



**HAL**  
open science

# Complexity of the Multilvel Critical Node Problem

Adel Nabli, Margarida Carvalho, Pierre Hosteins

► **To cite this version:**

Adel Nabli, Margarida Carvalho, Pierre Hosteins. Complexity of the Multilvel Critical Node Problem. Journal of Computer and System Sciences, 2022, 127, pp122-145. 10.1016/j.jcss.2022.02.004 . hal-03894879

**HAL Id: hal-03894879**

**<https://hal.science/hal-03894879v1>**

Submitted on 12 Dec 2022

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

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

---

# COMPLEXITY OF THE MULTILEVEL CRITICAL NODE PROBLEM

---

**Adel Nabli**      **Margarida Carvalho**

CIRRELT and Département d'Informatique et de Recherche Opérationnelle  
 Université de Montréal  
 adel.nabli@umontreal.ca  
 carvalho@iro.umontreal.ca

**Pierre Hosteins**

Évaluation des Systèmes de Transports Automatisés et de leur Sécurité (ESTAS)  
 Université Gustave Eiffel  
 pierre.hosteins@ifsttar.fr

## 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 or intractability of those problems vary. Our findings contribute with new NP-complete,  $\Sigma_2^p$ -complete and  $\Sigma_3^p$ -complete problems.

## 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) [5]. 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  $\Omega$  and a 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 *infect* subject to a budget limit  $\Phi$  and a 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  $\Lambda$  and a cost  $\{c_v\}_{v \in V}$ . Infected vertices propagate the infection to their neighbourhood, except to vaccinated or protected vertices. 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. We assume that all parameters of the problem are non-negative integers. The game description can be succinctly given by the following trilevel program:

$$(MCN) \quad \max_{z \in \{0,1\}^{|V|}} \min_{y \in \{0,1\}^{|V|}} \max_{\alpha \in [0,1]^{|V|}} \sum_{v \in V} b_v \alpha_v \quad (1a)$$

$$\sum_{v \in V} \hat{c}_v z_v \leq \Omega \quad \sum_{v \in V} h_v y_v \leq \Phi$$

$$\sum_{v \in V} c_v x_v \leq \Lambda \quad (1b)$$

$$\alpha_v \leq 1 + z_v - y_v \quad \forall v \in V \quad (1c)$$

$$\alpha_v \leq \alpha_u + x_v + z_v \quad \forall (u, v) \in A, \quad (1d)$$

where  $D = \{v \in V : z_v = 1\}$ ,  $I = \{v \in V : y_v = 1\}$  and  $P = \{v \in V : x_v = 1\}$ . See [5] for further details on this mathematical programming formulation.

**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 [5]. Furthermore, it contributes to the understanding of sequential combinatorial games within the polynomial hierarchy and it motivates the focus on potentially  $\Omega(2^{2^{|\mathcal{V}|}})$  algorithms, heuristic methods or novel solution definitions. Table 1 summarizes our results; unitary cases assume that all costs and benefits are 1.

DECISION VERSIONS	Undirected Graphs		Directed Graphs	
	UNITARY CASE <i>Section 3</i>	WEIGHTED CASE <i>Section 4</i>	UNITARY CASE <i>Sections 5 &amp; 6</i>	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] $\Sigma_2^p$ -complete	[13] NP-hard	[18] $\Sigma_2^p$ -complete
VACCINATION-ATTACK	[4] NP-complete	[9] $\Sigma_2^p$ -complete	[14] $\Sigma_2^p$ -complete	[19] $\Sigma_2^p$ -complete
MCN	[5] NP-hard	[10] $\Sigma_3^p$ -complete	[15] $\Sigma_2^p$ -hard	[20] $\Sigma_3^p$ -complete

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].

**Paper Organization.** In Section 2, we revise the literature associated with MCN, allowing to position our contribution in the context of critical node 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, either on directed or 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 [4, 21] to computational biology [9, 39].

**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.* [28] for a recent survey. Indeed, many of these belong to the class of problems called *Node-Deletion Problems* [28]. They consist in deleting the smallest subset of vertices from a graph so that the induced subgraph satisfies a certain property  $\pi$ . Lewis and Yannakakis [30] have shown that if  $\pi$  is *nontrivial* and *hereditary*, then the subsequent node 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, has been shown to be NP-hard in the strong sense thanks to this argument [36]. Moreover, some CNDP problems remain NP-hard even on particular graph classes [1, 28]. For example, the original *Critical Node Problem* (CNP) [4] 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 [1]. Several works try 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 [15] 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*) [36], the largest pairwise connectivity among all components [27] or an extension of pairwise connectivity based on the length of shortest paths in the remaining graph [2]. A stochastic version of the pairwise CNDP with node

attack failure was studied over trees in [22] and found to be strongly NP-hard, even with unit connection costs. The CNDP has also been studied on other specially structured graphs, such as series-parallel graphs [36, 2], graphs with bounded treewidth [1], proper interval graphs [27] or bipartite permutation graphs [26].

**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 [4]. Even though in this particular case, it has been shown that the multi-objective formulation is not equivalent to the original one [40], splitting the objective in two is sometimes possible. For example, Furini *et al.* [18] 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*). Whereas some interdiction games such as the *network flow interdiction* are NP-complete [41], others such as the *binary knapsack interdiction problem* [14, 11] or the *maximum clique interdiction game* [19] have been shown to be  $\Sigma_2^p$ -complete, shading light on the intrinsic relationship between this class of problems and the second level of the polynomial hierarchy.

However, the unitary undirected version of MCN, as originally introduced by Baggio *et al.* [5], 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 [17] and was studied more recently in [6] 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.* [10] to study the defense of critical infrastructure against malicious attacks.

**Defender-Attacker-Defender.** Although the general DAD has been claimed to be NP-hard in [32], complexity results for trilevel combinatorial optimization problems are scarce. In [24], a new proof that *Trilevel Linear Programming* is  $\Sigma_2^p$ -hard is provided, building upon the results in [8, 16, 23] showing that the *Multilevel Linear Programming* problem with  $L + 1$  levels is  $\Sigma_L^p$ -hard. In fact, the decision version of MCN problem can be formulated as "given 3 integer budgets  $\Omega, \Phi, \Lambda$ , 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  $\Sigma_3^p$ -complete by Meyer, Stockmeyer and Wrathall [33, 42], making one expect the MCN to be complete for this class.

