Optimization of network protection against virus spread

Eric Gourdin

Orange Labs, Issy-les-Moulineaux, France

Email: eric.gourdin@orange-ftgroup.com

Jasmina Omic

and Piet Van Mieghem

Faculty of Electrical Engineering

Mathematics and Computer Science

Delft University of Technology

P.O. Box 5031, 2600 GA Delft

Abstract—The effect of virus spreading in a telecommunication network, where a certain curing strategy is deployed, can be captured by epidemic models. In the N-intertwined model proposed and studied in [1], [2], the probability of each node to be infected depends on the curing and infection rate of its neighbors. In this paper, we consider the case where all infection rates are equal and different values of curing rates can be deployed within a given budget, in order to minimize the overall infection of the network. We investigate this difficult optimization together with a related problem where the curing budget must be minimized within a given level of network infection. Some properties of these problems are derived and several solution algorithms are proposed. These algorithms are compared on two real world network instances, while Erdös-Rényi graphs and some special graphs such as the cycle, the star, the wheel and the complete bipartite graph are also addressed.

I. Introduction

Security attacks and error propagation are one of the major issues when designing resilient networks. Computer and telecommunication networks are increasingly complex and interdependent. As a consequence, the security of each host depends heavily on the level of protection of its neighbors and, more generally, of all other hosts in the network. The choice, for every host, of an efficient protection strategy is hence a very complex problem.

Security breaches are very different in nature, to name a few: the spread of viruses and worms in the Internet, as well as social engineering compromises and direct exploitation of a host's vulnerability. An exposed host becomes a new source of infection, and data it contains about other hosts are new penetration points. We term a breach as a virus, and interdependency of systems is modeled as an infection process. In order to overcome such threats, hosts are equipped with protection and curing tools, such as antivirus software, with its signature quality and the speed of response to new virus strains. Each host is checked and secured with a certain frequency, however the protection depends on available resources.

The N-intertwined epidemic model, proposed and studied in [2], [1], is used as an approach to a general risk analysis framework. The system (computer or program) is modeled as a black box with interfaces to the other components and interactions between components are modeled by Markov theory. Protection of these boxes can be viewed as the "curing" of malfunctions.

While spreading processes in networks have been extensively studied (*Susceptible Infected Susceptible* (SIS) epidemics in [3], [4], [5], [6] and the influence of the topology on the spreading process in [7], [8], [9], [2], [1]), optimization of protection in such networks have hardly been considered. The optimization of network protection for the SIS model at the threshold was considered by Borgs et al. in [10]. The conclusion that the amount of protection depends on the node degree is a starting point of our research. We analyze the network above the threshold and determine that protection proportional to the node degree is not the optimal solution, but rather an upper bound. The optimization problem is a difficult fractional programming problem (which can be related to an NP-complete problem). Relying on heuristic approaches, we manage to solve these optimization problem for two real-size network instances. It appears that the set of optimal solutions, for various values of a parameter, are aligned, suggesting some hidden property in the model. Further, we analyzed some simple graphs, such as the cycle, the star, the wheel and the complete bipartite graph which, due to symmetry, reduce to a one variable function optimization. We determined the difference between the optimum and the upper bound for a complete bipartite graph with 1000 nodes.

II. N-intertwined Model

In this section, we briefly review the N-intertwined epidemic model [1]. In contrast to all published SIS-type of models, the N-intertwined model only makes 1 (mean-field) approximation in the exact SIS model and is applicable to all graphs. A network is represented by an undirected graph $G = (\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ ($|\mathcal{N}| = N$) is the set of nodes or vertices and $\mathcal{L}$ ($|\mathcal{L}| = L$) is the set of edges or links. The networks is specified by a symmetric adjacency matrix A: $a_{ij} = a_{ji} = 1$ iff there is an edge between nodes i and j (and otherwise, the value is 0). A node is not considered connected to itself, i.e., $a_{ii} = 0$. We denote $\mathcal{N}(i) = \{j \in \mathcal{N} : a_{ij} = 1\}$ the neighborhood of $i \in \mathcal{N}$. If $(x_i)_{i \in \mathcal{N}}$ is a vector of variables and S a subset of $\mathcal{N}$, we will sometimes denote the sum $\sum_{i \in S} x_i$ by $x(S)$.

The state of a node i is specified by a Bernoulli random variable $X_i \in \{0, 1\}$: $X_i = 0$ for a healthy node and $X_i = 1$

for an infected node. A node i at time t can be in one of the two states: *infected*, with probability $v_i(t) = \Pr[X_i(t) = 1]$ or *healthy*, with probability $1 - v_i(t)$. We assume that the curing process per node i is a Poisson process with rate δ_i, and that the infection rate per link is a Poisson process with rate β which is imminent for all nodes and thus constant in the network. For a node i, we can formulate the following differential equation

$$\frac{dv_i(t)}{dt} = \beta(1 - v_i(t)) \sum_{j=1}^{N} a_{ij} v_j(t) - \delta_i v_i(t) \qquad (1)$$

In words, the probability of a node being infected depends on the probability $(1 - v_i(t))$ that it is not infected, multiplied with the probability $a_{ij} v_j(t)$ that a neighbor j is infected and that it tries to infect the node i with the rate β. We denote the set of curing rates for a network by the vector $\Delta = [\delta_1 \ \delta_2 \ .. \ \delta_N]^T$.

The completely homogeneous setting, where all δ_i are equal was studied in [1]. On the contrary, both the curing rates δ_i and the infection rates β_i can be potentially different for each $i \in \mathcal{N}$. Detailed derivations in this heterogeneous setting are given in [2]. Here, our aim is to determine the influence of the protection distribution in a network, where $\beta_i = \beta$ for all i and where the curing rates δ_i may vary for each $i \in \mathcal{N}$.

