



Complexity of the multilevel critical node problem

Adel Nabli^a, Margarida Carvalho^{a,*}, Pierre Hosteins^b

^a CIRRELT and Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, QC, Canada

^b COSYS-ESTAS, Université Gustave Eiffel, Villeneuve d'Ascq, France



ARTICLE INFO

Article history:

Received 23 November 2020

Received in revised form 21 January 2022

Accepted 9 February 2022

Available online 23 February 2022

Keywords:

Trilevel programming

Defender-attacker-defender

Critical node problem

Polynomial hierarchy

ABSTRACT

In this work, we analyze a sequential game played in a graph called the Multilevel Critical Node problem (MCN). A defender and an attacker are the players of this game. The defender starts by preventively interdicting vertices (vaccination) from being attacked. Then, the attacker infects a subset of non-vaccinated vertices and, finally, the defender reacts with a protection strategy. We provide the first computational complexity results associated with MCN and its subgames. Moreover, by considering unitary, weighted, undirected, and directed graphs, we clarify how the theoretical tractability of those problems vary. Our findings contribute with new NP-complete, Σ_2^P -complete and Σ_3^P -complete problems. Furthermore, for the last level of the game, the protection stage, we build polynomial time algorithms for certain graph classes.

© 2022 Elsevier Inc. All rights reserved.

1. Introduction

Multilevel Critical Node. Graphs are powerful mathematical structures that enable us to model real-world networks. The problem of breaking the connectivity of a graph has been extensively studied in combinatorial optimization since it can serve to measure the robustness of a network to disruptions. In this work, we will focus on the Multilevel Critical Node problem (MCN) [1]. Let $G = (V, A)$ be graph with a set V of vertices and a set A of arcs. In MCN there are two players, designated by defender and attacker, whose individual strategies are given by a selection of subsets of V . The game goes as follows: first, the defender selects a subset of vertices $D \subseteq V$ to *vaccinate* subject to a budget limit Ω and a $\text{cost}\{\hat{c}_v\}_{v \in V}$; second, the attacker observes the vaccination strategy, and selects a subset of vertices $I \subseteq V \setminus D$ to (directly) *infect* subject to a budget limit Φ and a $\text{cost}\{h_v\}_{v \in V}$; and third, the defender observes the infection strategy, and selects a subset of vertices $P \subseteq V \setminus I$ to *protect* subject to a budget limit Λ and a $\text{cost}\{c_v\}_{v \in V}$. A directly or indirectly infected vertex v propagates the infection to a vertex u , if $(v, u) \in A$ and u is neither a vaccinated or a protected vertex. The goal of the defender is to maximize the benefit b_v of saved vertices (i.e., not infected), while the attacker aims to minimize it.¹ We assume that all parameters of the problem are non-negative integers. The game description can be succinctly given by the following mixed integer trilevel program:

* Corresponding author.

E-mail addresses: adel.nabli@umontreal.ca (A. Nabli), carvalho@iro.umontreal.ca (M. Carvalho), pierre.hosteins@ifsttar.fr (P. Hosteins).

¹ We note that the terms vaccination, infection and protection in MCN are used in a broad sense and are not intended to accurately reflect epidemiological concepts.

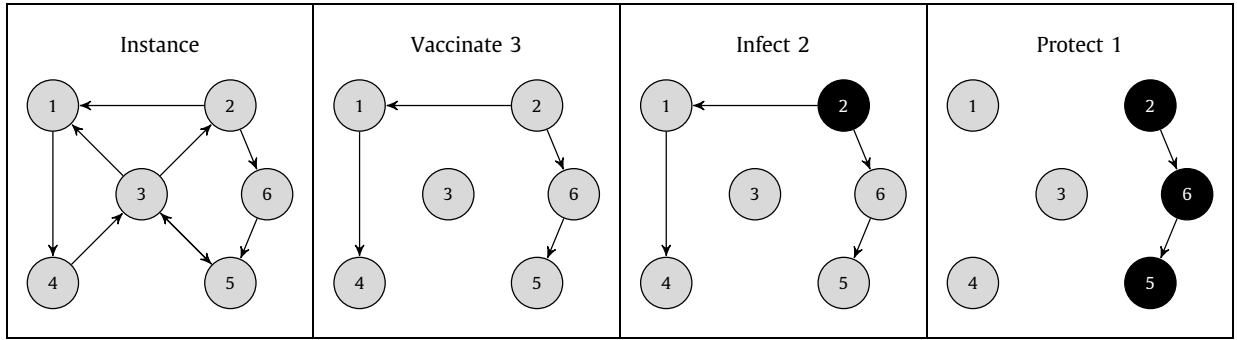


Fig. 1. Example of an MCN game with unitary costs and benefits, and budgets $\Omega = \Phi = \Lambda = 1$. We removed the vaccinated and protected vertices as an infection cannot pass through them (see Property 3.1). Vertices {1, 3, 4} are saved and {2, 5, 6} are infected.

$$\begin{aligned}
 (MCN) \quad & \max_{z \in \{0,1\}^{|V|}} \quad \min_{y \in \{0,1\}^{|V|}} \quad \max_{\substack{x \in \{0,1\}^{|V|} \\ \alpha \in [0,1]^{|V|}}} \sum_{v \in V} b_v \alpha_v \\
 & \sum_{v \in V} \hat{c}_v z_v \leq \Omega \quad \sum_{v \in V} h_v y_v \leq \Phi \\
 & \text{subject to } \sum_{v \in V} c_v x_v \leq \Lambda \\
 & \alpha_v \leq 1 + z_v - y_v, \quad \forall v \in V \quad (1a) \\
 & \alpha_v \leq \alpha_u + x_u + z_v, \quad \forall (u, v) \in A, \quad (1b)
 \end{aligned}$$

where z, y, x and α are decision vectors which coordinates are z_v, y_v, x_v and α_v for each $v \in V$. The first maximization is controlled by the defender whose selection of vertices to be vaccinated is encoded by the binary vector z and it must respect the budget constraint; we denote the set of vaccinated vertices by $D = \{v \in V : z_v = 1\}$. The minimization that follows is managed by the attacker with the directly infected vertices represented by the binary vector y and subject to the budget constraint; we denote the set of directly infected vertices by $I = \{v \in V : y_v = 1\}$. Finally, the last maximization is decided by the defender who controls the protected vertices, given by the binary vector x , and respects the associated budget constraint; we denote the set of protected vertices by $P = \{v \in V : x_v = 1\}$. Remark that in the last optimization, auxiliary variables α are included. They allow to mimic the propagation of the infection with $\alpha_v = 1$ if vertex v is saved; otherwise, 0. The right-hand-side of each Constraint (1a) is inferior to 1 (i.e., zero) if and only if vertex v is not vaccinated and it is attacked. For each Constraint (1b), its right-hand-side is inferior to 1 (i.e., zero) if and only if vertex v is not vaccinated neither protected and it has an adjacent vertex u that is infected (i.e., $\alpha_u = 0$). The effect of these constraints plus the fact that the last optimization maximizes a non-negative weighted sum of the variables α , guarantees the existence of an optimal solution with α binary and corresponding to the correct set of saved vertices. In multilevel optimization, the first stage (in MCN, the vaccination stage) is called the upper level or first level, the second stage is called the second level, and so on, with the last stage being also designated by lower level. Note that in each stage of a multilevel program, the subsequent stages are assumed to be solved to optimality. See [1] for further details on this mathematical programming formulation and Fig. 1 for an illustration of the game.

The description above makes it more clear that MCN is an example of a disruption problem. Its applicability to minimize the damage of fake news spreading on social networks and contagious malware in cybersecurity is reported in [1]. It should be remarked that MCN simplifies these real-world problems, e.g., the attacker is assumed to have full information about the network. Hence, MCN can be interpreted as a robust solution model, where the worst-case attacking scenario is always assumed to occur.

Contributions. To the best of our knowledge, this work is the first providing a computational complexity classification of the decision version of MCN, as well as, of its subgames. Namely, we investigate the subgames (i) PROTECT, where given D and I , the defender seeks the optimal protection strategy, (ii) ATTACK, where given D and no protection budget, the attacker determines the optimal infection strategy, (iii) ATTACK-PROTECT, where given D , the attacker computes the optimal infection strategy, and (iv) VACCINATION-ATTACK, where given no budget for protection, the defender finds the optimal vaccination strategy. This fundamental contribution sheds light on the practical difficulties dealt in [1]. Furthermore, it contributes to the understanding of sequential combinatorial games within the polynomial hierarchy and it motivates the focus on potentially $\Omega(2^{2^{|V|}})$ algorithms, heuristic methods or novel solution definitions. Table 1 summarizes our results for general graphs; unitary cases assume that all costs and benefits are 1, and undirected graphs assume that infection can traverse an edge in both directions. We stress the incorrectness of the following intuitive claim for multilevel optimization problems: if a subgame is C -hard for some complexity class C , then the associated game is at least C -hard. Note that in a multilevel optimization problem, like the MCN, the ultimate goal is to find the optimal first level decision. Hence, if for example in

Table 1

Computational complexity of the decision versions of the subproblems in MCN. Entries in gray correspond to results that follow as corollaries. In increasing order, we have: [4] \implies [5], [1] \implies [6], [12] \implies [13], [14] \implies [15], and [6-10] \implies [16-20].

DECISION VERSIONS	Undirected graphs		Directed graphs	
	UNITARY CASE Section 3	WEIGHTED CASE Section 4	UNITARY CASE Sections 5 and 6	WEIGHTED CASE
PROTECT	[1] NP-complete	[6] NP-complete	[11] NP-complete	[16] NP-complete
ATTACK	[2] Polynomial	[7] NP-complete	[12] NP-complete	[17] NP-complete
ATTACK-PROTECT	[3] NP-hard	[8] Σ_2^P -complete	[13] NP-hard	[18] Σ_2^P -complete
VACCINATION-ATTACK	[4] NP-complete	[9] Σ_2^P -complete	[14] Σ_2^P -complete	[19] Σ_2^P -complete
MCN	[5] NP-hard	[10] Σ_3^P -complete	[15] Σ_2^P -hard	[20] Σ_3^P -complete

the MCN, we had always $\Omega = |V|$, then we would know directly that all vertices are saved, even if the attack problem is theoretically intractable. This supports the interest of understanding the individual complexity of each subgame of MCN.

We also contribute with an algorithmic analyzes of PROTECT by exploring graph classes where it becomes polynomially solvable.

Paper organization. In Section 2, we revise the literature associated with MCN, allowing to position our contribution in the context of critical node problems, interdiction games and defender-attacker-defender problems. In Section 3, we focus on the case where graphs are undirected and each vertex benefit and cost is unitary. Section 4 adds the possibility of having non-unitary parameters, while Section 5 generalizes the game to directed graphs. Finally, Section 6 investigates structural properties of special graph classes that can be explored to make at least PROTECTION polynomially solvable, both on directed and undirected graphs.

2. Related literature

Assessing the vulnerability of complex infrastructures such as networks is of the utmost importance in practice. One way to measure the robustness of a given network is to study its connectivity properties, for which many metrics exist. With respect to a fixed metric, vertices often play different roles in the graph, with varying levels of importance. The most important vertices are qualified as *critical*. Thus, the problem of detecting subsets of critical vertices with respect to some connectivity measure is of great interest, either for defensive or for offensive purposes, and with applications in domains ranging from network immunization [2,3] to computational biology [4,5].

Critical Node Detection Problems (CNDP). The CNDPs have been extensively studied, with names varying with the connectivity metric to optimize and the constraints of the problem. Many of its studied versions have been shown to be NP-complete on general graphs; see Lalou et al. [6] for a recent survey. Indeed, many of these belong to the class of problems called *Node-Deletion Problems*. They consist in deleting the smallest subset of vertices from a graph so that the induced subgraph satisfies a certain property π . Lewis and Yannakakis [7] showed that if π is *nontrivial* and *hereditary*, then the subsequent vertex deletion problem is NP-hard. In particular, *MinMaxC*, the problem of finding a set of vertices D from a graph G with a budget constraint $|D| \leq \Omega$ such that the removal of D minimizes the size of the largest connected component in the remaining graph, was shown to be NP-hard in the strong sense thanks to this argument [8]. Moreover, some CNDP problems remain NP-hard even on particular graph classes [9,6]. For example, the original *Critical Node Problem* (CNP) [2] which seeks to minimize the *pairwise connectivity* of the graph by removing a limited number of vertices remains NP-hard on split or bipartite graphs [9]. Several works tried to clarify the frontier between polynomial and NP-hard instances for different variants of the CNDP. The version based on pairwise connectivity over trees is studied in [10] where it is found to be polynomial with unit connection costs and strongly NP-hard otherwise. Many other versions of the CNDP were studied in details over trees, such as the versions based on the cardinality of the largest component (*MinMaxC*) and the number of connected components (*MaxNum*) [8], the largest pairwise connectivity among all components [11] or an extension of pairwise connectivity based on the length of shortest paths in the remaining graph [12]. A stochastic version of the pairwise CNDP with node attack failure was studied over trees in [13] and found to be NP-hard, even with unit connection costs. The CNDP was also studied on other specially structured graphs, such as series-parallel graphs [12,8], graphs with bounded treewidth [9], proper interval graphs [11] or bipartite permutation graphs [14].

On the algorithmic side, many approaches have been proposed to solve the CNDP. The state of the art in terms of formulations with Mixed Integer Linear Programs (MILPs) is represented by the works in [15,16], which build on the model of Veremyev et al. [17]. Specifically, the compact MILP in [15] obtains the same linear relaxation strength as the one of Di Summa et al. [18] which has an exponential number of constraints. Karakose and McGarvey [16] use a reformulation based on the uncapacitated multi-commodity flow interdiction model to obtain a strong linear relaxation. These models allow to speed up the convergence sometimes by more than one order of magnitude with respect to [17] and to solve optimally instances with up to approximately a thousand vertices. However, this is true only for certain types of graphs, mainly those which are very sparse. Graphs of the *small-world* type with no articulation point are usually much harder to solve,

including through the use of metaheuristics. Several heuristics have been devised to tackle different versions of the CNDP: the algorithms which have proven to be competitive on (some part of) a benchmark of real and synthetic instances are based on evolutionary frameworks [19–21], local-search based metaheuristics [22], randomized greedy approaches [23,24] or a GRASP with exterior path relinking [25]. These algorithms have been tested on instances with up to 23,000 vertices and almost 100,000 edges. A couple of works have proposed algorithms for the CNDP based on known approximation techniques such as the randomized rounding of the linear relaxation solution or the region growing algorithm [26,27]. In a similar spirit, an approximation algorithm to the cardinality-constrained critical node detection problem is proposed in [28]. It should be mentioned that in the case of the CNDP, no approximation algorithm can provide a hard theoretical guarantee since it has been proved to be inapproximable within any approximation factor [9]. We refer the interested reader to the recent literature surveys of [6,29] on exact and heuristic algorithms for solving various versions of the CNDP.

Interdiction games. In several CNDP, although the optimization problem is formulated with a natural single objective, the task is inherently constituted of several ones. In the CNP, minimizing the pairwise connectivity maximizes the number of connected components in the residual graph, while simultaneously minimizing the variance in the component sizes [2]. Even though in this particular case, it has been shown that the multi-objective formulation is not equivalent to the original one [30], splitting the objective in two is sometimes possible. For example, Furini et al. [31] exhibited the hidden bilevel structure of the *Capacitated Vertex Separator problem* by formulating it as a two player Stackelberg game in which a leader interdicts the network by removing some of its vertices and a follower determines the maximum connected component in the remaining graph, highlighting the link between CNDP problems and *Interdiction Games*. Interdiction games on networks are a special family of two-player zero-sum Stackelberg games in which a leader interdicts parts of the network (*arcs or vertices*) subject to a budget limitation in order to maximize the disruption of the follower's objective who solves an optimization problem on the remaining graph (e.g., *the maximum flow or the maximum clique*). The most studied interdiction games where the follower's problem is a linear program, such as the *Network Flow Interdiction* [32], the *Shortest Path Interdiction* [33] and their extensions, are NP-complete. Some of the interdiction games with an integer program describing the follower were shown to be Σ_2^P -complete, e.g., the *Binary Knapsack Interdiction problem* [34,35] and the *Maximum Clique Interdiction game* [36,37], shading light on the intrinsic relationship between this class of problems and the second level of the polynomial hierarchy.

