every vertex has the same degree  $k$ . Figure 1 shows an example of a non regular graph on  $n = 7$  vertices and an example of a 3-regular graph on  $n = 12$  vertices.

In order to construct the Laplacian matrix associated to a graph  $G = (V, E)$ , let  $D$  be the diagonal matrix in which  $D(a, a)$  is the degree of vertex  $a$ . Define also the adjacency matrix  $A$  by

$$A(a, b) = \begin{cases} 1 & \text{if } (a, b) \in E \\ 0 & \text{otherwise.} \end{cases}$$

**Definition 2.1.** Given a graph  $G$  with adjacency matrix  $A$  and degree matrix  $D$  as defined above, the Laplacian matrix of  $G$  is given by

$$L := D - A.$$

In other words, the diagonal of  $L$  gives us the vertex degrees, and for  $a \neq b$ ,  $L(a, b) = -1$  if  $(a, b)$  is an edge of  $G$  and  $L(a, b) = 0$  otherwise. To further understand this matrix  $L$ , one can look at the associated quadratic form: given a column vector  $x \in \mathbb{R}^n$ , the Laplacian quadratic form is given by  $x^t L x$ , which is shown to be

$$x^t L x = \sum_{(a,b) \in E} (x(a) - x(b))^2, \tag{2.1}$$

after a simple explicit calculation. So  $x^t L x$  can be seen as a discrete version of the continuous Laplacian functional operator, given by the sum of second order derivatives. The spectrum (eigenvalues) of the continuous Laplacian operator is very useful for partial differential equations (PDE) analysis (see for instance [15]), it is thus natural to expect that the spectrum of its discretized version tells us something about the associated graph.

It will be convenient in Section 3.3 to associate weights to the edges of a graph. Let  $G = (V, E)$  be a graph with  $m$  edges and let  $w = (w_1, \dots, w_m) \in \mathbb{R}^m$  be a vector of weights. The weighted graph  $G = (V, E, w)$  is the graph  $(V, E)$  where edge  $i$  is assigned the weight  $w_i$ . If an edge  $(a, b) \in E$  has

weight  $w_{a,b}$ , the weighted adjacency matrix takes the value  $A(a,b) = w_{a,b}$ , and the entries of  $A$  not corresponding to edges are null. The degree of a vertex  $a \in V$  is defined by the sum of the degrees of its adjacent edges. In that case, the weighted Laplacian matrix  $L = D - A$  satisfies

$$x^t L x = \sum_{(a,b) \in E} w_{a,b} (x(a) - x(b))^2.$$

## 2.2 Spectral Theory

**Eigenvalues and Eigenvectors.** The fact that the spectrum of the Laplacian matrix gives non trivial information on the associated graph is an old idea, which has been extensively used for practical purposes in many domains such as computer science, quantum physics, chemistry, or information theory (see [10]). In particular, extremal eigenvalues of  $L$  provide topological information on the graph which can be used for cyber-risk assessment: we will focus here on the well know algebraic connectivity (Definition 2.2).

Note that by definition,  $L$  is an  $n$ -by- $n$ , real, symmetric matrix. Equality (2.1) shows that  $L$  is positive semidefinite. The spectral theorem implies that there exist real positive numbers  $\lambda_1, \dots, \lambda_n$  and  $n$  mutually orthogonal unit vectors  $\psi_1, \dots, \psi_n$  such that  $\psi_i$  is an eigenvector of  $L$  with eigenvalue  $\lambda_i$ , for  $i = 1, \dots, n$ . Without loss of generality, we order the eigenvalues in increasing order  $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ . Note that the sum of elements in each line of  $L$  is equal to 0, thus  $\lambda_1 = 0$  is an eigenvalue of  $L$ .

**Algebraic Connectivity.** Among all eigenvalues of the Laplacian, the following one is the most popular:

**Definition 2.2.** *The second smallest eigenvalue  $\lambda_2 \geq 0$  of  $L$  is called the algebraic connectivity. When needed, we write  $\lambda_2(G)$  to emphasize the dependence on the underlying graph  $G$ .*

This value, introduced by Fiedler [18], summarizes the "connectedness" of the graph. To illustrate this, let us mention a few known properties of the algebraic connectivity:

1.  $\lambda_2 = 0$  if and only if the graph is disconnected. Since we assumed that the graphs we are dealing with are connected, we will have  $\lambda_2 > 0$  throughout the paper.
2. Removing an edge from a connected graph cannot increase the algebraic connectivity. Said otherwise, the algebraic connectivity is monotone: if  $G$  and  $H$  are two graphs on the same vertex set, that do not share any edge in common, then

$$\lambda_2(G \cup H) \geq \lambda_2(G) + \lambda_2(H) \geq \lambda_2(G),$$

where  $G \cup H$  denotes the union of the graphs  $G$  and  $H$  (since  $G$  and  $H$  have the same vertex set,  $G \cup H$  is necessarily connected).

**Community Structure.** To better understand the algebraic connectivity and its properties, let us explain how it measures the community structure of a graph.

Assume first that  $G$  is a network that is separated into several distinct disconnected blocks, or communities. In that case, the Laplacian matrix is block-diagonal, and each block has its own spectrum, containing the value 0 associated to a constant eigenvector (the multiplicity of the eigenvalue 0 of  $G$  is the number of blocks). Now imagine that we connect this graph by introducing a small number of edges connecting the initially separated communities. Then the multiplicity of 0 becomes 1,

and small eigenvalues appear, whose associated eigenvectors have almost constant values along the vertices belonging to each initially separated subgraph. This justifies the use of spectral algorithms for classification into subgraph communities. A low value of  $\lambda_2$  corresponds to networks where it is possible to remove a few edges and obtain a clean cut into separated communities. On the other hand,  $\lambda_2(G)$  high means that  $G$  is not structured and separation into subgraphs is not clear.

Let  $K_n$  be the complete graph on  $n$  vertices, i.e. the graph where each vertex has degree  $n - 1$ , which means that each vertex is connected to every other vertex. A direct calculation (see for instance [6]) shows that  $\lambda_2(K_{10}) = 10$ . On the other hand, consider the graph  $G_{10}$  represented in Figure 3a, which also has 10 vertices: it is clear that it can be divided into 2 connected communities, and we have  $\lambda_2(G) = 0.298$ . The graph  $\tilde{G}$  in Figure 3b has more connections and satisfies  $\lambda_2(\tilde{G}) = 1$ .

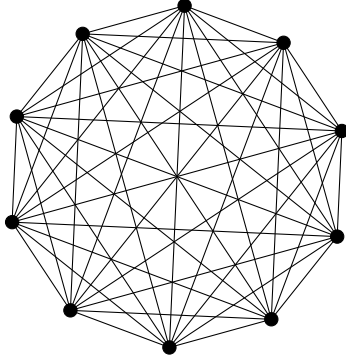
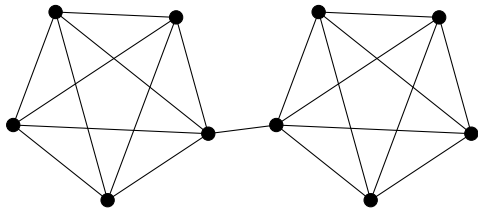
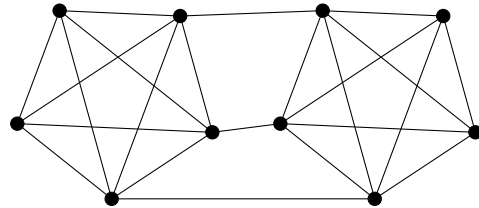


Figure 2: Complete graph  $K_{10}$  with  $\lambda_2(K_{10}) = 10$ .



