Attacking Power Indices by Manipulating Player Reliability

Gabriel Istrate*
West University of Timişoara and
e-Austria Research Institute
Timişoara, Romania
gabrielistrate@acm.org

Cosmin Bonchiş
West University of Timişoara and
e-Austria Research Institute
Timişoara, Romania
cosmin.bonchis@e-uvt.ro

Alin Brînduşescu
West University of Timişoara and
Elektrobit Automotive Romania
Braşov, Romania
alin.brindusescu@elektrobit.com

ABSTRACT

We investigate the manipulation of power indices in TU-cooperative games by stimulating (subject to a budget constraint) changes in the propensity of other players to participate to the game.

We display several algorithms that show that the problem is often tractable for so-called network centrality games and influence attribution games, as well as an example when optimal manipulation is intractable, even though computing power indices is feasible.

KEYWORDS

coalitional games;reliability extension;Shapley value;manipulation

ACM Reference Format:
Gabriel Istrate, Cosmin Bonchiş, and Alin Brînduşescu. 2019. Attacking Power Indices by Manipulating Player Reliability. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 9 pages.

1 INTRODUCTION

Control is a fundamental but difficult issue in multi-agent systems. A multi-agent society may be difficult to control due to the concurrence of several factors, that may interact and drive the dynamics in complex, unpredictable ways. Some of these factors could include uncertainty about agent involvement [6], coalition formation [13], the rules [17], the environment [48], about rewards [28], the presence (or lack) of synergies between players [39], etc.

A common type of control is *manipulation*[1], which often aims to change the power (index) of a given player by means of interventions in the settings or the dynamics of the agent society. Many types of manipulation have been considered in the literature, often in a computational social choice context. They include *identity* [2], *cloning* [18] and *quota* [49] manipulation in voting games, *collusion and mergers* [29], *sybil attacks* [45] and, finally, *multi-mode attacks* [21], just to name a few.

We contribute to this direction by studying yet another natural mechanism for manipulation: **changing the propensity of players to participate to the game.** This type of manipulation is quite frequent in real-life situations, a central example being voting - while parties cannot control with absolute certainty voter turnout on election day, they may employ tactics that aim to mobilize their supporters and deter participation of their opponents' voters[2]. Manipulation could be performed by a centralized actor (like in the voting example), or by a coalition of players [14], strategically modifying their behavior (in our case their reliabilities) in response to a perceived dominance of a player whose power index they wish to decrease.

The main impetus for our work was [9], where a model of strategic manipulation of player reliabilities was first investigated. Bachrach et al. [9] considered *max games*. In these games each player possesses a weight; the value of a coalition is the maximum weight of a component of the coalition. They proved a "no sabotage theorem" for (the reliability extension of) max-games with a common failure probability. They remarked that manipulating player reliabilities can be studied in principle for all coalitional games, and asked for further investigations of this problem, in settings similar to the one we consider, i.e. under costly player manipulation. Given the negative results for max-games [9] and the fact that computing power indices is often intractable [16], we concentrate mostly on proving *positive results*, showing that there exist natural scenarios where optimal attacks on power indices by manipulating players' reliabilities are easy to compute (and interesting). We hope that these positive results will encourage renewed interest (and research) on the scope and limits of reliability manipulation.

Contributions and outline In Section 2 we begin by informally stating the problem and justifying our choice of the two classes of coalitional games studied in this paper: *network centrality games* [1, 35, 43, 44] and *credit attribution games* [27, 38]. Even though credit attribution games may seem to be somewhat exotic/of limited use, their importance extends well-beyond scientometry: they were, in fact, anticipated, as *hypergraph games* (see [16] Section 3). The two games we consider from this class, *full credit* and *full obligation* games, are natural examples of read-once marginal contribution (MC) nets [19]. Full credit games are equivalent to the subclass of basic MC-nets [25] whose rules are *conjunctions of positive variables*; full obligation games correspond to generalized MC-nets whose rules consist of *disjunctions of positive variables.* Full obligation games can simulate *induced subgraph games* [16]. Full credit games capture an important subclass of coalitional skill games (CSG) [7, 8], that of CSG games with tasks consisting of a single skill.

Section 3 contains technical details and precise specifications of the models we investigate. We deal with two types of attacks: (node) *removal*, where we are allowed to remove (decrease to zero the reliability of) certain nodes, and *fractional attacks*, where reliability probabilities can be altered continuously.

*corresponding author

[1] We use the word with its wider, commonsense meaning, rather than the specialized one from voting theory [15]. Our usage encompasses both strategic behavior by an agent or coalition (voting theory "manipulation") and interventions by a chair or outside agent (such as control and bribery in voting [22]). We assume, however, that all such interventions are costly.

[2] Such scenarios are best modeled as *multichoice voting games* [23]. However, since such games are *multi-cooperative* (rather than cooperative) games [11], they fall outside of the scope of the present work, and will be dealt with in a subsequent paper.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

In Section 4 we give closed-form formulas for the Shapley values of the reliability extensions of network centrality and credit allocation games. Next we particularize these results to centrality games on specific network: first we show that no removal attack is beneficial; as for fractional attacks, we show that in the complete graph K_n or when attacking the center of the star graph S_n, a greedy approach works: one should increase the reliabilities of neighbors of the attacked node, in descending order of baseline reliabilities. When attacking a non-center player in S_n the result is similar, with the important exception that increasing the reliability of the center should precede all other moves. In contrast, the situation for the cycle graph C_n is more complex, involving all distance-two neighbors of the attacked node. A simple characterization is provided for the optimum as *the best of four fixed "greedy" solutions.* This characterization allows the determination of the optimum for all combinations of reliability values and budget.[3] An interesting, and unintuitive, qualitative feature of the result is that in the optimal attack *a non-neighbor of the attacked node could be targeted **before** some of the direct neighbors of the attacked node.*

In Section 6 we analyze full credit and full obligation games. Although these two games have the same Shapley value [27], we show that *they behave very differently with respect to attacks*: removal attacks are not beneficial for full credit games, NP-hard for full obligation games. Fractional attacks also behave differently, modifying probabilities in opposite directions. In a particular setting which includes the case of induced subgraph games we obtain greedy algorithms for both games, derived from expressing the problems as fractional knapsack problems. The determining quantities for the attack orders are (two different) cost-benefit measures.

2 PROBLEM STATEMENT AND CHOICE OF GAMES

The *power index attack problem*, the main problem of interest in this paper, has the following simple informal statement: we consider the reliability extension of a cooperative game. We are given a positive budget $B > 0$ and are allowed to modify reliabilities of all nodes, other than the targeted player x, as long as the total cost incurred is at most B. Which nodes should we target, and how should we change their reliabilities, in order to decrease as much as possible the Shapley value of node x?

A variant of the previous problem, called the *pairwise power index attack problem* and motivated by Example 2.1 below, is the following: we are given not one but *two* players x, y. The goal is to decrease as much (within the budget) the Shapley value of x, while not affecting at all the Shapley value of y. This restriction makes some nodes exempt from attacks: we are not allowed to change the reliabilities of players who contribute to the Shapley value of y.

Choice of games The problems described above could be investigated in all classes of TU-cooperative games, or compact representation frameworks. However, we feel that the most compelling cases are those where the computation of power indices, e.g. the Shapley value, of (the reliability extensions) of our games is tractable[4]. In other words *the intractability of manipulating a power index should not be a consequence of our inability to compute these indices.* In particular, we are interested in scenarios where computing power indices is easy, but computing an optimal attack on them is hard. Theorem 6.2 below provides such an example.

The appeal of studying attacks on node centrality in social networks is quite self-evident: game-theoretic concepts such as those considered in [1, 35, 43, 44] formalize appealing notions of leadership in social situations. They have been proposed as tools for identifying key actors, with applications e.g. to terrorist networks [32, 36]. In such a setting, a direct (physical) attack on a leading node may be infeasible. Instead, one could attempt to indirectly affect its status (centrality), by incentivizing some of its peers.