We stress that very few problems have been shown to be *naturally*  $\Sigma_3^p$ -complete in the literature up to now, in addition to infinite families of problems which have been shown to be  $\Sigma_L^p$  complete for any level  $L$  of the polynomial hierarchy (as, e.g., Satisfiability Problems, or Multi-level Linear Programming). The compendium of [35], whose last update dates back to 2008, describes 8 tri-level 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  $\Sigma_3^p$ -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 [34], deciding whether a propositional program has epistemic FLP (Faber, Leone and Pfeifer) answer sets [37], or checking the existence of max optimal outcomes over  $m$ CP-nets to study the aggregation of preferences over combinatorial domains in artificial intelligence [31]. To the best of our knowledge, and barring a few additional contributions which might have escaped us, there is but approximately a dozen proven natural  $\Sigma_3^p$ -complete problems in the literature, which makes it all the harder to derive  $\Sigma_3^p$ -completeness for a given tri-level problem. In this work, we add two more problems to the list of  $\Sigma_3^p$ -complete problems, i.e., a Knapsack Interdiction Problem and the Multi-level Critical Node Problem. Even though the set of proven  $\Sigma_2^p$ -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 below several new  $\Sigma_2^p$ -complete problems to this list.

### 3 Undirected graphs: the unitary case

In this section, we focus on undirected graphs  $G = (V, E)$ , *i.e.*, for each couples 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 undirected 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:

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

To ease our analysis, guided by the relationship between *Critical Node Detection Problems* and *Node-Deletion Problems*, we first write the immediate **Property 3.1** stating that *vaccinating* or *protecting* vertices has almost 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 resulting from 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  $\Lambda$  vertices  $P \subseteq V_a \setminus I$  that will maximise  $s(G_a, \emptyset, I, P)$ . 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:

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

We will show that finding such a  $P$  is NP-complete. We argue that it is a direct consequence of the results of [1] 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)* [4] is a related problem to ours. The setting is very similar to PROTECTION: 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:

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

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<sub>split</sub>:**

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 [1] 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.** [1] *CNP<sub>split</sub> is NP-hard.*

### 3.1.2 Complexity result

Next, we show that the decision version of PROTECT is NP-complete using a reduction from  $\text{CNP}_{split}$ . 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 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  $\text{CNP}_{split}$ . 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  $\bar{K} = \binom{K-1}{2}$ ). An example of such construction can be found in **Figure 1**.

Then, as there is only one attacked vertex, minimizing (3) on this instance of PROTECT corresponds to choose 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  $V_1 = \bar{V}_1 \cup \{u\}$ . Thus, we have that  $C_1 = \bar{C}_1 \cup \{u\}$  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  $\bar{G}$ . This finishes the proof that PROTECT is NP-hard.  $\square$

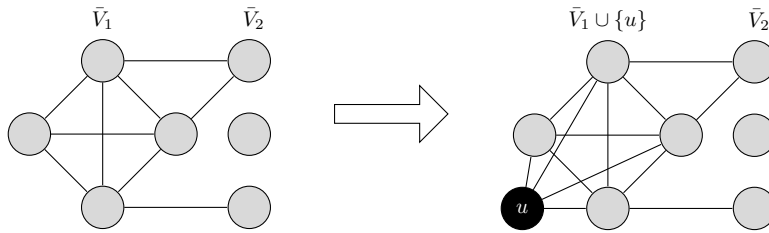


Figure 1: Example of construction of  $G_a$  from  $\bar{G}$

### 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  $\Phi, \Lambda$  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 [20], whose decision version is:

**DOMINATING SET:**

INSTANCE: A graph  $\bar{G} = (\bar{V}, \bar{E})$ , a positive integer  $B \leq |\bar{V}|$

QUESTION: Is there a subset  $U \subseteq \bar{V}$ ,  $|U| \leq B$ , such that  $\forall v \in \bar{V} \setminus U$ ,  $\exists u \in U$  such that  $(u, v) \in \bar{E}$ ?

**Theorem 3.4.** *ATTACK-PROTECT is NP-hard.*

*Proof.* Let us take a graph  $\bar{G} = (\bar{V}, \bar{E})$  and a positive integer  $B \leq |\bar{V}|$ . The instance of ATTACK-PROTECT is simply created by taking  $G_a = \bar{G}$ ,  $\Phi = B$ ,  $\Lambda = |\bar{V}| - \Phi - 1$  and  $K = \Phi + 1$ . In this configuration, we have a protection budget  $\Lambda$  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  $\Omega$  and  $\Phi$  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 3.5.** *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  $\Phi$  is to infect one vertex in each of the  $\Phi$  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  $\Phi$  largest connected components in the induced subgraph. Let's call this problem MINMAX $\Phi$ C:

**MINMAX $\Phi$ C:**

INSTANCE: A graph  $G = (V, E)$ , two non-negative integer budgets  $\Omega$  and  $\Phi$  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  $\Phi$  largest connected component in  $G[V \setminus D]$  is less than  $K$ ?

**Lemma 3.6.** *VACCINATION-ATTACK and MINMAX $\Phi$ C are equivalent problems.*

As Shen *et al.* [36] 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 $\Phi$ C is also NP-hard, which leads to the following corollaries:

**Corollary 3.7.** *VACCINATION-ATTACK is NP-complete.*

**Corollary 3.8.** *MCN is NP-hard.*

*Proof.* Given an instance of VACCINATION-ATTACK, there is a corresponding instance of MCN by taking the same  $G, \Omega, \Phi, K$  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,  $\text{MCN}_w$ , as well as 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  $v$ . Having introduced costs and benefits, this problem is thus intimately related to *Knapsack problems*, which we will use to demonstrate all of our complexity results in this part. First, we will highlight the direct relationship between  $\text{ATTACK}_w$  and  $\text{KNAPSACK}$ , which will get us the NP-completeness of this problem. Then, we will focus on the two bilevel sub-problems  $\text{VACCINATION-ATTACK}_w$  and  $\text{ATTACK-PROTECT}_w$  and prove they are  $\Sigma_2^p$ -complete thanks to a *Knapsack Interdiction problem*.