The structure of the games described above has been explored to design exact algorithms. For example, in [38] an exact method for the Binary Knapsack Interdiction problem was shown to efficiently solve instances with up to 500 items. In [37], efficient bounding techniques for the follower's problem together with a *branch-and-cut* approach, allow to tackle instances of the Maximum Clique Interdiction with up to 100,000 vertices within few seconds. There are some general exact methodologies. Fischetti et al. [39] devised a general branch-and-cut for the case where the follower solves an MILP. Experimental results reveal that their methodology is effective for medium-size *Multidimensional Binary Knapsack Interdiction problems*. In [40], the authors focused on interdiction games with submodular and monotone objectives. Again, a branch-and-cut algorithm is proposed and tested on instances of the *Weighted Maximal Covering Interdiction game* and of the *Bipartite Inference Interdiction game*. Finally, a heuristic approach for general interdiction games is presented in [41] with experiments revealing that it works well for different types of those games.

The unitary undirected version of MCN, as originally introduced by Baggio et al. [1], is not an interdiction problem per se but *contains* one. Indeed, the vaccination stage of the game focuses on identifying critical infrastructures in the network to interdict them preventively to counter an intentional attack, which falls into the framework of Network Interdiction problems. Nevertheless, the game does not finish with the attack: there is a third stage where the defender tries to isolate the propagation of the infection to maximize the unharmed fraction of the network. Finding a blocking strategy to limit the diffusion of an infection is related to the *Firefighter problem*, which has been shown to be NP-complete, even for trees of maximum degree three [42] and was studied more recently in [43] where the problem was shown to be NP-complete on split graphs and bipartite graphs but polynomial on graphs with bounded treewidth. Thus, the MCN problem combines two different paradigms in network protection, *prevention* and *blocking*, each being related to provably hard problems. The overall contraction leads to a trilevel optimization formulation for the MCN, making it fall under the *Defender-Attacker-Defender* (DAD) framework introduced by Brown et al. [44] to study the defense of critical infrastructure against malicious attacks. Such frameworks are the first step towards more complex situations where both players respectively interdict and protect network elements sequentially over time. An example of such works can be found in [45], where the attacker removes graph vertices one at a time while the defender randomly adds k edges between attack rounds. Another more recent example is [46] where the authorities imprison drug dealers to move up in the criminal hierarchy, committing investigation resources, over several time periods.

Defender-Attacker-Defender. Although the general DAD was claimed to be NP-hard in [47], complexity results for trilevel combinatorial optimization problems are scarce. In [48], a new proof that *Trilevel Linear Programming* is Σ_2^P -hard is provided, building upon the results in [49–51] showing that the *Multilevel Linear Programming* problem with $L + 1$ levels is Σ_L^P -hard. In fact, the decision version of MCN problem can be formulated as “given 3 integer budgets Ω, Φ, Δ , a graph G and an integer K , is there a vaccination D such that for all attacks I there exists a protection P saving at least K vertices?” Thus, there seems to be a link between the MCN and the 3-*alternating quantified satisfiability* problem which has been shown to be Σ_3^P -complete by Meyer, Stockmeyer and Wrathall [52,53], making one expect the MCN to be complete for this class.

We stress that very few problems have been shown to be naturally Σ_3^P -complete in the literature up to now, in addition to infinite families of problems which have been shown to be Σ_L^P -complete for any level L of the polynomial hierarchy (as, e.g., Satisfiability Problems, or Multilevel Linear Programming). The compendium of [54], whose last update dates back to 2008, describes eight Σ_3^P -complete problems including graph theory problems, problems over sets as well as number theory problems. Since this compendium was last updated, a handful of other problems have been demonstrated to be Σ_3^P -complete, in the domains of logic, knowledge representation and artificial intelligence. We can cite, e.g., the problem of *Binding Forms in First-Order Logic* [55], deciding whether a propositional program has epistemic FLP (Faber, Leone and Pfeifer) answer sets [56], or checking the existence of max optimal outcomes over mCP -nets to study the aggregation of preferences over combinatorial domains in artificial intelligence [57]. To the best of our knowledge, there is approximately a dozen proven natural Σ_3^P -complete problems in the literature, which makes it all the harder to derive Σ_3^P -hardness for a given trilevel problem. In this work, we add two more problems to the list of Σ_3^P -complete problems, the *Trilevel Knapsack Interdiction Problem* and the *Multilevel Critical Node Problem*. Even though the set of proven Σ_2^P -complete problems is larger by one order of magnitude, i.e., a little more than roughly a hundred such problems are known, we also add several new Σ_2^P -complete problems to this list.

There are methodological works tackling DADs that do not prove the associated theoretical computational complexity (not even for its subgames), e.g., [44,58,1,59–61]. In [58], a *backward sampling* mechanism is proposed for general DADs. Their algorithm is tested on problems where the lower level is polynomially solvable; e.g., for the graph problems in their testbed, grid graphs with up to a few thousand vertices can be solved within an hour of computational time. The typical technique used to solve a DAD is usually a tailored *cutting plane* algorithm; e.g. [1,59,60] address DADs in graphs whose lower level contains binary variables. The size of the instances that can be tackled by such approaches is of the order of a hundred vertices with a restrictive attacker’s budget (usually, less than 5). Similarly, the recent paper by Barbosa et al. [61] proposes a cutting plane algorithm. Such an approach hints that their problem is a DAD although the paper does not explicitly frame it as one: the authors formulate a bilevel model where the defender adds edges while anticipating that the attacker solves a CNDP; once an attack occurs, the computation of the graph connectivity can be modeled as an optimization problem. Concerning heuristics for DADs, Nabli and Carvalho [62] designed and tested a reinforcement learning approach to automatically build a greedy solution to MCN. Their heuristic is shown to provide solutions close to the optimum for those instances solved in [1].

3. Undirected graphs: the unitary case

In this section, we focus on undirected graphs $G = (V, E)$, i.e., for each couple of vertices $(u, v) \in V \times V$, if the arc (u, v) is in G , then (v, u) is also in the graph. We thus call E the set of edges. Here, we also consider unit benefits and costs, i.e., $\forall v \in V, \hat{c}_v = h_v = c_v = b_v = 1$. We introduce s , the function that, given a graph G , the vaccination strategy D , the attack strategy I and the protection strategy P , returns $s(G, D, I, P)$, the number of saved vertices in the end of the game. Thus, in this setting, the trilevel formulation of the problem is simply:

$$\max_{\substack{D \subseteq V \\ |D| \leq \Omega}} \min_{\substack{I \subseteq V \setminus D \\ |I| \leq \Phi}} \max_{\substack{P \subseteq V \setminus (I \cup D) \\ |P| \leq \Lambda}} s(G, D, I, P). \tag{2}$$

To ease our analysis, guided by the relationship between *Critical Node Detection Problems* and *Node-Deletion Problems*, we first write the immediate Property 3.1² stating that *vaccinating* or *protecting* vertices has the same effect as *removing* them from the graph with respect to s . Starting from $G = (V, E)$ and a subset $W \subseteq V$, we denote by $G[V \setminus W]$ the graph induced by the deletion of the vertices in W and its incident edges.

Property 3.1. *Given G, D, I, P , we have that*

$$s(G, D, I, P) = s(G[V \setminus (D \cup P)], \emptyset, I, \emptyset) + |D| + |P|.$$

What Property 3.1 actually says is that the infected vertices in G are the ones in the connected components of $G[V \setminus (D \cup P)]$ where there is at least one attacked vertex in I .

We will start by classifying the computational complexity of PROTECT, followed by the one of ATTACK-PROTECT, and, finally, VACCINATION-ATTACK. From the latter, we obtain the complexity of ATTACK, and the minimum complexity of MCN.

3.1. The PROTECTION problem

In PROTECT, the defender is given D and I and seeks to find an optimal P . Thus, thanks to Property 3.1, we can assume that the game takes place in $G_a = G[V \setminus D]$ for this last move: the defender wants to find at most Λ vertices $P \subseteq V_a \setminus I$ that

² It is easy to see that Property 3.1 holds for general directed weighted versions with $s(G, D, I, P) = s(G[V \setminus (D \cup P)], \emptyset, I, \emptyset) + \sum_{v \in D} b_v + \sum_{v \in P} b_v$ and $s(G, D, I, P)$ equal to the benefit associated with the saved vertices in the end of the game.

will maximize $s(G_a, \emptyset, I, P)$. For a given choice of P , we introduce $C_1(P), \dots, C_{N(P)}(P)$, the $N(P)$ connected components in the graph $G_a[V_a \setminus P]$. Hence, the objective of the defender being to find P minimizing the number of infected vertices $f(P)$, we can define it as:

$$f(P) = \sum_{i=1}^{N(P)} |C_i(P)| \times \mathbb{1}_{C_i(P) \cap I \neq \emptyset}. \tag{3}$$

We will show that finding such a P is NP-complete. We argue that it is a direct consequence of the results of [9] showing that the *Critical Node Problem* is NP-hard on split graphs.

3.1.1. The Critical Node Problem on split graphs

The *Critical Node Problem (CNP)* [2] is a related problem to ours. The setting is very similar to PROTECT: we have an undirected graph $\bar{G} = (\bar{V}, \bar{E})$, an integer budget B , and we want to find a subset \bar{P} of vertices to remove that minimizes the *pairwise connectivity* of the residual subgraph $\bar{G}[\bar{V} \setminus \bar{P}]$ under the constraint of having $|\bar{P}| \leq B$. If we denote by $\bar{C}_1(\bar{P}), \dots, \bar{C}_{N(\bar{P})}(\bar{P})$ the $N(\bar{P})$ connected components of $\bar{G}[\bar{V} \setminus \bar{P}]$, the measure we want to minimize is:

$$g(\bar{P}) = \sum_{i=1}^{N(\bar{P})} \binom{|\bar{C}_i(\bar{P})|}{2} \tag{4}$$

where each term in the sum is the pairwise connectivity of $\bar{C}_i(\bar{P})$. Here, we will focus more particularly on *split graphs*. A split graph is a graph $\bar{G} = (\bar{V}, \bar{E})$ whose vertices \bar{V} can be split in two sets \bar{V}_1 and \bar{V}_2 , \bar{V}_1 forming a clique and \bar{V}_2 an independent set. Thus, the decision problem for this particular case of the CNP is:

CNP_{split}:
 INSTANCE: A split graph $\bar{G} = (\bar{V}_1, \bar{V}_2; \bar{E})$, a non-negative integer budget $B \leq |\bar{V}|$ and a non-negative integer \bar{K} .
 QUESTION: Is there a subset $\bar{P} \subseteq \bar{V}$, $|\bar{P}| \leq B$ such that $g(\bar{P}) \leq \bar{K}$?

As [9] noted, in this setting there is at most one connected component of the residual subgraph $\bar{G}[\bar{V} \setminus \bar{P}]$ that contains more than one vertex. Moreover, it is easy to see that if this *nontrivial connected component* exists, it necessarily contains a subclique of $\bar{G}[\bar{V}_1]$. More than that, it is the only connected component of $\bar{G}[\bar{V} \setminus \bar{P}]$ containing vertices from \bar{V}_1 . Thus, we can name \bar{C}_1 the connected component containing vertices of \bar{V}_1 (in the case of $\bar{P} \supseteq \bar{V}_1$, then \bar{C}_1 is either a singleton from \bar{V}_2 or is empty and our reasoning still holds). Then, minimizing (4) is equivalent to minimize $|\bar{C}_1|$. But finding the subset of vertices \bar{P} to remove to do that has been shown to be NP-hard:

Lemma 3.2 ([9]). *CNP_{split} is NP-hard.*

3.1.2. Complexity result

Next, we show that the decision version of PROTECT is NP-complete using a reduction from CNP_{split}. The decision problem is the following:

PROTECT:
 INSTANCE: A graph $G_a = (V_a, E_a)$, a set of attacked vertices $I \subseteq V_a$, a non-negative integer budget $\Lambda \leq |V_a| - |I|$ and a non-negative integer K .
 QUESTION: Is there a subset $P \subseteq V_a \setminus I$, $|P| \leq \Lambda$ such that the number of infected vertices $f(P) \leq K$?

Note that the question can be equivalently re-written with the inequality $s(G_a, \emptyset, I, P) \geq |V_a| - K$.

Theorem 3.3. *PROTECT is NP-complete.*

Proof. It is easy to see that PROTECT is NP as determining the objective value only requires finding the connected components of $G_a[V_a \setminus P]$ which can be done in linear time using a depth-first search (DFS).

To complete the proof, we exhibit an immediate reduction from CNP_{split}. Let us take an instance of this problem, i.e. a split graph $\bar{G} = (\bar{V}_1, \bar{V}_2; \bar{E})$, a non-negative integer budget B and a non-negative integer \bar{K} . Given that, we build a graph G_a by growing by one the size of the clique $\bar{G}[\bar{V}_1]$ with the addition of a vertex u . Thus, $V_a = \bar{V}_1 \cup \{u\} \cup \bar{V}_2$ and E_a is obtained by taking \bar{E} and adding an edge $(u, \bar{v}_1) \forall \bar{v}_1 \in \bar{V}_1$. In fact, the new graph is still a split graph $G_a = (\bar{V}_1 \cup \{u\}, \bar{V}_2; E_a)$. Finally, the corresponding instance of PROTECT is given by G_a , $I = \{u\}$, $\Lambda = B$ and $K = \lfloor \frac{1}{2}(3 + \sqrt{8\bar{K} + 1}) \rfloor$ (obtained by solving

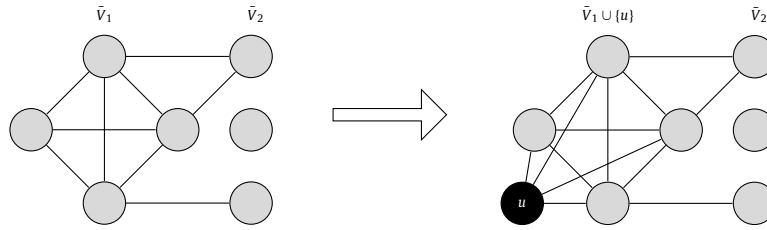


Fig. 2. Example of construction of G_a from \tilde{G} .

$\bar{K} = \binom{K-1}{2}$). An example of such construction can be found in Fig. 2. Then, as there is only one attacked vertex, minimizing (3) on this instance of PROTECT corresponds to choosing a P that minimizes the size of the unique connected component to which u belongs in $G_a[V_a \setminus P]$. Let's name C_1 this connected component. But as u belongs to the clique part of the split graph G_a , C_1 is also the unique connected component of $G_a[V_a \setminus P]$ containing vertices from $V_1 = \tilde{V}_1 \cup \{u\}$. Thus, we have that $C_1 = \tilde{C}_1 \cup \{u\}$ and $g(P) = \binom{f(P)-1}{2}$. Hence, finding P that minimizes f on G_a is equivalent to finding P that minimizes g on \tilde{G} . This finishes the proof that PROTECT is NP-hard. \square

Remark 3.4. In [43], it is shown that the FIREBREAK problem is NP-complete. This problem is equivalent to PROTECT when $|I| = 1$. Hence, their result can be used to establish Theorem 3.3. Nevertheless, given that our reductions differ significantly and were obtained independently, we decided to present our alternative reduction.