Relevant examples of targeting nodes in order to affect power indices arose (implicitly) in even earlier work [38], that attempted to develop coalitional models of credit allocation in scientific work. The following is a version of the example in [38]:

*Example 2.1. Two scientists A, B are compared with respect to their publication record[5]. All their papers have exactly one co-author. Figure 1 displays this information as a graph, listing for each author pair, the number of publications they have authored and the number of citations. If using the Hirsch index, it would seem that candidate A has a better track record than candidate B. But if we discard publications both of them have co-written with "famous scientist Y" (that is, **remove** Y **and its publications from consideration**), then their relative ranking would be reversed.*

The authors of [38] attempted to use the Shapley value of a game based on the Hirsch index for credit allocation. An ulterior, more general and cleaner game-theoretic approach is [27]. The author defines several *credit allocation games*, and uses their (identical) Shapley values as a measure of individual publication record. Slightly modified versions of this measure have (regrettably) actually been used in some countries to set minimum publication thresholds for access and promotion to academic positions, e.g. the minimal standards in Romania.

In such a context one could naturally ask the following question: *what are the top k coauthors that account for most of a scientists' publication record?* When using the game-theoretic framework for scientific credit from [27], this is equivalent to finding the k coauthors whose removal (together with the joint papers) causes the scientist's' Shapley value to decrease the most.

Collaborations may, however, be genuinely productive or just bring to one of the scientists the benefits of association with leading scientists[6]. The Shapley value approach of [27] does not distinguish between these two scenarios, as it gives equal credit to all authors of a joint paper, irrespective of "leadership status". Recent work, e.g. Hirsch's *alpha index* [24], has attempted to quantify "scientific leadership". It is possible to define a measure based on the reliability extension of credit allocation games that factors out the "well connectedness" of an individual from its score[7]. Given such a measure,

[3] The precise formula for the optimum is cumbersome, hence deferred to the full version.

[4] This requirement disqualifies many natural candidate games such as *weighted voting games* [16, 33], as well as most subclasses of *coalitional skill games* [3]

[5] We **do not** condone and caution against the real-life use of such crude quantitative metrics for tasks like the one described in this example or our models.

[6] One could argue, of course, that such an association itself reflects positively on the scientist. But the opposite argument, that prestige drives scientific inequality, has recently been substantiated by real data [37] and is, at the very least, hard to ignore.

[7] The measure computes appropriate values of reliability probabilities, the lower the probability the more of a "scientific leader" a coauthor is; we are currently investigating the practicality of such an approach.

the previous question, that of finding the top-k co-authors is still interesting, as it *identifies the most (genuinely) fruitful collaborations of a given author, irrespective of status.* This is modeled by the power index attack problem in credit allocation games.

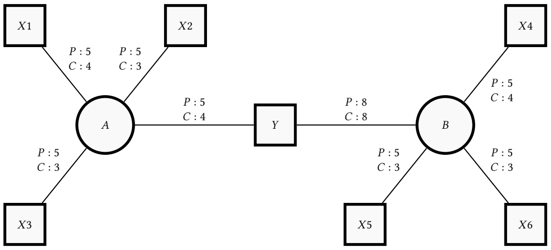

Figure 1: The scenario (from [38]) in Example 2.1

3 TECHNICAL DETAILS

We will be working in the framework of Algorithmic Cooperative Game Theory, see [12] for a readable introduction.

We will make use of notation $f|_a^b$ as a shorthand for $f(b) - f(a)$. Given graph $G = (V,E)$ and vertex $v \in V$, we will denote by $N(v)$ the set of neighbors of V and by $\widehat{N(v)} = \{v\} \cup N(v)$. Given $S \subseteq V$, we denote by $\delta(S)$ the set of nodes $y \in V \setminus S$ such that there exists $x \in S$, $(x,y) \in E$. We generalize the setting above to the case when G is a *weighted graph*, i.e. there exists a weight function $w : E \rightarrow \mathbb{R}_{+}$. Given set $S \subseteq V$ and integer $r \geq 1$ we define $B_w(S,r)$, *the ball of radius r around S*, to be the set $B_w(S,r) = \{x \in V : (\exists y \in S)$ s.t. $d_w(x,y) \leq r\}$. We may omit w from this notation when it is simply the graph distance in G. Also, given "cutoff" distance d_{cut} we define $N_{cut}(x) = B(\{x\}, d_{cut})$.

We will deal with cooperative games with transferable utility, that is pairs $\Gamma = (N,v)$ where N is a set of *players* and $v : \mathcal{P}(S) \rightarrow \mathbb{R}_{+}$ is a *value function* under the partial sets of S. If $S \subseteq N$ is a set of players, $v(S)$ is the value that players in coalition S can guarantee for themselves irrespective of the other players' participation.

Although we could prove similar results for other power indices, e.g. the Banzhaf value, in this paper we restrict ourselves to the *Shapley value.* This index measures the portion of the grand coalition value $v(N)$ that a given player $x \in N$ could fairly request for itself. It has the formula [12] $Sh[v](x) = \frac{1}{n!} \cdot \sum_{\pi \in S_n} [v(S_\pi^x \cup \{x\}) - v(S_\pi^x)]$, where $S_\pi^x = \{\pi[i] | \pi[i]$ precedes x in $\pi\}$ and S_n is the set of permutation.

We are concerned with two classes of cooperative games. The first one arose from efforts to define game-theoretic notions of *network centrality* [1, 35, 43, 44]. We define these games as follows:

- Game Γ_{NC_1} is specified by its value function $v_{NC_1}(S) = |S \cup \delta(S)|$.
- Given integer $k \geq 1$, game Γ_{NC_2} is specified by its value function $v_{NC_2}(S) = |S \cup \{x \notin S$ s.t. $|N(x) \cap S| \geq k\}|$.
- In game Γ_{NC3} graph G is *weighted.* We are also given a positive "cutoff distance" d_{cut}. We give the characteristic function v_{NC_3} by $v_{NC_3}(S) = |B(S, d_{cut})|$.

A second class of games, related to the example in [38] is that of *influence-attribution games*, formally defined by Karpov [27]. A *credit-attribution game* is formalized by a set of authors $N = \{1,...,n\}$ and a set of publications $P = \{P_1,...,P_m\}$. Each paper P_j is naturally endowed with a set of *authors* $Auth_j \subseteq N$ and a *quality score* $w_j \in \mathbb{R}_{+}$. In real-life scenarios the quality measure could be 1 (i.e. we simply count papers), a score based on the ranking of the venue the paper was published in, the number of its citations, or even some iterative, PageRank-like variant of the above measures.

- *The full credit game* Γ_{FC} is specified by its value function $v_{FC}(S)$ which is simply the sum of weights of papers whose authors' list contains **at least one member from** S**.**
- *The full obligation game* Γ_{FO} is specified by its value function $v_{FO}(S)$ which is the sum of weights of papers whose authors **are all members of** S**.**

Denote by Pap_x the set of papers of x, and by $CA(x)$ the set of co-authors of x, i.e. the set of players l for which there exists a $k \in Pap_x \cap Pap_l$. If $l \in CA(x)$ denote by $C(x,l) = \sum_{k \in Pap_x \cap Pap_l} w_k$ *the joint contribution* of x, l.

Reliability extension and attack models We will be working within the framework of *reliability extension* of games, first defined in [6] and further investigated in [9]. The *reliability extension* of cooperative game $G = (N,v)$ with parameters $(p_1, p_2, \ldots, p_n)$ is the cooperative game $\Gamma = (N, \overline{v})$ with $\overline{v}(S) = \sum_{T \subseteq S} v(T) \cdot \Pi_{T,S}$, where $\Pi_{T,S} = (\prod_{i \in T} p_i) \cdot (\prod_{i \in S \setminus T} (1 - p_i))$.