(a) A graph  $G$ , such that  $\lambda_2(G) = 0.298$ .



(b) A graph  $\tilde{G}$ , such that  $\lambda_2(\tilde{G}) = 1$ .

Figure 3: The right hand graph  $\tilde{G}$  is "more connected" than the left hand graph  $G$ .

**Expansion Properties.** Let us end this section by mentioning a quantity which is controlled by  $\lambda_2$  (in the sense of inequality (2.2)), the so-called isoperimetric ratio (or Cheeger constant, or minimal expansion rate), that also measures connectedness properties.

Let  $S$  be a subset of vertices of the network. The frontier  $\partial(S)$  of  $S$  is the set of edges with one end in  $S$  and the other end outside  $S$ :

$$\partial(S) := \{(a, b) \in E \mid a \in S, b \notin S\}.$$

One way of measuring how well  $S$  is connected to the rest of the graph is to count the number of edges connecting  $S$  to the rest of the graph and compare it with either the number of vertices in  $S$  or the number of vertices in  $V \setminus S$ . More precisely, define

$$\theta(S) := \frac{|\partial(S)|}{\min(|S|, |V \setminus S|)}.$$



The isoperimetric ratio  $\theta_G$  is then given by the minimum value of  $\theta(S)$  over all subsets  $S$  containing at most half of the vertices, i.e.:

$$\theta_G := \min \left\{ \theta(S), |S| \leq \frac{n}{2} \right\}.$$

Then the following inequality relates  $\lambda_2$  with  $\theta_G$ :

$$\frac{\lambda_2}{2} \leq \theta_G. \quad (2.2)$$

In the case of a  $k$ -regular graph, the so-called Cheeger's inequality [7] says that  $\theta_G \leq \sqrt{2k\lambda_2}$ . From inequality (2.2), we know that graphs with large algebraic connectivity also have large isoperimetric ratio, and hence high connectedness relative to this measure.

### 3 Model and Assumptions

In this section, we set up an insurance demand model, which is based on the underlying graph characteristics.

#### 3.1 SIS Epidemic Model

**Inhomogeneous SIS Model.** In the previous section, we have only described intrinsic topological properties of graphs, independently of any probabilistic model. We introduce now a particular probabilistic model of epidemic spread, which is known as the Susceptible-Infected-Susceptible (SIS) model. It is a time dynamic Markovian model, in which the state at time  $t$  of edge  $i$ , denoted  $X_i(t)$ , is a Bernoulli random variable:  $X_i(t) = 0$  when edge  $i$  is "healthy", but susceptible to the infection, and  $X_i(t) = 1$  when edge  $i$  is infected, meaning that the connection going through edge  $i$  is interrupted. We assume that only when infected, edge  $i$  can infect each edge  $j$  sharing a common vertex with rate  $\beta_{ij}$ , which means that the infection process from edge  $i$  to edge  $j$  is a Poisson process with intensity  $\beta_{ij} > 0$ .

Each edge  $i$  can be cured, i.e. its state  $X_i$  jumps from 1 to 0, and the curing process is also a Poisson process with intensity  $\delta_i > 0$ . All these Poisson infection and curing processes are assumed independent.

More precisely, throughout the paper, we work on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ , where the filtration  $\mathbb{F} = (\mathcal{F})_{t \geq 0}$  is right-continuous and  $\mathcal{F}_0$  contains the  $\mathbb{P}$ -negligible sets. The continuous-time Markov process  $X_t := (X_1(t), \dots, X_m(t))$  has càdlàg trajectories and state space  $E = \{0, 1\}^m$ . In the SIS epidemics model, we assume that  $X$  is a Feller process with infinitesimal generator  $\mathcal{G} : C(E) \rightarrow \mathbb{R}$  given by

$$\mathcal{G}f(x) = \sum_{i=1}^m \left\{ (1 - x_i) \sum_{j=1}^m \tilde{a}_{ij} \beta_{ij} x_j + \delta_i x_i \right\} (f(x^i) - f(x)), \quad x \in E, f \in C(E), \quad (3.1)$$

where  $x_j^i = x_j$  for  $I \neq j$  and  $x_i^i = 1 - x_i$ ,  $(\tilde{a}_{ij})_{a \leq i, j \leq m}$  are the entries of the edge-adjacency matrix of the graph, and  $C(E)$  denotes the set of real functions on  $E$ .

**Remark 3.1.** *In order to focus on communication interruption, we present here the SIS epidemic model on edges, instead of the more common SIS model on vertices of a graph. The SIS model on edges of a graph  $G$  is equivalent to the standard SIS model on the vertices of the line graph of  $G$ .*

The finite-dimensional distributions of the process  $X$  can be analyzed using a system of coupled ordinary differential equations (ODEs), as follows. The state  $X_i(t)$  of an edge  $i$  is a time-dependent Bernoulli random variable that can change from 1 to 0 with curing rate  $\delta_i$  and from 0 to 1 with infection rate  $\sum_{j=1}^m \tilde{a}_{ij}\beta_{ij}X_j(t)$ , that depends on the "neighbour" Bernoulli variables. The variation during a time interval  $\Delta t$  is given by

$$\frac{X_i(t + \Delta t) - X_i(t)}{\Delta t} = (1 - X_i(t)) \sum_{j=1}^m \tilde{a}_{ij}\beta_{ij}X_j(t) - \delta_i X_i(t).$$

Formally, taking the expectation in the previous equality and the limit when  $\Delta t$  goes to 0, we get, denoting  $v_i(t) = \mathbb{P}(X_i(t) = 1)$ ,

$$\frac{dv_i(t)}{dt} = \sum_{j=1}^m \tilde{a}_{ij}\beta_{ij}v_j(t) - \sum_{j=1}^m \tilde{a}_{ij}\beta_{ij}\mathbb{E}[X_i(t)X_j(t)] - \delta_i v_i(t). \quad (3.2)$$

Equation (3.2) is derived in a precise way as a consequence of Kolmogorov's forward equation.

**Mean Field Approximation.** The system of ODEs (3.2) is not closed. Indeed, the ODE for  $v_i$  depends on pair correlations  $\mathbb{E}[X_i(t)X_j(t)]$  for every edge  $j$  that shares a common vertex with  $i$ . Similarly, a system of ODEs can be written for pair correlations that depends on triples, and so on, up to the term  $\mathbb{E}[X_1(t) \dots X_m(t)]$ . In total, the closed system contains  $2^m$  equations. This is reminiscent of the cardinality of the state space  $E$  of  $X$ , equal to  $2^m$ , which is rapidly prohibitive in practice.

In order to overcome this, the idea of the  $N$ -Intertwined Mean Field Approximation (NIMFA) developed in [38, 37] is to replace all terms  $\mathbb{E}[X_i(t)X_j(t)]$  in (3.2) by  $\mathbb{E}[X_i(t)]\mathbb{E}[X_j(t)]$ , thus assuming null correlation. This is referred to as closing the model at the level of pairs and this leads to an upper bound for the probabilities  $v_i(t)$ ,  $i = 1, \dots, m$ . We refer to [12, 16] for details and analysis of higher order approximation schemes. The obtained NIMFA system of non linear ODEs is given by

$$\frac{dv_i(t)}{dt} = \left( \sum_{j=1}^m \tilde{a}_{ij}\beta_{ij}v_j(t) \right) (1 - v_i(t)) - \delta_i v_i(t), \quad i = 1, \dots, m. \quad (3.3)$$

The idea of the NIMFA is to replace a system of  $2^m$  linear equation by a closed system of  $m$  non linear ODEs, at the price of exactness. The system (3.3) can be solved numerically to get an approximation of (3.2), for a fixed terminal time  $T$  ([22]).