Note that the NP-completeness of  $\text{PROTECT}_w$  is immediate from the previous section. Hence, we start with complexity results to  $\text{ATTACK}_w$ , then  $\text{ATTACK-PROTECT}_w$ , and  $\text{VACCINATION-ATTACK}_w$ , and, to conclude,  $\text{MCN}_w$ . 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 $\text{ATTACK}_w$ 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  $\Phi$ , and as there is no protection 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:

**ATTACK<sub>w</sub>:**

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  $\Phi$ , 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  $K$ ?

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 [25]:

**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<sub>w</sub> is strictly equivalent to KNAPSACK.*

*Proof.* It is easy to see that  $\text{ATTACK}_w \in \text{NP}$ . Indeed, 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 values associated with each infected vertex to verify that it is greater or equal to  $K$ .

First we prove that each instance of  $\text{ATTACK}_w$  reduces to an instance of  $\text{KNAPSACK}$ . Given an instance of  $\text{ATTACK}_w$ , it is straightforward to see that it is sufficient to infect the node  $v$  with lowest infection cost  $h_v$  of a given maximal connected component to infect the whole component and collect the reward  $b$  of each node inside said component. If  $\text{Comp}(G_a)$  represents the set of maximal connected components of  $G_a$ , to each connected component  $C \in \text{Comp}(G_a)$  we can then 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  $\text{KNAPSACK}$  instance where each connected component of  $G_a$  is mapped to an object of the knapsack, with the same total profit and cost. If  $\Phi = B$  and  $K = \bar{K}$ , we have built an equivalent  $\text{KNAPSACK}$  instance for any instance of  $\text{ATTACK}_w$ .

Conversely, if we start from an instance of  $\text{KNAPSACK}$ , we construct an instance of  $\text{ATTACK}_w$  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  $\text{ATTACK}_w$  on undirected graphs and  $\text{KNAPSACK}$  can be reduced to each other, both problems are equivalent.  $\square$

**Corollary 4.2.** *ATTACK<sub>w</sub> 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  $\text{ATTACK}_w$  which has no edge,  $\text{ATTACK}_w$  is NP-complete on trivial graphs.  $\square$

#### 4.2 The $\text{ATTACK-PROTECT}_w$ 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 of the knapsack. That way, the complexity results we devise also hold for trees.

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

##### **ATTACK-PROTECT<sub>w</sub>:**

INSTANCE: A graph  $G_a = (V_a, E_a)$ , a non-negative integer  $K$ , two non-negative integer budgets  $\Phi$  and  $\Lambda$ ,  $\forall v \in V_a$  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  $\text{ATTACK-PROTECT}_w$  is  $\Sigma_2^p$ -complete, we use the *Bilevel Interdiction Knapsack Problem* introduced by DeNegre [14] and proven to be  $\Sigma_2^p$ -complete in [11]. 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 its profit, subject to a constraint capping the maximum profit obtainable for 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.**  $\text{ATTACK-PROTECT}_w$  is strongly  $\Sigma_2^p$ -complete, even if the graph is a tree.

*Proof.* First,  $\text{ATTACK-PROTECT}_w$  is in  $\Sigma_2^p$  since this decision problem is exactly of the form  $\exists I \forall P Q(I, P)$ .

Next, we prove the problem  $\Sigma_2^p$ -hardness. Let us begin by noting that we can restrict the instances of KIP to the ones where  $\bar{K}$  and  $B$  are strictly inferior to  $\sum_{o \in O} p_o$ , otherwise, KIP is trivial to solve. This remark is used in the second part of this proof.

Starting from an instance of BIK, we construct an instance of  $\text{ATTACK-PROTECT}_w$  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 [Figure 2](#). Finally, we set  $\Phi = A + 1$ ,  $\Lambda = B$  and  $K = \bar{K}$ .

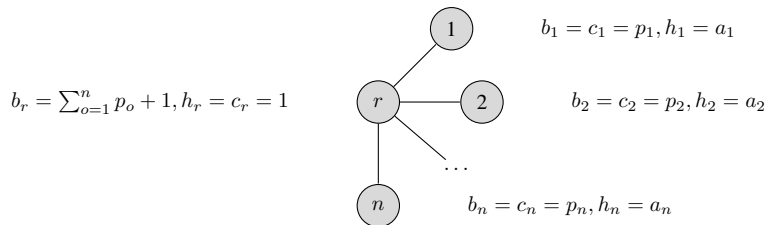


Figure 2: Graph reduction from BIK to  $\text{ATTACK-PROTECT}_w$  when  $O = \{1, 2, \dots, n\}$ .

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  $\text{ATTACK-PROTECT}_w$ , the attacker can select the subset of vertices  $I = \{r\} \cup \{v_o : o \in O_l\}$  with a feasible attacking cost

$\sum_{v \in I} h_v = 1 + \sum_{o \in O_I} a_o \leq A + 1 = \Phi$ . Now, the defender can only protect vertices in  $\{v_o : o \notin O_I\}$  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  $\Lambda$ . 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  $\text{ATTACK-PROTECT}_w$  is a *Yes* instance.

Now suppose that  $\text{ATTACK-PROTECTION}_w$  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  $\text{ATTACK-PROTECTION}_w$  *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  $\sum_{v \in P} c_v = \sum_{o: v_o \in P} p_o \leq \Lambda = B$ , the total benefit for the defender is  $\sum_{v \in P} b_v = \sum_{o: v_o \in P} p_o \leq \bar{K} - 1$ . Consequently, BIK is also a *Yes* instance.

This completes the proof that  $\text{ATTACK-PROTECT}_w$  is  $\Sigma_2^p$ -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  $\text{ATTACK-PROTECT}$  subgame. Since a star graph is a tree, the result stated in the theorem holds.  $\square$

### 4.3 The VACCINATION-ATTACK<sub>w</sub> problem

Using a similar reduction to the one in the proof of [Theorem 4.3](#), we show that the  $\text{VACCINATION-ATTACK}_w$  on weighted graphs is  $\Sigma_2^p$ -complete. As in the unitary case, this is equivalent to studying  $\text{MCN}_w$  problems where we set  $\Lambda = 0$ . The decision version of the problem is:

#### VACCINATION-ATTACK<sub>w</sub>:

INSTANCE: A graph  $G = (V, E)$ , a non-negative integer  $K$ , two non-negative integer budgets  $\Omega$  and  $\Phi$ ,  $\forall v \in V$  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.**  $\text{VACCINATION-ATTACK}_w$  is strongly  $\Sigma_2^p$ -complete, even if the graph is a tree.