A useful result about these quantities is:

Claim 1. *Let* $S \subseteq W$*. We have*

$$\frac{\partial \Pi_{S,W}}{\partial p_j} = \begin{cases} \Pi_{S \setminus j, W \setminus j} \ \textit{if } j \in S, \\ -\Pi_{S, W \setminus j} \ \textit{if } j \in W \setminus S, \textit{ and} \\ 0, \ \textit{if } j \notin W \end{cases}$$

We will consider in the sequel the following two attack models:

(1). *fractional attack:* In this type of attack every node j different from the attacked node x has a *baseline reliability* $p_j^* \in (0,1]$. We are allowed to manipulate the reliability of each such node $j \neq x$ by changing it from p_j^* to an arbitrary value p. To do so we will incur, however, a cost $u_j(p)$. We assume that cost function $u_j(\cdot)$ is defined and has an unique zero[8] at $p = p_j^*$, is decreasing and linear on $[0, p_j^*]$ and increasing and linear on $[p_j^*, 1]$ (Figure 2). That is: for every player $j \neq x$ there exist values $L_j, R_j > 0$ such that

$$u_j(p) = \begin{cases} L_j(p_j^* - p), & \text{if } p < p_j^*, \\ 0, & \text{if } p = p_i^*, \\ R_j(p - p_j^*), & \text{if } p > p_j^*. \end{cases}$$

(2). *removal attack:* In this type of attack we are only allowed to change the reliability of any node j (different from the targeted node x) from p_j^* to 0. To do so will incur a cost c_j.

A basis for fractional attacks The following simple result will be used to analyze fractional attacks in network centrality games:

Lemma 3.1 ("IMPROVING SWAPS"). *Let* D *be an open set in* $\mathbb{R}^n$*, let* $x = (x_1, \ldots, x_n) \in D$ *and* $f : D \rightarrow \mathbb{R}$ *be an analytic function. Assume* $1 \leq i,j \leq n$ *are indices such that* $\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} > \frac{\partial f(x_1,\ldots,x_n)}{\partial x_j}$*. Define* $x_{i,j}(\epsilon) = (x_k(\epsilon))$*, with*

$$x_k(\epsilon) = \begin{cases} x_k + \epsilon, & \textit{if } k = j, \\ x_k - \epsilon, & \textit{if } k = i, \\ x_k, & \textit{otherwise.} \end{cases} \quad (1)$$

[8] There is no cost for keeping the baseline reliability.

Then there exists $\epsilon_0 > 0$ such that function $g : [0, \epsilon_0] \to \mathbb{R}$, $g(\epsilon) = f(x_{i,j}(\epsilon))$ is monotonically decreasing.

In other words, to minimize function f one could decrease the variables with the largest partial derivative, while symmetrically increasing a smaller one.

Proof. By the chain rule $g'(0) = \sum\limits_{k=1}^{n} \frac{\partial f(x_1,\ldots,x_n)}{\partial x_k} \frac{\partial x_k(\epsilon)}{\partial \epsilon}|_{\epsilon=0}$

$$= \frac{\partial f(x_1,\ldots,x_n)}{\partial x_j} - \frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} < 0.$$

Since g' is continuous, g' is strictly negative on some interval $[0, \epsilon_0]$. The result follows. □

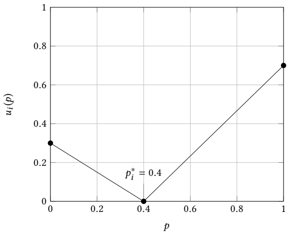

Figure 2: Shape of utility functions in fractional attacks.

4 CLOSED-FORM FORMULAS

The basis for our manipulation of network centralities is the following characterization of the Shapley value of the reliability extension:

Theorem 4.1. *The Shapley values of the reliability extensions of network centrality games $\Gamma_{NC_1}, \Gamma_{NC_2}, \Gamma_{NC_3}$ have the formulas:*

$$Sh[\overline{v_{NC_1}}](x) = p_x \sum_{\substack{y \in \widehat{N(x)} \\ S \subseteq \widehat{N(y)} \setminus x}} \frac{1}{|S|+1} \Pi_{S, \widehat{N(y)} \setminus x}$$

$$Sh[\overline{v_{NC_2}}](x) = p_x [\sum_{y \in N(x)} \sum_{\substack{S \subseteq \widehat{N(y)} \setminus x \\ |S| \geq k-1}} \frac{(|S|+1-k)}{|S|(|S|+1)} \Pi_{S, \widehat{N(y)} \setminus x} +$$

$$+ \sum_{S \subseteq N(x)} \frac{k}{|S|+1} \Pi_{S, N(x)}]$$

$$Sh[\overline{v_{NC_3}}](x) = p_x \sum_{\substack{y \in \widehat{N}(x) \\ S \subseteq \widehat{N_{cut}(y)} \setminus x}} \frac{1}{|S|+1} \Pi_{S, \widehat{N_{cut}(y)} \setminus x}$$

As for credit atribution games, the corresponding result is

Theorem 4.2. *The Shapley values of the reliability extensions of Γ_{FC}, Γ_{FO} with probabilities $(p_1, p_2, \ldots, p_n)$ have the formulas*

$$Sh[\overline{v_{FC}}](x) = p_x \cdot \sum_{k \in Pap_x} w_k \cdot [\sum_{S \subseteq Auth_k \setminus \{x\}} \frac{\Pi_{\emptyset, S}}{(n_k - |S|)\binom{n_k}{|S|}}] \quad (2)$$

where $Auth_k$ is the set of coauthors of paper k and $n_k = |Auth_k|$, and

$$Sh[\overline{v_{FO}}](x) = \sum_{k \in Pap_x} \frac{w_k}{n_k} \cdot \Pi_{Auth_k, Auth_k} \quad (3)$$

5 ATTACKING NETWORK CENTRALITIES

The next result follows from Theorem 4.1 and Claim 1:

Corollary 5.1. *In the reliability extensions of the centrality games $\Gamma_{NC_1}, \Gamma_{NC_2}, \Gamma_{NC_3}$, the Shapley values of player 1 are monotonically decreasing functions of distance-two neighbors' reliabilities (and do not depend on other players).*

Proof. Deferred to the full version. □

The previous corollary shows that for network centrality games no removal attack is beneficial:

Theorem 5.2. *No removal attack on the centrality of a player in games $\Gamma_{NC_1}, \Gamma_{NC_2}, \Gamma_{NC_3}$ can decrease its Shapley value.*

Fractional attacks on specific networks Given that removal attacks are not beneficial, we now turn to fractional attacks. The objective of this section is to show that the analysis of optimal fractional attacks is often feasible. Since the graphs in this section are fairly symmetric, we will assume (for these examples) that the slopes of all utility curves are identical. That is, there exist positive constants L, R such that if $i \neq j$ are different agents then $L_i = L_j = L$ and $R_i = R_j = R$ (though, of course, baseline probabilities p_i^* and p_j^* may differ). The graphs we are going to be concerned with are the complete graph K_n, the star graph S_n (where node 1 is either the center or an outer node) and the n-cycle C_n (Figure 3).

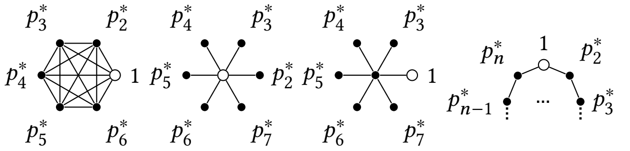

Figure 3: Target topologies for fractional attacks.

Note that, when $G = K_n$ or $G = S_n$, pairwise Shapley value attacks are trivially impossible: indeed, these graphs have diameter at most two. Since all distance-two neighbors influence the Shapley value of a given player, all nodes are exempt from attacks.

On the other hand, for these topologies it turns out that the best attack on Shapley value of player x is to increase the reliabilities of its neighbors in the descending order of their baseline reliabilities:

Theorem 5.3. *Let G be either the complete graph K_n with n vertices. or the star graph with n vertices S_n centered at node $x = 1$. To optimally attack the centrality of x in the reliability extension of Γ_{NC_1} use the following algorithm:*

- *Consider nodes $2, \ldots, n$ in the decreasing order of their baseline reliabilities, breaking ties arbitrarily. $p^*_{sorted(2)} \geq p^*_{sorted(3)} \geq \ldots \geq p^*_{sorted(n)}$.*
- *While the budget allows it, increase to one (if not already equal to 1) the probabilities $p_{sorted(i)}$, starting with $i = 2$ and successively increasing i.*
- *If the budget no longer allows increasing $p_{sorted(i)}$ to one, increase it as much as possible.*
- *Leave all other probabilities to their baseline values.*

If, on the other hand, $G = S_n$ centered, say, at node 2, to optimally attack the centrality of node $x = 1$, the algorithm changes as follows:

- Consider nodes $2, \ldots, n$ in the following order: node 2, followed by nodes $3, \ldots, n$ sorted in decreasing order of their baseline reliabilities $p^*_{sorted(3)} \geq \ldots \geq p^*_{sorted(n)}$, breaking ties arbitrarily. Denote the new order by Q.
- Follow the previous greedy protocol, increasing baseline probabilities up to one (if allowed by the budget) according to the new ordering Q.