Instead, the model that we consider in this paper is the steady state vector  $(v_1(\infty), \dots, v_m(\infty))$  associated to (3.3), that is obtained by assuming that the left hand side of (3.3) is null, which gives

$$v_i = \frac{\beta \sum_{j=1}^m \tilde{a}_{ij}v_j}{\delta_i + \beta \sum_{j=1}^m \tilde{a}_{ij}v_j}, \quad i = 1, \dots, m, \quad (3.4)$$

where  $v_i$  denotes the value  $v_i(\infty)$ , and where we took a constant value  $\beta_{ij} = \beta$ . The null vector is an obvious solution to (3.4), and it is also the steady state solution of the exact SIS model (3.2). The approximation (3.3) can give rise to a non zero solution of (3.4), called an *endemic state*. Sufficient conditions for the existence of such an endemic state are given in [31] in terms of a generalized Laplacian matrix.

In the next subsection, we are going to write an optimization problem, that consists in choosing  $(\delta_1, \dots, \delta_m)$  in order to maintain high connectivity of the graph when it is in the endemic state. Note that we do not need to include the endemic state existence conditions provided in [31] in our optimization problem, since it will be embedded in the budget constraint (see Remark 3.4(v)).

### 3.2 Demand Model

Suppose that an economic agent, typically a company, faces a potential malicious cyber attack on a given network, represented by a graph  $G_0 = (V_0, E_0)$  at time  $t = 0$ . If a cyber attack occurs, then it can interrupt the communications going through infected edges. Assumption 3.2 below says that buying insurance protection helps reconnecting the broken edges faster. Indeed, *Stand Alone Cyber contracts* usually contain, among others, clauses stipulating the implementation of repairing actions, through the intervention of a specified cyber security company.

**Assumption 3.1.** *We assume that the cyber attack spreads in the network according to an inhomogeneous SIS model, with constant infection rate  $\beta > 0$ .*

The fact that  $\beta$  is constant is just for simplicity. All the results below readily extend to inhomogeneous infection rates.

The decision variables are the curing rates  $\delta_i$ ,  $i = 1, \dots, m$ , controlled by the agent.

**Assumption 3.2.** *The level of cyber insurance purchased influences directly the curing rates on edges.*

This last assumption can also be interpreted in terms of prevention: a company can indeed engage in prevention efforts to reduce the risk of communication interruption. It can for instance invest in passwords protection, data encryption, firewalls, back-ups, upgrades, application patching, users training, or merely network mapping and inventory, which are all examples of *cyber hygiene* measures. It can also use more sophisticated processes, like honeypots, that consist in having servers meant to deceive potential attackers, making them believe that they have taken control of the machine, where in fact they are being identified and neutralized.

**Insurance Cost.** These measures are expected to increase the recovery rates. Whether they are interpreted as pure prevention or as insurance, these measures come with a cost, that we describe now. Since in our model, the vector  $\delta$  is directly controlled by the agent, it is tempting to use a cost function  $c(\delta)$ , classically assumed increasing and convex. However, in order to be closer to standard insurance models and practice, we choose to use functionals of the expectation, such as the expectation principle with a loading factor, to represent the insurance premium.

Assume that in the endemic state, the interruption of communication going through edge  $i$  generates a loss  $Z_i$  for the company, where  $(Z_k)_{k \geq 1}$  is a sequence of independent and non negative random variables (not necessarily identically distributed). Let  $X(\delta)$  denote the total loss of the graph, when the curing rates are given by a vector  $\delta \in \mathbb{R}^m$ , and let  $\delta_0$  be the vector of curing rates before any insurance purchase. Improving the curing rates from  $\delta_0$  to  $\delta$  via insurance decreases the average loss by  $\mathbb{E}[X(\delta_0)] - \mathbb{E}[X(\delta)]$ . Note that the effect of insurance on the graph total loss, i.e. moving from  $X(\delta_0)$  to  $X(\delta)$  is highly non linear, and less obvious than traditional proportional or non proportional insurance or reinsurance contracts.

**Assumption 3.3.** *The insurance cost function  $c : \mathbb{R}_+^m \rightarrow \mathbb{R}_+$  is given by*

$$c(\delta) := \varphi \{ \mathbb{E}[X(\delta_0)] - \mathbb{E}[X(\delta)] \}, \quad (3.5)$$

where  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is non decreasing.

The expected value principle with a loading factor  $\theta > 0$  corresponds to the particular case where  $\varphi(x) = (1 + \theta)x$ . Note that  $c(\delta)$  is non decreasing in each  $\delta_i$ , but the convexity or concavity of this function is not guaranteed (see [31], part IV). Thus in this graph epidemics context, the expected value principle is already a non trivial operation.

**Remark 3.2.** *Note that to compute the total loss  $X(\delta)$ , we take a picture of the graph at infinite time horizon in the endemic state and only consider the infected edges at that horizon. We thus ignore the trajectories that lead to the endemic state, in particular, we ignore the edges that have been infected and then cured. Taking this into consideration would lead to a path dependent cost, and to a stochastic control problem that is out of the scope of this paper. We rather focus on the endemic state situation, that can be interpreted as cascading infections.*

**Optimization Criteria.** Let us now discuss the optimization criteria. The agent's goal is to maintain the network as connected as possible in the endemic state. As discussed in Section 2, one natural way to measure a graph's connectivity is given by the second smallest eigenvalue of the associated Laplacian matrix. Recall that the initial graph  $G_0 = (V, E_0)$  is fixed and that the malicious attack is modeled as a SIS epidemic on the graph. At a given time  $t > 0$ , the state of the network is represented by a random graph  $G_t = (V, E_t)$  where  $E_t \subset E_0$  denotes the set of remaining active edges at time  $t$ . As mentioned in Section 3.1, the steady state associated with the exact SIS model corresponds to the graph  $\tilde{G}_\infty = (V, \tilde{E}_\infty)$  where  $\tilde{E}_\infty = E_0$ , i.e. the state where all infected edges have been cured and the infection probabilities are all equal to zero. However, the NIMFA approximation gives rise to a possible endemic state, on a random graph  $G_\infty = (V, E_\infty)$ , where a positive fraction of edges remain infected, and where infection probabilities are constant and solve the problem (3.4).

We consider the Laplacian matrix  $L_\infty$  associated to  $G_\infty$ , and its expectation

$$L := \mathbb{E}[L_\infty],$$

which is defined as the componentwise expectation. We suppose that the agent seeks to maximize the algebraic connectivity  $\lambda_2(L)$  computed from  $L$ . The optimization problem is then:

$$\begin{aligned} & \underset{\delta \in \mathcal{A}}{\text{maximize}} && \lambda_2(L) \\ & \text{subject to} && c(\delta) \leq B, \end{aligned} \tag{3.6}$$

where  $B > 0$  is a given total available budget, and  $\mathcal{A}$  is a set of admissible curing rates. Note that  $L$  depends on  $\delta$  in a highly non linear way. The algebraic connectivity is a concave function of  $L$  (see [5], Section 2.1), but it is neither concave nor convex in the vector  $\delta$ . Furthermore, the map  $\delta \mapsto c(\delta)$  is not convex in general, and hence problem (3.6) is far from a convex optimization problem in the variable  $\delta$ , but as shown in Theorem 3.1, it is possible recast it as a convex problem.