In the steady-state, where $\frac{dv_i(t)}{dt} = 0$ for each node $i \in \mathcal{N}$ and $v_{i\infty} = \lim_{t \to \infty} v_i(t)$, we deduce from (1) that

$$v_{i\infty} = \frac{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty}}{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty} + \delta_i} \qquad (2)$$

For any given curing vector Δ, this system of equations has $2N$ solutions with one positive solution and one solution equal to 0 [1]. The positive solution gives the probability of nodes being in the infected state. The fraction $y(t)$ of infected nodes at any given time t can be calculated as a sum of probabilities that the nodes are infected, thus, in the steady-state $y_\infty = \frac{1}{N} \sum_{j=1}^{N} v_{j\infty}$.

In the pure homogeneous case (same δ for each node), the fraction of infected nodes as a function of the curing rate per node is depicted in Figure 1. The model shows a threshold value $\delta = \delta_c$. For values of δ larger than δ_c, the epidemic extinguishes and the number of infected nodes in the steady-state is 0. In the case of different curing rates per node, the epidemic threshold occurs [2] when

$$\beta_c = \frac{1}{\lambda_{\max}(A_\delta)} \qquad (3)$$

where $\lambda_{\max}(A_\delta)$ is the largest eigenvalue of the matrix $A_\delta = diag(\frac{1}{\delta_i})A$. There are many different matrices A_δ with the same largest eigenvalue, and as a consequence, there are many different curing rate 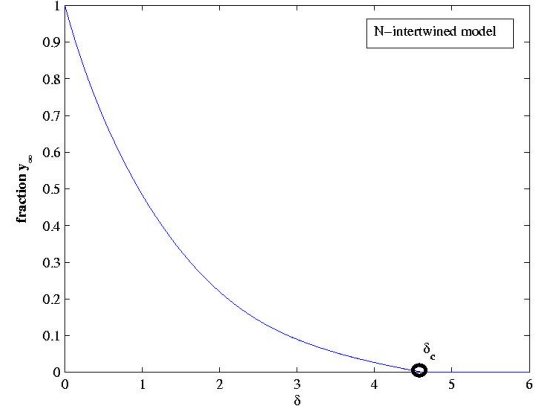vectors Δ that result in the same threshold β_c. If we have $\beta < \beta_c$, the infection dies out, and for $\beta >$

Fig. 1. Fraction y_∞ of infected nodes as a function of the curing rate δ (when all curing rates are equal).

β_c the epidemic persists with the average number of infected nodes equal to y_∞. If all the curing rates are the same $\delta_1 = \delta_2 = .. = \delta_N = \delta$, the threshold is given by $\frac{\beta_c}{\delta_c} = \frac{1}{\lambda_{\max}(A)}$. For $\beta_c = 1$, the critical curing rate [1] is

$$\delta_c = \lambda_{\max}(A) \qquad (4)$$

For example, the largest eigenvalue of a path graph is $\lambda_{\max}(A) \simeq 2 = \delta_c$, while that of a star topology is $\lambda_{\max}(A) = \sqrt{N-1} = \delta_c$. Although both trees have the same number of nodes and links, the level of protection in the homogeneous case required for a star is significantly higher than for a path topology.

III. PROBLEM STATEMENT

Given a limited curing "budget" $\sum_{i \in \mathcal{N}} \delta_i$, we are interested to find a curing strategy that minimizes the steady-state infection vector $V_\infty = [v_{1\infty} \ v_{2\infty} \ .. \ v_{N\infty}]^T$ of the network. Since we are considering steady-state conditions, we will, from now on, denote v_i instead of $v_{i\infty}$. We make the dependence of the steady-state probability on the curing rates of all the nodes, explicitly and write $v_i(\Delta) \in [0, 1]$, where $\Delta = [\delta_1 \ \delta_2 \ .. \ \delta_N]^T$. We formulate the following problem:

$$(P_1) : \min \quad f_1(\Delta) = \sum_{j \in \mathcal{N}} v_j(\Delta) \qquad (5)$$

$$\text{s.t.} \qquad \sum_{i \in \mathcal{N}} \delta_j = 2L\alpha\beta \qquad (6)$$

$$0 \leq \delta_j \leq \delta_c, \qquad j \in \mathcal{N}. \qquad (7)$$

where α is a parameter in $[0, 1]$ and the constant right-hand-side of the constraint (6) is proportional to the number L of links in the graph.

For every α, we denote the optimal value of problem P_1 by $f_1^*(\alpha)$. When $\alpha = 0$, (6) shows that all δ_j must be equal to 0, and (2) yields $v_j(\Delta) = 1$, for all $j \in \mathcal{N}$, so that $f_1^*(0) = N$.

In general, the probability of infection v_i cannot be found from (2) as an explicit function of Δ. Hence, given a vector Δ, the value $f_1(\Delta)$ cannot be obtained analytically, which complicates the resolution of problem P_1. Fortunately, given

a vector Δ, v_i can be approximated to any desired level of accuracy by a continued fraction expansion (see [1], [2]), or can also be obtained using a fixed point algorithm.

A. An Inverse Optimization Problem

Although v_i cannot be elegantly expressed as an explicit function of Δ, the reverse is not true and any δ_i can be written from (2) as a function of the infection probability vector $V = [v_1\ v_2\ ..\ v_N]$,

$$\delta_i(V) = \frac{1 - v_i}{v_i}\ \beta \sum_{j \in \mathcal{N}} a_{ij} v_j \equiv \frac{1 - v_i}{v_i} \times \beta v(\mathcal{N}(i)). \quad (8)$$

Problem P_1 can be rephrased into an inverse problem, where the sum of curing rates is to be minimized whereas the probability of infection is bounded:

$$(P_2): \min \quad f_2(V) = \sum_{j \in \mathcal{N}} \delta_j(V) \quad (9)$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{N}} v_j = N\alpha \quad (10)$$

$$0 \leq \delta_j(V) \leq \delta_c, \quad j \in \mathcal{N}. \quad (11)$$

where δ_c is set to $\frac{2L}{N}$ and $\alpha \in [0, 1]$. We will denote Π the hyperplane represented by the equality constraint (10) and Ω the set resulting from the bounding constraints (11).