3.2. The ATTACK-PROTECT problem

We showed that solving the last level of MCN is NP-complete, now we will prove that ATTACK-PROTECT is also NP-hard. In this bilevel problem, we are taking the side of the attacker: the aim is to find the attack that will maximize the number of infected vertices after protection. The decision version of the problem is:

ATTACK-PROTECT:
 INSTANCE: A graph $G_a = (V_a, E_a)$, two non-negative integer budgets Φ, Λ such that $\Phi + \Lambda \leq |V_a|$ and a non-negative integer $K \leq |V_a|$
 QUESTION: Is there a subset $I \subseteq V_a, |I| = \Phi$ such that $\forall P \subseteq V_a \setminus I, |P| \leq \Lambda$, the number of infected vertices $f(P) \geq K$?

We will use a reduction from the Dominating Set problem, a known NP-complete problem [63], whose decision version is:

DOMINATING SET:
 INSTANCE: A graph $\tilde{G} = (\tilde{V}, \tilde{E})$, a positive integer $B \leq |\tilde{V}|$
 QUESTION: Is there a subset $U \subseteq \tilde{V}, |U| \leq B$, such that $\forall v \in \tilde{V} \setminus U, \exists u \in U$ such that $(u, v) \in \tilde{E}$?

Theorem 3.5. ATTACK-PROTECT is NP-hard.

Proof. Let us take a graph $\tilde{G} = (\tilde{V}, \tilde{E})$ and a positive integer $B \leq |\tilde{V}|$. The instance of ATTACK-PROTECT is simply created by taking $G_a = \tilde{G}, \Phi = B, \Lambda = |V_a| - \Phi - 1$ and $K = \Phi + 1$. In this configuration, we have a protection budget Λ which is exactly one less than the number of vertices that are not attacked. Thus, if all the protection budget is spent, there is only one vertex u in the graph that is neither attacked nor protected. Therefore, if u becomes infected after protection (i.e. $f(P) = K = \Phi + 1$), that means that the protection strategy did not manage to save one unit of budget while saving all the other vertices, meaning that the other vertices were all in direct contact with at least one attacked one (if it was not the case, one unit of budget could have been saved by protecting all the neighbors of the vertex that is not in direct contact with I). As u also becomes infected, it also means that it is adjacent to one vertex in I . Thus, finding I such that $\forall P, f(P) \geq K$ means that I is a dominating set of size B , which concludes the proof. \square

3.3. The VACCINATION-ATTACK problem

In this part, we will ignore the fact that there is a protection stage at the end. This is a particular case of MCN since it is equivalent to studying it with protection budget $\Lambda = 0$. We will show that the bilevel problem VACCINATION-ATTACK is NP-complete. The decision problem is the following:

VACCINATION-ATTACK:
 INSTANCE: A graph $G = (V, E)$, two non-negative integer budgets Ω and Φ such that $\Omega + \Phi \leq |V|$ and a non-negative integer K .
 QUESTION: Is there a subset $D \subseteq V$, $|D| \leq \Omega$ such that $\forall I \subseteq V \setminus D$ with $|I| \leq \Phi$, the number of infected vertices $|V| - s(G, D, I, \emptyset) \leq K$?

First, we argue that in this configuration, finding the optimal attack following a given vaccination can be done in polynomial time.

Lemma 3.6. *VACCINATION-ATTACK \in NP. Moreover, ATTACK can be solved in polynomial time.*

Proof. Given a vaccinated set D , we want to verify that all the possible subsequent attacks cannot infect more than K vertices. To do that, it suffices to find the best attack, i.e., solve the Attacker optimization problem, and check whether or not it complies with the inequality. But, as we highlighted it with Property 3.1, the graph on which the attack phase takes place is $G_a = G[V \setminus D]$ and the saved vertices in the end are exactly the ones in the connected components of G_a that do not contain any attacked vertex. Thus, the best attack possible given G_a and budget Φ is to infect one vertex in each of the Φ largest connected components of G_a . This can be done in linear time using a DFS. Hence, VACCINATION-ATTACK \in NP. \square

In fact, this proof showed that VACCINATION-ATTACK is actually equivalent to another problem: finding a subset of vertices D to remove from G that minimizes the sum of the sizes of the Φ largest connected components in the induced subgraph. Let's call this problem MINMAX Φ C:

MINMAX Φ C:
 INSTANCE: A graph $G = (V, E)$, two non-negative integer budgets Ω and Φ such that $\Omega + \Phi \leq |V|$ and a non-negative integer K .
 QUESTION: Is there a subset $D \subseteq V$, $|D| \leq \Omega$ such that the sum of the sizes of the Φ largest connected component in $G[V \setminus D]$ is less than K ?

Lemma 3.7. *VACCINATION-ATTACK and MINMAX Φ C are equivalent problems.*

As Shen et al. [8] argued that MINMAXC, the problem that only seeks to minimize the size of the largest connected component in the residual graph, is NP-hard, we have as a direct consequence that MINMAX Φ C is also NP-hard, which leads to the following corollaries:

Corollary 3.8. *VACCINATION-ATTACK is NP-complete.*

Corollary 3.9. *MCN is NP-hard.*

Proof. Given an instance of VACCINATION-ATTACK, there is a corresponding instance of MCN by taking the same G, Ω, Φ, K and by setting $\Lambda = 0$. \square

4. Undirected graphs: the weighted case

In this section, we study the version of MCN presented in problem (1) restricted to undirected graphs. We will use the subscript w to denote the weighted version, MCN_w , as well as for its subgames. In this problem, given a graph $G = (V, E)$, each vertex $v \in V$ is associated with a benefit b_v and cost parameters \hat{c}_v, h_v and c_v , respectively the cost of vaccinating, attacking and protecting vertex v . While in the unitary case all these parameters, b_v, \hat{c}_v, h_v and c_v , take value 1, in a weighted instance, they can take any non-negative integer value. First, note that the NP-completeness of PROTECT $_w$ is immediate from the previous section.

Having introduced costs and benefits, our game and its subgames are intimately related to *Knapsack problems*, which we will use to demonstrate all of our complexity results in this part. We will start by highlighting the direct relationship between ATTACK $_w$ and KNAPSACK, which will get us the NP-completeness of this problem. Then, we will focus on the two bilevel sub-problems VACCINATION-ATTACK $_w$ and ATTACK-PROTECT $_w$ and prove they are Σ_2^P -complete thanks to a *Knapsack Interdiction problem*. To conclude, we show that MCN_w is Σ_3^P -complete. We will observe that the introduction of non-unitary parameters offers sufficient flexibility to go a level up in the polynomial hierarchy in comparison with the unitary undirected cases.

4.1. The ATTACK $_w$ problem

In the attack phase, the vaccination already took place so we effectively work on G_a , which is the result of the deletion of the vaccinated vertices from the original graph. We are given a non-negative attack budget Φ , and as there is no protection

ATTACK_w:
 INSTANCE: An undirected graph $G_a = (V_a, E_a)$, a non-negative integer cost h_v and value b_v for each vertex $v \in V$, a non-negative integer budget Φ , and a non-negative integer number K .
 QUESTION: Is there a subset of vertices $I \subseteq V_a$ to attack, with cost $\sum_{v \in I} h_v \leq \Phi$ such that the sum of the benefits of the resulting infected vertices in G_a is greater or equal to K ?

phase afterwards, we set $\Lambda = 0$. The goal is thus to harvest the most benefit possible by infecting vertices subject to a budget limit. The decision version of the problem is then:

To make evident the NP-completeness of the problem, we simply state the decision version of the *Knapsack problem*, one of the Karp’s 21 NP-complete problems [64]:

KNAPSACK:
 INSTANCE: Finite set U , for each $u \in U$, a positive integer size $a_u \in \mathbb{N}$ and a positive integer profit $p_u \in \mathbb{N}$, and two positive integers B and \bar{K} .
 QUESTION: Is there a subset $U' \subseteq U$ such that $\sum_{u \in U'} a_u \leq B$ verifying $\sum_{u \in U'} p_u \geq \bar{K}$?

Theorem 4.1. *ATTACK_w is equivalent to KNAPSACK.*

Proof. First, we prove that each instance of *ATTACK_w* reduces to an instance of *KNAPSACK*. Given an instance of *ATTACK_w*, it is straightforward to see that it is sufficient to infect the vertex v with lowest infection cost h_v of a given connected component to infect the whole component and collect the benefit b of each vertex included in that component. If $N(G_a)$ represents the set of connected components of G_a , to each connected component $C \in N(G_a)$ we can assign a total profit $b_C = \sum_{v \in C} b_v$ and infection cost $h_C = \min_{v \in C} h_v$. We can then straightforwardly build a *KNAPSACK* instance where the set $N(G_a)$ is mapped to U , $a_u = h_C$ and $p_u = b_C$ for $C \in N(G_a)$, and $B = \Phi$ and $\bar{K} = K$.

Conversely, if we start from an instance of *KNAPSACK*, we construct an instance of *ATTACK_w* by setting $V_a = U$, $E_a = \emptyset$, $K = \bar{K}$, $\Phi = B$, and $\forall v \in V_a$, $h_v = a_v$, $b_v = p_v$. In this configuration, G_a having no edges, the attacked vertices are exactly the infected ones in the end, and the goal of the attacker is equivalent to filling up a knapsack with limited capacity by choosing which vertices to attack.

Given that both *ATTACK_w* and *KNAPSACK* can be reduced to each other, both problems are equivalent. \square

Remark that given an attack I , finding the subsequent infected vertices can be done in linear time thanks to a DFS. Then, it suffices to sum the cost of the vertices in I to verify the budget constraints and to sum the benefits associated with the infected vertices to verify that it is greater or equal to K . Hence, *ATTACK_w* \in NP and thus:

Corollary 4.2. *ATTACK_w on undirected graphs is weakly NP-complete, even on trivial graphs.*

Proof. Since it well known that *KNAPSACK* is weakly NP-complete, the result follows from the above theorem. Moreover, since any instance of *KNAPSACK* reduces to an instance of *ATTACK_w* which has no edges, *ATTACK_w* is NP-complete on trivial graphs. \square

4.2. The *ATTACK-PROTECT_w* problem

In the proof of Theorem 4.1, we highlighted how a *KNAPSACK* instance can be directly transformed into a weighted graph with no edges. In this section, as well as in the next one, we will use a similar transformation, but add one additional *root vertex* to our construction in order to build a star graph: one root vertex connected with an edge to each of the other vertices, each one representing an item $u \in U$ of the knapsack. That way, the complexity results we devise hold for trees, a very particular class of graphs where frequently theoretically intractable problems become polynomially solvable.

As before, the vaccination having already been done, we start from G_a , the graph where the vaccinated vertices have been removed.

ATTACK-PROTECT_w:
 INSTANCE: A graph $G_a = (V_a, E_a)$, a non-negative integer K , two non-negative integer budgets Φ and Λ , $\forall v \in V_a$ two non-negative integer costs h_v , c_v and a non-negative integer benefit b_v .
 QUESTION: Is there a subset $I \subseteq V_a$, with cost $\sum_{v \in I} h_v \leq \Phi$ such that $\forall P \subseteq V_a \setminus I$ with cost $\sum_{v \in P} c_v \leq \Lambda$, the sum of the benefit of the saved vertices is strictly less than K ?

In order to show that ATTACK-PROTECT_w is Σ_2^P -complete, we use the *Bilevel Interdiction Knapsack Problem* introduced by DeNegre [34] and proven to be Σ_2^P -complete in [35]. In this problem, two players, a leader and a follower, can select items in the same set of objects O . First, the leader packs some items into her knapsack, then the follower chooses among the remaining ones. The aim of the leader is to interdict a subset of items, subject to a capacity constraint, in order to minimize the total profit of the follower. The objective of the follower is to maximize her profit, subject to a constraint capping the maximum profit obtainable by her. The decision problem is then:

BILEVEL INTERDICTION KNAPSACK (BIK):
 INSTANCE: A set of items O such that each $o \in O$ has a positive integer weight a_o and a positive integer profit p_o , a positive integer maximum weight capacity A for the leader, a positive integer maximum profit B for the follower, and a positive integer $\bar{K} \leq B$.
 QUESTION: Is there a subset $O_l \subseteq O$ of items for the leader to select, with $\sum_{o \in O_l} a_o \leq A$, such that every subset $O_f \subseteq O \setminus O_l$ with $\sum_{o \in O_f} p_o \leq B$ that the follower can create has a total profit $\sum_{o \in O_f} p_o < \bar{K}$?

Theorem 4.3. ATTACK-PROTECT_w is strongly Σ_2^P -complete, even if the graph is a tree.

Proof. First, ATTACK-PROTECT_w is in Σ_2^P since this decision problem is exactly of the form $\exists I \forall P Q(I, P)$, where $Q(I, P)$ is a proposition that can be evaluated in polynomial time (i.e., it verifies the attack and protection budget constraints, as well as, the benefit of the saved vertices).

Next, we prove the problem Σ_2^P -hardness. Let us begin by noting that we can restrict the instances of BIK to the ones where \bar{K} and B are strictly inferior to $\sum_{o \in O} p_o$, otherwise, BIK reduces to KNAPSACK. This remark is used in the second part of this proof.

Starting from an instance of BIK, we construct an instance of ATTACK-PROTECT_w as follows. We first build a star graph $G_a = (V_a, E_a)$ with a root vertex r and a vertex v_o for each $o \in O$ linked to r through an edge (r, v_o) . We set $b_r = \sum_{o \in O} p_o + 1$ and $h_r = c_r = 1$. We also set $b_{v_o} = c_{v_o} = p_o$ and $h_{v_o} = a_o$ for each $o \in O$. See Fig. 3. Finally, we set $\Phi = A + 1$, $\Lambda = B$ and $K = \bar{K}$.

Suppose first that BIK is a Yes instance. Then, there is a set of items $O_l \subseteq O$ of total weight $\sum_{o \in O_l} a_o \leq A$ such that for all $O_f \subseteq O \setminus O_l$ feasible for the follower, it holds $\sum_{o \in O_f} p_o \leq \bar{K} - 1$. Consequently, in the ATTACK-PROTECT_w , the attacker can select the subset of vertices $I = \{r\} \cup \{v_o : o \in O_l\}$ with a feasible attacking cost $\sum_{v \in I} h_v = 1 + \sum_{o \in O_l} a_o \leq A + 1 = \Phi$. Now, the defender can only protect vertices in $\{v_o : o \notin O_l\}$ and since the central vertex of the star graph is infected, the saved vertices will be the protected ones. The aim of the defender is therefore to select the subset of vertices of maximum total benefit with respect to the protection budget Λ . This is exactly the follower's problem in BIK. Hence, since BIK is an Yes instance, the defender (follower in BIK) cannot attain a benefit (profit in BIK) equal or greater to $K = \bar{K}$ through a feasible action. Therefore, the ATTACK-PROTECT_w is a Yes instance.

Now suppose that ATTACK-PROTECT_w is a Yes instance. Thus, there exists an attack strategy $I \subseteq V_a$ such that there is **no** feasible subset $P \subseteq V_a \setminus I$ of protected vertices leading to a total benefit greater or equal to K for the defender. As $\Phi \geq 1$, it is obvious that the attacker will attack at least the central vertex r , otherwise, the defender would pick it and achieve a benefit superior to K (recall that $K = \bar{K} < \sum_{o \in O} p_o$), contradicting ATTACK-PROTECT_w Yes instance. Hence, the attacker is left with budget $\Phi - h_r = A$. Once the central vertex is attacked, only the other vertices subsequently protected will not be infected. Therefore, the rest of the attack budget A is spent on a subset of vertices of $\{v_o \in V_a : o \in O\}$ and it ensures that for any $P = \{v_o \in V_a : o \in O \setminus I\}$ with $\sum_{v \in P} c_v = \sum_{o: v_o \in P} p_v \leq \Lambda = B$, the total benefit for the defender is $\sum_{v \in P} b_v = \sum_{o: v_o \in P} p_v \leq \bar{K} - 1$. Consequently, BIK is also a Yes instance.