The set  $\mathcal{A}$  of admissible controls on  $\delta$  is just taken to be  $(0, +\infty)^m$ . One can also set lower bounds  $\delta^0 = (\delta_1^0, \dots, \delta_m^0)$  that represent curing rates when no insurance is purchased. In that case,  $\mathcal{A}$  is the set of vectors  $\delta$  such that  $\delta \geq \delta^0$ . Since this does not qualitatively change the model, nor add any particular difficulty, we just set  $\delta^0 \equiv 0$ .

**Remark 3.3.** *We consider here the average Laplacian matrix and analyze the associated algebraic connectivity. One could also consider the random algebraic connectivity  $\tilde{\lambda}_2$  computed from  $L_\infty$  (or from  $L_t$ , for a fixed time  $t$ ) and then maximize  $\mathbb{E}[\tilde{\lambda}_2]$ . To do so, one needs at least some information on the distribution of  $\tilde{\lambda}_2$ , which is difficult in general. One possibility is to approximate the distribution of  $\tilde{\lambda}_2$  when the network size is large by using known results on the distribution of the eigenvalues of large random matrices. In the case of Laplacian matrices of random graphs, analytic results are given in [8] and [13], but, to the best of our knowledge, no results for the specific case of the SIS model exist yet in the literature.*

### 3.3 A Solution

To solve the non convex problem (3.6), the idea is to reformulate it as a problem on the vector variable  $v$  first, and then deduce from it a representation of an optimal  $\delta$ . Recall that to each edge  $i$  is associated a non negative random variable  $Z_i$ , representing the company's monetary loss if communication through edge  $i$  is interrupted. Let  $c_i$  denote the expectation of  $Z_i$ . The following theorem is the main result of this paper:

**Theorem 3.1.** *An optimal solution to problem (3.6) is given by*

$$\delta_i^* = \left( \beta \sum_{j=1}^m \tilde{a}_{ij} v_j^* \right) \frac{1 - v_i^*}{v_i^*}, \quad (3.7)$$

where  $\bar{v}^* = (1 - v_1^*, \dots, 1 - v_m^*)$  is the  $w$  component of a solution to the following convex optimization problem in the variables  $\gamma \in \mathbb{R}$ ,  $\mu \in \mathbb{R}$  and  $w \in \mathbb{R}^m$ :

$$\begin{aligned} & \sup \quad \gamma \\ & \text{subject to} \quad L(w) - \gamma I + \mu \mathbf{1}\mathbf{1}^t \succcurlyeq 0 \\ & \quad \quad \quad 0 \leq w \leq 1 \\ & \quad \quad \quad \varphi \left( \sum_{j=1}^m c_j w_j \right) \leq B, \end{aligned} \quad (3.8)$$

where  $\mathbf{1}^t := (1, \dots, 1) \in \mathbb{R}^m$ ,  $L(w)$  denotes the Laplacian matrix of the weighted graph  $(V, E_0, w)$ , and where the notation  $M \succcurlyeq 0$  means that the matrix  $M$  is positive semidefinite.

Let us discuss this theorem, before proving it.

- (i) First, note that it is expected that the optimal vector  $\delta^*$  should be increasing with the infection rate  $\beta$ . This theorem says that it is even linear in  $\beta$ .
- (ii) From Step 3 in the proof of this theorem, we know that problem (3.8) can be rephrased as choosing the weights  $w$  of the graph  $(V, E_0, w)$  in order to maximize the algebraic connectivity. We also know from Lemma 3.1 below that the weights on each edge of the graph correspond to non infection probabilities. Theorem 3.1 then says that to solve (3.6), it is enough to choose first the optimal non infection probabilities  $(1 - v_i^*)_{1 \leq i \leq m}$ , and then the associated curing rates, such that  $(v_i^*)_{1 \leq i \leq m}$  satisfies the steady state equations (3.4).
- (iii) From (3.7), we see that the optimal insurance demand  $\delta_i^*$  depends on the infection probability  $v_i^*$  of edge  $i$  and on the infection probabilities of the neighbouring edges. The vector  $(1 - v^*)$ , solution to (3.8), highly depends on the whole network topology. Hence this theorem emphasizes the hierarchization of the edges induced by the optimal demand vector  $\delta^*$ .
- (iv) This theorem makes it possible to compute optimal insurance demand parameters for large graphs, containing up to several thousands edges. Indeed, the numerical procedure used to solve (3.8) is scalable, and several well known interior point methods can be used to solve it in polynomial time (see Section 4 for details). For even larger graphs, containing several hundred thousands edges, gradient algorithms are available, allowing to solve (3.8) numerically (see Section 5 in [5]).

- (v) Using the gradient descent algorithms mentioned in (iv), adding a convex constraint in (3.8) does not add much difficulty to obtain a numerical solution. In particular, other cost functions are possible, as long as they are convex, as a function of  $w$ .

The following elementary lemma clarifies the dependence of  $\mathbb{E}[L_\infty]$  on  $v$ .

**Lemma 3.1.**  $L = \mathbb{E}[L_\infty]$  is the Laplacian matrix associated to the weighted graph  $G = (V, E_0, \bar{v})$ , where each edge  $\ell$  is given the weight  $\bar{v}_\ell = 1 - v_\ell$ , with  $(v_1, \dots, v_m)$  satisfying (3.4).

**Proof.** Every non diagonal entry  $L_{ij}$  of  $L_\infty$  is equal to  $-1$  if edge  $(i, j)$  is active (i.e. non infected) and is equal to  $0$  otherwise. Thus  $\mathbb{E}[L_{ij}]$  is nothing else than the negative of the infection probability of edge  $(i, j)$  in the endemic state, equal to  $-(1 - v_{ij}) = -\bar{v}_{ij}$ . By definition, the diagonal entry of a given line  $i$  of  $L_\infty$  is equal to  $\sum_{j=1}^m L_{ij}$  and so

$$L_{ii} = (\mathbb{E}[L_\infty])_{ii} = \sum_{j=1}^m \mathbb{E}[L_{ij}] = \sum_{j=1}^m \bar{v}_{ij}.$$

These values are by definition the entries of the Laplacian associated to the weighted graph  $(V, E_0, \bar{v})$ .  $\square$

Let us now compute the cost function  $c(\delta)$  as a function of  $v$ .

**Lemma 3.2.** *The expected loss can be written as follows:*

$$\mathbb{E}[X(\delta_0)] - \mathbb{E}[X(\delta)] = \sum_{j=1}^m c_j \bar{v}_j.$$

*In particular, the cost function  $c$  given in (3.5) is a convex function of  $v$ .*

**Proof.** For each edge  $i$  in  $\{1, \dots, m\}$ , let  $X_i^\infty$  be a Bernoulli random variable, equal to  $1$  if edge  $i$  is infected in the endemic state and equal to  $0$  otherwise. The total loss  $X(\delta)$  can be written as

$$X(\delta) = \sum_{j=1}^m X_j^\infty Z_j.$$

Using the independence of the sequence  $(Z_k)_{k \geq 1}$  and the SIS Markov process  $X$ , we get

$$\mathbb{E}[X(\delta)] = \sum_{j=1}^m \mathbb{E}[X_j^\infty] \mathbb{E}[Z_j] = \sum_{j=1}^m v_j c_j.$$

Since no edge is infected in the initial graph  $(V, E_0)$ , we get

$$\mathbb{E}[X(\delta_0)] - \mathbb{E}[X(\delta)] = \sum_{j=1}^m c_j - \sum_{j=1}^m v_j c_j = \sum_{j=1}^m c_j \bar{v}_j.$$

$\square$

**Proof. of Theorem 3.1.** Step 1: Consider the following eigenvalue maximization problem:

$$\begin{aligned} & \sup \quad \lambda_2(L(w)) \\ & \text{subject to} \quad 0 \leq w \leq 1 \\ & \quad \quad \quad \varphi \left( \sum_{j=1}^m c_j w_j \right) \leq B \end{aligned} \tag{3.9}$$

in the variable  $w \in \mathbb{R}^m$ . We first prove that (3.9) is a convex problem (i.e. maximizing a concave function under convex constraints).