*Proof.* As before,  $\text{VACCINATION-ATTACK}_w$  is in  $\Sigma_2^p$  since this decision problem is exactly of the form  $\exists D \forall I Q(D, I)$ .

Now, we establish the problem  $\Sigma_2^p$ -hardness. We start from an instance of BIK, defined in the previous section, and we then construct an instance of  $\text{VACCINATION-ATTACK}_w$  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 [Figure 3](#). 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  $K = \bar{K}$ .

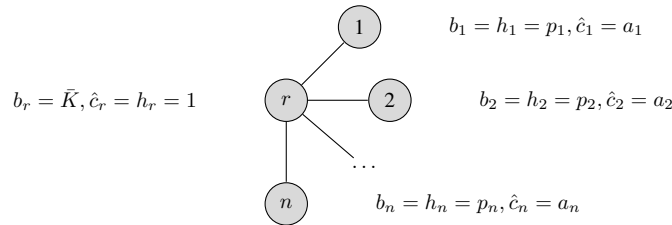


Figure 3: Graph reduction from BIK to  $\text{VACCINATION-ATTACK}_w$  when  $O = \{1, 2, \dots, n\}$ .

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  $\Sigma_2^p$ -complete even on trees.  $\square$

**Corollary 4.5.**  $\text{MCN}$  is  $\Sigma_2^p$ -hard, even if the graph is a tree.

#### 4.4 The $\text{MCN}_w$ problem

In this section we show that the decision problem  $\text{MCN}_w$  is  $\Sigma_3^p$ -complete.

**$\text{MCN}_w$ :**

INSTANCE: A graph  $G = (V, E)$ , a non-negative integer  $K$ , three non-negative integer budgets  $\Omega$ ,  $\Phi$  and  $\Lambda$ ,  $\forall v \in V$  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 ( $B_3 \cap 3\text{CNF}$ ), known to be  $\Sigma_3^p$ -complete problem [38, 42], in order to prove that the generalization of BIK to a trilevel, the *Trilevel Interdiction Knapsack* (TIK), is  $\Sigma_3^p$ -complete. Then, TIK is used to demonstrate that  $\text{MCN}_w$  is  $\Sigma_3^p$ -complete.

**3-ALTERNATING QUANTIFIED SATISFIABILITY ( $B_3 \cap 3\text{CNF}$ ):**

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 a 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 4.6.** *TIK is  $\Sigma_3^p$ -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  $\Sigma_3^p$ .

Next, we use a reduction from the  $B_3 \cap 3\text{CNF}$  which is very much in line with the reduction from 3-SAT to Subset Sum presented in [13, 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 Figure 4 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  $u \in Y$ .

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  $B_3 \cap CNF$  (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.

□

**Theorem 4.7.**  $MCN_w$  is  $\Sigma_3^P$ -complete, even on trees.

*Proof.*  $MCN_w$  is clearly in  $\Sigma_3^P$ .

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  $b_r = K$ .

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

Figure 4: 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.

- 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 Figure 5 for an illustration of our reduction.

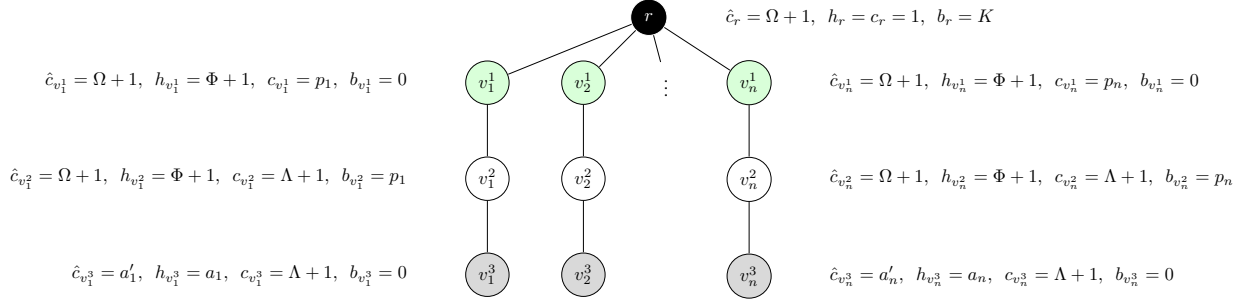


Figure 5: Graph reduction from TIK to  $MCN_w$  when  $O = \{1, 2, \dots, n\}$ . The only vertices resulting in positive benefit are the ones in white. The vertices in gray can be vaccinated and directly attacked. The vertices in green can be protected. The vertex in black can be attacked (and protected).

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  $\Sigma_2^P$ -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  $\Sigma_2^P$ -complete. The difficulty on dealing with this subgame is related to the lack of  $\Sigma_2^P$ -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 KIP (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 [12]. Another example, 2-CNF-ALTERNATING QUANTIFIED SATISFIABILITY, to be introduced in Section 5.2, and which is  $\Sigma_2^P$ -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  $\Phi$ . 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  $\Phi$  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 [25].

### 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.**  $\text{ATTACK}_{dir}$  is NP-complete, even on directed acyclic graphs.

*Proof.*  $\text{ATTACK}_{dir} \in \text{NP}$  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  $\text{ATTACK}_{dir}$  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 [Figure 6](#). We set  $\Phi = |U|$ ,  $K = |U| \times (|U| + |C|) + |C|$  and argue that answering  $\text{ATTACK}_{dir}$  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  $\Phi$  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  $\text{ATTACK}_{dir}$  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  $\alpha^*$  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  $\alpha_{up}$  on  $\alpha^*$  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  $\alpha_{up}$ , we proved that no strategy not infecting all the paths can infect  $K$  vertices.