For every α, the optimal value of problem P_2 is denoted by $f_2^*(\alpha)$.

When $\alpha = 1$, since every v_j is bounded by 1, the constraint implies that $v_j = 1$ for all $j \in \mathcal{N}$ and, hence, that $f_2^*(1) = 0$. When $\alpha = 0$, then $v_j = 0$ for all $j \in \mathcal{N}$, and it is shown in [11] that $\lim_{\alpha \to 0} f_2^*(0) = 2L\beta$.

The objective function of problem P_2 is now explicit, but it is still non-convex, and potentially possessing several local optima. More precisely, this problem belongs to the class of fractional programming problems [12]. As many other global optimization problems, fractional programming problems involving a sum of ratios are difficult problems, because many local minima can exist. Freund et al. [13] have shown that such problems can be reformulated as combinatorial NP - complete problems. Several authors [14], [15] proposed algorithms for solving linear sum of ratios functional program.

One approach towards local minima are the so-called KKT (Karush-Kuhn-Tucker) conditions. These first order necessary optimality conditions are usually expressed by the Lagrangean $L(V; \mu_2)$ of the problem

$$L(V; \mu_2) = \sum_{j \in \mathcal{N}} \delta_j(V) + \mu_2 \left(N\alpha - \sum_{j \in \mathcal{N}} v_j \right). \quad (12)$$

where μ_2 is the Lagrangean multiplier associated with the single equality constraint of the problem. A couple of primal-dual vectors (V^*, μ_2^*) is a solution of problem P_2 if it satisfies the first order conditions. From these conditions, we deduce for every $i \in \mathcal{N}$ and $\beta = 1$ that

$$\mu_2^* = \frac{\partial f_2}{\partial v_i}(V) = \beta \left(-d_i + \sum_{j \in \mathcal{N}(i)} \left(\frac{1}{v_j^*} - \frac{v_j^*}{v_i^{*2}} \right) \right), \quad (13)$$

As a consequence, at every local optima V^* of problem P_2 (and hence also for the global minimum), we have that:

$$\frac{\partial f_2}{\partial v_i}(V^*) = \frac{\partial f_2}{\partial v_j}(V^*), \ \forall i, j \in \mathcal{N}. \quad (14)$$

Another consequence of (13) is

Lemma 1: For every pair (V^*, μ_2^*) satisfying the first order necessary conditions, it holds that $\mu_2^* < 0$.

Proof: Consider an optimal solution V^* that satisfies the first order necessary conditions and let $v_{min}^* = \min_{i \in \mathcal{N}} v_i^* > 0$ with corresponding index i_{min}. Since (13) is valid for any $i \in \mathcal{N}$, it is valid for i_{min} and

$$\frac{v_j^*}{v_{min}^{*2}} \geq \frac{v_j^*}{v_j^{*2}} = \frac{1}{v_j^*}$$

immediately shows that $\mu_2^* \leq -d_{i_{min}} < 0$. $\blacksquare$
Finally, we conclude with

Theorem 1: The optimal value of problem P_2 is a strictly decreasing function of α

$$\frac{\partial f_2^*}{\partial \alpha}(\alpha) < 0. \quad (15)$$

Proof: According to the *envelope theorem* (which refers to perturbation functions in optimization [16]) and given an optimal solution $V^*(\alpha)$, the variations of the global optimum value $f_2^*(\alpha)$ with respect to α are

$$\frac{\partial f_2^*}{\partial \alpha}(\alpha) = \frac{\partial L}{\partial \alpha}(V^*(\alpha); \mu_2^*) = N\mu_2^*. \quad (16)$$

By Lemma 1, $\mu_2^* < 0$ and Theorem 1 follows immediately. $\blacksquare$

B. Bounds on the optimal value $f_2^*(\alpha)$

As for any minimization problem, any feasible solution of P_2 provides an upper-bound on $f_2^*(\alpha)$. Among all possible feasible solutions, the local minimizers are those that provide the better bounds. Consider, for instance, the vector $V_1(\alpha) = \alpha \mathbb{1}$ where $\mathbb{1}$ is the N-dimensional vector $(1, \ldots, 1)$.

Theorem 2: For $\alpha > 0$, the vector $V_1(\alpha)$ is a feasible solution of problem P_2 and $f_2^*(\alpha) \leq 2L(1 - \alpha)\beta$. Moreover, if the graph is regular (all degrees are equal), then $V_1(\alpha)$ is a local minimum.

Proof: First, this solution is indeed feasible, because it satisfies the equality constraint. Moreover, for every i, we have $\delta_i = (1 - \alpha)\beta d_i \leq L$. Hence, it also satisfies the second set of inequality constraints such that $f_2(V_1(\alpha)) = \sum_{i \in \mathcal{N}} (1 - \alpha)\beta d_i = 2L(1 - \alpha)\beta$ is the value of the upper-bound on the optimal value of problem P_2.

The gradient of f_2 evaluated at $V = V_1(\alpha)$ (with $\alpha > 0$) is:

$$\frac{\partial f_2}{\partial v_i}(V_1(\alpha)) = -\beta d_i \leq 0. \quad (17)$$

The local direction for improving f_2 (since we are minimizing) around $V_1(\alpha)$ is hence the vector $(d_i)_{i \in N}$, which, for a regular graph with degree r, reduces to $r\mathbb{1}$. Since this vector is orthogonal to the hyperplane Π, $V_1(\alpha)$ is indeed a local

minimum. Alternatively, we can immediately check that $V_1(\alpha)$ satisfies the first order necessay conditions with $\mu_2^* = -\beta d$ and hence the result holds. ∎

As soon as the graph is not regular, there is a direction of improvement for f_2 along the hyperplane Π and $V_1(\alpha)$ is then no longer a local minimum.