Since  $\varphi$  is a convex function, both constraints in (3.9) are clearly convex. Let us prove that  $w \mapsto \lambda_2(L(w))$  is a concave map. Since by definition, all rows of  $L(w)$  sum to 0, the vector  $\mathbf{1}$  is an eigenvector of  $L(w)$  associated to the smallest eigenvalue 0. By the Courant-Fisher Theorem, we have

$$\lambda_2(L(w)) = \inf_{\|u\|=1, \mathbf{1}^t u = 0} u^t L(w) u = \inf_{\|u\|=1, \mathbf{1}^t u = 0} \sum_{(a,b) \in E_0} w_{a,b} (u(a) - u(b))^2,$$

which shows that  $\lambda_2(L(w))$  is a pointwise infimum of linear functions of  $w$ , and hence is concave in  $w$ . Note that this implies that  $w \mapsto \lambda_2(L(w))$  is continuous, and since the constraint set in (3.9) is compact, a solution  $w^*$  of problem (3.9) exists (uniqueness is however not guaranteed).

Step 2: Any choice of  $\delta$  induces non infection probabilities  $\bar{v}(\delta)$ , and from Lemma 3.1, we know that  $L(\delta) = \mathbb{E}[L_\infty]$  is the graph Laplacian with weights  $\bar{v}(\delta)$ . By the optimality of  $w^*$ ,  $\lambda_2(L(w^*)) \geq \lambda_2(L(\delta))$  for any  $\delta \in \mathcal{A}$ . Define  $\delta^*$  by:

$$\delta_i^* := \left( \beta \sum_{j=1}^m \tilde{a}_{ij} (1 - w_j^*) \right) \frac{w_i^*}{1 - w_i^*}, \quad 1 \leq i \leq m,$$

and let  $\bar{v}^* = 1 - v^* := w^*$ . Then by construction, the vector of curing rates  $\delta^*$  is such that the associated non infection probabilities are given by  $\bar{v}^*$ . Also by construction, the vector  $v^*$  solves the steady state equations (3.4) with curing rates  $\delta^*$ . Finally, by Lemma 3.2, given the constraints in (3.9),  $\delta^*$  satisfies the budget constraints from (3.6), which means that  $\delta^*$  solves (3.6).

Step 3: It remains to prove that solving (3.9) is equivalent to solving (3.8), which is standard in eigenvalue optimization theory. We provide an elementary proof here for completeness of the paper. Let  $y_1, \dots, y_n$  be eigenvectors of  $L(w)$  with eigenvalues  $0 = \lambda_1 \leq \dots \leq \lambda_n$ , such that  $\|y_i\| = 1$  and  $y_i^t y_j = 0$  for  $i \neq j$ . The spectral decomposition of  $L(w)$  writes:

$$L(w) = 0 \times \mathbf{1}\mathbf{1}^t + \lambda_2 y_2 y_2^t + \lambda_n y_n y_n^t.$$

Using these eigenvectors, we can write the spectral decomposition of  $M := L(w) - \gamma I + \mu \mathbf{1}\mathbf{1}^t$  appearing in (3.8) as follows:

$$M = (0 - \gamma + \mu) \mathbf{1}\mathbf{1}^t + (\lambda_2 - \gamma) y_2 y_2^t + (\lambda_n - \gamma) y_n y_n^t.$$

$M$  is positive semidefinite if and only if all its eigenvalues are non negative. So the positive semidefinite constraint in (3.8) can be written  $\mu \geq \gamma, \lambda_2 \geq \gamma, \dots, \lambda_n \geq \gamma$ , which is equivalent to  $\mu \geq \gamma$  and  $\lambda_2 \geq \gamma$ . Then by maximizing  $\gamma$ , we maximize  $\lambda_2$ . The proof is now complete.  $\square$

**Remark 3.4.** (i) A solution to problem (3.9) need not be unique. This implies that there may be different vectors  $\delta^*$  achieving optimal algebraic connectivity in (3.6).

(ii) The optimal value of problem (3.9) is positive: we can indeed choose  $w$  with arbitrary small positive values to obtain  $\lambda_2(w) > 0$ . This implies that with the optimal choice  $\delta^*$ , we ensure that the graph will remain connected after an endemic malicious attack.

(iii) When  $\varphi$  appearing in the cost function is invertible, the last constraint in (3.9) can be written  $\sum_{j=1}^m c_j w_j \leq \varphi^{-1}(B)$ , which makes (3.9) a semidefinite programming problem (SDP). This is the case when we consider the expectation principle with a loading factor.

(iv) Note that in problem (3.8), we authorize entries of  $w$  to be equal to 1, which implies that we authorize some entries of the optimal vector  $\delta$  to be  $+\infty$ . If for some  $i$ ,  $\delta_i = +\infty$ , then the curing process jumps instantly, the corresponding edge is never infected anymore, and the endemic infection probability of this edge is indeed given by  $v_i = 1 - w_i = 0$ . Note also that this feature of the model is not ruled out by the budget equation in problem (3.6) in the variable  $\delta$ , since this constraint only involves expectations. One can include the explicit constraint  $\delta < +\infty$  in problem (3.6), in that case the constraint  $w < 1$  has to be added to (3.8), which makes the supremum in both problems (3.8) and (3.6) not necessarily attained, which only happens in practice when the available budget  $B$  is large.

(v) Assume for simplicity that  $\varphi$  is invertible. If the available budget  $B$  is large, in the sense that  $\sum_{j=1}^m c_j \leq \varphi^{-1}(B)$ , then the optimal solution to (3.9) is obviously  $w \equiv 1$ , which means that it is possible to reach the steady state where all edges are cured. To be able to analyze the solution in the endemic state, we can realistically assume that  $\sum_{j=1}^m c_j > \varphi^{-1}(B)$  since in that case, at least one entry in a solution  $w^*$  of (3.9) is strictly less than one. The numerical solutions presented in the next section implicitly include this assumption, which merely says that the available budget for prevention or insurance is less than the average cost if all edges are infected.

**Probabilistic Interpretation.** In [36], the authors introduce the Fastest Mixing Markov Process Problem (FMMP for short). Consider a symmetric continuous time Markov process  $Y$  on the vertices  $V$  of the graph  $G = (V, E_0)$ . The edges set  $E_0$  represents the allowed transitions. To each edge  $(a, b) \in E_0$  is associated a weight  $w_{ab} \geq 0$ , equal to the transition rate of the process  $Y$  between vertices  $a$  and  $b$ . The distribution  $\pi(t) \in \mathbb{R}^n$  of  $Y$  at time  $t$  can be described with the weighted Laplacian  $L(w)$  associated to  $(V, E_0, w)$ :

$$\frac{d\pi(t)}{dt} = -L\pi(t).$$

The long time behaviour of  $\pi(t)$  is determined by the algebraic connectivity  $\lambda_2 = \lambda_2(L(w))$ . More precisely, the uniform distribution  $\mathbf{1}/n$  is an equilibrium distribution for  $Y$ , and the rate of convergence of  $\pi(t)$  to  $\mathbf{1}/n$  is governed by  $\lambda_2$ :

$$\sup_{\pi(0)} \|\pi(t) - \mathbf{1}/n\|_{\text{TV}} \leq \frac{\sqrt{n}}{2} e^{-\lambda_2 t},$$

where the supremum is taken over initial distributions  $\pi(0)$  and where  $\|\cdot\|_{\text{TV}}$  is the total variation distance between probability measures. To make  $\pi(t)$  converge as fast as possible towards the equilibrium distribution, the FMMP consists in maximizing the algebraic connectivity, under a budget constraint on the transition rates  $w_{ab}$ . We refer the reader to [36, 4] for details and further physical interpretations.