Thus, as there is exactly  $\Phi$  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  $\text{ATTACK}_{dir}$ , 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.  $\square$

**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  $\text{VACCINATION-ATTACK}_{dir}$ .

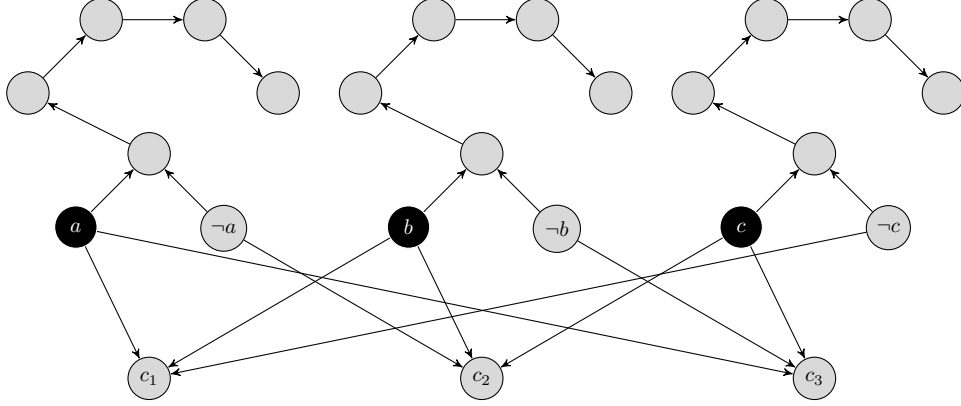


Figure 6: Example of construction of  $G_a$  from the boolean expression in CNF with 3 literals in each clause  $E = (a \vee b \vee \neg c) \wedge (\neg a \vee b \vee c) \wedge (a \vee \neg b \vee c)$ . We have  $U = \{a, b, c\}$  and  $|C|= 3$ . Taking  $I = \{v_a, v_b, v_c\}$  is optimal.

## 5.2 The VACCINATION-ATTACK<sub>dir</sub> problem

Our demonstration of NP-completeness for ATTACK<sub>dir</sub> inspires our proof for the  $\Sigma_2^P$ -completeness of VACCINATION-ATTACK<sub>dir</sub>. The formulation of this decision problem is

### VACCINATION-ATTACK<sub>dir</sub>:

INSTANCE: A graph  $G = (V, A)$ , two non-negative integer budgets  $\Omega$  and  $\Phi$  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  $\Sigma_2^P$ -complete [33]. If the Boolean formula studied in  $B_2$  is in DNF with 3 literals per clause, then the problem is still  $\Sigma_2^P$ -complete [42]. 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 [24]:

### 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.** VACCINATION-ATTACK<sub>dir</sub> is  $\Sigma_2^P$ -complete.

*Proof.* From the formulation in the form of  $\exists D \forall I Q(D, I)$ , we deduce that VACCINATION-ATTACK<sub>dir</sub>  $\in \Sigma_2^P$ .

To show that it is  $\Sigma_2^P$ -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  $k_y$ .



An example of this construction can be found in **Figure 7**. 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  $\Phi$ . 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  $\text{ATTACK}_{dir}$  (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.  $\square$

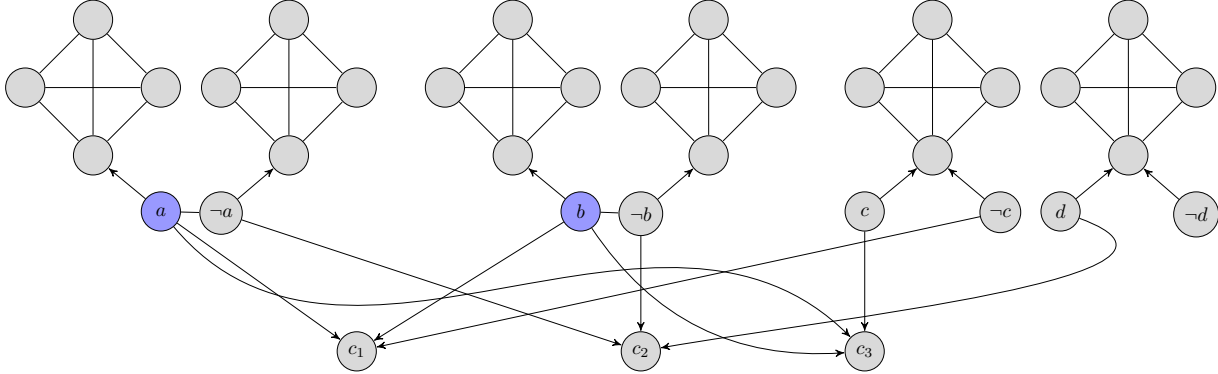


Figure 7: Example of construction of  $G$  from the boolean expression in CNF with 3 literals in each clause  $E = (a \vee b \vee \neg c) \wedge (\neg a \vee \neg b \vee d) \wedge (a \vee c \vee b)$ . Here,  $X = \{a, b\}$  and  $Y = \{c, d\}$ . Taking  $D = \{v_a, v_b\}$ , i.e obliging both  $a$  and  $b$  to be *False* makes it impossible to satisfy  $E$ .

**Corollary 5.4.**  $\text{MCN}_{dir}$  is  $\Sigma_2^P$ -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 are restricting our attention to the induced graph obtained by only considering non-saved vertices when there is no protection.

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 trees. 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 [29, 7]. In section 6.2, we describe the problem properties for DAGs, making it simpler to show that  $\text{PROTECTION}_{dir}$  is NP-complete. We terminate this section by showing that the optimal protection strategy can be determined in polynomial time for arborescences.

## 6.1 PROTECT over trees

We show next that unweighted PROTECT over undirected trees is polynomial, which sheds light on the hardness status of the unweighted tri-level MCN Problem over unweighted trees. 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 [6] 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 (MSO) 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.