This completes the proof that ATTACK-PROTECT_w is Σ_2^P -complete. Moreover, since the BIK was shown to be NP-complete even for unary encoding, we can conclude that no pseudopolynomial-time algorithm exists to solve the ATTACK-PROTECT subgame. Since a star graph is a tree, the result stated in the theorem holds. \square

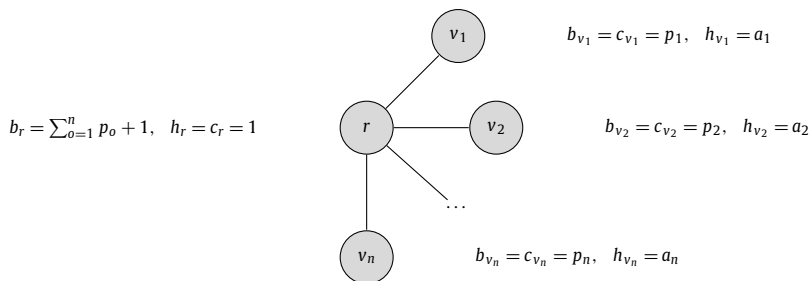


Fig. 3. Graph reduction from BIK to ATTACK-PROTECT_w when $O = \{1, 2, \dots, n\}$.

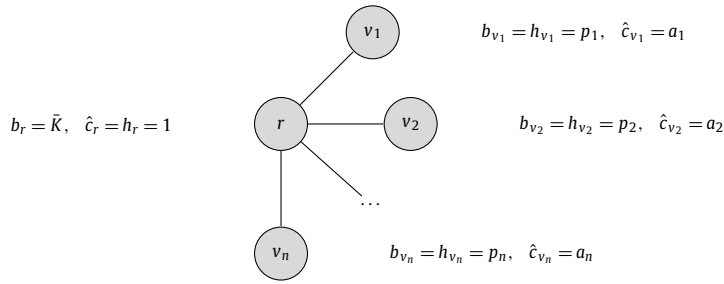


Fig. 4. Graph reduction from BIK to VACCINATION-ATTACK_w when $O = \{1, 2, \dots, n\}$.

4.3. The VACCINATION-ATTACK_w problem

Using a similar reduction to the one in the proof of Theorem 4.3, we show that the VACCINATION-ATTACK_w on weighted graphs is Σ_2^p -complete. As in the unitary case, this is equivalent to studying MCN_w problems where we set $\Lambda = 0$. The decision version of the problem is:

VACCINATION-ATTACK_w:
 INSTANCE: A graph $G = (V, E)$, a non-negative integer K , two non-negative integer budgets Ω and Φ , $\forall v \in V$ two non-negative integer costs \hat{c}_v, h_v and a non-negative integer benefit b_v .
 QUESTION: Is there a subset $D \subseteq V$, with cost $\sum_{v \in D} \hat{c}_v \leq \Omega$ such that $\forall I \subseteq V \setminus D$ with cost $\sum_{v \in I} h_v \leq \Phi$, the sum of the benefit of the infected vertices is strictly less than K ?

Theorem 4.4. VACCINATION-ATTACK_w is strongly Σ_2^p -complete, even if the graph is a tree.

Proof. As before, VACCINATION-ATTACK_w is in Σ_2^p since this decision problem is exactly of the form $\exists D \forall I Q(D, I)$ is a proposition that can be evaluated in polynomial time.

Now, we establish the problem Σ_2^p -hardness. We start from an instance of BIK, defined in the previous section, and we then construct an instance of VACCINATION-ATTACK_w as follows. First, we build a star graph $G = (V, E)$ with a central vertex r and $|O|$ leaf vertices v_o with $o \in O$. See Fig. 4. We add an edge (r, v_o) for each such leaf vertex. The central vertex has benefit $b_r = \bar{K}$ and costs $\hat{c}_r = h_r = 1$. Each leaf vertex v_o with $o \in O$ has a benefit $b_{v_o} = p_o$, cost for the defender $\hat{c}_{v_o} = a_o$ and cost for the attacker $h_{v_o} = p_o$. Finally, we fix $\Omega = A + 1$, $\Phi = B$ and $K = \bar{K}$.

This is exactly the setting of BIK and one can easily complete the proof of equivalence of the two decision instances following a path very similar to the proof of Theorem 4.3.

Finally, the reduction used a star graph which is a particular case of a tree. Hence, the problem is Σ_2^p -complete even on trees. \square

4.4. The MCN_w problem

In this section we show that the decision problem MCN_w is Σ_3^p -complete.

MCN_w:
 INSTANCE: A graph $G = (V, E)$, a non-negative integer K , three non-negative integer budgets Ω, Φ and Λ , $\forall v \in V$ three non-negative integer costs \hat{c}_v, h_v and c_v , and a non-negative integer benefit b_v .
 QUESTION: Is there a subset $D \subseteq V$, with cost $\sum_{v \in D} \hat{c}_v \leq \Omega$ such that $\forall I \subseteq V \setminus D$ with cost $\sum_{v \in I} h_v \leq \Phi$, there is $P \subseteq V \setminus I$ with cost $\sum_{v \in D} c_v \leq \Lambda$ such that the sum of the benefit of the saved vertices is greater or equal to K ?

In order to achieve our ultimate goal, we take the 3-Alternating Quantified Satisfiability problem ($B_3 \cap 3CNF$), known to be Σ_3^p -complete problem [65,53], in order to prove that the generalization of BIK to a trilevel, the Trilevel Interdiction Knapsack (TIK), is Σ_3^p -complete. Then, TIK is used to demonstrate that MCN_w is Σ_3^p -complete.

3-ALTERNATING QUANTIFIED SATISFIABILITY ($B_3 \cap 3CNF$):
 INSTANCE: Disjoint non-empty sets of variables X, Y and Z , and a Boolean expression E over $U = X \cup Y \cup Z$ in conjunctive normal form with at most 3 literals in each clause $c \in C$.
 QUESTION: Is there a 0-1 assignment for X so that for all 0-1 assignments of Y there is a 0-1 assignment of Z such that E is satisfied?

TRILEVEL INTERDICTION KNAPSACK (TIK):

INSTANCE: A set of items O such that each $o \in O$ has two positive integer weights a'_o and a_o and a positive integer profit p_o , two positive integer maximum weight capacities A' and A , a positive integer maximum profit B and a positive integer goal $\bar{K} \leq B$.

QUESTION: Is there a subset $O_1 \subseteq O$ of items, with $\sum_{o \in O_1} a'_o \leq A'$, such that every subset $O_2 \subseteq O \setminus O_1$, with $\sum_{o \in O_2} a_o \leq A$, there is a subset $O_3 \subseteq O \setminus O_2$, with $\sum_{o \in O_3} p_o \leq B$, such that $\sum_{o \in O_3} p_o \geq \bar{K}$ holds?

Theorem 4.5. *TIK is Σ_3^P -complete.*

Proof. The statement of TIK is of the form $\exists O_1 \forall O_2 \exists O_3 Q(O_1, O_2, O_3)$, directly implying that it is in Σ_3^P .

Next, we use a reduction from the $B_3 \cap 3CNF$ which is very much in line with the reduction from 3-SAT to Subset Sum presented in [66, Theorem 34.15]:

- For each variable $u \in U$, we create two items o_u and $o_{\bar{u}}$, one for each possible 0-1 assignment of u . We designate by $O_U = \{o_u : u \in U\}$ and $O_{\bar{U}} = \{o_{\bar{u}} : u \in U\}$ the two sets of items of size $|U|$.
- For each clause $c \in C$, (i) if c has 1 literal, we create one item o_c^1 , (ii) if c has 2 literals, we create two items o_c^1 and o_c^2 , and (iii) if c has 3 literals, we create three items o_c^1 , o_c^2 and o_c^3 . We designate by O_C the set of items associated with C .
- Weights, profits, maximum capacities, maximum profit and goal will be given by digits of size $|X| + |Y| + |Z| + |C| + 1$ in base 10. Hence, each digit position is labeled by a variable or a clause: the first $|C|$ positions (least significant numbers) are labeled by the clauses, then the next $|X|$ positions are labeled by the variables X , then the next $|Y|$ positions are labeled by the variables Y , then the next $|Z|$ positions are labeled by the variables Z , and, finally, the last position is labeled as *forbidden*.
 - For each $u \in U$, the two corresponding items o_u and $o_{\bar{u}}$ have weights and profits as described next. The weights and profits $a'_{o_u}, a_{o_u}, p_{o_u}, a'_{o_{\bar{u}}}, a_{o_{\bar{u}}}$ and $p_{o_{\bar{u}}}$ have digit 1 in the position labeled by the variable U and 0 in the positions labeled by other variables; the remaining digits are zero for $a'_{o_u}, a_{o_u}, a'_{o_{\bar{u}}}$ and $a_{o_{\bar{u}}}$. In particular, for all $o \in O_U \cup O_{\bar{U}}$, it holds $a'_{o_u} = a_{o_u}$ and $a'_{o_{\bar{u}}} = a_{o_{\bar{u}}}$.
If the literal u appears in clause $c \in C$, then p_{o_u} has digit 1 in the position labeled as c , and 0 otherwise. Similarly, if the literal $\neg u$ appears in clause $c \in C$, $p_{o_{\bar{u}}}$ has digit 1 in the position labeled by c , and 0 otherwise. Finally, for all $o \in O_U \cup O_{\bar{U}}$, p_{o_u} and $p_{o_{\bar{u}}}$ have digit 0 in the position labeled as forbidden.
 - For each $c \in C$, the associated items have weights and profits as follows. If c has one literal, $a'_{o_c^1}$ and $a_{o_c^1}$ have 1 in the position labeled as forbidden and 0 elsewhere; $p_{o_c^1}$ has digit 3 in the position labeled as c and 0 elsewhere. If c has two literals, $a'_{o_c^1}, a'_{o_c^2}, a_{o_c^1}$ and $a_{o_c^2}$ have 1 in the position labeled as forbidden and 0 elsewhere; $p_{o_c^1}$ and $p_{o_c^2}$ have digit 3 and 2, respectively, in the position labeled as c and 0 elsewhere. If c has three literals, $a'_{o_c^1}, a'_{o_c^2}, a'_{o_c^3}, a_{o_c^1}, a_{o_c^2}$ and $a_{o_c^3}$ have 1 in the position labeled as forbidden and 0 elsewhere; $p_{o_c^1}, p_{o_c^2}$ and $p_{o_c^3}$ have digit 3, 2 and 1, respectively, in the position labeled as c and 0 elsewhere.
 - The weight capacity A' has 1s for all digits with labels in X and 0s elsewhere. Hence, O_1 cannot contain items from $\{o_u, o_{\bar{u}} : u \in Z \cup Y\} \cup O_C$.
 - The weight capacity A has 1s for all digits with labels in Y , 2s for all digits with labels in X and 0s elsewhere. Hence, O_2 cannot contain items from $\{o_u, o_{\bar{u}} : u \in Z\} \cup O_C$.
 - The maximum profit B has 1s for all digits with labels in $X \cup Z$, 2s for all digits with labels in Y , 4s for all digits with labels in C , and 0s elsewhere. Hence, O_3 can take any item (as long as not interdicted by O_2).
 - We make \bar{K} is equal to B , except for the digits with labels Y , where it is 1.

See Fig. 5 for an illustration of our reduction.

Let $B_3 \cap 3CNF$ be a Yes instance. Then, take in O_1 the items o_u such that $u \in X$ is 1 and the items $o_{\bar{u}}$, otherwise. Clearly, this choice of O_1 respects the maximum weight A' . By construction, given this O_1 , the best O_2 will take all items associated with X and not taken by O_1 , as it does not interfere with the budget left for the items associated with Y . Furthermore, the optimal O_2 will also take exactly one of the items o_u or $o_{\bar{u}}$ for $u \in Y$:

- The two items associated with the most significant digit whose label is in Y cannot be taken simultaneously in O_2 as it would violate the weight capacity A . In fact, exactly one of these items must be taken, as otherwise O_3 would select them both, making the achievement of the profit \bar{K} only dependent on the items associated with the Z ; consequently, the goal would be achieved.
- The two items associated with the second most significant digit whose label is in Y cannot be taken simultaneously, since we already know that one of the items associated with the most significant digit in Y is taken which would result in a violation of the weight capacity A . Hence, reasoning as before, O_2 will take exactly of the items associated with the second most significant digit in Y .
- The reasoning above propagates until the least significant digit labeled in Y . We conclude that the best O_2 will have exactly one of the items o_u or $o_{\bar{u}}$ for $u \in Y$.

O	forbidden	Z			Y		X		C		
		d	c	b	a	c ₁	c ₂	c ₃			
o_a	$a'_{oa} = a_{oa}$	0	0	0	0	1	0	0	0		
	p_{oa}	0	0	0	0	1	1	0	1		
$o_{\bar{a}}$	$a'_{o\bar{a}} = a_{o\bar{a}}$	0	0	0	0	1	0	0	0		
	$p_{o\bar{a}}$	0	0	0	0	1	0	1	1		
o_b	$a'_{ob} = a_{ob}$	0	0	0	1	0	0	0	0		
	p_{ob}	0	0	0	1	0	1	0	1		
$o_{\bar{b}}$	$a'_{o\bar{b}} = a_{o\bar{b}}$	0	0	0	1	0	0	0	0		
	$p_{o\bar{b}}$	0	0	0	1	0	0	1	0		
o_c	$a'_{oc} = a'_{oc}$	0	0	1	0	0	0	0	0		
	p_{oc}	0	0	1	0	0	0	0	1		
$o_{\bar{c}}$	$a'_{o\bar{c}} = a'_{o\bar{c}}$	0	0	1	0	0	0	0	0		
	$p_{o\bar{c}}$	0	0	1	0	0	1	0	1		
o_d	$a'_{od} = a_{od}$	0	1	0	0	0	0	0	0		
	p_{od}	0	1	0	0	0	0	1	0		
$o_{\bar{d}}$	$a'_{o\bar{d}} = a_{o\bar{d}}$	0	1	0	0	0	0	0	0		
	$p_{o\bar{d}}$	0	1	0	0	0	0	0	0		
$o_{c_1}^1$	$a'_{oc_1^1}$	1	0	0	0	0	0	0	0		
	$a_{oc_1^1}$	1	0	0	0	0	0	0	0		
	$p_{oc_1^1}$	0	0	0	0	0	3	0	0		
$o_{c_1}^2$	$a'_{oc_1^2}$	1	0	0	0	0	0	0	0		
	$a_{oc_1^2}$	1	0	0	0	0	0	0	0		
	$p_{oc_1^2}$	0	0	0	0	0	2	0	0		
$o_{c_1}^3$	$a'_{oc_1^3}$	1	0	0	0	0	0	0	0		
	$a_{oc_1^3}$	1	0	0	0	0	0	0	0		
	$p_{oc_1^3}$	0	0	0	0	0	1	0	0		
$o_{c_2}^1$	$a'_{oc_2^1}$	1	0	0	0	0	0	0	0		
	$a_{oc_2^1}$	1	0	0	0	0	0	0	0		
	$p_{oc_2^1}$	0	0	0	0	0	0	3	0		
$o_{c_2}^2$	$a'_{oc_2^2}$	1	0	0	0	0	0	0	0		
	$a_{oc_2^2}$	1	0	0	0	0	0	0	0		
	$p_{oc_2^2}$	0	0	0	0	0	0	2	0		
$o_{c_2}^3$	$a'_{oc_2^3}$	1	0	0	0	0	0	0	0		
	$a_{oc_2^3}$	1	0	0	0	0	0	0	0		
	$p_{oc_2^3}$	0	0	0	0	0	0	1	0		
$o_{c_3}^1$	$a'_{oc_3^1}$	1	0	0	0	0	0	0	0		
	$a_{oc_3^1}$	1	0	0	0	0	0	0	0		
	$p_{oc_3^1}$	0	0	0	0	0	0	0	3		
$o_{c_3}^2$	$a'_{oc_3^2}$	1	0	0	0	0	0	0	0		
	$a_{oc_3^2}$	1	0	0	0	0	0	0	0		
	$p_{oc_3^2}$	0	0	0	0	0	0	0	2		
$o_{c_3}^3$	$a'_{oc_3^3}$	1	0	0	0	0	0	0	0		
	$a_{oc_3^3}$	1	0	0	0	0	0	0	0		
	$p_{oc_3^3}$	0	0	0	0	0	0	0	1		
	A'	0	0	1	0	1	1	0	0	0	
	A	0	0	1	1	2	2	0	0	0	
	B	0	1	2	1	1	4	4	4	4	
	\bar{K}	0	1	1	1	1	4	4	4	4	

Fig. 5. Example of construction of TIK from an instance $B_3 \cap 3CNF$ with $E = (a \vee b \vee \neg c) \wedge (\neg a \vee \neg b \vee d) \wedge (a \vee c \vee b)$, where $X = \{a, b\}$, $Y = \{c\}$, $Z = \{d\}$ and the clauses are labeled from left to right.