However, in the particular case where $\alpha = 1$, $V_1(\alpha)$ is again a local minimum. This is basically due to the fact that there is no direction u pointed within Ω such that the scalar product $< u, \nabla f_2(\mathbb{1}) >$ is positive. In other words, all improving direction lead outside Ω. Note that at this particular solution of $P_2(\alpha = 1)$ we also have a locally convex function. Indeed, the Hessian matrix of f_2 can then be written as

$$H_f(V) = 2V^T A \times \mathrm{diag}(V^{-3}) - \mathrm{diag}(V^{-2}) \times A - A \times \mathrm{diag}(V^{-2}). \tag{18}$$

The Hessian matrix evaluated at the unit vector $\mathbb{1}$ reduces to:

$$H_f(\mathbb{1}) = 2D - 2A = 2Q, \tag{19}$$

where D is the diagonal matrix where each entry d_{jj} is equal to $d(j)$ the degree of node $j \in V$ and Q is the Laplacian matrix of a graph G. Since the Laplacian matrix is positive semi-definite, it follows that f_2 is indeed convex at $V = \mathbb{1}$.

When the graph is not regular, starting from $V_1(\alpha)$, a direction of improvement of function f_2 can be computed as the orthogonal projection of the gradient $\nabla f(V_1(\alpha))$ onto the hyperplane Π. The j-th component of this projected vector is $-\beta d_j + \frac{2L}{N}$. As a direct consequence, the function f will decrease the most along directions where the degree is maximal.

Starting from $V_1(\alpha)$ or from any other vector in $\Omega \cap \Pi$, a gradient or projected-gradient method can be used to iteratively improve the current value of the objective function and finally converge to some local minimum of Problem P_2.

IV. Relationship between problems P_1 and P_2

Both problems P_1 and P_2 involve essentially two sums: denote $S_V = \sum_{j \in \mathcal{N}} v_j$ and $S_\Delta = \sum_{j \in \mathcal{N}} \delta_j$. In problem P_1, the sum S_V is minimized while the sum S_Δ is constrained to be equal to the constant value $2L\alpha_1$. In problem P_2, the sum S_Δ is minimized while the sum S_V is constrained to be equal to the constant value $N\alpha_2$. Hence, it is legitimate to investigate the set of vectors $(S_V, S_\Delta) \in \mathbb{R}^2$ such that both vectors V and Δ are linked by the system (2). More precisely, we are interested in the vectors that are optimal solutions of either P_1 or P_2. We have:

$$P_1 : f_1^*(\alpha_1) = \min_{\Omega_\Delta} \{ S_V(\alpha_1, \Delta) : S_\Delta = 2L\alpha_1 \},$$

and

$$P_2 : f_2^*(\alpha_2) = \min_{\Omega_V} \{ S_\Delta(\alpha_2, V) : S_V = N\alpha_2 \}.$$

We have some interesting results linking both problems:

Lemma 2: Any feasible solution of one of the two problems P_1 and P_2 leads to a feasible solution for the other problem.

Proof: : Consider first a solution Δ_1, feasible for problem $P_1(\alpha_1)$. We know that there exists a solution V_2 of system (2) such that $0 \leq v_{2i} \leq 1$ for all i. It follows that $0 \leq \sum_{i \in \mathcal{N}} v_{2i} \leq N$ and hence, there exists $\alpha_2 \in [0, 1]$ such that this sum equals $N\alpha_2$. Given a feasible solution V_2 of P_2, the solution Δ_1 such that $\delta_{1i} = \beta \frac{1 - v_{1i}}{v_{1i}} v_1(\mathcal{N}(i))$ for all i is then feasible for P_1 with $\alpha_1 = \frac{1}{2L} \sum_{i \in \mathcal{N}} \delta_{1i}$. ∎

As a direct consequence of this lemma, an optimal solution of one problem yields a feasible solution of the other problem. We can even demonstrate a stronger result:

Theorem 3: Any optimal solution of one of the two problems P_1 and P_2 leads to an optimal solution of the other problem.

Proof: : We will only prove that an optimal solution of P_2 can be deduced from an optimal solution of P_1 (the other proof is analogous). Consider a given $\alpha_1 \in [0, 1]$, an optimal solution $\Delta^*(\alpha_1)$ of problem $P_1(\alpha_1)$ and the corresponding optimal value $f_1^*(\alpha_1)$. We then consider the problem P_2 with parameter:

$$\alpha_2 = \frac{1}{N} f_1^*(\alpha_1).$$

We know that the solution $v_j = v_j(\Delta^*(\alpha_1))$ implicitly obtained from system (2), is a feasible solution of problem $P_2(\alpha_2)$ with a corresponding value $f_2(\alpha_2) = 2L\alpha_2\beta$ (see Figure 2). If this solution is not optimal then we have $f_2^*(\alpha_2) < f_2(\alpha_2) = 2L\alpha_2\beta$. Denoting $\tilde{\alpha}_1 = \frac{1}{2L\beta} f_1^*(\alpha_1)$, we have $\tilde{\alpha}_1 < \alpha_1$ and both values of α yield the same optimal value. Since f_2^* is a strictly decreasing function of α_2, there would exist $\tilde{\alpha}_2 < \alpha_2$ such that $f_2(\tilde{\alpha}_2) = f_2(\alpha_2)$ and hence a $f_1(\alpha_1) < f_1^*(\alpha_1)$ which is in contradiction with the fact that $f_1^*(\alpha_1)$ is the optimal value. ∎

As a consequence, assume that, for two distinct values $\alpha_2^a < \alpha_2^b$, one knows the optimal values of the corresponding problems P_2, namely $f_2^*(\alpha_2^a) > f_2^*(\alpha_2^b)$. Letting

$$\alpha_1^\ell = \frac{1}{2L\beta} f_2^*(\alpha_2^\ell), \text{ for } \ell = a, b,$$

then one has the result that all the optimal values of the problems $P_1(\alpha_1)$ with $\alpha_1 \in [\alpha_1^b, \alpha_1^a]$ lie in the interval $[\alpha_2^a, \alpha_2^b]$. This useful property can be used to design an approximation algorithm to solve problem P_1, assuming some algorithm is available to solve problem P_2.