Similar statements hold for game Γ_{NC_2}, and for Γ_{NC_3} for large enough values of parameter d_{cut}.

In the previous examples the optimal attack involved a determined node targeting order, which privileged direct neighbors and could depend on baseline reliabilities but was independent of the value of the budget. None of this holds in general: as the next result shows, on graph C_n the optimum can be computed by taking the best of *four* node targeting orders. The optimum may lack the two previously discussed properties of optimal orders:

- in optimal attacks one should sometimes target a distance-two neighbor (3 or n-1) *before* targeting both of $x = 1$'s neighbors (2 and n, see Figure 3).
- the order (among the four) that characterizes the optimum may depend on the budget value B as well. Formally:

Theorem 5.4. *Let P,Q,R,S be the vectors $[2,n,n-1,3]$, $[2,n-1,n,3]$, $[n,3,2,n-1]$, $[n,2,3,n-1]$, respectively. Let $Sol_P, Sol_Q, Sol_R, Sol_S$ be the configurations obtained by increasing in turn (as much as possible, subject to the budget B) the reliabilities of nodes $2,3,n-1,n$ in the order(s) specified by P,Q,R,S, respectively. Then*

a. *The best of $Sol_P, Sol_Q, Sol_R, Sol_S$ is an optimal attack on the centrality of $x = 1$ in game Γ_{NC_1} on the cycle graph C_n.*
b. *There exist values of $p_2^*, p_3^*, p_{n-1}^*, p_n^*$ s.t. Sol_P is optimal for all values of B (by symmetry a similar statement holds for Sol_S).*
c. *There exist values of $p_2^*, p_3^*, p_{n-1}^*, p_n^*$ and an nonempty open interval I for the budget B such that Sol_Q is an optimum for all $B \in I$ (by symmetry a similar statement holds for Sol_R).*

6 ATTACKS IN CREDIT ATTRIBUTION

In this section we study removal attacks in credit attribution games. Interestingly, while the Shapley values have identical formulas in Γ_{FC}, Γ_{FO} [27], the two games are **not** similar with respect to attacks. Indeed, similarly to the case of network centrality, we have:

Theorem 6.1. *No removal attack can decrease the Shapley value of a given player in a full credit attribution game.*

Proof. At first, this seems counterintuitive, as it would seem to contradict Example 2.1. The answer is that *this example does not correspond to the full credit game, but to the full obligation one*: in game Γ_{FC} a player does **not** lose credit for a paper due to removal of a coauthor; in fact its Shapley value will increase, since the credit for the paper divides among fewer coauthors. It is in Γ_{FO} where players may lose credit as a result of coauthor removal. □

This difference between Γ_{FC} and Γ_{FO} is evident with respect to attacks: As the next result shows, in full-obligation games, finding optimal removal attacks can simulate a well-known hard combinatorial problem:

Theorem 6.2. *The* budgeted maximum coverage problem *(which is NP-complete) reduces to minimizing the Shapley value of a given player in the full-obligation game (under removal attacks).*

Proof. Deferred to the full version. □

Fractional attacks The following is a simple consequence of the formulas in Theorem 4.2 and Claim 1 shows that *optimal attacks are different in games Γ_{FC} and Γ_{FO} irrespective of the topology of the coauthorship hypergraph*: in the first case we need to increase the reliability of x's coauthors, in the other case we aim to decrease it:

Theorem 6.3. *In the reliability extensions of the credit allocation games Γ_{FC}, Γ_{FO} the Shapley value of player x is a decreasing (respectively increasing) function of coauthors' reliabilities (and does not depend on other players).*

Optimal attacks can be explicitly described in the particular scenario when, just as in Example 2.1, each paper has exactly two authors (a situation that corresponds, under the full obligation model, to induced subgraph games). It turns out that *the relevant quantity is the ratio between the score of coauthors' joint contribution with the attacked node and its marginal cost:*

Theorem 6.4. *To optimally decrease the Shapley value of node x in game Γ_{FC} in the two-author special case:*

(a). Sort the coauthors l of x in the decreasing order of the fractions $\frac{C(x,l)}{R(l)}$, breaking ties arbitrarily.
*(b). While the budget allows it, for $i = 1, \ldots |CA(x)|$, **increase** to 1 the probability of the i'th most valuable coauthor.*
(c). If the budget does not allow increasing the probability of the i'th coauthor up to 1, increase it as much as possible.
(d). Leave all other probabilities to their baseline values.

Corollary 6.5. *In the setting of Theorem 6.4, to optimally solve the pairwise Shapley value attack problem for x, y, run the algorithm in the Theorem only on those z that are coauthors of x but not of y.*

As for game Γ_{FO}, the optimal attack is symmetric. Since we are decreasing probabilities, we will be using fractions $\frac{C(x,l)}{L(l)}$ instead:

Theorem 6.6. *To optimally decrease the Shapley value of node x in the full obligation game Γ_{FO} in the two-author special case:*

(a). Sort the coauthors of x in the decreasing order of the fractions $\frac{C(x,l)}{L(l)}$, breaking ties arbitrarily.
*(b). While the budget allows it, for $i = 1, \ldots |CA(x)|$, **decrease** to 0 the probability of the i'th most valuable coauthor.*
(c). If the budget does not allow decreasing the probability of the i'th coauthor up to 0, decrease it as much as possible.
(d). Leave all other probabilities to their baseline values.

Corollary 6.7. *In the setting of Theorem 6.6, to solve the pairwise Shapley value attack problem for players x, y, run the algorithm in the Theorem only on those z that are coauthors of x but not of y.*

7 PROOF HIGHLIGHTS

In this section we present some of the proofs of our results. Some other proofs are included in the Appendix, others are deferred to the full version of the paper, to be posted on arxiv.org:

7.1 Proof of Theorem 4.1

We prove the formula for the first game only. Similarly to [35], the proofs for the other two games are completely analogous, and deferred to the full version. Define, for $y \in V$, $W \subseteq V$

$$f_y(W) = \begin{cases} 1, & \text{if } y \notin W \cup \delta(W), \\ 0, & \text{otherwise.} \end{cases}$$

A simple case analysis proves that, for every $W \subseteq V$, $v_{NC_1}(W \cup \{x\}) - v_{NC_1}(W) = \sum_{y \in \widehat{N(x)}} f_y(W)$. We therefore have

$$Sh[\overline{v_{NC_1}}](x) = E_{\pi \in S_n}[\overline{v_{NC_1}}(S_\pi^x \cup \{x\}) - \overline{v_{NC_1}}(S_\pi^x)] = E_{\pi \in S_n}[p_x \cdot \sum_{W \subseteq S_\pi^x} [v_{NC_1}(W \cup \{x\}) - v_{NC_1}(W)] \cdot \Pi_{W,S_\pi^x}] = p_x E_{\pi \in S_n} \sum_{W \subseteq S_\pi^x} \sum_{y \in \widehat{N(x)}} f_y(W) \cdot \Pi_{W,S_\pi^x} = p_x E_{\pi \in S_n} \sum_{y \in \widehat{N(x)}} \sum_{W \subseteq S_\pi^x} f_y(W) \cdot \Pi_{W,S_\pi^x}$$

We now introduce two notations that will help us reinterpret the previous sum: given $W \subseteq V$, denote by $Alive(W)$ the set of nodes in W that are *alive* under the reliability extension model. Also, given permutation $\pi \in S_n$ and $W \subseteq V$, denote by $First_\pi(W)$ the element of W that appears first in enumeration π. With these notations

$$Sh[\overline{v_{NC_1}}](x) = p_x \sum_{y \in \widehat{N(x)}} Pr_{\pi \in S_n}[y \notin Alive(S_\pi^x) \cup \delta(Alive(S_\pi^x))]$$

$$= p_x \sum_{y \in \widehat{N(x)}} Pr_{\pi \in S_n}[x = First_\pi(\widehat{N(y)} \cap Alive(V)) | x \in Alive(V)]$$

If $S = (\widehat{N(y)} \setminus \{x\}) \cap Alive(V)$ then the conditional probability that x is $First_\pi(S \cup \{x\})$, given that x is alive, is $\frac{1}{|S|+1}$. We thus get the desired formula.