**Property 6.1.** *PROTECT with unit weights is polynomial over graphs with bounded treewidth.*

*Proof.* The key factor is to reformulate our problem in terms of an MSO-formula  $\varphi$  based on set variables, which captures the graph structure of the problem, and an evaluation relation  $\psi$  over a set of integer variables, which captures the “number” aspect of the problem. In order to do so, we will define as in [6] two sets  $P$  and  $X$  where  $P$  is the set of protected nodes that separate the infected nodes of set  $I$  from the saved nodes  $X$ . 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 nodes  $x$  and  $y \in V$ . The definitions of  $\varphi$  and  $\psi$  for PROTECT are the following:

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

Through the above expression for  $\varphi$ , we make it so that the sets  $P$  and  $X$  have a null intersection and that any neighbour of a node in  $X$  must either be 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 [3] implies that the problem can be solved in  $O(f(w)n)$  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 [29]. 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 tractable algorithms. In this section we propose a DP algorithm to solve the unweighted 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 [15] for the pairwise CNDP over trees.

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 Figure 8 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 Figure 8, 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:

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

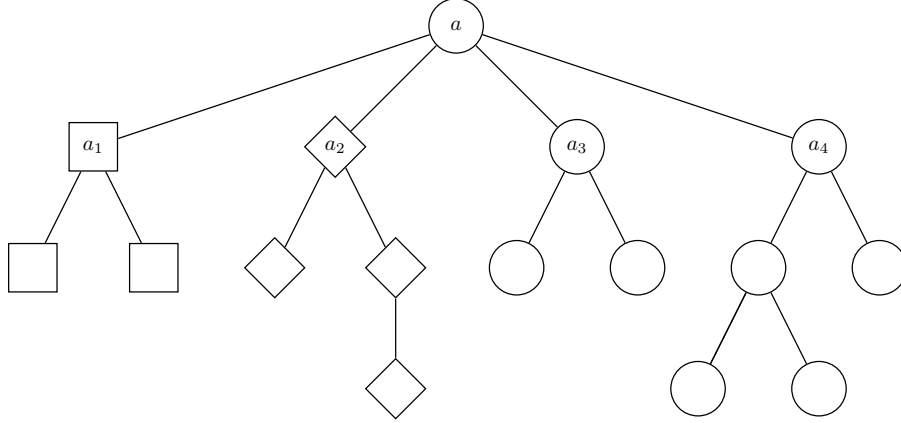


Figure 8: Example of a tree with subtree  $\mathcal{T}_{a_2}$  represented by diamond shaped vertices and subtree  $\mathcal{T}_{a_3 \rightarrow a_4}$  represented by round shaped vertices.

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

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:

$$F_a(c, m, \sigma) = \begin{cases} 0 & \text{if } c = 0, m = 0 \text{ and } \sigma = 1 \\ 1 & \text{if } c = 1, m = 0 \text{ and } \sigma = 0 \text{ or } c = 0, m = 1 \text{ and } \sigma = 0; \\ -\infty & \text{otherwise} \end{cases} \quad (5)$$

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

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 infected vertex inside  $\mathcal{T}_{a_s}$  ( $\sigma' = 0$ ). The second case deals with either a vaccinated or infected  $a$  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:

$$F_a(c, m, \sigma) = G_{a_1}(c, m, \sigma) \quad \text{for a non-leaf vertex } a \in V. \quad (7)$$

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

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

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.

**Property 6.2.** PROTECT over trees with unit protection costs and prizes and given sets of infected and vaccinated vertices 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  $\sigma$  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 tri-level problem over unweighted trees cannot be  $\Sigma_3^p$ -hard. As an observation, the same algorithm can be devised when vertices have integer protection costs  $c_i$  and remains polynomial. By inverting the roles of the value of the recursion functions and the budget  $c$ , we can devise a polynomial algorithm whenever there are integer protection weights  $c_i$  and unit prizes  $b_i = 1$ . When both types of weights are integer, the algorithms become pseudo-polynomial and the problem becomes weakly NP-hard.

## 6.2 PROTECTION<sub>dir</sub> over directed acyclic graphs

We will show that an optimal protection strategy can be restricted to *candidate* vertices for any directed acyclic graph.

**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 strongly connected saved vertices, is called *candidate*. Denote by  $\mathcal{C}$  the set of candidate vertices.

In other words, a candidate vertex  $v$  has no predecessor that implies saving  $v$ . See Figure 10a for an illustration:  $\mathcal{C} = \{1, 2, 3, 9\}$ ; 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  $\{2, 3, 4, 5, 6, 7, 8\}$ .

**Lemma 6.3.** Let  $G = (V, A)$  be a directed acyclic graph. Given  $I$  and  $\Lambda$ , 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 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 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 Figure 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 Figure 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.

**Remark 6.4.** Definitions 6.1 and 6.2, and Lemma 6.3 do not need the condition that  $G$  is a directed acyclic graph. Hence, they extend to any directed graph and thus, undirected graphs. However, since we know that the protection problem is NP-complete for general directed graphs, for sake of simplicity, we decided to write the full section assuming DAGs.

**Theorem 6.5.**  $\text{PROTECT}_{dir}$  is NP-complete, even for directed acyclic graphs.

*Proof.* The statement of  $\text{PROTECT}_{dir}$  is exactly the one of  $\text{PROTECT}$  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  $\text{CNP}_{split}$  to  $\text{PROTECT}_{dir}$ , showing its NP-hardness. Given an instance of  $\text{CNP}_{split}$ , we build the following graph  $G = (V, A)$ :

- For each  $v \in \bar{V}_1$ , 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  $v \in \bar{V}_2$ , we replicate it in  $G$ , and for each edge  $(r, v) \in \bar{E}$  with  $v \in \bar{V}_2$ , 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  $v \in \bar{V}_1$ , 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 Figure 9 for an illustration of the reduction.

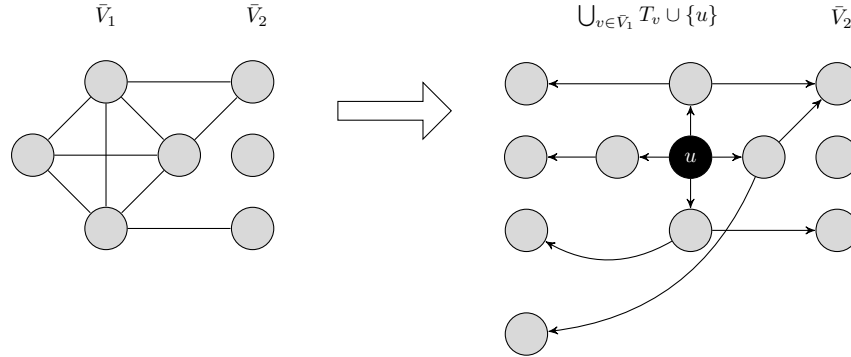


Figure 9: Example of construction of  $G$  from  $\bar{G}$ .