V. Some particular graphs

When dealing with specially structured graphs, we can use the fact that symmetrical elements or parts of the graph behave in a similar fashion. For instance in the steady-state, all nodes in a cycle C_N with N nodes will incur the same infection and have the same protection rate. As a consequence, the problem size reduces from N to 1 and can be treated analytically. In this Section, we investigate some special cases of graphs for which the problem can be solved analytically.
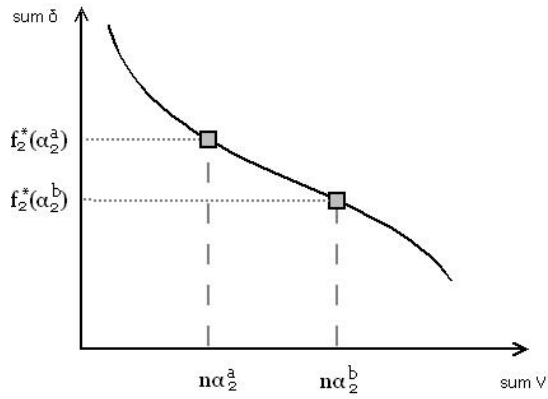
Fig. 2.   Illustration of the decreasing function $f_2^*(\alpha)$.

A. Cycles

In a cycle C_N each node $i \in \{1, \ldots, N\}$ is linked to its two neighbors $(i-1)$ and $(i+1) \mod N$ (see Figure 3(a)). By symmetry, all infections rates are equal: $v_i = v(cte)$. In problem P_2, the constraint (10) translates into $v = \alpha$. As a consequence, the sum of curing rates is a linear function of α: $f_2^{cycle}(V) = 2N(1-\alpha)$.
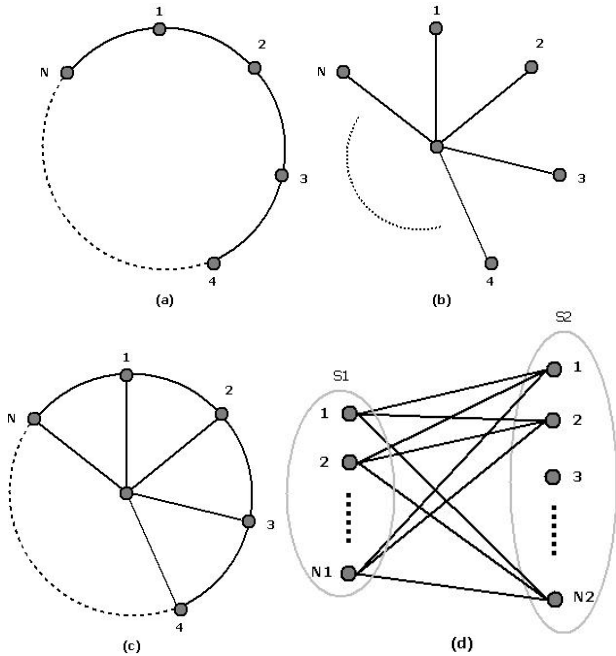

Fig. 3.   Some particular graphs: (a) cycle, (b) star, (c) wheel, (d) complete bipartite graph.

B. Stars

A star $K_{1,N}$ has a central node v_0 connected to all other vertices $v_i, i \in \{1, \ldots, N\}$ (see Figure 3(b)). Except the central node, all other vertices play a symmetrical role. As a consequence, we have: $v_i = v(cte), \{1, \ldots, N\}$. The curing rates are then given by:

$$\delta_0 = \frac{1 - v_0}{v_0} Nv, \qquad (20)$$

$$\delta_i = \delta(cte) = \frac{1 - v}{v} v_0, \ i = 1, \ldots, N. \qquad (21)$$

It follows that:

$$f_2^{star}(V) = \delta_0 + N\delta = N\left(\frac{v}{v_0} + \frac{v_0}{v} - v - v_0\right). \qquad (22)$$

Using the fact that $v_0 + Nv = (N+1)\alpha$, we can then derive an expression of $f_2^{star}(V)$ as a function of v (and α).

C. Wheels

A wheel centered around a node v_0 and with N additional vertices is the edge-wise union of a cycle and a star (see Figure 3(c)). Using the same symmetry arguments, we can easily compute the curing rates:

$$\delta_0 = \frac{1 - v_0}{v_0} Nv, \qquad (23)$$

$$\delta_i = \delta(cte) = \frac{1 - v}{v}(v_0 + 2v), \ i = 1, \ldots, N. \qquad (24)$$

It is interesting to note that, the total curing level $f_2^{wheel}(V)$ can be expressed as the sum of the two previous functions:

$$f_2^{wheel}(V) = f_2^{cycle}(V) + f_2^{star}(V). \qquad (25)$$

D. Complete bipartite graphs

A complete bipartite graph K_{N_1, N_2} consists of two disjoint sets S_1 and S_2 ($|S_1| = N_1$ and $|S_2| = N_2$), such that all nodes in S_1 are connected to all nodes in S_2 (see Figure 3(d)).