7.2 Proof of Theorem 4.2

Denote, for a set of authors C, by $Pap_C = \cup_{l \in C} Pap_l$ the set of papers with at least one author in C. We decompose function v_{FC} as $v_{FC}(\cdot) = \sum_k w_k v_k(\cdot)$ where

$$v_k(C) = \begin{cases} 1, \text{if } k \in Pap_C \\ 0, \text{otherwise.} \end{cases} \tag{4}$$

$$\text{Thus } v_{FC}(C) = \sum_{R \subseteq C} v_{FC}(R)\Pi_{R,C} = \sum_{R \subseteq C} \Pi_{R,C} \sum_k w_k v_k(R) = = \sum_k \sum_{R \subseteq C} \Pi_{R,C} w_k v_k(R) = \sum_k w_k \overline{v_k}(C)$$

which means that we can decompose $\overline{v_{FC}} = \sum_k w_k \overline{v_k}$, and the Shapley value of $\overline{v_{FC}}$ decomposes as well $Sh(\overline{v_{FC}}) = \sum_k w_k \cdot Sh(\overline{v_k})$, and similarly for v_{FO}. On the other hand

$$Sh[\overline{v_k}](x) = \frac{1}{n!} \sum_{\pi \in S_n} [\overline{v_k}(S_\pi^x \cup \{x\}) - \overline{v_k}(S_\pi^x)]$$

Given set A of authors,

$$\overline{v_k}(A \cup \{x\}) - \overline{v_k}(A) = \sum_{R \subseteq A \cup \{x\}} v_k(R)\Pi_{R,A\cup\{x\}} - \sum_{R \subseteq A} v_k(R)\Pi_{R,A}$$

$$= (1 - p_x) \sum_{R \subseteq A \setminus x} v_k(R)\Pi_{R,A\setminus\{x\}} + p_x \sum_{R \subseteq A \setminus x} \Pi_{R,A} v_k(R \cup \{x\})$$

$$- \sum_{R \subseteq A} \Pi_{R,A} v_k(R) = p_x \cdot \sum_{R \subseteq A} \Pi_{R,A}[v_k(R \cup \{x\}) - v_k(R)]$$

Now $v_k(R \cup \{x\}) - v_k(R)$ is 1 if $k \in Pap_x \setminus Pap_R$, 0 otherwise. For $k \notin Pap_x$, $\overline{v_k}(A \cup \{x\}) - \overline{v_k}(A) = 0$. Otherwise $\overline{v_k}(A \cup \{x\}) - \overline{v_k}(A) = p_x \cdot \sum_{\substack{R \subseteq A \\ k \notin Pap_R}} \Pi_{R,A}$.

We can interpret this quantity as the probability that the live subset of A does not cover k, but x is live and does. Applying this to the Shapley value we infer that $Sh[\overline{v_k}](x)$ is the probability that in a random permutation π the live subset of S_π^x does not cover k, but x is live and does.

Full credit model: There are $n_k!$ permutations Ξ of indices in $Auth_k$, each of them equally likely when π is a random permutation in S_n. Given subset $S \subset Auth_k \setminus \{x\}$, the probability that Ξ starts with S followed by x is $\frac{|S|!(n_k - |S| - 1)!}{n_k!}$. To make x pivotal for paper k, none of the agents in S must be live. This happens with probability $\Pi_{\emptyset,S}$. Given the above argument, we have

$$Sh[\overline{v_k}](x) = p_x \cdot \sum_{S \subseteq Auth_k \setminus \{x\}} \frac{(|S|)!(n_k - |S| - 1)!}{n_k!} \cdot [\prod_{l \in S}(1 - p_l)]$$

$$= p_x \cdot \sum_{S \subseteq Auth_k \setminus \{x\}} \frac{\Pi_{\emptyset,S}}{(n_k - |S|)\binom{n_k}{|S|}}, \text{ hence}$$

$$Sh[\overline{v_{FC}}](x) = p_x \cdot \sum_{k \in Pap_x} w_k \cdot [\sum_{S \subseteq Auth_k \setminus \{x\}} \frac{\Pi_{\emptyset,S}}{(n_k - |S|)\binom{n_k}{|S|}}] \tag{5}$$

which is what we had to prove.

Full obligation model: For x to be pivotal for paper k, x and all its coauthors in $Auth_k$ must all be live, and all elements of $Auth_k \setminus x$ must appear before x in ordering π. This happens with probability $\frac{1}{n_k} \cdot \Pi_{Auth_k,Auth_k}$.

7.3 Proof of Theorem 5.3

First of all, the following claim holds for all graphs G:

Claim 2. *The minimum of function $z \to Sh[\overline{v_{NC_1}}](1)|_z$ exists and is reached on some profile (p_i) with $p_i^* \le p_i \le 1$.*

Proof. Function $z \to Sh[\overline{v_{NC_1}}](1)|_z$ is continuous and the set $[0,1]^n$ is compact, so the minimum is reached. Assuming some $p_j < p_j^*$, we could increase p_j up to p_j^*, reducing total cost. This does not increase (and perhaps further decreases) the Shapley value. □

Next, we (jointly) prove cases $G = K_n$ and $G = S_n$ with $x = 1$ being a center, since the proofs are practically identical. The remaining case ($K = S_n$, $x = 1$ not a center) is deferred to the Appendix. We start with the following

Lemma 7.1. *For $G = K_n$ or $G = S_n$, $j \ne l \in V(G) \setminus 1$ and any probability profile $p = (p_1, \ldots, p_n) \in (0,1]^n$,*

$$sign\Big(\frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial p_j}|_p - \frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial p_l}|_p\Big) = sign(p_j - p_l)$$

Proof. Deferred to the full version. □

We first prove that in the optimal solution on these graphs no two variables could assume equal values, unless both equal to the endpoints of their restricting intervals:

Lemma 7.2. *In the setting of Theorem 5.3, suppose $p = (p_1, \ldots, p_n)$ is such there is are indices $2 \leq i \neq j \leq n$ with $0 < p_i = p_j < 1$. Then there exists $\epsilon_0 > 0$ such that for every $\epsilon \in [-\epsilon_0, \epsilon_0]$, $\epsilon \neq 0$, $Sh[\overline{v_{NC_1}}](1)|_{p_{i,j}(\epsilon)} < Sh[\overline{v_{NC_1}}](1)|_p$, (where $p_{i,j}(\epsilon)$ is defined as in equation (1)).*

Proof. Deferred to the full version. □

Now we prove:

Claim 3. *In the optimal solution there is at most one index i_1 with $p_{i_1} \in (p^*_{i_1}, 1)$. In other words, in the optimal solution some probabilities are increased up to 1, some ae left unchanged to their baseline values, and at most one variable is increased to a value less than 1.*

Proof. Suppose there were two different indices $i_1 \neq i_2$. We must have $p_{i_1} = p_{i_2}$, or, by Lemma 3.1, one could decrease the Shapley value by increasing the larger one and symmetrically decreasing the smaller one. But this is impossible, due to Lemma 7.2. □

Note that the greedy solution Γ has the structure from Claim 3 and that any permutation of OPT on variables $p_2, \ldots, p_n$ has the same Shapley value as OPT (since K_n, S_n have this symmetry).

We compare the vectors Γ, OPT, both sorted in decreasing order. Our goal is to show that these sorted versions are equal. Without loss of generality, we may assume that OPT creates the same ordering on variables as the p_i^*'s (and Γ), when considered in decreasing sorted order (we break ties, if any, in the same way). Indeed, if there were indices i, j such that $p^*_{sorted(i)} \geq p^*_{sorted(j)}$ but $p_{sorted(i)} < p_{sorted(j)}$ then, since $p_{sorted(j)} > p_{sorted(i)} \geq p^*_{sorted(i)} \geq p^*_{sorted(j)}$, we could simply swap values $p_{sorted(i)}$ and $p_{sorted(j)}$ and obtain another legal, optimal solution.