Finally, O_3 will contain O_1 and all the items associated with Y not in O_2 . This makes the rest of the items selection for O_3 completely equivalent to variable assignment in Z for $B_3 \cap 3CNF$ (precisely, the standard reduction from 3-SAT to Subset Sum). Therefore, TIK is a Yes instance.

Next, suppose that TIK is a Yes instance. Certainly, an optimal O_1 must have exactly one of the items o_u and $o_{\bar{u}}$ for $u \in X$, otherwise, O_2 could interdict some o_u and $o_{\bar{u}}$, making the goal \bar{K} impossible to be achieved. As argued before, an optimal reaction O_2 to O_1 will select the items associated with X not in O_1 .

Assign 1 to $u \in X$ such that $o_u \in O_1$, and 0 otherwise. For any valid assignment of the variables in Y , the correspondence in TIK is the following: if $u \in Y$ is 1, add $o_{\bar{u}}$ to O_2 , otherwise add o_u . This forces O_3 to select for each $u \in Y$, o_u if u is 1 and $o_{\bar{u}}$ if u is 0; otherwise, the goal \bar{K} is not attained. Since, by hypothesis, TIK is a Yes instance, for those O_1 and O_2 , there is O_3 such that the profit \bar{K} is exactly achieved which implies that there is an assignment of Z such that E is satisfied. \square

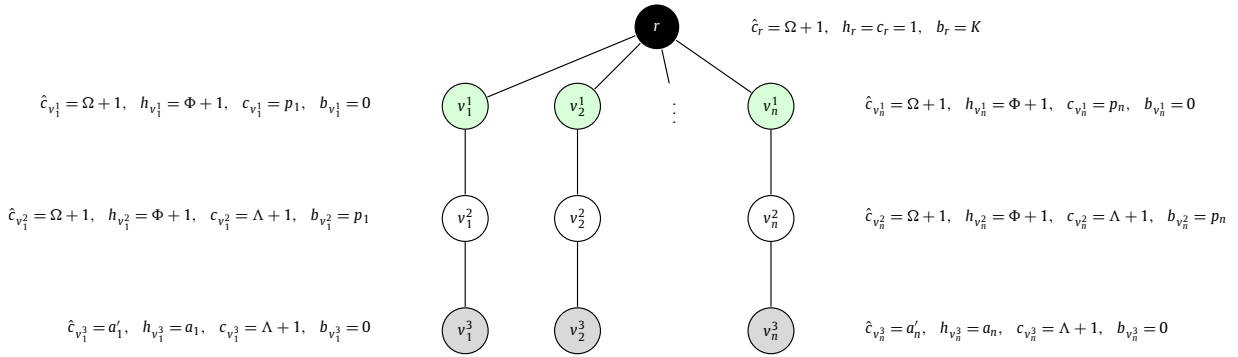


Fig. 6. Graph reduction from TIK to MCN_w when $O = \{1, 2, \dots, n\}$. The only vertices resulting in positive benefit are the ones in white. The vertices in gray can be vaccinated and directly attacked. The vertices in green can be protected. The vertex in black can be attacked (and protected). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Theorem 4.6. MCN_w is Σ_3^P -complete, even on trees.

Proof. MCN_w is clearly in Σ_3^P .

Next, from an instance of TIK, we construct the following instance of MCN_w :

- Let $\Omega = A'$, $\Phi = A + 1$, $\Lambda = B$ and $K = \bar{K}$.
- For each item $o \in O$ create three vertices v_o^1 , v_o^2 and v_o^3 with
 - $\hat{c}_{v_o^1} = \Omega + 1$, $h_{v_o^1} = \Phi + 1$, $c_{v_o^1} = p_o$ and $b_{v_o^1} = 0$; this vertex is only available for the protection set P ;
 - $\hat{c}_{v_o^2} = \Omega + 1$, $h_{v_o^2} = \Phi + 1$, $c_{v_o^2} = \Lambda + 1$ and $b_{v_o^2} = p_o$; this vertex cannot be vaccinated, directly infected or protected;
 - $\hat{c}_{v_o^3} = a_o'$, $h_{v_o^3} = a_o$, $c_{v_o^3} = \Lambda + 1$ and $b_{v_o^3} = 0$; this vertex is only available for the vaccination set D and for the direct infection set I ;
- Create a vertex r with $\hat{c}_r = \Phi + 1$, $h_r = 1$, $c_r = 1$ and $b_r = K$.
- For each item $o \in O$, add the edges (r, v_o^1) , (v_o^1, v_o^2) and (v_o^2, v_o^3) .

See Fig. 6 for an illustration of our reduction.

The key ingredients of this reduction are the following: (i) independently of the vaccination strategy, an optimal attack will always include the vertex r , (ii) hence, the only way to collect a positive benefit p_o is by ensuring that vertex v_o^2 is saved, (iii) the latter is only possible if v_o^3 is vaccinated and v_o^1 is protected or if v_o^3 is not attacked and v_o^1 is protected. These observations allow to show that TIK is a Yes instance if and only if MCN_w is a Yes instance. The remainder of the proof follows a similar reasoning to the previous proofs for the weighted games. \square

5. Directed graphs

In this section, we consider directed graphs $G = (V, A)$ and restrict costs and benefits to be unitary. We use the subscript dir for these problem versions. Clearly, these problems inherit the complexity of their unitary undirected versions, as they are more general. In fact, we were able to go a level up in the polynomial hierarchy for some of its subgames in comparison with the unitary undirected cases. In this section, we first prove that the $ATTACK_{dir}$ is NP-complete, and then demonstrate that $VACCINATION-ATTACK_{dir}$ is Σ_2^P -complete. Later, in Section 6, we present special properties of $PROTECT_{dir}$ that allow us to easily prove NP-completeness for directed acyclic graphs and polynomiality for arborescences.

It should be remarked that we do not address $ATTACK-PROTECT_{dir}$ and thus, it remains open whether it is Σ_2^P -complete. The difficulty on dealing with this subgame is related to the lack of Σ_2^P -hard problems involving unitary parameters or a division on the two players decision variables: in $ATTACK-PROTECT_{dir}$ all parameters are 1 and all vertices can be subject to infection or protection. On the other hand, as an example, non-trivial instances of BIK (presented in Section 4.2) should have weights not all 1, otherwise it becomes polynomially solvable as it can be reduced to its continuous version and, consequently, efficiently solved [67]. Another example, 2-CNF-ALTERNATING QUANTIFIED SATISFIABILITY, to be introduced in Section 5.2, and which is Σ_2^P -complete, demands each player to control distinct sets of variables. For $VACCINATION-ATTACK_{dir}$, we were able to bypass this challenge but an analogous trick does not seem easily adaptable for $ATTACK-PROTECT_{dir}$.

5.1. The $ATTACK_{dir}$ problem

First, we study the *Attack problem* on directed graphs, $ATTACK_{dir}$. We are given a directed graph G_a resulting from the deletion of the vaccinated vertices from the original graph, and an integer budget Φ . In this setting, there is no protection phase, i.e. $\Lambda = 0$. The decision version of the problem is:

ATTACK_{dir}:

INSTANCE: A directed graph $G_a = (V_a, A_a)$, a non-negative integer budget $\Phi \leq |V_a|$, and a non-negative integer K .

QUESTION: Is there a subset of vertices $I \subseteq V_a$, $|I| \leq \Phi$ such that the number of infected vertices in G_a is greater or equal to K ?

We saw that in the undirected case, this problem is solvable in linear time, the best strategy being to infect the Φ largest connected components of G_a . But in the directed case, the infection is only allowed to propagate itself according to the direction of the arcs, which makes the problem of choosing the right set of vertices to attack NP-complete. We will use a reduction from the 3-Satisfiability problem, which is one of the Karp's 21 NP-complete problems [64].

3-SATISFIABILITY (3-SAT):

INSTANCE: Set U of variables, Boolean expression E over U in conjunctive normal form with exactly 3 literals in each clause $c \in C$.

QUESTION: Is there a 0-1 assignment for the variables in U that satisfies E ?

Theorem 5.1. *ATTACK_{dir} is NP-complete, even on directed acyclic graphs.*

Proof. *ATTACK_{dir} ∈ NP* as, given a set of attacked vertices I , checking whether the set of infected vertices is greater than K is easily done using a DFS.

To prove that *ATTACK_{dir}* is NP-hard, we take an instance of 3-SAT. We build a directed acyclic graph G_a as follows:

- For each variable $u \in U$, we create two vertices v_u and $v_{\bar{u}}$, one for each possible 0-1 assignment of u . We call $V_U = \{v_u; u \in U\}$ and $V_{\bar{U}} = \{v_{\bar{u}}; u \in U\}$ the two sets of vertices of size $|U|$. For each variable u , we also create a directed path p_u of length $|C| + |U| - 1$, with an in-going arc from both v_u and $v_{\bar{u}}$ at the beginning of the path.
- For each clause $c \in C$, we create a vertex $v_c \in V_C$.
- From each vertex $v_u \in V_U$, we draw an arc (v_u, v_c) to every clause in which the positive literal u appears. Similarly, we draw an arc $(v_{\bar{u}}, v_c)$ from each $v_{\bar{u}} \in V_{\bar{U}}$ to every clause in which the negative literal $\neg u$ appears.

An example of this construction can be found in Fig. 7. We set $\Phi = |U|$, $K = |U| \times (|U| + |C|) + |C|$ and argue that answering *ATTACK_{dir}* on this instance is the same as answering 3-SAT.

Indeed, suppose that 3-SAT is a *Yes* instance, *i.e.* there is a 0-1 assignment to the variables in U such that every clause in E is true. Taking this assignment, by attacking v_u if u is set to be 1 and $v_{\bar{u}}$ otherwise, we attack exactly Φ vertices in G_a . Moreover, each path p_u is infected, and for each pair $(v_u, v_{\bar{u}})$, there is exactly one vertex infected due to the direction of the arcs. Finally, as E is true, each clause c is true, which translates into the fact that each v_c in the graph G_a is infected. Overall, there are exactly $|U| + |U| \times |p_u| + |C| = |U| \times (|U| + |C|) + |C|$ vertices infected in the graph.

Conversely, we prove that if *ATTACK_{dir}* is a *Yes* instance, *i.e.*, there is a feasible attack I^* on G_a leading to at least $K = |U| \times (|U| + |C|) + |C|$ vertices infected, then E is satisfiable and the corresponding 0-1 assignment can be read in I . Let I^* be such an attack strategy. First, we remark that the largest possible set of infected vertices should contain all the vertices V_{p_u} of each path p_u : it is possible to infect them all as $\Phi = |U|$ and due to their size equal to $|C| + |U| - 1$, we can prove that not infecting all of them results in a sub-optimal solution. Indeed, suppose that for one u' we do not infect any of the vertices $V_{p_{u'}}$ of the path $p_{u'}$. Let α^* be the maximum number of vertices we can infect without infecting $p_{u'}$. As $p_{u'}$ is not infected, $v_{u'}$ and $v_{\bar{u}'}$ cannot be either. Thus, an easy upper bound α_{up} on α^* is obtained by saying that every vertex of the graph is infected, except for the ones in $\{v_{u'}, v_{\bar{u}'}\} \cup V_{p_{u'}}$. Then,

$$\begin{aligned} \alpha^* &\leq \alpha_{up} = (|U| - 1) \times |p_u| + 2(|U| - 1) + |C| \\ &= (|U| - 1) \times (|U| + |C| - 1) + 2|U| - 2 + |C| \\ &= |U|^2 + |U| \times |C| - 2|U| - |C| + 1 + 2|U| - 2 + |C| \\ &= |U| \times (|U| + |C|) - 1. \end{aligned}$$

As we assumed that the optimal attack I^* infected at least $K = |U| \times (|U| + |C|) + |C|$ vertices, which is strictly greater than α_{up} , we proved that no strategy not infecting all the paths can infect K vertices.

Thus, as there is exactly Φ different paths, we should attack exactly one element in each set of vertices $\{v_u, v_{\bar{u}}\} \cup V_{p_u}$: if we attacked more than one, then the remaining budget would not allow to attack all the paths. As attacking v_u or $v_{\bar{u}}$ leads to a strictly greater number of infected vertices than infecting a vertex in p_u , there is no harm in assuming that no vertex inside the p_u is in I^* . This implies that $I^* \subset V_U \cup V_{\bar{U}}$. At this point, there are at least $|p_u| \times |U| + |U| = |U| \times (|U| + |C|)$ vertices infected. Since we supposed that we had a *Yes* instance to *ATTACK_{dir}*, there must be $K = |U| \times (|U| + |C|) + |C|$ infected vertices, which implies that all vertices in V_C are infected. Thus, 3-SAT is a *Yes* instance and I^* is a 0-1 assignment of U that makes E true, concluding the proof. \square

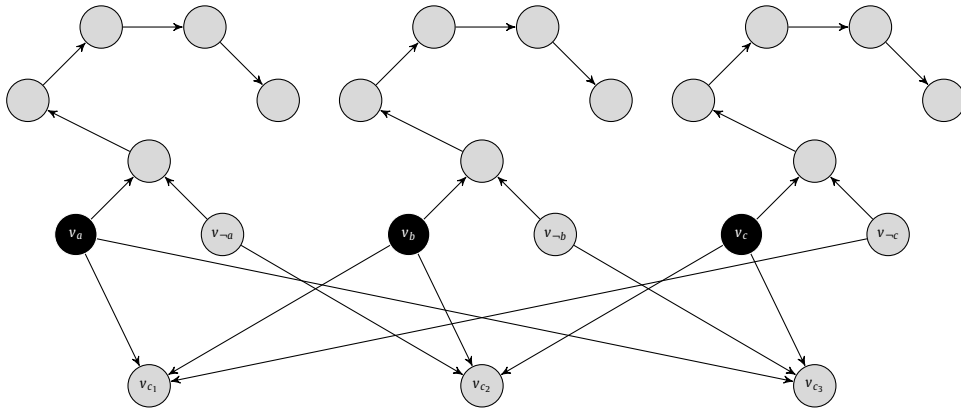


Fig. 7. Example of construction of G_a from the boolean expression in CNF with 3 literals in each clause $E = (a \vee b \vee \neg c) \wedge (\neg a \vee b \vee c) \wedge (a \vee \neg b \vee c)$. We have $U = \{a, b, c\}$ and $|C| = 3$. Taking $I = \{v_a, v_b, v_c\}$ is optimal.

Remark 5.2. Note that the proof of Theorem 5.1 holds if p_u is replaced by a complete graph with $|C| + |U| - 1$ vertices (the length of the path). This observation will be useful for the reduction used in $\text{VACCINATION-ATTACK}_{dir}$

5.2. The $\text{VACCINATION-ATTACK}_{dir}$ problem

Our demonstration of NP-completeness for ATTACK_{dir} inspires our proof for the Σ_2^P -completeness of $\text{VACCINATION-ATTACK}_{dir}$. The formulation of this decision problem is

VACCINATION-ATTACK_{dir}:
 INSTANCE: A graph $G = (V, A)$, two non-negative integer budgets Ω and Φ such that $\Omega + \Phi \leq |V|$ and a non-negative integer K .
 QUESTION: Is there a subset $D \subseteq V$, $|D| \leq \Omega$ such that $\forall I \subseteq V \setminus D$ with $|I| \leq \Phi$, the number of infected vertices $|V| - s(G, D, I, \emptyset) \leq K$?