Due to the symmetry, we can assume that nodes from the same set will have the same protection rate. Thus, the set of N equations in the steady state, reduces to only two explicit equations, derived in [1],

$$v_1 = \frac{N_1 N_2 - \delta_1 \delta_2}{N_1(N_2 + \delta_1)}; \ v_2 = \frac{N_1 N_2 - \delta_1 \delta_2}{N_2(N_1 + \delta_2)}. \qquad (26)$$

For $N_1 = 1$ and $N_2 = N$, the relations (26) simplify to those of the star in Sec. V-B.

From (26), we find that

$$
\begin{aligned}
f_1(\Delta) &= \sum_{j=1}^{N} v_j = N_1 v_1 + N_2 v_2 \\
&= (N_1 N_2 - \delta_1 \delta_2)\left(\frac{1}{N_2 + \delta_1} + \frac{1}{N_1 + \delta_2}\right)
\end{aligned}
$$

Using $f_2(V) = \sum_{j=1}^{N} \delta_j = N_1 \delta_1 + N_2 \delta_2$ yields

$$
\begin{aligned}
f_1(\Delta) &= \left(N_1 N_2 + \frac{N_1}{N_2}\delta_1^2 - \frac{1}{N_2}\delta_1 f_2(V)\right) \\
&\times \left(\frac{1}{N_2 + \delta_1} + \frac{1}{N_1 + \frac{f_2(V)}{N_2} - \frac{N_1}{N_2}\delta_1}\right)
\end{aligned}
$$

which shows that $f_1(\Delta)$ is a function of δ_1, given that $f_2(V)$ is constant (see (6)). Moreover, this expression illustrates that $f_1(\Delta)$ is approximately linear in $f_2(V)$, provided the last

fraction is negligibly small. In any case, with $0 \leq \delta_j \leq \delta_c$, we have the bounds

$$\frac{1}{N_1} \leq \frac{1}{N_1 + \frac{f_2(V)}{N_2} - \frac{N_1}{N_2}\delta_1} \leq \frac{1}{N_1 + \delta_c}$$

which, indeed, confirms that the sensitivity on $f_2(V)$ is small for any δ_1, thus also for the optimum δ_1^* (which is the purpose of our attempts here).
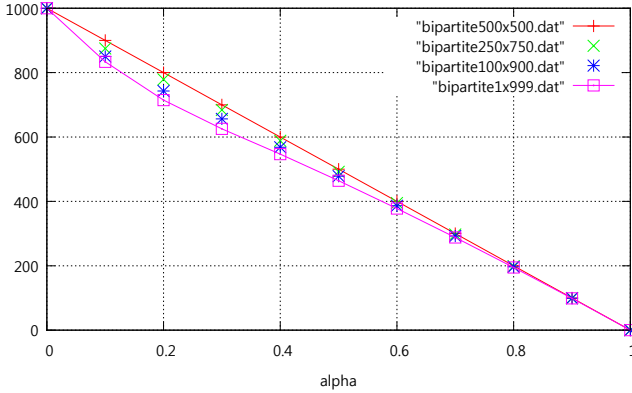

Fig. 4. Each curve represents the optimal value f_1^* as a function of α for several values of the couple (N_1, N_2).

In figure 4, the optimal solution curves for problem P_1 are represented for several values of the couple (N_1, N_2). The total number of nodes $N_1 + N_2 = 1000$ is constant. The curve for the symmetrical case $N_1 = N_2 = 500$ is the line corresponding to $V = \alpha \mathbb{1}$. This line is also an upper-bound for the other cases and it corresponds to the approximate solution where the protection level of each node is taken proportional to its degree. A slight distortion of the curve appears when $N_1 < N_2$ with a maximum for $N_1 = 1$ and $N_2 = 199$, corresponding to a star graph. In this extreme case, the approximate solution based on node degrees can have a gap up to 12% from the optimal solution.

Finally, we end this section by providing a general argument that supports the observed almost perfect linearity in numerical evaluations in Sec. VI. Using the bound in [2]

$$1 - \frac{1}{\min_{1 \leq k \leq N} \frac{\gamma_k}{\delta_k}} \leq y_\infty \leq 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1 + \frac{\gamma_i}{\delta_i}}$$

where $\gamma_i = \sum_{j \in \text{ neighbor}(i)} \beta_j$ is total infection rate of node i, incurred by all neighbors towards node i, we find, *for any graph*, the linear upper bound

$$f_1(\Delta) \leq N - \frac{1}{\max_{1 \leq k \leq N} \frac{\gamma_k}{\delta_k}} f_2(V) \qquad (27)$$

Unfortunately, it is difficult in general to show that minimizing $f_1(\Delta)$ with respect to the set of $\{\delta_i\}_{1 \leq i < N}$ subject to $f_2(V) = 2L\alpha\beta$ almost results in equality in (27).