If Γ were different from OPT, since Greedy increases the largest variables first, there must be variables x, y such that $\Gamma_x \geq \Gamma_y$, $\Gamma_x > p_x$ and $\Gamma_y < p_y$. Since Γ and OPT have the same ordering of variables, we also must have in fact $p_x \geq p_y$, i.e. $1 \geq \Gamma_x > p_x \geq p_y > \Gamma_y \geq p_y^*$. But then, using either Lemma 3.1 (if $p_x \neq p_y$) or Lemma 7.2 (otherwise) we could further improve OPT by increasing p_x and symmetrically decreasing p_y, a contradiction.

7.4 Proof of Theorem 5.4

A simple computation shows that for $G = C_n$

$$Sh[\overline{v_{NC_1}}](1) = p_1\left(\frac{p_2p_n + p_2p_3 + p_{n-1}p_n}{3} - \frac{p_3 + p_{n-1}}{2} - p_2 - p_n + 3\right).$$

As p_1 does not influence any attack on itself, w.l.o.g. we will assume $p_1 = 1$. We need to minimize the above quantity, subject to

$$p_2 + p_3 + p_{n-1} + p_n = B + p_2^* + p_3^* + p_{n-1}^* + p_n^*, p_i^* \leq p_i \leq 1.$$

We now prove a result somewhat similar to Claim 3. However, now we will only interdict certain patterns.

Claim 4. *In an optimal solution it is not possible that $p_k^* < p_k < 1$, $p_l^* < p_l < 1$ when:*

a. *$k = 2$, $l = n - 1$ (and, symmetrically, $k = 3$, $l = n$). In fact, in this case we have the stronger implication $p_{n-1} > p^*_{n-1} \Rightarrow p_2 = 1$. Symetrically, $p_3 > p_3^* \Rightarrow p_n = 1$.*
b. *$k = 2$, $l = n$.*
c. *$k = 2$, $l = 3$ (and, symmetrically, $k = n$, $l = n - 1$.) In the case when $\frac{p_3+p_n}{3} \leq \frac{p_2}{3} - \frac{1}{2}$ we have the stronger implication $p_3 > p_3^* \Rightarrow p_2 = 1$. Symetrically, in the case when $\frac{p_2+p_{n-1}}{3} \leq \frac{p_n}{3} - \frac{1}{2}$, $p_{n-1} > p^*_{n-1} \Rightarrow p_n = 1$.*

Proof. Suppose there were two such indices k, l. We must also have $\frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial x_k} = \frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial x_l}$, otherwise we could decrease the Shapley value using Lemma 3.1. We reason in all cases by contradiction:

a. We prove directly the stronger result. Suppose $p_2 < 1$. We have $\frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial x_2} = \frac{p_3+p_{n-1}}{3} - 1 \leq \frac{p_3}{3} - \frac{2}{3} < \frac{p_3}{3} - \frac{1}{2} = \frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial x_{n-1}}$. So we can apply Lemma 3.1 to p_2 and p_{n-1}, further decreasing the Shapley value as we increase p_2 and decrease p_{n-1}.

b. Equality of partial derivatives can be rewritten as $p_2 + p_n = p_3 + p_{n-1}$. An easy computation (which uses this equality) shows that $Sh[\overline{v_{NC_1}}](1)|_p^{p_{n,2}(\epsilon)} = -\frac{\epsilon^2}{3}$. But then it means that one could further decrease the Shapley value of player 1, hence we are not at an optimum, a contradiction.

c. As in the proof of a. $\frac{p_3+p_n-p_2}{3} - \frac{1}{2} = \frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial x_2} - \frac{\partial Sh[\overline{v_{NC_1}}](1)}{\partial x_3} = 0$, otherwise we could use Lemma 3.1 with p_2, p_3 to decrease the Shapley value. An easy computation (which uses this equality) shows that $Sh[\overline{v_{NC_1}}](1)|_p^{p_{3,2}(\epsilon)} = \frac{\epsilon(p_n-p_2+p_3)}{3} - \frac{\epsilon}{2} - \frac{\epsilon^2}{3} = -\frac{\epsilon^2}{3} < 0$. But then one could further decrease the Shapley value of 1, a contradiction.

□

We use Claim 4 to prove Theorem 5.4:

a. The conclusion of this claim is that the only case when there could exist two values p_k, p_l strictly between their baseline values and 1 is $k = 3, l = n - 1$ (or vice-versa), a case when we must further have $p_2 = p_n = 1$. Thus the optimal solution is the best of the configurations obtained by greedily increasing probabilities (up to 1, if the budget will allow it) in one of the orders $[2, n, 3, n-1], [2, n, n-1, 3], [2, n-1, n, 3], [n, 3, 2, n-1], [n, 2, 3, n-1], [n, 2, n-1, 3]$. An easy computation shows that the first two orders are equally good for all possible budget values B, and so are the last two. So, in the end we only have to compare the four orders P, Q, R, S to find an optimum, proving the first part of the theorem.

b,c: Symmetry between 2,3 and n,n-1 reduces the proof of these two points to analyzing the "winners" among $Sol_P, Sol_Q, Sol_R, Sol_S$, and proving that, under suitable conditions, it belongs either to $\{Sol_P, Sol_S\}$ (point b.) or to $\{Sol_Q, Sol_R\}$ (point c.).

If we start by increasing p_2 by ϵ, the Shapley value decreases by $\epsilon(1 - \frac{p_3^*+p_{n-1}^*}{3})$. We will call the number $1 - \frac{p_3^*+p_{n-1}^*}{3}$ the *speed of the decrease*. It is maintained while p_2 increases from p_2^* to 1, i.e. over a *segment* (interval) of *size* $1 - p_2^*$. There are four segments, corresponding to the four variables being increased. The table in Figure 4 summarizes the effect of variable increases on the decrease of the Shapley value of node 1. Using this table it is easy to compare the four permutations with respect to this decrease:

P versus Q: Since they use the same variable, $\Delta_P = \Delta_Q$ throughout the first segment. At the (common) end of the third segment, a

Perm,	Sp1	sz1	Sp2	sz2	Sp3	sz3	Sp4	sz4
P	$1-\frac{p_3^*+p_n^*}{3}$	$1-p_2^*$	$\frac{2-p_{n-1}^*}{3}$	$1-p_n^*$	1/6	$1-p_{n-1}^*$	1/6	$1-p_3^*$
Q	$1-\frac{p_3^*+p_n^*}{3}$	$1-p_2^*$	$\frac{1}{2}-\frac{p_n^*}{3}$	$1-p_{n-1}^*$	1/3	$1-p_n^*$	1/6	$1-p_3^*$
R	$1-\frac{p_2^*+p_{n-1}^*}{3}$	$1-p_n^*$	$\frac{1}{2}-\frac{p_2^*}{3}$	$1-p_3^*$	1/3	$1-p_2^*$	1/6	$1-p_{n-1}^*$
S	$1-\frac{p_2^*+p_{n-1}^*}{3}$	$1-p_n^*$	$\frac{2-p_3^*}{3}$	$1-p_2^*$	1/6	$1-p_3^*$	1/6	$1-p_{n-1}^*$

Figure 4: Dynamics of the decrease of the Shapley value.

simple computation yields $\Delta_P - \Delta_Q = 0$, and since P, Q use identical fourth segments, $\Delta_P = \Delta_Q$ throughout their fourth segment.

As for the second/third segments, if $p_n^* < 1$ and $p_{n-1}^* - p_n^* > \frac{1}{2}$ then throughout the common portion of the second segment $\Delta_P < \Delta_Q$. Afterwards the difference will start shrinking, and will become positive after a certain value $\lambda_{P,Q}$ where $\Delta_P = \Delta_Q$. Note that at the end of the second segment of Q, $\Delta_P - \Delta_Q = \frac{1-p_{n-1}^*}{6} \geq 0$, so $\lambda_{P,Q}$ is in the second segment of P and the third of Q.
To determine $\lambda_{P,Q}$ write $\lambda_{P,Q} = 1 - p_2^* + 1 - p_{n-1}^* + \mu_{P,Q}$. We have: $\frac{2-p_{n-1}^*}{3}(1 - p_{n-1}^* + \mu_{P,Q}) = (\frac{1}{2} - \frac{p_n^*}{3})(1 - p_{n-1}^*) + \frac{\mu_{P,Q}}{3}$, or $\mu_{P,Q} = p_{n-1}^* - p_n^* - \frac{1}{2}$ hence $\lambda_{P,Q} = \frac{3}{2} - p_2^* - p_n^*$.