If  $Y$  represents one unit of information randomly moving along the vertices of the graph, the agent's problem consists in making the information reach all the vertices of the graph as fast as possible, even in an endemic situation. This indeed matches well with the idea of maintaining the graph as connected as possible.

Finally, note that the optimal demand vector  $\delta^*$  given in (3.7) depends on the underlying graph topology via the fastest mixing problem on the graph. This fact is not obvious from the formulation (3.6) of the problem.



## 4 Numerical Analysis

In this section, we consider particular graph examples and numerically solve problem (3.6) on those graphs.

Throughout this section, we will assume that  $\varphi$  appearing in the last constraint of (3.8) is invertible, which makes the latter a semidefinite programming problem (SDP). Theorem 3.1 asserts that it is enough to solve the SDP (3.8), which is a particular convex optimization problem. It is well known that SDPs can be solved very efficiently using interior points methods, that were initially developed for quadratic and linear programming. Linear Problems (LPs) are particular cases of SDPs and many features of LPs extend to the case of SDPs. We refer the reader to [39, 32] for a detailed presentation and analysis of SDPs and their large variety of applications.

SDPs include in their constraint set a so-called *Linear Matrix Inequality* (LMI), i.e. an inequality of the form  $\sum_{k=1}^N x_k M_k \succcurlyeq 0$ , where  $M_1, \dots, M_N$  are given symmetric matrices,  $(x_1, \dots, x_N) \in \mathbb{R}^N$  are the optimization variables and  $M \succcurlyeq 0$  means that  $M$  is positive semidefinite, i.e.  $u^t M u \geq 0$  for every  $u \in \mathbb{R}^N$ . The inequality  $L(w) - \gamma I + \mu \mathbf{1}\mathbf{1}^t \succcurlyeq 0$  appearing in (3.8) is indeed a LMI, since  $L(w)$  can be written as a linear combination

$$L(w) = \sum_{(a,b) \in E_0} w_{ab} L^{ab},$$

where  $L^{ab}$  is a symmetric  $n \times n$  matrix such that  $L_{a,a}^{ab} = L_{b,b}^{ab} = 1$ ,  $L_{a,b}^{ab} = L_{b,a}^{ab} = -1$  and all the other terms are equal to 0.

Many popular SDP solvers, such as CSDP, SDPA, SDPNAL, SDPT3, SeDuMi etc are available as free and open softwares, many of them being compatible with several scientific computing platforms such as Matlab, Python or *R*. For the examples presented below, we use the Matlab library CVX [11], combined with the SDPT3 solver.

Throughout the rest of this section, the depicted graphs will be randomly generated, using the Erdős-Rényi model: the number  $n$  of nodes is fixed, and there is an edge connecting two given vertices with fixed probability  $p$ , independently of the other edges.

Let us start with a small sized graph example, represented in Figure 4. This graph has 8 nodes and 13 edges which are labeled with the optimal  $\delta$ , solution to (3.6). This is a realization of an Erdős-Rényi random graph with  $p = 0.5$ . To obtain the optimal  $\delta$ , we solve the following SDP, assuming that  $\varphi$  is invertible:

$$\begin{aligned} & \sup \quad \gamma \\ & \text{subject to} \quad L(w) - \gamma I + \mu \mathbf{1}\mathbf{1}^t \succcurlyeq 0 \\ & \quad \quad \quad 0 \leq w \leq 1 \\ & \quad \quad \quad \sum_{j=1}^m c_j w_j \leq \varphi^{-1}(B), \end{aligned} \tag{4.1}$$

where  $\varphi^{-1}(B) = 2$ , allowing several different choices for  $\varphi$ , and where the expected loss  $c_j = 1$  for every  $j$ . Then we obtained the optimal  $\delta$  using formula (3.7), with a constant infection rate  $\beta = 1$ . The following table shows the optimal non infection probability associated to each edge, along with the optimal choice of  $\delta$ :

Edge	$w^*$	$\delta^*$
(1, 4)	0.2154	1.1149
(1, 7)	0.1231	0.2591
(1, 8)	0.1846	0.3762
(2, 4)	0.3077	1.3128
(2, 7)	0.0308	0.0586
(3, 4)	0.1846	0.7942
(3, 6)	0.0923	0.2065
(4, 5)	0.1846	0.7942
(4, 6)	0.1846	1.0032
(4, 7)	0.1846	0.8360
(5, 6)	0.0923	0.2065
(6, 8)	0.2154	0.4561

Note that the vector  $w^*$  sums to 2, which is coherent with the fact that the constraint in the SDP above is going to be saturated. Due to some symmetry on the graph,  $\delta^*$  contains several times the same value. The completely symmetric case corresponds to  $k$ -regular graphs, where all nodes have the same degree  $k$ , which implies that each edge has  $2k$  incident edges. In that case, all nodes and edges play the same role, problem (3.8) has a constant vector  $w^*$  as solution, and Equation (3.7) gives  $\delta_i = \beta 2k\bar{w}(1 - \bar{w})$  for every  $j$ , where  $\bar{w}$  is the constant value in the vector  $w^*$ .

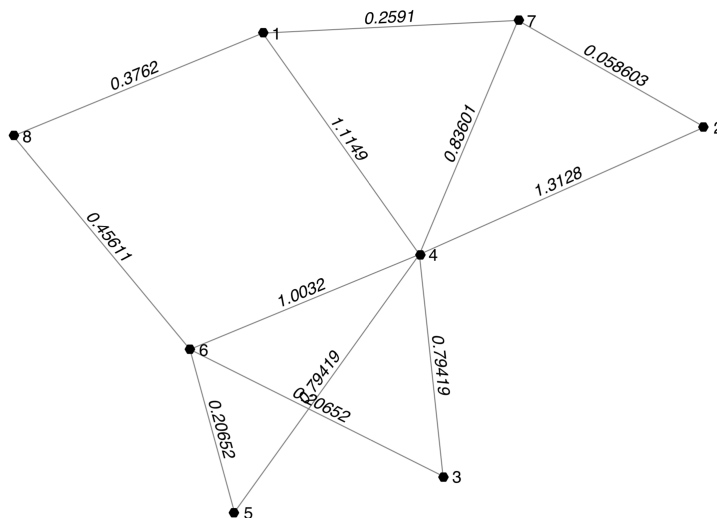


Figure 4: Erdős-Rényi random graph with  $(n, p) = (8, 0.5)$ .

We consider now a larger graph, with  $n = 30$  nodes and  $m = 102$  edges, obtained again as an Erdős-Rényi random graph, with  $(n, p) = (30, 0.2)$ . We solved the SDP (4.1), with  $\varphi^{-1}(B) = 5$  and  $c_j \equiv 1$ . Figure 5 shows this graph, where the dotted edges correspond to null values of  $w^*$  and hence null values of  $\delta^*$ . There are 10 such edges in this particular case, for which the optimal insurance demand should be zero, which can also be interpreted as the fact that the optimal prevention effort associated to these edges should be small compared to the rest of the edges. The same graph is represented in Figure 6, where the width of each edge is proportional to the optimal value  $\delta^*$ . For a fixed budget  $B$ , this tells *where* the budget should be invested in terms of insurance or prevention, in order to optimize the connectivity of the graph after a malicious attack.