First, note that  $\mathcal{C}$  of  $G$  is  $\{t_v^1 : v \in \bar{V}_1\} \cup \bar{V}_2$ , 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  $\Lambda = B \geq \bar{V}_1$ , then the instance of  $\text{CNP}_{split}$  is trivial. Therefore, we can assume  $\Lambda = B < \bar{V}_1$  and thus, it holds  $P^* \subset \{t_v^1 : v \in \bar{V}_1\}$ . 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 DAG  $G = (V, A)$  is an arborescence if its underlying undirected graph is a tree (forest) and there is a single vertex (root) that has a unique directed path from it to all other vertices.

In arborescence, it is direct the determination of  $\mathcal{C}$ . Since all vertices in  $V \setminus I$  have in-degree 1, either they are protected by their predecessor, and thus 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 Figure 10c. We can prove that in this case a greedy approach leads to optimality.

**Lemma 6.6.** *Given  $G = (V, A)$ ,  $I$  and  $\Lambda$ , 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  $\Lambda$  candidate vertices of largest value 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  $P$ .  $\square$

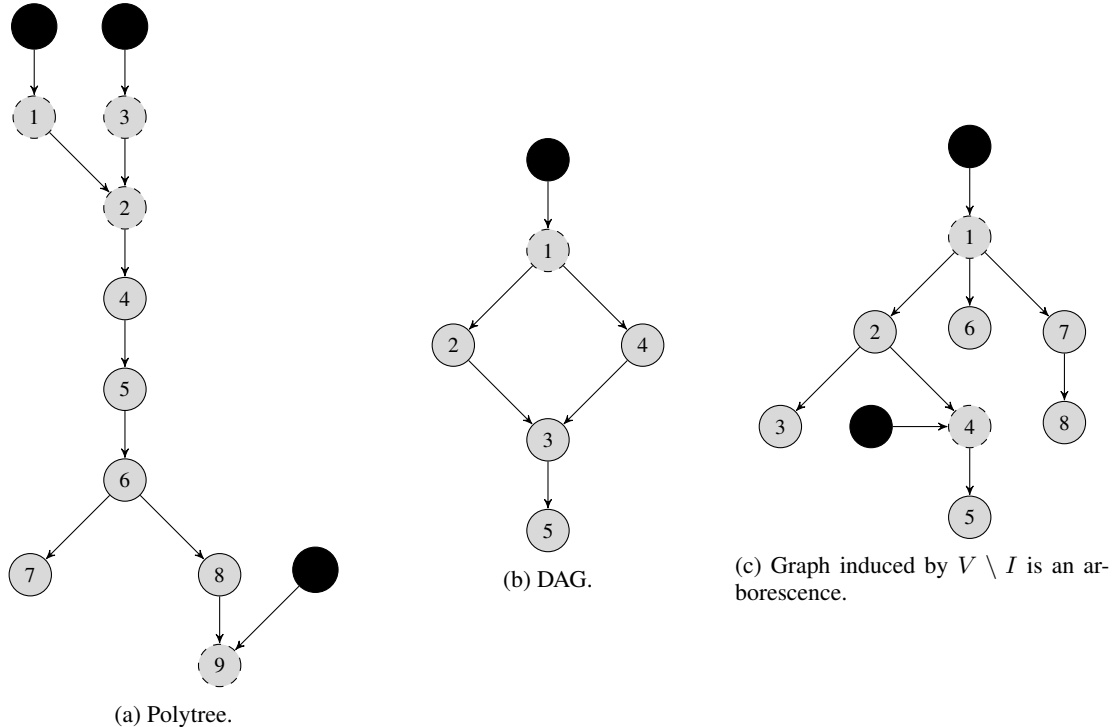


Figure 10: The set  $I$  is represented by black vertices and candidate vertices are dashed.

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 6.6 does not extend to the undirected case.

**Remark 6.7.** *Note that in Lemmata 6.3 and 6.6, we did not use the fact that  $b_v = 1$ . Thus, it also holds when vertices benefits are not unitary.*