We will use a reduction from a variant of the *2-Alternating Quantified Satisfiability problem* (B_2). Historically, B_2 was the first problem shown to be Σ_2^P -complete [52]. If the Boolean formula studied in B_2 is in DNF with 3 literals per clause, then the problem is still Σ_2^P -complete [53]. Thus, if we consider expressions in CNF with 3 literals per clause, instead of seeking to satisfy the Boolean formula, we should state the question as formulated in [48]:

2-CNF-ALTERNATING QUANTIFIED SATISFIABILITY (B_2^{CNF}):
 INSTANCE: Disjoint non-empty sets of variables X and Y , Boolean expression E over $U = X \cup Y$ in conjunctive normal form with exactly 3 literals in each clause.
 QUESTION: Is there a 0-1 assignment for X so that there is no 0-1 assignment for Y such that E is satisfied?

Theorem 5.3. $\text{VACCINATION-ATTACK}_{dir}$ is Σ_2^P -complete.

Proof. From the formulation in the form of $\exists D \forall I Q(D, I)$, we deduce that $\text{VACCINATION-ATTACK}_{dir} \in \Sigma_2^P$.

To show that it is Σ_2^P -hard, we take an instance of B_2^{CNF} . We build G in a similar fashion to how G_a was built in the proof of the Theorem 5.1, the main difference being the use of cliques instead of paths. However, to differentiate the variables in X from the ones in Y , we slightly change the construction:

- For each variable $x \in X$, we create two vertices v_x and $v_{\bar{x}}$, one for each possible 0-1 assignment of x . We call V_X and $V_{\bar{X}}$ the sets of v_x and $v_{\bar{x}}$. We also create two cliques k_x and $k_{\bar{x}}$ of $|C| + |Y| - 1$ vertices V_{k_x} and $V_{k_{\bar{x}}}$.
- For each variable $y \in Y$, we create two vertices v_y and $v_{\bar{y}}$, one for each possible 0-1 assignment of y . Let V_Y and $V_{\bar{Y}}$ be these two sets of vertices, and $V_U = V_X \cup V_Y$, $V_{\bar{U}} = V_{\bar{X}} \cup V_{\bar{Y}}$. We also create a clique k_y of size $|C| + |Y| - 1$.
- For each clause $c \in C$, we create a vertex $v_c \in V_C$.
- From each vertex $v_u \in V_U$, we draw an arc (v_u, v_c) to every clause in which the positive literal u appears. Similarly, we draw an arc $(v_{\bar{u}}, v_c)$ from each $v_{\bar{u}} \in V_{\bar{U}}$ to every clause in which the negative literal $\neg u$ appears.
- From every v_x , we draw an arc to one node in k_x , and do the same thing with $v_{\bar{x}}$ and $k_{\bar{x}}$. We also draw an undirected edge between each v_x and $v_{\bar{x}}$.
- Finally, from each v_y and each $v_{\bar{y}}$, we draw an arc to one node in k_y .

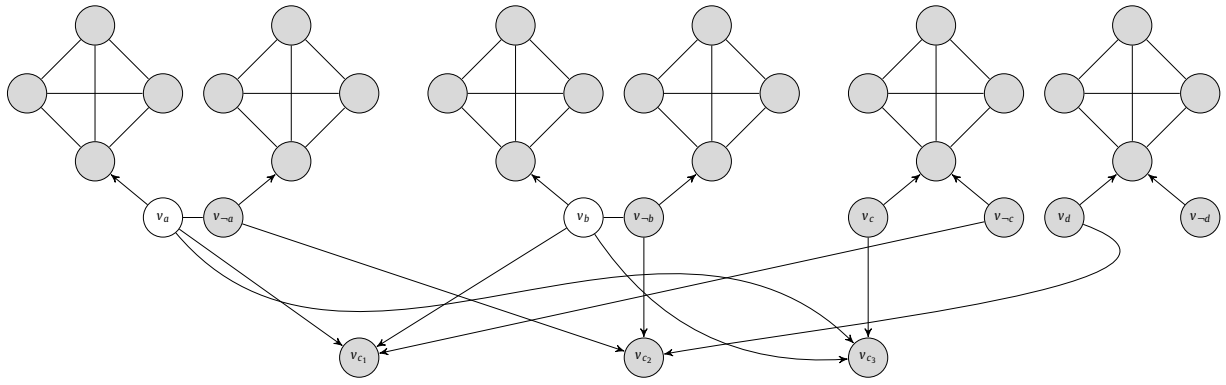


Fig. 8. Example of construction of G from the boolean expression in CNF with 3 literals in each clause $E = (a \vee b \vee \neg c) \wedge (\neg a \vee \neg b \vee d) \wedge (a \vee c \vee b)$. Here, $X = \{a, b\}$ and $Y = \{c, d\}$. Taking $D = \{v_a, v_b\}$, i.e., obliging both a and b to be *False* makes it impossible to satisfy E .

An example of this construction can be found in Fig. 8. We set $\Omega = |X|$, $\Phi = |X| + |Y|$, $K = (|X| + |Y|) \times (|Y| + |C|) + |C| - 1$ and argue that answering $\text{VACCINATION-ATTACK}_{dir}$ on this instance is the same as answering B_2^{CNF} .

Indeed, if we are given a solution to a *Yes* instance of B_2^{CNF} , then by vaccinating the vertices corresponding to the opposite of the 0-1 assignment of X , we oblige the attacker to infect the vertices corresponding to the truth values for X . From there, by following the same reasoning as before, it is easy to see that the *Yes* instance of B_2^{CNF} leads to a *Yes* instance of $\text{VACCINATION-ATTACK}_{dir}$, i.e. the attacker cannot infect more than K vertices.

Conversely, we show that a set D^* corresponding to a solution of a *Yes* instance of $\text{VACCINATION-ATTACK}_{dir}$ is a solution to a *Yes* instance of B_2^{CNF} . The first thing to notice is that given that the vaccination budget is $\Omega = |X|$, that the size of the cliques k_x and $k_{\bar{x}}$ is equal to $|C| + |Y| - 1$ and that each clique can be disconnected from the graph by spending only one unit of vaccination budget, we necessarily have that the best vaccination strategy $D^* \subset \bigcup_{x \in X} \{v_x, v_{\bar{x}}\}$. Next, we show that the defender would be worse off if she decides to vaccinate both $v_{x'}$ and $v_{\bar{x}'}$ for some $x' \in X$ instead of vaccinating exactly one of each member of $\{v_x, v_{\bar{x}}\}$. In the best case scenario, in addition to the vertices already vaccinated, deciding to vaccinate the two members of a pair will allow her to protect $|C| - 1$ nodes in V_C (it is not possible to remove all the arcs between the $V_U \cup V_{\bar{U}}$ and the V_C as we suppose that $Y \neq \emptyset$, thus at least one clause contains a variable from Y). But by doing so, as $\Omega = |X|$, the defender will also not protect at all a group of vertices $\{v_{x''}, v_{\bar{x}''}\} \cup V_{k_{x''}} \cup V_{k_{\bar{x}''}}$. Thus, the attacker can then spend only one unit of her own budget to attack all of this group, a quantity of infected vertices that otherwise would have been obtained by spending two units of his budget Φ . Thus, defending the two members of $\{v_{x'}, v_{\bar{x}'}\}$ spared one unit of budget for the attacker, which she can then use to attack one of the disconnected cliques of size $|C| + |Y| - 1 > |C| - 1$. Thus, making such a move for the defender is strictly worse than not doing it and D^* contains exactly one vertex from each $\{v_x, v_{\bar{x}}\}$.

After this stage, it is easy to see that the best move for the attacker is to attack all of the $D^* \setminus (V_x \cup V_{\bar{x}})$, and for the variables in Y , the situation reduces to the one we already discussed with ATTACK_{dir} (note that it is always more interesting for the attacker to spend her budget on attacking the v_y and $v_{\bar{y}}$ than the disconnected cliques as it will always infect more vertices). Hence, in the end, if the attacker did not manage to infect strictly more than $(|Y| + |X|) \times (|Y| + |C|) + |C| - 1$ vertices, it means that at least one clause is false, which concludes the proof. \square

Corollary 5.4. MCN_{dir} is Σ_2^P -hard.

6. PROTECTION: tractability limits

In this section, we will concentrate on optimal protection strategies given I (directly infected vertices). Without loss of generality, in what follows, we restrict our attention to the induced graph obtained by considering only non-saved vertices when there is no protection (the remaining are already saved, even without no protected vertices).

The motivation to provide a closer look to the protection problem in the unitary cases (undirected and directed graphs) is based on the fact that their NP-hardness was established for split graphs, while for the weighted case it was proven even for trivial graphs. Such results do not clarify the problem complexity for trees, or even graphs of bounded treewidth, neither for directed acyclic graphs (DAGs), polytrees and arborescences. Frequently, NP-complete problems on graphs become polynomially solvable on such graph classes. In section 6.1, we describe a dynamic programming approach for trees to determine the optimal protection solution in polynomial time. We also connect our problem complexity with the results in monadic second-order logic for tree-decomposable graphs [68,43]. In section 6.2, we describe the problem properties for DAGs, making it simple to show that PROTECTION_{dir} is NP-complete. We conclude this section by showing that the optimal protection strategy can be determined in polynomial time for arborescences.

6.1. PROTECT over trees

We next focus on PROTECT; recall that it is the case of undirected graphs with unitary costs and benefits. Results for a special version of this problem where only one vertex is infected, aka the FIREFIGHTER PROBLEM, already exist in the literature. The recent work of [43] establishes that for this special version of PROTECT, the decision version of the problem can be solved in linear time over graphs of bounded treewidth, through the use of a reformulation in *Extended Monadic Second Order* (EMSO) logic. We first extend this result to the case of an arbitrary number of infected vertices to show that PROTECT is solvable in polynomial time over graphs of bounded treewidth.

Lemma 6.1. PROTECT can be solved in polynomial time over graphs with constant bounded treewidth.

Proof. The key factor is to reformulate our problem in terms of an MSO-formula φ based on set variables, which captures the graph structure of the problem, and an evaluation relation ψ over a set of integer variables, which captures the “number” aspect of the problem. In order to do so, we will define as in [43] two sets P and X where P is the set of protected vertices that separate the infected vertices of set I from the saved vertices which are not protected X . Apart from the classic universal quantifier and logical connectives, we need to make use of a binary relation $adj(x, y)$ to assess the adjacency of two vertices x and $y \in V$: $adj(x, y)$ true if $(x, y) \in E$, and false otherwise. The definitions of φ and ψ for PROTECT are the following:

$$\begin{aligned} \varphi &= (\forall v(v \in I \Rightarrow (v \notin P) \wedge (v \notin X))) \wedge (\forall x(x \in P) \Rightarrow (x \notin X)) \wedge \\ &(\forall x \forall y((x \in X) \wedge (adj(x, y)) \wedge (y \notin P)) \Rightarrow (y \in X)). \\ \psi &= (|P| \leq \Lambda) \wedge (|X| + |P| \geq K), \end{aligned}$$

where the aim is to save at least K vertices. Through the above expression for φ , it is established that the sets P and X have an empty intersection and that any neighbor of a vertex in X must be either in P or X itself, i.e., the set P is a separator for the sets I and X . Since the above definitions respect the limitations of MSO logic formulations, a theorem due to [69] implies that the problem can be solved in $O(f(w) \cdot |V|)$ where f is a function of the treewidth w of the graph. Consequently, we can conclude that PROTECT can be solved in polynomial time for graphs whose treewidth is bounded by a constant. \square

Even though the above theorem is powerful from a theoretical perspective, it is of little practical use, as underlined in [68]. Indeed, the function $f(w)$ in the worst case complexity formula grows extremely fast with w and the algorithm suffers from space problems in practical implementations. Therefore, *Dynamic Programming* (DP) is often used to provide more efficient algorithms. In this section we propose a DP algorithm to solve the optimization problem associated with PROTECT on trees. We consider a recursion scheme that works with growing subtrees, starting from the leaves and climbing up to the root vertex, solving the optimization problem on each subtree recursively and merging them as needed at each step of the recursion. This recursion scheme bears similarities with the scheme proposed in [10] for the pairwise CNDP over trees. In the following, we label a vertex as *attacked* when it has been directly infected by the attacker while the term *infected* is used both for directly attacked vertices and indirectly infected ones, and *secondary infected* vertices is used only for vertices indirectly infected after the initial attack.

For further analysis, we denote by \mathcal{T}_a the subtree of tree \mathcal{T} rooted at vertex $a \in V$, and by a_i with $i \in \{1, \dots, s\}$ the children of a . We define as $\mathcal{T}_{a_i \rightarrow s}$ the subtree constituted by $\{a\} \cup_{j=i, \dots, s} \mathcal{T}_{a_j}$. An example of a tree \mathcal{T} rooted at vertex a is depicted in Fig. 9 where subtree \mathcal{T}_{a_2} is represented by diamond shaped vertices while subtree $\mathcal{T}_{a_3 \rightarrow 4}$ is represented by round shaped vertices. All recursions in our dynamic programming approaches are based on traversing the tree in postorder (i.e. from the leaves to the root) and from the right part of each tree level to the left one. For example in Fig. 9, once the recursive functions are computed and saved for subtrees \mathcal{T}_{a_3} and \mathcal{T}_{a_4} , we will compute the recursion functions associated to $\mathcal{T}_{a_3 \rightarrow 4}$ by merging the results for both subtrees in both situations when a is vaccinated, infected, protected or neither of these possibilities. We consider that tree \mathcal{T} is rooted at vertex r .

We introduce the following recursion functions:

$$\begin{aligned} F_a(c, m, \sigma) &:= \text{maximum number of saved vertices in subtree } \mathcal{T}_a \text{ when } c \text{ vertices have been protected, } m \text{ unprotected} \\ &\text{vertices in } \mathcal{T}_a \text{ are linked to } a \text{ by an unprotected but non-infected path (including } a \text{ itself) and } \sigma = 1 \text{ if an} \\ &\text{attacked vertex in } \mathcal{T}_a \text{ is linked to } a \text{ by an unprotected path (including } a \text{) and } \sigma = 0 \text{ otherwise.} \\ G_{a_i}(c, m, \sigma) &:= \text{maximum number of saved vertices in subtree } \mathcal{T}_{a_i \rightarrow s} \text{ when } c \text{ vertices have been protected, } m \text{ unprotected} \\ &\text{vertices in } \mathcal{T}_{a_i \rightarrow s} \text{ are linked to } a \text{ by an unprotected but non-infected path (including } a \text{ itself) and } \sigma = 1 \text{ if an} \\ &\text{attacked vertex in } \mathcal{T}_{a_i \rightarrow s} \text{ is linked to } a \text{ by an unprotected path (including } a \text{) and } \sigma = 0 \text{ otherwise.} \end{aligned}$$

Using the previously described functions, we can define the following recursions. The initial conditions for each leaf vertex a and rightmost subtree \mathcal{T}_{a_s} are as follows:

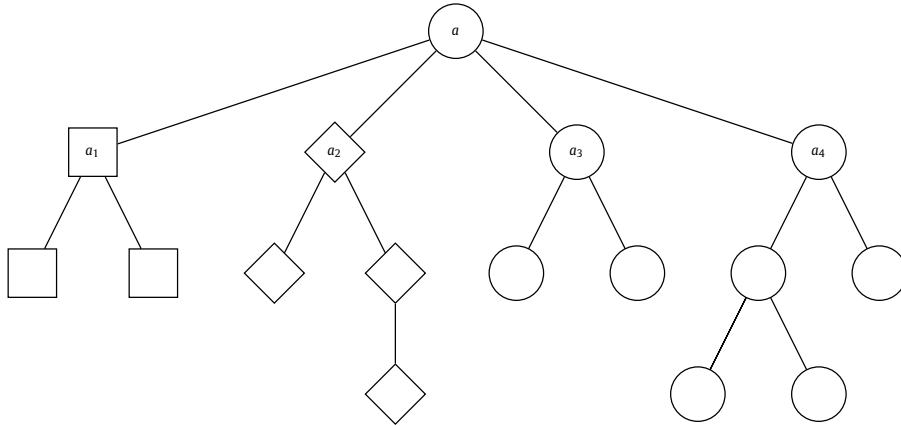


Fig. 9. Example of a tree with subtree \mathcal{T}_{a_2} represented by diamond shaped vertices and subtree $\mathcal{T}_{a_3 \rightarrow 4}$ represented by round shaped vertices.

$$F_a(c, m, \sigma) = \begin{cases} 0 & \text{if } (c = 0 \wedge m = 0 \wedge \sigma = 1) \\ 1 & \text{if } (c = 1 \wedge m = 0 \wedge \sigma = 0) \vee (c = 0 \wedge m = 1 \wedge \sigma = 0); \\ -\infty & \text{otherwise (i.e., infeasible configurations)} \end{cases} \quad (5)$$

$$G_{a_s}(c, m, \sigma) = \begin{cases} \max \{ F_{a_s}(c - 1, m', \sigma') + m'(1 - \sigma') : m' = 0, \dots, |\mathcal{T}_{a_s}|; \sigma' = 0, 1 \} \\ \quad \text{if } a \text{ is protected } (m = \sigma = 0) \\ \max \{ F_{a_s}(c, m', \sigma') + m'(1 - \sigma')(1 - \sigma) : m' = 0, \dots, |\mathcal{T}_{a_s}|; \sigma' = 0, 1 \} \\ \quad \text{if } a \text{ is vaccinated } (m = \sigma = 0) \text{ or attacked } (m = 0, \sigma = 1) \\ F_{a_s}(c, m - 1, \sigma) \\ \quad \text{if } a \text{ is neither vaccinated, protected or attacked.} \end{cases} \quad (6)$$

In Eq. (6), the first case deals with a protected a vertex and all m' unprotected vertices below a_s are saved if a_s is not linked to an attacked vertex inside \mathcal{T}_{a_s} ($\sigma' = 0$). The second case deals with either a vaccinated or attacked a vertex so that the budget c needs not be updated going from \mathcal{T}_{a_s} to $\mathcal{T}_{a_s} \cup \{a\}$. The last case deals with an unattacked and unprotected a vertex and parameter m is incremented as the subtree is enlarged by vertex a .