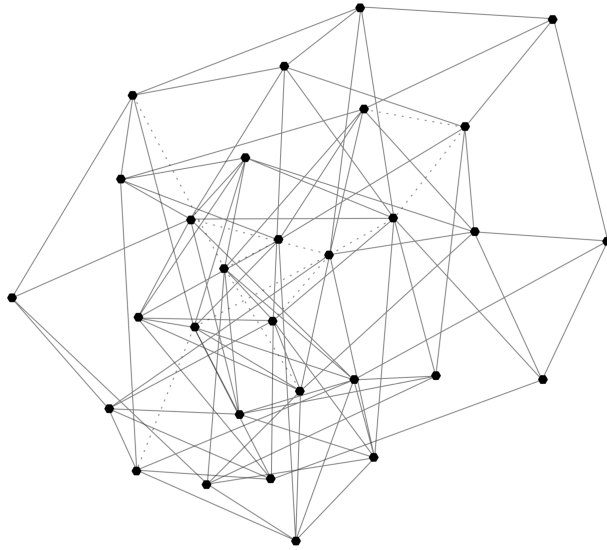


Figure 5: Graph with  $m = 102$  edges. The dotted edges have optimal  $\delta$  equal to 0.

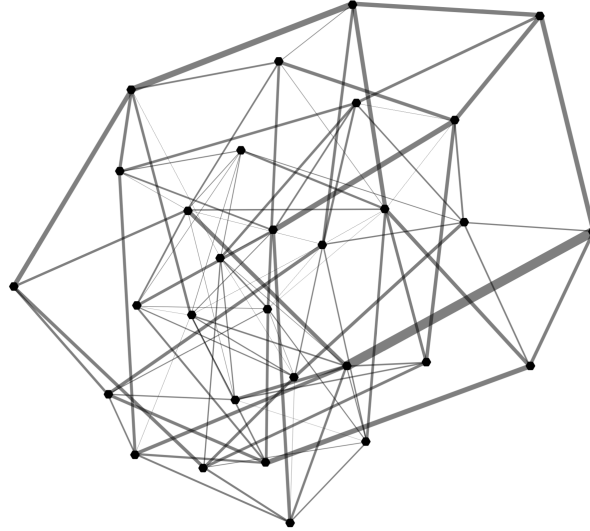


Figure 6: Graph with  $m = 102$  edges. Each edge width is proportional to the optimal  $\delta$ .

For larger graphs, there is still a non negligible proportion of values of  $\delta^*$  which are equal to 0. The graph depicted in Figure 7 contains 100 nodes and 1004 edges. There are 92 red edges, corresponding to null values of  $\delta^*$ . Figure 8 shows the same graph, where the edge widths are proportional to  $\delta^*$ . Note that interior well connected edges are associated to small values of  $\delta^*$  whereas exterior edges can correspond to maximal values of  $\delta^*$ . In particular, the lower left node in Figure 8 is less connected than the others (it has degree 5), but is incident to the edges with higher  $\delta^*$  values. This is due to the fact that this less connected node could be easily disconnected from the graph, and we would end up with an algebraic connectivity equal to its minimal value 0.

To end this section, we consider a very large graph, with  $n = 500$  nodes and  $m = 24978$  edges. The distribution of the values of the vector  $\delta^*$  are given in Figure 9b. Again, we see a large number of small values: in particular, 1042 of these values are equal to 0. Numerically solving the optimization problem (3.8) for this very large graph took approximatively 125 minutes of CPU time on a 2.4 GHz laptop from 2013.

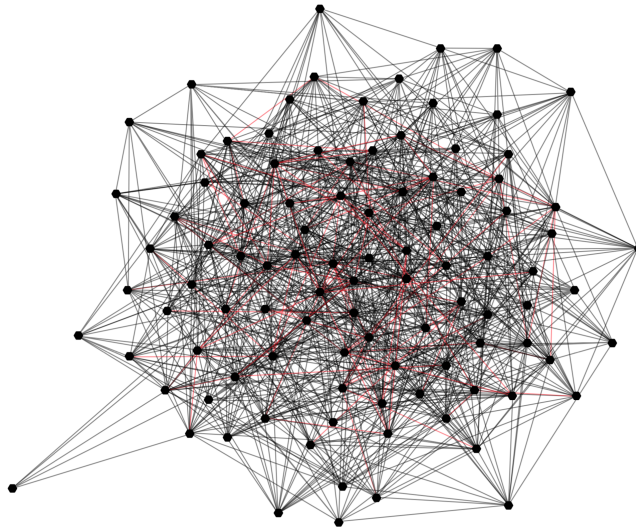


Figure 7: Graph with  $m = 1004$  edges. Red edges have null  $\delta^*$  values.

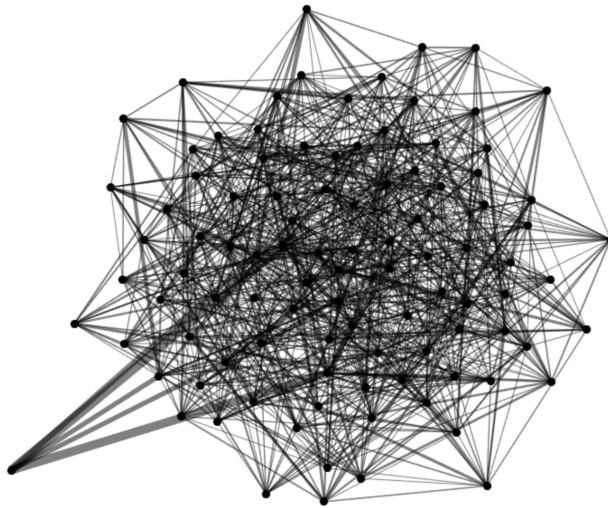
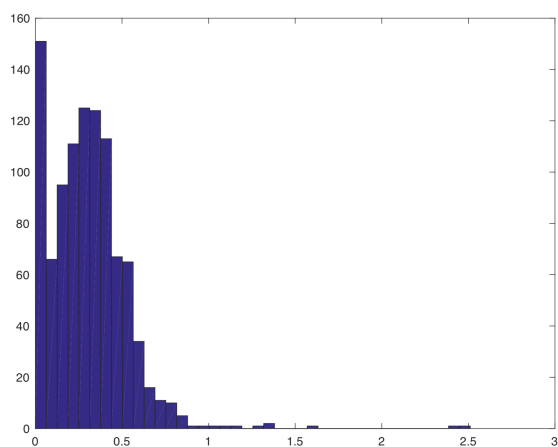


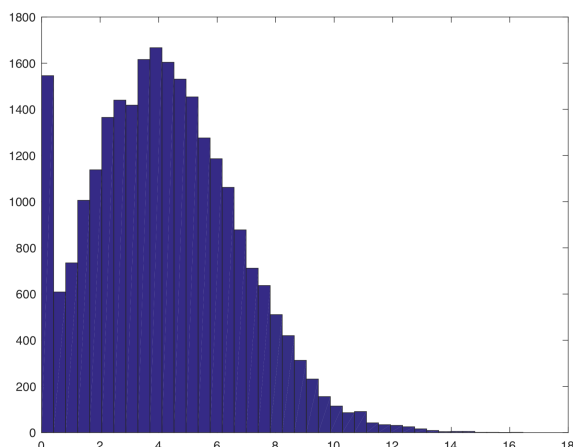
Figure 8: Graph with  $m = 1004$  edges. Each edge width is proportional to the optimal  $\delta$ .