VI. NUMERICAL RESULTS

In this section, we give the results obtained, using several optimization algorithms, on two network instances: the ARPAnet topology with 20 nodes and 31 links and the Cost266 instance from the SNDlib [17] made of 37 nodes and 57 links (see Figure 5). These two networks are typical US and European backbones.
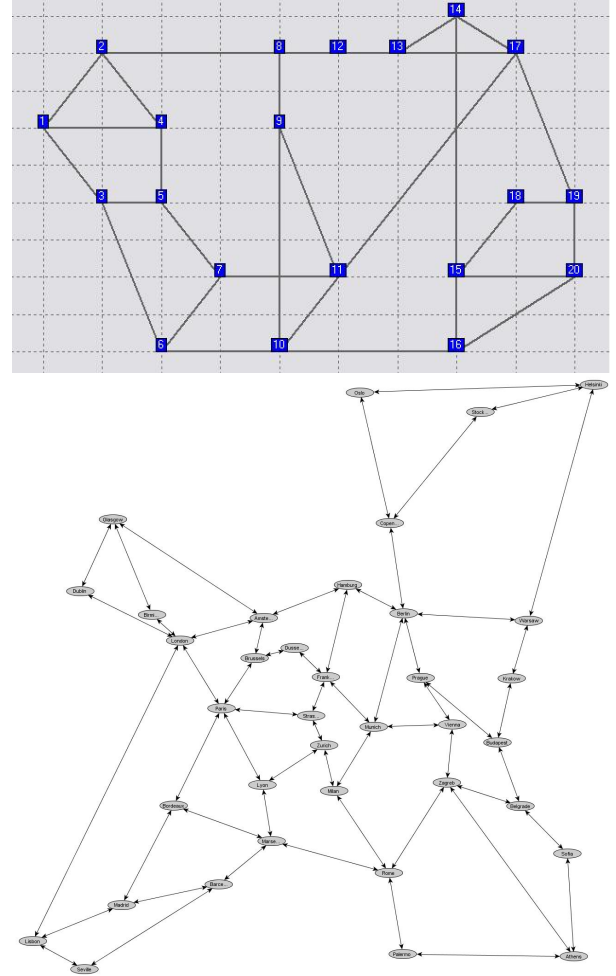

Fig. 5. The ARPAnet and Cost266 instances

The results are depicted in Figure 6. For each instance, problem P_2 was solved for several values of α_2 using a branch-and-bound algorithm. This algorithm is designed to provide exact optimal solutions, but only for very small instances. We have hence limited the number of iterations resulting to a computing time of one to two minutes (we observed that giving more time to the algorithm would not improve significantly the solution). Since the problem of interest is P_1, we have used these results on P_2 to try to infer the right α_2 that would provide an optimal solution of P_1 with a given value of α_1 (according to theorem 3). These dots named "Pb2 inverse" on the figures are hence very closed to the set of dots "Pb2" obtained by optimizing problem P_2. Hence, these dots correspond to pairs $(\sum v_j, \sum \delta_j)$ where

one sum is optimized and the other held constant, but are also valid optimal solutions for the other problem. It clearly appears that this approach outperforms significantly a simple random search"Pb1 random" that was used to solve problem P_1 directly on the ARPAnet instance. Finally, the solution where all protection levels are taken proportional to the node degrees was also investigated ("Pb1 degree"), using a fixed point algorithm to deduce the corresponding infection levels.
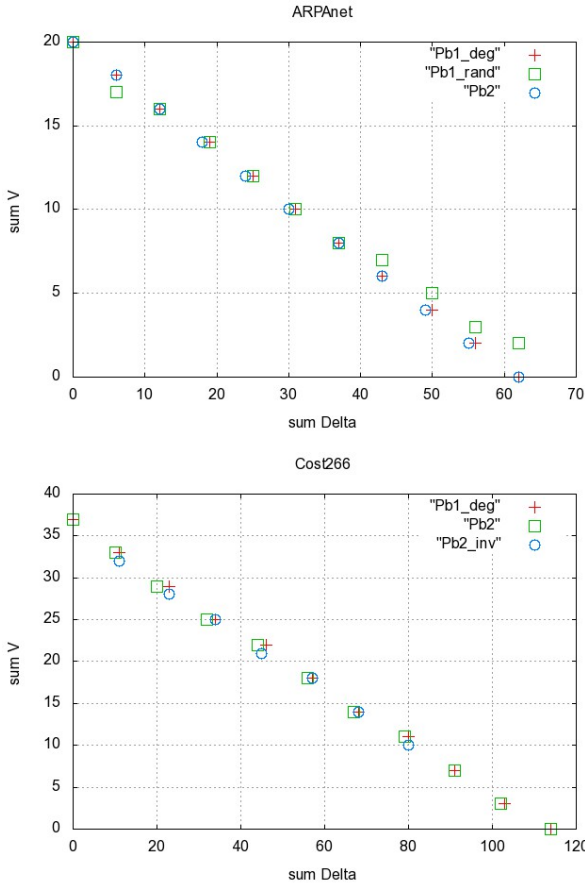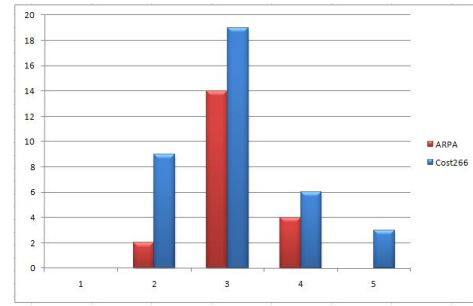

Fig. 7.   Distribution of degrees for the ARPAnet and Cost266 graphs

The results shown in figure 4 illustrate the fact that $f_1^*(\Delta)$ (and as a direct consequence, $f_2^*(V)$ are not always affine functions of α. However, the strategy assigning the protection proportional to the degree seems to be a very good approximation, except in highly non-symmetrical cases, such as, for instance the star network ($N_1 = 1, N_2 = 999$). In this case, there can be a gap up to 12 % (for $\alpha = 0.2$). However, the sensitivity to the degree distribution seems to be relatively weak, since the gap reduce to 8 % for the bipartite graph with $N_1 = 100$ and $N_2 = 900$. With more "standard" distribution degrees (such as the ones of ARPAnet and Cost266 shown in figure 7), the alignment of dots is excellent.

Additional numerical experiments have been performed on randomly generated graphs. Two series of 10 Erdös-Rényi graphs have been generated, respectively with 100 and 200 nodes. The probability of existence of an edge was set to $p = 2p_c = \log N/N$ (p_c is the critical disconnectivity threshold), thus having a very strong guaranty that the generated graphs are connected. The problems P_2 have been considered for various values of α. The resulting curves are depicted in Figures 8 and 9.