The following equations handle the general case, for vertices which are neither leaf vertices or the root of rightmost subtrees:

$$F_a(c, m, \sigma) = G_{a_i}(c, m, \sigma) \quad \text{for a non-leaf vertex } a \in V. \quad (7)$$

For each non-leaf vertex $a \in V$ and $i < s$: if a is attacked

$$G_{a_i}(c, 0, 1) = \max \{ F_{a_i}(c', m', \sigma') + G_{a_{i+1}}(c - c', 0, 1) : c' = 0, \dots, |\mathcal{T}_{a_i}|; m' = 0, \dots, |\mathcal{T}_{a_i}|; \sigma' = 0, 1 \}, \quad (8a)$$

if a is protected (either from vaccination or protection)

$$G_{a_i}(c, 0, 0) = \max \{ F_{a_i}(c', m', \sigma') + G_{a_{i+1}}(c - c', 0, 0) + m'(1 - \sigma') : c' = 0, \dots, |\mathcal{T}_{a_i}|; m' = 0, \dots, |\mathcal{T}_{a_i}|; \sigma' = 0, 1 \}, \quad (8b)$$

otherwise, if a is neither protected nor infected

$$G_{a_i}(c, m, \sigma) = \max \{ F_{a_i}(c', m', \sigma') + G_{a_{i+1}}(c - c', m - m', \sigma'') + m'(1 - \sigma)\delta_{ar} : c' = 0, \dots, |\mathcal{T}_{a_i}|; m' = 0, \dots, |\mathcal{T}_{a_i}|; \sigma', \sigma'' = 0, 1; \sigma = \max\{\sigma', \sigma''\} \}. \quad (8c)$$

Equation (8a) focuses on the case where vertex a is infected. In this case, no additional vertex is saved as all vertices below a which were not infected in \mathcal{T}_a and with an unprotected path to a will be infected themselves, therefore the total number of saved vertices is the sum of already saved vertices from the two merged subtrees. Equation (8b) regards the case of a protected a vertex, either through earlier vaccination or through protection. In this case, the vertices under a_i who were unprotected, linked to a_i by an unprotected path and who are not in contact with an infected vertex through an unprotected path are confirmed saved and added to the cost function, additionally to the already saved vertices from both subtrees. Finally, Equation (8c) deals with the last case where a is neither infected nor protected in any way. In this case,

the cost of the objective function is updated by the number of unprotected vertices linked to a_i by an unprotected path, but only in the case that a is the root vertex r ($\delta_{ar} = 1$ if $a = r$ and 0 otherwise) and a is not linked to an infected vertex through an unprotected path ($\sigma = 0$). Otherwise, we cannot ensure that the unprotected vertices below a will be saved in the optimal solution.

The optimal value for the problem is given by the quantity

$$\max \{F_r(c, m, \sigma) : c = 0, \dots, \Lambda; m = 0, \dots, n; \sigma = 0, 1\}$$

where r is the root vertex of the tree, since it represents the maximum number of saved vertices for each protection budget and the solution can be recovered by backtracking. Considering the proposed dynamic program, we can state the following proposition.

Theorem 6.2. PROTECT over trees admits a polynomial time algorithm with time complexity $O(n^5)$.

Proof. The number of functions $F_a(\cdot)$ and $G_a(\cdot)$ to compute for each value of c , m and σ is bounded by $2n^2$. The recursion steps involved in Equation (8c) are bounded by $2n^2$ operations at most. Considering all vertices n , the running time of the dynamic programming algorithm is thus bounded by $O(n^5)$. \square

Since the lower level of the problem over trees is polynomial, the MCN over trees cannot be Σ_3^P -hard. Following a classic trick for DP algorithm, see e.g. [10], a similar algorithm can be devised when vertices have protection costs and unit benefits, which remains polynomial. When both types of weights are integer, the algorithms become pseudo-polynomial and the problem becomes weakly NP-hard.

6.2. PROTECTION_{dir} over directed acyclic graphs

We will show that an optimal protection strategy can be restricted to *candidate* vertices.

Definition 6.1. In a directed graph $G = (V, A)$, a vertex $v \in V \setminus I$ that can be reached from a vertex of I by a directed path and whose isolated protection results in a maximal set of saved vertices, is called candidate. Denote by \mathcal{C} the set of candidate vertices.

In other words, a candidate vertex v has no predecessor whose protection implies saving v . See Fig. 10 for an illustration on popular graph sub-classes of DAGs. In the case of Fig. 10a, $\mathcal{C} = \{1, 2, 3, 9\}$; e.g., vertex 5 is not a candidate, since its protection saves vertices $\{6, 7, 8\}$, but this is also guaranteed by saving vertex 2 instead, resulting in the maximal set of saved vertices $\{2, 3, 4, 5, 6, 7, 8\}$. In Fig. 10b, protecting vertex 1 suffices to save all the remaining non attacked vertices. Finally, in Fig. 10c, the successors of the attacked vertices are exactly the set of candidates.

Lemma 6.3. Let $G = (V, A)$ be a directed graph. Given I and Λ , there is an optimal protection strategy $P \subseteq \mathcal{C}$.

Proof. Let $P \subseteq V \setminus I$ be an optimal protection strategy such that exists $v \in P \setminus \mathcal{C}$. Then, by the definition of candidate, there is a vertex $u \in \mathcal{C}$ whose isolated protection implies saving v , as well as, all the vertices that v alone was saving. Hence, a feasible protection strategy can be obtained by removing v from P and adding u to P : note that either the used budget is maintained, if $u \notin P$, or decreased, if $u \in P$. Let this strategy be denoted by $\tilde{P} = (P - \{v\}) \cup \{u\}$.

By contradiction, suppose that \tilde{P} is not optimal: there is some vertex r that was saved in P but not in \tilde{P} . In fact, we can conclude that under P , r was saved due to v being saved (protected) and possibly due to some other vertices in $P \setminus \{v\} \subseteq \tilde{P}$. However, under \tilde{P} , v is also saved, as well, as the vertices in $P \setminus \{v\}$. Consequently, r is saved in \tilde{P} , resulting in a contradiction. \square

Furthermore, we can compute the *value* of candidate vertices.

Definition 6.2. For each $v \in \mathcal{C}$, the value of v is denoted by p_v and it corresponds to the number of saved vertices if v is the only protected vertex.

In the example of Fig. 10a, $p_1 = 1$, $p_2 = 6$, $p_3 = 1$ and $p_9 = 1$. However, note that this analysis does not make the problem trivial: in Fig. 10a, if $\Lambda = 2$, the optimal protection cannot be computed in a greedy way, i.e., protecting vertices 1 and 2 is not optimal; the only optimal solution is to protect vertices 1 and 3.

Theorem 6.4. PROTECT_{dir} is NP-complete, even for directed acyclic graphs.

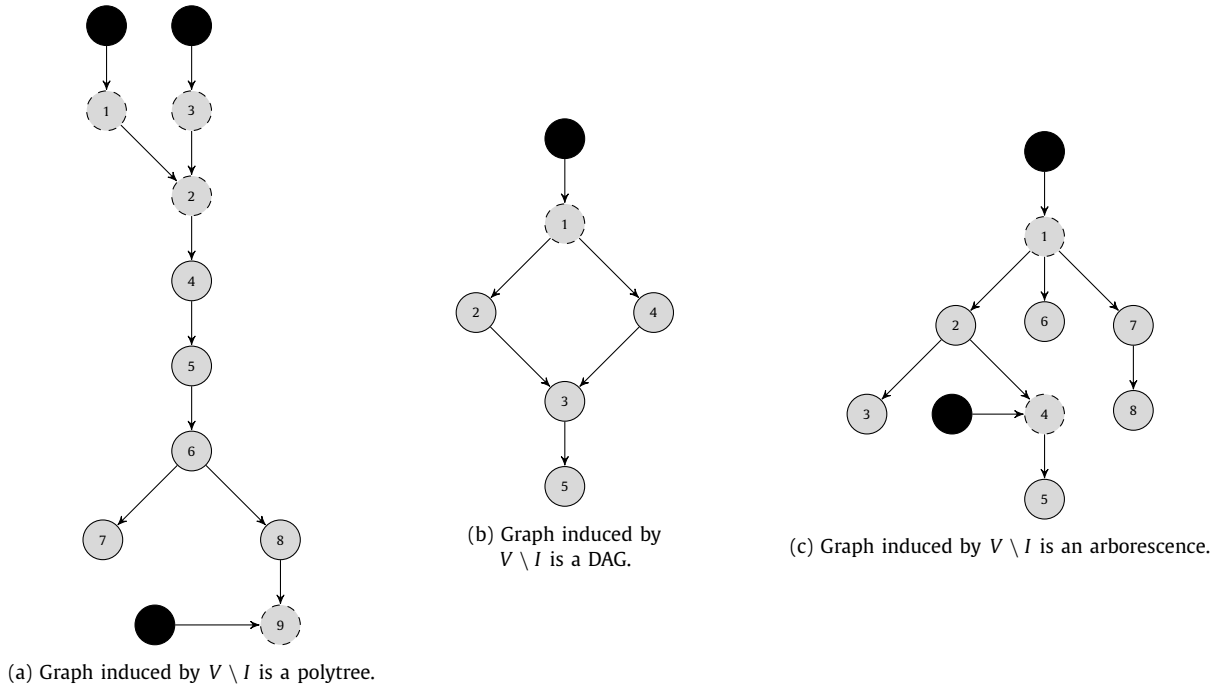


Fig. 10. The set I is represented by black vertices and candidate vertices are dashed.

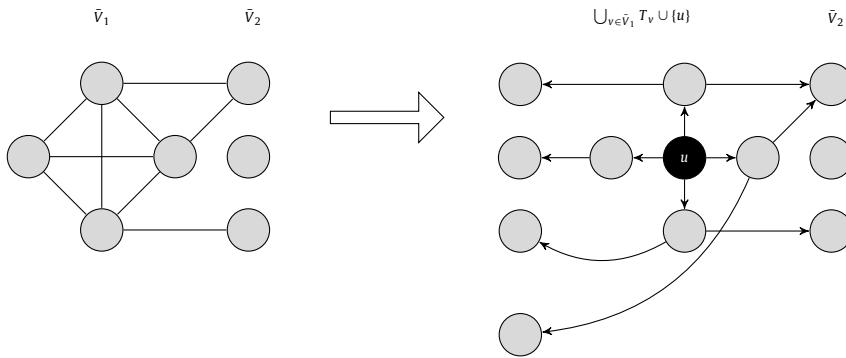


Fig. 11. Example of construction of G from \tilde{G} .

Proof. The statement of PROTECT_{dir} is exactly the one of PROTECT in Section 3, except that the graph is directed. For sake of simplicity, we drop the subscript a from G_a .

The problem is clearly in NP as given the protection P , the number of infected can be determined in polynomial time through a DFS.

Next, we reduce CNP_{split} to PROTECT_{dir} , showing its NP-hardness. Given an instance of CNP_{split} , we build the following graph $G = (V, A)$:

- For each $v \in \tilde{V}_1$, we create the set of vertices $T_v = \{t_v^1, t_v^2\}$ in G , and the arc (t_v^1, t_v^2) .
- For each $v \in \tilde{V}_2$, we replicate it in G , and for each edge $(r, v) \in \tilde{E}$ with $v \in \tilde{V}_2$, the arc (t_r^1, v) is added in G .
- Finally, we add the only attacked vertex u to G and connect it with each t_v^1 for $v \in \tilde{V}_1$, through the arc (u, t_v^1) .

To complete the reduction it remains to set $\Lambda = B$ and $K = \lfloor 2 + \sqrt{8\bar{K} + 1} \rfloor$ (obtained by solving $\bar{K} = \binom{K-1}{2}$). See Fig. 11 for an illustration of the reduction.

First, note that C of G is $\{t_v^1 : v \in \tilde{V}_1\} \cup \tilde{V}_2$, where the vertices in the first set have value at least 2, and the ones in the second have value 1. Hence, it is clear that the best protection strategy will prioritize the vertices t_v^1 . In fact, we can argue that only those vertices can be in an optimal protection strategy. If $\Lambda = B \geq \tilde{V}_1$, then the instance of CNP_{split} is trivial. Therefore, we can assume $\Lambda = B < \tilde{V}_1$ and thus, it holds $P^* \subset \{t_v^1 : v \in \tilde{V}_1\}$. Consequently, choosing the optimal P^* means

to minimize the vertices in T_v , for $v \in \bar{V}_1$, and in \bar{V}_2 that are connected to u . By construction, those vertices connected with u correspond to a connected component \bar{C}_1 in \bar{G} . Thus, P^* minimizes the size of $\bigcup_{v \in \bar{C}_1} \{t_v^2\} \cup \{u\} \cup \bar{C}_1$. The remaining of the proof follows an analogous reasoning to the proof of Theorem 3.3. \square

6.2.1. Arborescence

In this section we restrict the protection problem to the case where the graph induced by $V \setminus I$ is an arborescence.

Definition 6.3. A directed acyclic graph $G = (V, A)$ is an arborescence if its underlying undirected graph is a tree and there is a single vertex (root) that has a unique directed path from it to all other vertices.

In arborescence, the determination of \mathcal{C} is straightforward. Since all vertices in $V \setminus I$ have in-degree 1, either they are protected by their predecessor, and thus are not a candidate, or they are direct successors of vertices in I . Therefore, \mathcal{C} is the set of all successors of vertices in I . For an illustration see Fig. 10c, where the vertices in $\mathcal{C} = \{1, 4\}$ have in-degree 1. We can prove that in this case a greedy approach leads to optimality.