The conclusion is that $\Delta_P \geq \Delta_Q$ for all budgets if $p_{n-1}^* - p_n^* \leq \frac{1}{2}$. Otherwise $\Delta_P \geq \Delta_Q$, except for $B \in I_{P,Q} := (1 - p_2^*, \frac{3}{2} - p_2^* - p_n^*)$. Similar conclusions hold for comparing S versus R.

P versus S: At the (common) end of their second segment $\Delta_P - \Delta_S = (1 - p_n^*)(\frac{p_2^*-1}{3}) + (1 - p_2^*)(\frac{1-p_n^*}{3}) = 0$. So $\Delta_P = \Delta_S$, and this prevails throughout the third and fourth segments.

As for the first and second segment, $\Delta_P - \Delta_S \leq 0$ if $p_3^* + p_n^* \geq p_2^* + p_{n-1}^*$, $\Delta_P - \Delta_S \geq 0$ if $p_3^* + p_n^* \leq p_2^* + p_{n-1}^*$. Hence $\Delta_P \leq \Delta_S$ for all budgets if $p_3^* + p_n^* \geq p_2^* + p_{n-1}^*$. Otherwise $\Delta_P \geq \Delta_S$.

Summing up:

- If $p_{n-1}^* - p_n^* < 1/2$, $p_3^* - p_2^* < 1/2$, $p_3^* + p_n^* \leq p_2^* + p_{n-1}^*$, then $\Delta_P \geq \Delta_Q$, $\Delta_S \geq \Delta_R$, $\Delta_P \geq \Delta_S$ for all budgets, so P is optimal. If the last condition is reversed then S is optimal.

- If $p_{n-1}^* - p_n^* > 1/2$, $p_3^* - p_2^* > 1/2$ then $\Delta_P \leq \Delta_Q$ on $I_{P,Q}$, $\Delta_S \leq \Delta_R$ on $I_{S,R}$. So the best of Q, R is an optimum on $I_{P,Q} \cap I_{S,R}$. Since Q, R are piecewise linear functions, one of them is better than the other one on an open interval.

7.5 Proof sketch of Theorems 6.4 and 6.6

The two proof are very similar, so we only present the one of Theorem 6.4. Particularizing formula 2 to the case of induced subgraph games, we infer that the Shapley value of player x has the formula $Sh[\overline{v_{FC}}](x) = p_x \cdot \sum_{l \in CA(x)} C(x,l) \cdot \frac{2-p_l}{2}$ (*).

We claim that minimizing $Sh[\overline{v_{FC}}](x)$ is equivalent to solving the following fractional knapsack problem:

$$\begin{cases} max[\sum_{l \in CA(x)} C(x,l)(1-p_l^*) \cdot y_l] \\ \sum_{l \in CA(x)} R_l(1-p_l^*) \cdot y_l = \sum_{l \in CA(x)} R_l \cdot (1-p_l^*) - B. \\ 0 \leq y_l \leq 1, \forall l \in CA(x) \end{cases} \tag{6}$$

Indeed, by formula (*) it is only efficient to increase the reliability probabilities of x's authors from p_l^* to some $p_l \in [p_l^*, 1]$. If we introduce variables $y_l \in [0,1]$ by equation $1 - y_l = \frac{p_l - p_l^*}{1-p_l^*}$, (or, equivalently, $y_l = \frac{1-p_l}{1-p_l^*}$), the cost of such move is $R_l \cdot (p_l^* - p_l) = R_l \cdot (1 - y_l)(1 - p_l^*)$. The total costs must add up to B, so $\sum_{l \in CA(x)} R_l \cdot (1 - y_l)(1 - p_l^*) = B$, which is equivalent to system (6). The minimization of the Shapley value is easily seen to correspond to the maximization of the objective function of (6).

Now it is well-known that the greedy algorithm that considers variables y_l in decreasing order of their cost/benefit ratio finds an optimal solution to problem (6). Reinterpreting this result in our language we get the algorithm described in Theorem 6.4.

8 RELATED WORK[9]

First of all, *network interdiction* (see e.g. [41, 42]) is a well-established theme in combinatorial optimization. Our removal model can be seen as a special case of node interdiction.

Results on the *reliability extension* of a cooperative game [5, 6, 9, 10, 34] are naturally related. So is the rich literature on *manipulation*, both in non-cooperative and coalitional settings [2, 21, 29, 45–47, 49] and *bribery* [20] in voting. Our framework covers both scenarios, that in which an external perpetrator bribes agents to change their reliabilities, and that in which this is done by a coalition of agents.

A lot of work has been devoted recently to measuring and characterizing *synergies between players* in multi-agent settings [30, 31, 39]. Synergies between players in cooperative games are obviously relevant to the theme of this paper: synergic agents' participation to coalitions increases the Shapley value of the given agent. The nature of some of our results (Theorems 5.3, 6.4 and 6.6), that target nodes in a fixed order, provide a concrete way for ranking synergies between these nodes and the attacked one.

9 CONCLUSIONS AND OPEN ISSUES

Our results have uncovered a rich typology of optimal attacks on players' power indices: Sometimes no attack is beneficial. Sometimes, the optimal attack is intractable, even when computing the power indices is feasible. For fractional attacks, in many cases (but not always) greedy-type approaches provide an optimal strategy. An open question raised by our work is the complexity of fractional attacks in general full-obligation credit attribution games. Motivated by Theorem 6.2 we believe that even this version is intractable. On the other hand we would like to see our framework applied to more settings. They include bicooperative games [11], generalized MC-nets [19], etc. Of special interest are cases when computing the Shapley value is easy, e.g. voting games with super-increasing weights [4], flow games on series-parallel networks [19], or games with bounded dependency degree [26].

As for relative attacks, we propose studying a more realistic *bicriteria optimization* version of the problem [40]: decrease as much as possible the Shapley value of node x while not affecting the Shapley value of node y by more than a certain amount D.

Finally, the related problem of *increasing* the power index of a given node subject to budget constraints is also worth investigating.

ACKNOWLEDGEMENTS

This work was supported by a grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P4-ID-PCE-2016-0842, within PNCDI III.

[9] For reasons of space this section is only sketched.

REFERENCES

[1] Karthik V. Aadithya and Balaraman Ravindran. 2010. Game theoretic network centrality: exact formulas and efficient algorithms. In *Proceedings of the 2010 International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*. 1459–1460.

[2] Haris Aziz, Yoram Bachrach, Edith Elkind, and Mike Paterson. 2011. False-name manipulations in weighted voting games. *Journal of Artificial Intelligence Research* 40 (2011), 57–93.

[3] Haris Aziz, Oded Lachish, Mike Paterson, and Rahul Savani. 2009. Power indices in spanning connectivity games. In *International Conference on Algorithmic Applications in Management*. Springer, 55–67.

[4] Yoram Bachrach, Yuval Filmus, Joel Oren, and Yair Zick. 2016. Analyzing Power in Weighted Voting Games with Super-Increasing Weights. In *International Symposium on Algorithmic Game Theory*. Springer, 169–181.

[5] Yoram Bachrach, Ian Kash, and Nisarg Shah. 2012. Agent failures in totally balanced games and convex games. In *International Workshop on Internet and Network Economics*. Springer, 15–29.

[6] Yoram Bachrach, Reshef Meir, Michal Feldman, and Moshe Tennenholtz. 2011. Solving Cooperative Reliability Games. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI'11)*. AUAI Press, 27–34.

[7] Yoram Bachrach, David Parkes, and Jeffrey S. Rosenschein. 2013. Computing cooperative solution concepts in coalitional skill games. *Artificial Intelligence* 204 (2013), 1–21.

[8] Yoram Bachrach and Jeffrey S Rosenschein. 2008. Coalitional skill games. In *Proceedings of the 2007 international joint conference on Autonomous agents and multiagent systems-Volume 2*. 1023–1030.