Fig. 6.   Solutions obtained on the ARPAnet and Cost266 instances

First note that the dots corresponding to the random sampling strategy are sometimes quite far from the "optimal" dots, showing that this approach might be relatively inefficient, depending on the values of α. The main observation is that all these dots seem to be perfectly aligned in both instances. This suggest that the set of optimal values has some strong property that we were not yet able to prove in the general case. A second striking observation is that the solutions proportional to the node degrees are very close to the optimal ones (or at least, to those obtained by a more involved resolution approach). This observation seems to generalize, in a certain sens, the result proved in the case of regular graphs, that protection proportional to the node degree is generally a good answer to the infection problem.

These observations are somewhat tempered by results obtained analytically in the case of complete bipartite graphs.
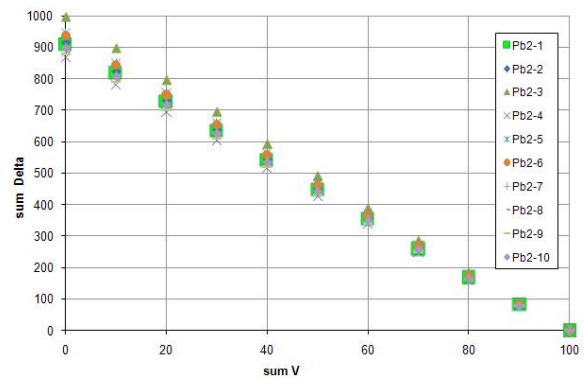

Fig. 8.   Solutions obtained for Problem P_2 on 10 Erdös-Rényi instances (with 100 nodes)

Again, the almost perfect alignment of the dots confirms that, in a vast majority of cases, the optimal functions $f_1^*(\Delta)$ and $f_2^*(V)$ behave almost as affine functions of α. The precise sensitivity of this results to the distribution of degrees in the graph should be investigated more closely but analytical
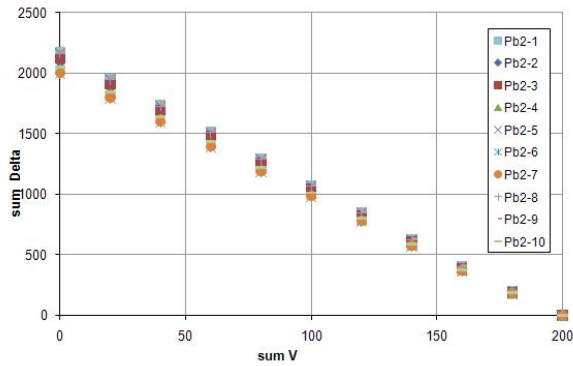
Fig. 9.　Solutions obtained for Problem P_2 on 10 Erdös-Rényi instances (with 100 nodes)

conclusions might be quite difficult to derive since these functions are not known analytically themselves.

VII. CONCLUSIONS

We have proposed and analyzed two closely related optimization problems that naturally arise in the context of virus attacks spreading over a telecommunication network where a curing strategy is deployed.We have shown how these two problems are linked and the fact that they share the same set of optimal solutions. Numerical experiments where optimal solutions are perfectly aligned, seem even to indicate that the relationship between the problems is even stronger. The natural approach consisting in setting a curing strategy proportional to the degree of each node seems to be quite effective, although not always optimal.

ACKNOWLEDGMENT

REFERENCES

[1] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 1–14, 2009.
[2] P. Van Mieghem and J. Omic, "In-homogeneous virus spread in networks," Technical report 20080801, TUDelft, Tech. Rep., 2008.
[3] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, p. 378, 2000.
[4] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," *Security and Privacy, IEEE Symposium on*, vol. 0, p. 343, 1991.
[5] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys Rev Lett*, vol. 86, no. 14, pp. 3200–3203, Jan. 2001.
[6] Y. Wang, D. Chakrabarti, C. Faloutsos, C. Wang, and C. Wang, "Epidemic spreading in real networks: An eigenvalue viewpoint," in *In SRDS*, 2003, pp. 25–34.
[7] G. Canright and K. Engoe-Monsen, "Spreading on networks: a topographic view," in *Proceedings of the European Conference on Complex Systems*, nov 2005.
[8] A. J. Ganesh, L. Massoulié, and D. F. Towsley, "The effect of network topology on the spread of epidemics," in *INFOCOM*. IEEE, 2005, pp. 1455–1466.
[9] M. Garetto and W. Gong, "Modeling malware spreading dynamics," in *In Proceedings of IEEE INFOCOM*, 2003, pp. 1869–1879.
[10] C. Borgs, J. Chayes, A. Ganesh, and A. Saberi, "How to distribute antidote to control epidemics," *Random Structure Algorithms*, vol. 37, no. 2, pp. 204–222, 2010.
[11] J. Omic, A. Orda, and P. Van Mieghem, "Protecting against network infections: A game theoretic perspective," in *INFOCOM*. IEEE, 2009.
[12] H. Frenk and S. Schaible, "Fractional programming," in *Encyclopedia of Optimization*, 2009, pp. 1080–1091.
[13] R. W. Freund and F. Jarre, "Solving the sum-of-ratios problem by an interior-point method," *Journal of Global Optimization*, vol. 19, no. 1, pp. 83–102, 2001.
[14] J. E. Falk and S. W. Palocsay, "Image space analysis of generalized fractional programs," *Journal of Global Optimization*, vol. 4, no. 1, pp. 63–88, 1994.
[15] H. Konno and K. Fukaishi, "A branch and bound algorithm for solving low rank linear multiplicative and fractional programming problems," *Journal of Global Optimization*, vol. 18, no. 3, pp. 283–299, 2000.
[16] J. Gauvin and R. Janin, "Directional derivative of the value function in parametric optimization," *Annals of Operations Research*, vol. 27, 1990.
[17] "Sndlib: Survivable network design library." [Online]. Available: SNDlib.zip.de