Lemma 6.5. Given $G = (V, A)$, I and Δ , if the graph induced by $V \setminus I$ is an arborescence, then an optimal protection can be determined in polynomial time, specifically, $O(|V| \log(|V|))$. Moreover, if the induced graph is a set of arborescences, the result also holds.

Proof. We start by showing that a greedy procedure runs in time $O(|V| \log(|V|))$.

As previously observed, for arborescences, the set of candidate vertices is easy to compute: it is the set of all successors of I .

Next, the calculation of p_v for each $v \in \mathcal{C}$ can be performed through a depth-first-search that records the saved vertices by candidates. This requires $O(|V|)$ since the graph is an arborescence.

Finally, the Δ candidate vertices of largest values are protected. This requires to order the vertices accordingly with $\{p_v\}_{v \in \mathcal{C}}$. Thus, the greedy method runs in $O(|V| \log(|V|))$.

Next, we show that the described method provides an optimal protection. Let P be the obtained protection through the greedy method. The key idea to prove the optimality of P is essentially due to the fact that in an arborescence, \mathcal{C} is simply the set of all successors of I , otherwise, if we have a vertex of in-degree at least 2, we do not have an arborescence. Thus, the protection strategy P cannot imply the protection of some candidate not in P . This shows the optimality of P . \square

Note that in trees (undirected graphs), it does not hold that \mathcal{C} is the set of successors of the vertices in I . Hence, Lemma 6.5 does not extend to the undirected case.

Remark 6.6. Note that in Lemmata 6.3 and 6.5, we did not use the fact that $b_v = 1 \forall v \in V$. Thus, it also holds when vertices' benefits are not unitary.

CRedit authorship contribution statement

Adel Nabli: Conceptualization, Methodology, Writing.

Margarida Carvalho: Conceptualization, Methodology, Writing, Supervision, Funding acquisition.

Pierre Hosteins: Conceptualization, Methodology, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the Institut de Valorisation des Données and Fonds de recherche du Québec through the FRQ-IVADO Research Chair in Data Science for Combinatorial Game Theory, and the Natural Sciences and Engineering Research Council of Canada through the discovery grant 2019-04557.

References

- [1] A. Baggio, M. Carvalho, A. Lodi, A. Tramontani, Multilevel approaches for the critical node problem, *Oper. Res.* 69 (2021) 486–508.
- [2] A. Arulsevan, C.W. Commander, L. Eleftheriadou, P.M. Pardalos, Detecting critical nodes in sparse graphs, *Comput. Oper. Res.* 36 (7) (2009) 2193–2200, <https://doi.org/10.1016/j.cor.2008.08.016>.
- [3] J. He, H. Liang, H. Yuan, Controlling infection by blocking nodes and links simultaneously, in: N. Chen, E. Elkind, E. Koutsoupias (Eds.), *Internet and Network Economics*, Springer, Berlin, Heidelberg, 2011, pp. 206–217.

- [4] V. Boginski, C.W. Commander, Identifying Critical Nodes in Protein-Protein Interaction Networks, Clustering Challenges in Biological Networks, World Scientific, 2009, pp. 153–167.
- [5] V. Tomaino, A. Arulseelan, P. Veltri, P.M. Pardalos, Studying Connectivity Properties in Human Protein-Protein Interaction Network in Cancer Pathway, Springer US, Boston, MA, 2012, pp. 187–197.
- [6] M. Lalou, M.A. Tahraoui, H. Kheddouci, The critical node detection problem in networks: a survey, *Comput. Sci. Rev.* 28 (2018) 92–117.
- [7] J.M. Lewis, M. Yannakakis, The node-deletion problem for hereditary properties is NP-complete, *J. Comput. Syst. Sci.* 20 (2) (1980) 219–230.
- [8] S. Shen, J.C. Smith, R. Goli, Exact interdiction models and algorithms for disconnecting networks via node deletions, *Discrete Optim.* 9 (3) (2012) 172–188.
- [9] B. Addis, M. Di Summa, A. Grosso, Identifying critical nodes in undirected graphs: complexity results and polynomial algorithms for the case of bounded treewidth, *Discrete Appl. Math.* 161 (16) (2013) 2349–2360.
- [10] M. Di Summa, A. Grosso, M. Locatelli, Complexity of the critical node problem over trees, *Comput. Oper. Res.* 38 (2011) 1766–1774.
- [11] M. Lalou, M.A. Tahraoui, H. Kheddouci, Component-cardinality-constrained critical node problem in graphs, *Discrete Appl. Math.* 210 (2016) 150–163, <https://doi.org/10.1016/j.dam.2015.01.043>.
- [12] R. Aringhieri, A. Grosso, P. Hosteins, R. Scatamacchia, Polynomial and pseudo-polynomial time algorithms for different classes of the Distance Critical Node Problem, *Discrete Appl. Math.* 253 (2019) 103–121.
- [13] P. Hosteins, R. Scatamacchia, The Stochastic Critical Node Problem over trees, *Networks* 76 (3) (2020) 381–401.
- [14] M. Lalou, H. Kheddouci, A polynomial-time algorithm for finding critical nodes in bipartite permutation graphs, *Optim. Lett.* 13 (2019) 1345–1364.
- [15] K. Pavlikov, Improved formulations for minimum connectivity network interdiction problems, *Comput. Oper. Res.* 97 (2018) 48–57.
- [16] G. Karakose, R. McGarvey, Optimal detection of critical nodes: improvements to model structure and performance, *Netw. Spat. Econ.* 19 (2019) 1–26.
- [17] A. Veremyev, O.A. Prokopyev, E.L. Pasiliao, An integer programming framework for critical elements detection in graphs, *J. Comb. Optim.* 28 (2014) 233–273.
- [18] M. Di Summa, A. Grosso, M. Locatelli, Branch and cut algorithms for detecting critical nodes in undirected graphs, *Comput. Optim. Appl.* 53 (2012) 649–680.
- [19] R. Aringhieri, A. Grosso, P. Hosteins, R. Scatamacchia, A general evolutionary framework for different classes of Critical Node Problems, *Eng. Appl. Artif. Intell.* 55 (2016) 128–145, <https://doi.org/10.1016/j.engappai.2016.06.010>.
- [20] Y. Zhou, J. Hao, F. Glover, Memetic search for identifying critical nodes in sparse graphs, *IEEE Trans. Cybern.* 49 (10) (2019) 3699–3712.
- [21] Y. Zhou, J.-K. Hao, Z.-H. Fu, Z. Wang, X. Lai, Variable population memetic search: a case study on the critical node problem, *IEEE Trans. Evol. Comput.* 25 (1) (2021) 187–200.
- [22] R. Aringhieri, A. Grosso, P. Hosteins, R. Scatamacchia, Local search metaheuristics for the critical node problem, *Networks* 67 (3) (2016) 209–221, <https://doi.org/10.1002/net.21671>.
- [23] B. Addis, R. Aringhieri, A. Grosso, P. Hosteins, Hybrid constructive heuristics for the critical node problem, *Ann. Oper. Res.* 238 (1) (2016) 637–649, <https://doi.org/10.1007/s10479-016-2110-y>.
- [24] W. Pullan, Heuristic identification of critical nodes in sparse real-world graphs, *J. Heuristics* 21 (5) (2015) 577–598.
- [25] D. Purevsuren, G. Cui, M. Chu, N.N.H. Win, Hybridization of grasp with exterior path relinking for identifying critical nodes in graphs, *IAENG Int. J. Comput. Sci.* 44 (2017).
- [26] M. Ventresca, D. Aleman, A derandomized approximation algorithm for the critical node detection problem, *Comput. Oper. Res.* 43 (2014) 261–270.
- [27] M. Ventresca, D. Aleman, A region growing algorithm for detecting critical nodes, in: COCOA'14, 2014, pp. 593–602.
- [28] M. Ventresca, D. Aleman, A randomized algorithm with local search for containment of pandemic disease spread, *Comput. Oper. Res.* 48 (2014) 11–19, <https://doi.org/10.1016/j.cor.2014.02.003>.
- [29] J. Rezaei, F. Zare-Mirakabad, S.A. MirHassani, S.-A. Marashi, EIA-CNDP: an exact iterative algorithm for critical node detection problem, *Comput. Oper. Res.* 127 (2021) 105138, <https://doi.org/10.1016/j.cor.2020.105138>.
- [30] M. Ventresca, K.R. Harrison, B.M. Ombuki-Berman, The bi-objective critical node detection problem, *Eur. J. Oper. Res.* 265 (3) (2018) 895–908.
- [31] F. Furini, I. Ljubic, E. Malaguti, P. Paronuzzi, Casting light on the hidden bilevel combinatorial structure of the k-vertex separator problem, in: OR-19-6, DEI, 2019, University of Bologna.
- [32] R. Wood, Deterministic network interdiction, *Math. Comput. Model.* 17 (2) (1993) 1–18.
- [33] E. Israeli, R.K. Wood, Shortest-path network interdiction, *Networks* 40 (2) (2002) 97–111, <https://doi.org/10.1002/net.10039>.
- [34] S. DeNegre, Interdiction and discrete bilevel linear programming, Ph.D. thesis, Lehigh University, 2011.
- [35] A. Caprara, M. Carvalho, A. Lodi, G.J. Woeginger, A study on the computational complexity of the bilevel knapsack problem, *SIAM J. Optim.* 24 (2014) 823–838.
- [36] V. Rutenburg, Propositional truth maintenance systems: classification and complexity analysis, *Ann. Math. Artif. Intell.* 10 (1994) 207–231, <https://doi.org/10.1016/j.ejor.2019.02.028>.
- [37] F. Furini, I. Ljubic, S. Martin, P. San Segundo, The maximum clique interdiction problem, *Eur. J. Oper. Res.* 277 (2019) 112–127, <https://doi.org/10.1016/j.ejor.2019.02.028>.
- [38] F. Della Croce, R. Scatamacchia, An exact approach for the bilevel knapsack problem with interdiction constraints and extensions, *Math. Program.* (2020) 1–33.
- [39] M. Fischetti, I. Ljubić, M. Monaci, M. Sinnl, Interdiction games and monotonicity, with application to knapsack problems, *INFORMS J. Comput.* 31 (2) (2019) 390–410, <https://doi.org/10.1287/ijoc.2018.0831>.
- [40] K. Tanınmış, M. Sinnl, A branch-and-cut algorithm for submodular interdiction games, arXiv:2103.15788, 2021.
- [41] M. Fischetti, M. Monaci, M. Sinnl, A dynamic reformulation heuristic for generalized interdiction problems, *Eur. J. Oper. Res.* 267 (1) (2018) 40–51, <https://doi.org/10.1016/j.ejor.2017.11.043>.
- [42] S. Finbow, A. King, G. MacGillivray, R. Rizzi, The firefighter problem for graphs of maximum degree three, *Discrete Math.* 307 (16) (2007) 2094–2105.
- [43] K.D. Barnetson, A.C. Burgess, J. Enright, J. Howell, D.A. Pike, B. Ryan, The firebreak problem, *Networks* 77 (2021) 372–382.
- [44] G. Brown, M. Carlyle, J. Salmerón, R. Wood, Defending critical infrastructure, *Interfaces* 36 (2006) 530–544, <https://doi.org/10.1287/inte.1060.0252>.
- [45] M. Ventresca, Network robustness versus multi-strategy sequential attack, *J. Complex Netw.* 3 (1) (2015) 126–146.
- [46] F. Enayaty-Ahangar, C.E. Rainwater, T.C. Sharkey, A logic-based decomposition approach for multi-period network interdiction models, *Omega* 87 (2019) 71–85.
- [47] P. Martin, Tri-level optimization models to defend critical infrastructure, Master's thesis, Naval Postgraduate School, Monterey, 2007.
- [48] B. Johannes, New classes of complete problems for the second level of the polynomial hierarchy, Ph.D. thesis, Technischen Universität, Berlin, 2011.
- [49] C. Blair, The computational complexity of multi-level linear programs, *Ann. Oper. Res.* 34 (1) (1992) 13–19, <https://doi.org/10.1007/BF02098170>.
- [50] T. Dudás, B. Klinz, G.J. Woeginger, The computational complexity of multi-level bottleneck programming problems, in: A. Migdalas, P.M. Pardalos, P. Várbrand (Eds.), *Multilevel Optimization: Algorithms and Application*, Springer US, Boston, MA, 1998, pp. 165–179.
- [51] R.G. Jeroslow, The polynomial hierarchy and a simple model for competitive analysis, *Math. Program.* 32 (2) (1985) 146–164, <https://doi.org/10.1007/BF01586088>.
- [52] A.R. Meyer, L.J. Stockmeyer, The equivalence problem for regular expressions with squaring requires exponential space, in: 13th Annual Symposium on Switching and Automata Theory, SWAT 1972, 1972, pp. 125–129.

- [53] C. Wrathall, Complete sets and the polynomial-time hierarchy, *Theor. Comput. Sci.* 3 (1) (1976) 23–33.
- [54] M. Schaefer, C. Umans, Completeness in the polynomial-time hierarchy a compendium, *SIGACT News* 33 (01 2002).
- [55] F. Mogavero, G. Perelli, Binding forms in first-order logic, in: S. Kreutzer (Ed.), 24th EACSL Annual Conference on Computer Science Logic, CSL 2015, in: *Leibniz International Proceedings in Informatics (LIPIcs)*, vol. 41, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2015, pp. 648–665.
- [56] Y.-D. Shen, T. Eiter, Evaluating epistemic negation in answer set programming, *Artif. Intell.* 237 (2016) 115–135.
- [57] E. Malizia, More complexity results about reasoning over (m)cp-nets, in: *Proc. of the 17th International Conference on Autonomous Agents and Multi-agent Systems, AAMAS 2018*, 2018.
- [58] L. Lozano, J.C. Smith, A backward sampling framework for interdiction problems with fortification, *INFORMS J. Comput.* 29 (2017) 123–139.
- [59] M. Carvalho, X. Klimentova, K. Glorie, A. Viana, M. Constantino, Robust models for the kidney exchange problem, *INFORMS J. Comput.* 33 (3) (2021) 861–881, <https://doi.org/10.1287/ijoc.2020.0986>.
- [60] D. Blom, C. Hojny, B. Smeulders, A benders-type approach for robust optimization of kidney exchanges under full recourse, arXiv:2105.08565, 2021.
- [61] F. Barbosa, A. Agra, A. de Sousa, The minimum cost network upgrade problem with maximum robustness to multiple node failures, *Comput. Oper. Res.* 136 (2021) 105453, <https://doi.org/10.1016/j.cor.2021.105453>.
- [62] A. Nabli, M. Carvalho, Curriculum learning for multilevel budgeted combinatorial problems, in: H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, H. Lin (Eds.), *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 7044–7056.
- [63] M.R. Garey, D.S. Johnson, *Computers and Intractability, a Guide to the Theory of NP-Completeness*, W. H. Freeman & Co., USA, 1990.
- [64] R.M. Karp, *Reducibility Among Combinatorial Problems*, Springer US, Boston, MA, 1972, pp. 85–103.
- [65] L.J. Stockmeyer, A.R. Meyer, Word problems requiring exponential time (preliminary report), in: *Proceedings of the Fifth Annual ACM Symposium on Theory of Computing, STOC '73*, Association for Computing Machinery, New York, NY, USA, 1973, pp. 1–9.
- [66] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, 3rd edition, The MIT Press, 2009.
- [67] M. Carvalho, A. Lodi, P. Marcotte, A polynomial algorithm for a continuous bilevel knapsack problem, *Oper. Res. Lett.* 46 (2) (2018) 185–188, <https://doi.org/10.1016/j.orl.2017.12.009>.
- [68] A. Langer, F. Reidl, P. Rossmanith, S. Sikdar, Practical algorithms for mso model-checking on tree-decomposable graphs, *Comput. Sci. Rev.* 13 (C) (2014) 39–74, <https://doi.org/10.1016/j.cosrev.2014.08.001>.
- [69] S. Arnborg, J. Lagergren, D. Seese, Easy problems for tree-decomposable graphs, *J. Algorithms* 12 (2) (1991) 308–340.