## Conclusion

We considered in this paper the point of view of a company that needs to protect a given network against communication interruption, and has a fixed total budget for that task. We determined *where in the network* it is optimal to invest in prevention measures or in insurance, by optimally choosing the curing rates associated to each edge of the graph, given a SIS underlying epidemic model. As an optimization criteria, we retained the fact to maintain the graph as connected as possible, after an attack that became endemic, by maximizing the algebraic connectivity of the graph. We provided a solution, based on standard techniques from eigenvalues optimization theory. The numerical analysis shows that indeed, the optimal choice strongly depends on the network topology. In particular, it is optimal for that criteria to concentrate the protection efforts on a subset of the edges, that our method allows to clearly identify, even for very large graphs.



(a)  $\delta^*$  values, for the graph represented in Fig 8.



(b)  $\delta^*$  values, for a graph with 24978 edges.

Figure 9: Distribution of the values in  $\delta^*$

## References

- [1] BESSY-ROLAND, Y., BOUMEZOUED, A., AND HILLAIRET, C. Multivariate hawkes process for cyber insurance.
- [2] BÖHME, R. Towards insurable network architectures. *it-Information Technology Methoden und innovative Anwendungen der Informatik und Informationstechnik* 52, 5 (2010), 290–293.
- [3] BÖHME, R. *The economics of information security and privacy*. Springer, 2013.
- [4] BOYD, S. Convex optimization of graph laplacian eigenvalues. In *Proceedings of the International Congress of Mathematicians* (2006), vol. 3, pp. 1311–1319.
- [5] BOYD, S., DIACONIS, P., AND XIAO, L. Fastest mixing markov chain on a graph. *SIAM review* 46, 4 (2004), 667–689.
- [6] BROUWER, A. E., AND HAEMERS, W. H. *Spectra of graphs*. Springer Science & Business Media, 2011.
- [7] CHEEGER, J. A lower bound for the smallest eigenvalue of the laplacian, problems in analysis (papers dedicated to salomon bochner, 1969), 1970.
- [8] CHUNG, F., LU, L., AND VU, V. The spectra of random graphs with given expected degrees. *Internet Mathematics* 1, 3 (2004), 257–275.
- [9] CVETKOVIĆ, D., AND SIMIĆ, S. Graph spectra in computer science. *Linear Algebra and its Applications* 434, 6 (2011), 1545–1562.
- [10] CVETKOVIC, D. M., AND GUTMAN, I. *Selected topics on applications of graph spectra*. Matematički institut SANU Beograd, 2011.
- [11] CVX. Matlab software for disciplined convex programming. <http://cvxr.com/cvx/>. Accessed: 2020-07-24.

- [12] DEVRIENDT, K., AND VAN MIEGHEM, P. Unified mean-field framework for susceptible-infected-susceptible epidemics on networks, based on graph partitioning and the isoperimetric inequality. *Physical Review E* 96, 5 (2017), 052314.
- [13] DING, X., JIANG, T., ET AL. Spectral distributions of adjacency and laplacian matrices of random graphs. *The annals of applied probability* 20, 6 (2010), 2086–2117.
- [14] EIOPA. Understanding cyber insurance - a structured dialogue with insurance companies.
- [15] EVANS, L. C. *Partial differential equations*, vol. 19. American Mathematical Soc., 2010.
- [16] FAHRENWALDT, M. A., WEBER, S., AND WESKE, K. Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin: The Journal of the IAA* 48, 3 (2018), 1175–1218.
- [17] FARKAS, S., LOPEZ, O., AND THOMAS, M. Cyber claim analysis through generalized pareto regression trees with applications to insurance.
- [18] FIEDLER, M. Algebraic connectivity of graphs. *Czechoslovak mathematical journal* 23, 2 (1973), 298–305.
- [19] FIEDLER, M. Absolute algebraic connectivity of trees. *Linear and Multilinear Algebra* 26, 1-2 (1990), 85–106.
- [20] GORDON, L. A., AND LOEB, M. P. The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)* 5, 4 (2002), 438–457.
- [21] GORDON, L. A., LOEB, M. P., ZHOU, L., ET AL. Investing in cybersecurity: insights from the gordon-loeb model. *Journal of Information Security* 7, 02 (2016), 49.
- [22] HADJICHRYSANTHOU, C., AND SHARKEY, K. J. Epidemic control analysis: designing targeted intervention strategies against epidemics propagated on contact networks. *Journal of theoretical biology* 365 (2015), 84–95.
- [23] JEVTIĆ, P., AND LANCHIER, N. Dynamic structural percolation model of loss distribution for cyber risk of small and medium-sized enterprises for tree-based lan topology. *Insurance: Mathematics and Economics* 91 (2020), 209–223.
- [24] KELLEY, T., AND CAMP, L. J. Online promiscuity: Prophylactic patching and the spread of computer transmitted infections. In *The Economics of Information Security and Privacy*. Springer, 2013, pp. 157–180.
- [25] KNAUER, U. Algebraic graph theory: morphisms, monoids and matrices, volume 41 of, 2011.
- [26] LA, R. J. Role of network topology in cybersecurity. In *53rd IEEE Conference on Decision and Control* (2014), IEEE, pp. 5290–5295.
- [27] LELARGE, M., AND BOLOT, J. Economic incentives to increase security in the internet: The case for insurance. In *IEEE INFOCOM 2009* (2009), IEEE, pp. 1494–1502.
- [28] LEWIS, A. S. Convex analysis on the hermitian matrices. *SIAM Journal on Optimization* 6, 1 (1996), 164–177.
- [29] LIGGETT, T. M. *Interacting particle systems*, vol. 276. Springer Science & Business Media, 2012.

- [30] MAROTTA, A., MARTINELLI, F., NANNI, S., ORLANDO, A., AND YAUTSIUKHIN, A. Cyber-insurance survey. *Computer Science Review* 24 (2017), 35–61.
- [31] MIEGHEM, P., AND OMIK, J. In-homogeneous virus spread in networks. *arXiv preprint arXiv:1306.2588* (2014).
- [32] NEMIROVSKI, A. Interior point polynomial time methods in convex programming. *Lecture notes* (2004).
- [33] OVERTON, M. L. Large-scale optimization of eigenvalues. *SIAM Journal on Optimization* 2, 1 (1992), 88–120.
- [34] PASTOR-SATORRAS, R., CASTELLANO, C., VAN MIEGHEM, P., AND VESPIGNANI, A. Epidemic processes in complex networks. *Reviews of modern physics* 87, 3 (2015), 925.
- [35] PASTOR-SATORRAS, R., AND VESPIGNANI, A. Epidemic dynamics and endemic states in complex networks. *Physical Review E* 63, 6 (2001), 066117.
- [36] SUN, J., BOYD, S., XIAO, L., AND DIACONIS, P. The fastest mixing markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM review* 48, 4 (2006), 681–699.
- [37] VAN MIEGHEM, P. The n-intertwined sis epidemic network model. *Computing* 93, 2-4 (2011), 147–169.
- [38] VAN MIEGHEM, P., OMIK, J., AND KOOLIJ, R. Virus spread in networks. *IEEE/ACM Transactions On Networking* 17, 1 (2008), 1–14.
- [39] VANDENBERGHE, L., AND BOYD, S. Semidefinite programming. *SIAM review* 38, 1 (1996), 49–95.
- [40] XU, M., AND HUA, L. Cybersecurity insurance: Modeling and pricing. *North American Actuarial Journal* 23, 2 (2019), 220–249.