[9] Yoram Bachrach, Rahul Savani, and Nisarg Shah. 2014. Cooperative max games and agent failures. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 29–36.

[10] Yoram Bachrach and Nisarg Shah. 2013. Reliability weighted voting games. In *International Symposium on Algorithmic Game Theory*. Springer, 38–49.

[11] J.M. Bilbao, J.R. Fernandez, A. Losada Jiménez, and E Lebrón. 2000. Bicooperative games. *Cooperative games on combinatorial structures*. (2000), 131–295.

[12] George Chalkiadakis, Edith Elkind, and Michael Wooldridge. 2011. Computational Aspects of Cooperative Game Theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning* (2011).

[13] Georgios Chalkiadakis, Evangelos Markakis, and Craig Boutilier. 2007. Coalition formation under uncertainty: Bargaining equilibria and the Bayesian core stability concept. In *Proceedings of the 2007 international joint conference on Autonomous agents and multiagent systems*. 412–419.

[14] Vincent Conitzer, Tuomas Sandholm, and Jérôme Lang. 2007. When are elections with few candidates hard to manipulate? *Journal of the A.C.M.* 54, 3 (2007), 14.

[15] Vincent Conitzer and Toby Walsh. 2016. *Handbook of Computational Social Choice*. Cambridge University Press, Chapter Barriers to Manipulation in Voting.

[16] Xiaotie Deng and Christos H. Papadimitriou. 1994. On the complexity of cooperative solution concepts. *Mathematics of Operations Research* 19, 2 (1994), 257–266.

[17] Edith Elkind and Gábor Erdélyi. 2012. Manipulation under voting rule uncertainty. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. 627–634.

[18] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. 2011. Cloning in elections: Finding the possible winners. *Journal of Artificial Intelligence Research* 42 (2011), 529–573.

[19] Edith Elkind, Leslie Ann Goldberg, Paul W. Goldberg, and Michael Wooldridge. 2009. A tractable and expressive class of marginal contribution nets and its applications. *Mathematical Logic Quarterly* 55, 4 (2009), 362–376.

[20] Piotr Faliszewski, Edith Hemaspaandra, and Lane A. Hemaspaandra. 2006. The complexity of bribery in elections. In *Proceedins of the AAAI*, Vol. 6. 641–646.

[21] Piotr Faliszewski, Edith Hemaspaandra, and Lane A. Hemaspaandra. 2011. Multimode control attacks on elections. *Journal of Artificial Intelligence Research* 40 (2011), 305–351.

[22] Piotr Faliszewski and Joerg Rothe. 2016. *Handbook of Computational Social Choice*. Cambridge University Press, Chapter Control and Bribery in Voting.

[23] Dan S. Felsenthal and Moshé Machover. 1997. Ternary voting games. *International journal of game theory* 26, 3 (1997), 335–351.

[24] Jorge E. Hirsch. 2018. h_α : An index to quantify an individual's scientific leadership. *arXiv preprint arXiv:1810.01605* (2018).

[25] Samuel Ieong and Yoav Shoham. 2005. Marginal contribution nets: a compact representation scheme for coalitional games. In *Proceedings of the 6th ACM conference on Electronic commerce*. ACM, 193–202.

[26] Ayumi Igarashi, Rani Izsak, and Edith Elkind. 2018. Cooperative Games With Bounded Dependency Degree. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*. 1063–1070.

[27] Alexander Karpov. 2014. Equal weights coauthorship sharing and the Shapley value are equivalent. *Journal of Informetrics* 8, 1 (2014), 71–76.

[28] Steven Ketchpel. 1994. Forming coalitions in the face of uncertain rewards. In *Proceedings of the AAAI*, Vol. 94. 414–419.

[29] Omer Lev, Maria Polukarov, Yoram Bachrach, and Jeffrey S. Rosenschein. 2013. Mergers and collusion in all-pay auctions and crowdsourcing contests. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. 675–682.

[30] Somchaya Liemhetcharat and Manuela Veloso. 2012. Modeling and learning synergy for team formation with heterogeneous agents. In *Proceedings of the 2012 International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. 365–374.

[31] Somchaya Liemhetcharat and Manuela Veloso. 2014. Weighted synergy graphs for effective team formation with heterogeneous ad hoc agents. *Artificial Intelligence* 208 (2014), 41–65.

[32] R. Lindelauf, H. Hamers, and B. Husslage. 2013. Cooperative game theoretic centrality analysis of terrorist networks: The cases of Jemaah Islamiyah and Al Qaeda. *European Journal of Operational Research* 229, 1 (2013), 230–238.

[33] Yasuko Matsui and Tomomi Matsui. 2001. NP-completeness for calculating power indices of weighted majority games. *Theoretical Computer Science* 263, 1-2 (2001), 305–310.

[34] Reshef Meir, Moshe Tennenholtz, Yoram Bachrach, and Peter Key. 2012. Congestion Games with Agent Failures.. In *Proceedings of the AAAI*, Vol. 12. 1401–1407.

[35] Tomasz P. Michalak, Karthik V. Aadithya, Piotr L Szczepanski, Balaraman Ravindran, and Nicholas R. Jennings. 2013. Efficient computation of the Shapley value for game-theoretic network centrality. *Journal of Artificial Intelligence Research* (2013), 607–650.

[36] Tomasz P. Michalak, Talal Rahwan, Oskar Skibski, and Michael Wooldridge. 2015. Defeating terrorist networks with game theory. *IEEE Intelligent Systems* 30, 1 (2015), 53–61.

[37] Allison C. Morgan, Dimitrios J. Economou, Samuel F. Way, and Aaron Clauset. 2018. Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Science* 7, 1 (19 Oct 2018), 40.

[38] Panagiotis Papapetrou, Aristides Gionis, and Heikki Mannila. 2011. A Shapley value approach for influence attribution. In *Machine Learning and Knowledge Discovery in Databases*. Springer, 549–564.

[39] Ariel D. Procaccia, Nisarg Shah, and Max Lee Tucker. 2014. On the Structure of Synergies in Cooperative Games.. In *Proceedings of the AAAI*. 763–769.

[40] R. Ravi, Madhav V. Marathe, S.S. Ravi, Daniel J. Rosenkrantz, and Harry B. Hunt III. 1993. Many birds with one stone: Multi-objective approximation algorithms. In *Proceedings of the twenty-fifth annual ACM Symposium on Theory of Computing*. 438–447.

[41] J. Cole Smith. 1997. Basic Interdiction Models. In *Wiley Encyclopedia of Operations Research and Management Science*. American Cancer Society.

[42] J. Cole Smith, Mike Prince, and Joseph Geunes. 2013. Modern network interdiction problems and algorithms. In *Handbook of combinatorial optimization*. Springer, 1949–1987.

[43] Rama N. Suri and Y. Narahari. 2008. Determining the top-k nodes in social networks using the Shapley value. In *Proceedings of the 2008 international conference on Autonomous agents and multi-agent systems*. 1509–1512.

[44] Mateusz K. Tarkowski, Tomasz P. Michalak, Talal Rahwan, and Michael Wooldridge. 2017. Game-theoretic Network Centrality: A Review. *arXiv preprint arXiv:1801.00218* (2017).

[45] Thibaut Vallée, Grégory Bonnet, Bruno Zanuttini, and François Bourdon. 2014. A study of sybil manipulations in hedonic games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 21–28.

[46] Marcin Waniek, Tomasz P. Michalak, Talal Rahwan, and Michael Wooldridge. 2017. On the Construction of Covert Networks. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS '17)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1341–1349. http://dl.acm.org/citation.cfm?id=3091282.3091310

[47] Marcin Waniek, Tomasz P. Michalak, Michael J. Wooldridge, and Talal Rahwan. 2018. Hiding individuals and communities in a social network. *Nature Human Behaviour* 2, 2 (2018), 139–147.

[48] Makoto Yokoo, Vincent Conitzer, Tuomas Sandholm, Naoki Ohta, and Atsushi Iwasaki. 2005. Coalitional games in open anonymous environments. In *Proceedings of AAAI*, Vol. 5. 509–514.

[49] Michael Zuckerman, Piotr Faliszewski, Yoram Bachrach, and Edith Elkind. 2012. Manipulating the quota in weighted voting games. *Artificial Intelligence* 180 (2012), 1–19.