## Acknowledgements

The authors wish to thank the support of 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] Bernardetta Addis, Marco Di Summa, and Andrea Grosso. Identifying critical nodes in undirected graphs: Complexity results and polynomial algorithms for the case of bounded treewidth. *Discrete Applied Mathematics*, 161(16):2349–2360, Nov 2013.
- [2] Roberto Aringhieri, Andrea Grosso, Pierre Hosteins, and Rosario Scatamacchia. Polynomial and pseudo-polynomial time algorithms for different classes of the Distance Critical Node Problem. *Discrete Applied Mathematics*, 253:103–121, 2019.
- [3] Stefan Arnborg, Jens Lagergren, and Detlef Seese. Easy problems for tree-decomposable graphs. *Journal of Algorithms*, 12(2):308 – 340, 1991.
- [4] Ashwin Arulselman, Clayton W. Commander, Lily Eleftheriadou, and Panos M. Pardalos. Detecting critical nodes in sparse graphs. *Computers & Operations Research*, 36(7):2193 – 2200, 2009.
- [5] Andrea Baggio, Margarida Carvalho, Andrea Lodi, and Andrea Tramontani. Multilevel approaches for the critical node problem. *Operations Research*, To appear, 2020.
- [6] Kathleen D. Barnetson, Andrea C. Burgess, Jessica Enright, Jared Howell, David A. Pike, and Brady Ryan. The firebreak problem. *Networks*, To Appear, 2020.
- [7] Kathleen D. Barnetson, Andrea C. Burgess, Jessica Enright, Jared Howell, David A. Pike, and Brady Ryan. The firebreak problem. *Networks*, n/a(n/a), 2020.
- [8] Charles Blair. The computational complexity of multi-level linear programs. *Annals of Operations Research*, 34(1):13–19, Dec 1992.
- [9] Vladimir Boginski and Clayton W. Commander. *Identifying Critical Nodes in Protein-Protein Interaction Networks*, pages 153–167. Clustering Challenges in Biological Networks. World Scientific, Feb 2009.
- [10] Gerald Brown, Matt Carlyle, Javier Salmerón, and R. Wood. Defending critical infrastructure. *Interfaces*, 36:530–544, 12 2006.
- [11] Alberto Caprara, Margarida Carvalho, Andrea Lodi, and Gerhard J. Woeginger. A study on the computational complexity of the bilevel knapsack problem. *SIAM Journal of Optimization*, 24:823–838, 2014.
- [12] Margarida Carvalho, Andrea Lodi, and Patrice Marcotte. A polynomial algorithm for a continuous bilevel knapsack problem. *Operations Research Letters*, 46(2):185 – 188, 2018.
- [13] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition, 2009.
- [14] S. DeNegre. *Interdiction and Discrete Bilevel Linear Programming*. PhD thesis, Lehigh University, 2011.
- [15] Marco Di Summa, Andrea Grosso, and Marco Locatelli. Complexity of the critical node problem over trees. *Computers and Operations Research*, 38:1766–1774, 2011.
- [16] Tibor Dudás, Bettina Klinz, and Gerhard J. Woeginger. The computational complexity of multi-level bottleneck programming problems. In Athanasios Migdalas, Panos M. Pardalos, and Peter Värbrand, editors, *Multilevel Optimization: Algorithms and Application*, pages 165–179, Boston, MA, 1998. Springer US.
- [17] Stephen Finbow, Andrew King, Gary MacGillivray, and Romeo Rizzi. The firefighter problem for graphs of maximum degree three. *Discrete Mathematics*, 307(16):2094–2105, Jul 2007.
- [18] Fabio Furini, Ivana Ljubic, Enrico Malaguti, and Paolo Paronuzzi. Casting light on the hidden bilevel combinatorial structure of the k-vertex separator problem. In *OR-19-6, DEI, University of Bologna*, 2019.
- [19] Fabio Furini, Ivana Ljubic, Sebastien Martin, and Pablo San Segundo. The maximum clique interdiction problem. *European Journal of Operational Research*, 277:112–127, 02 2019.
- [20] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., USA, 1990.

- [21] Jing He, Hongyu Liang, and Hao Yuan. Controlling infection by blocking nodes and links simultaneously. In Ning Chen, Edith Elkind, and Elias Koutsoupias, editors, *Internet and Network Economics*, pages 206–217, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [22] Pierre Hosteins and Rosario Scatamacchia. The Stochastic Critical Node Problem over trees. *Networks*, 76(3):381–401, 2020.
- [23] Robert G. Jeroslow. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming*, 32(2):146–164, Jun 1985.
- [24] Berit Johannes. *New Classes of Complete Problems for the Second Level of the Polynomial Hierarchy*. PhD thesis, Technischen Universität Berlin, 2011.
- [25] Richard M. Karp. *Reducibility among Combinatorial Problems*, pages 85–103. Springer US, Boston, MA, 1972.
- [26] Mohammed Lalou and Hamamache Kheddouci. A polynomial-time algorithm for finding critical nodes in bipartite permutation graphs. *Optimization Letters*, 13:1345–1364, 2019.
- [27] Mohammed Lalou, Mohammed Amin Tahraoui, and Hamamache Kheddouci. Component-cardinality-constrained critical node problem in graphs. *Discrete Applied Mathematics*, 210:150–163, 2016.
- [28] Mohammed Lalou, Mohammed Amin Tahraoui, and Hamamache Kheddouci. The critical node detection problem in networks: A survey. *Computer Science Review*, 28:92 – 117, 2018.
- [29] Alexander Langer, Felix Reidl, Peter Rossmanith, and Somnath Sikdar. Practical algorithms for mso model-checking on tree-decomposable graphs. *Comput. Sci. Rev.*, 13(C):39–74, November 2014.
- [30] John M. Lewis and Mihalis Yannakakis. The node-deletion problem for hereditary properties is np-complete. *Journal of Computer and System Sciences*, 20(2):219–230, Apr 1980.
- [31] Enrico Malizia. More complexity results about reasoning over (m)cp-nets. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018)*, 2018.
- [32] P.A. Martin. Tri-level optimization models to defend critical infrastructure. Master’s thesis, Naval Postgraduate School, Monterey, 01 2007.
- [33] A. R. Meyer and 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)*, pages 125–129, 1972.
- [34] Fabio Mogavero and Giuseppe Perelli. Binding Forms in First-Order Logic. In Stephan Kreutzer, editor, *24th EACSL Annual Conference on Computer Science Logic (CSL 2015)*, volume 41 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 648–665, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [35] Marcus Schaefer and Christopher Umans. Completeness in the polynomial-time hierarchy a compendium. *Sigact News - SIGACT*, 33, 01 2002.
- [36] Siqian Shen, J. Cole Smith, and Roshan Goli. Exact interdiction models and algorithms for disconnecting networks via node deletions. *Discrete Optimization*, 9(3):172–188, Aug 2012.
- [37] Yi-Dong Shen and Thomas Eiter. Evaluating epistemic negation in answer set programming. *Artificial Intelligence*, 237:115 – 135, 2016.
- [38] L. J. Stockmeyer and A. R. Meyer. Word problems requiring exponential time (preliminary report). In *Proceedings of the Fifth Annual ACM Symposium on Theory of Computing, STOC ’73*, page 1–9, New York, NY, USA, 1973. Association for Computing Machinery.
- [39] Vera Tomaino, Ashwin Arulsevan, Pierangelo Veltri, and Panos M. Pardalos. *Studying Connectivity Properties in Human Protein–Protein Interaction Network in Cancer Pathway*, pages 187–197. Springer US, Boston, MA, 2012.
- [40] Mario Ventresca, Kyle Robert Harrison, and Beatrice M. Ombuki-Berman. The bi-objective critical node detection problem. *European Journal of Operational Research*, 265(3):895–908, Mar 2018.
- [41] R. Kevin Wood. Deterministic network interdiction. *Mathematical and Computer Modelling*, 17(2):1–18, Jan 1993.
- [42] Celia Wrathall. Complete sets and the polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):23–33, Oct 1976.