ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240887

Learning and Optimizing with an SSB Representation of Intransitive Preferences on Sets

Hugo Gilbert[a], Mohamed Ouaguenouni[b] and Olivier Spanjaard[b]

[a]Université Paris-Dauphine, PSL University, CNRS, LAMSADE, Paris, France
[b]Sorbonne Université, CNRS, LIP6, Paris, France

Abstract. We propose a Skew-Symmetric Bilinear (SSB) model to represent intransitive preferences on subsets of a ground set of items. More precisely, the SSB model accounts for preference intensities between pairs of subsets. We provide a procedure to learn the parameters of the SSB model from a set of known pairwise preferences between subsets, managing to find a sparse model, and as simple as possible in terms of the degree of interaction between items. The SSB model can be viewed as a concise representation of a weighted tournament on subsets. We study the complexity of determining the winners according to various tournament rules. Numerical tests on synthetic and real-world data are carried out.

1 Introduction

A wide range of problems in Artificial Intelligence (AI) can be formulated as *subset selection* problems, namely to select a subset from a ground set of items. Examples of such problems include notably (but are not limited to) feature selection in learning tasks [26, 30], multiattribute decision making with binary attributes [2, 15], or course recommendation [34]. The definition of an "optimal" subset of items requires a formalism for expressing preferences over subsets. Furthermore, this preference model should be able to account for interactions (synergies) between items, known as the *portfolio effect* [8]. For example, on a backpacking trip, even if you prefer to take a bottle of water rather than an apple, you may prefer to take a bottle of water and an apple rather than two bottles of water [25].

Various works in AI have tackled this topic of representing preferences over sets. Brafman et al. [5] and Binshtok et al. [3] have considered a framework in which the items from which sets are built are associated with attributes, and the values of these attributes are what distinguish the different items. They define properties based on properties of items in a set (which makes it possible to express synergies between items in a set), and the preferences over sets are defined by comparing set properties, e.g., "I would like a set for which the number of items with such value for such attribute is at least (resp. at most, equal to) k". Desjardins and Wagstaff [13, 14] proposed an alternative attribute-based model in which preferences over the *diversity* of items in a set may be expressed, e.g., "I would like a set for which the values of the items for such attribute are evenly dispersed across the range of possible values" (as would be desirable for instance in a recommendation list). They also designed learning methods for their model, as well as optimization methods for solving the subset selection problem with preferences represented by their model. Guo and Gomes [25] have presented a *non-parametric* method for learning attribute-based preferences over sets, i.e., they do not assume a pre-defined parameterized preference function over sets of items. They learn a function which, given a new unseen ground set of items, predicts the subset of it that minimizes the expected value of a set similarity loss w.r.t. the optimal subset.

Domshlak and Joachims [15], Bigot et al. [2] and Gilbert et al. [22] studied a setting in which there is no attribute-based description of the items. Note that in the first two papers, not sets but binary vectors are compared, which is formally equivalent. Similarly to the expression of a Choquet integral in terms of Möbius masses in multicriteria decision making [see e.g., 23], the authors use utility parameters $w(S)$ to account for synergies between subsets S of items. More precisely, the utility $u(A)$ of a set A is defined as $u(A) = \sum_{S \subseteq A} w(S)$, and a set A is preferred to a set B if $u(A) > u(B)$. All three works aim to propose a method for learning such a utility function from a set of pairwise preference statements. First, Domshlak and Joachims proposed an SVM approach [11] for this purpose, relying on the *kernel trick* [36] for the efficiency of the method. For their part, Bigot et al. gave a probably approximately correct (PAC) method to learn a compact decomposition of the utility function into parameters $w(S)$. The method is polynomial-time if a constant bound is known on the size $|S|$ of the greatest subset S considered in the expression of the utility function. Finally, Gilbert et al. recently proposed a learning method that uses both a Gaussian process method [10] to learn the decomposition into parameters $w(S)$ and a robust ordinal regression method [24, 28] to predict preferences.

To the best of our knowledge, the question of taking intransitivities into account in preferences between sets has not been addressed in the AI literature to date. Yet it is well known that intransitivities in preferences can occur in many real-world situations, as illustrated by Condorcet's paradox in voting (a majority of voters may prefer A to B, B to C and C to A), which shows that there may be cycles in strict preferences [21], or by Luce's example of sugar grains in coffee [18, 31], which shows that indifference may be intransitive (think of a sugarless coffee in which you add grains of sugar one by one).

In an attempt to fill this gap, our contributions are the following: 1) We propose a new model for representing possibly intransitive preferences over sets (Section 2), largely inspired by the SSB (Skew-Symmetric Bilinear utility) model proposed by Fishburn [19] for comparing lotteries in the setting of decision under risk. We adapt the model to the comparison of sets, and we extend it to account for synergies between items. 2) We provide a method to learn the model parameters from a set of observed preference intensities over sets (Section 3). Our proposed method aims to produce a model as simple and compact as possible. We then apply it to synthetic and real-world data to assess the quality of the learned model. 3) We study the complexity of determining the winning sets according to different rules in the tournament on sets induced by the SSB model, establishing both negative (Section 4) and positive complexity results (Sections 5 and 6). We conclude by giving some research directions for future work.

2 The SSB Model with Synergies

We consider a set $\mathcal{E}$ of n elements, from which we derive a set $\mathcal{A} \subseteq 2^{\mathcal{E}}$ of N subsets. To express preferences on sets, a standard tool is the linear utility model, i.e., given a weighting function $w : \mathcal{E} \to \mathbb{R}$, the value of a set $A \in \mathcal{A}$ is $u(A) = \sum_{e \in A} w(e)$ and A is preferred (resp. indifferent) to a set B, denoted by $A \succ B$ (resp. $A \sim B$), iff $u(A) > u(B)$ (resp. $u(A) = u(B)$). As mentioned above, we consider in this work preferences that may present some intransitivities, which prevent from resorting to the standard utility model.

Example 1. *A classic example, often attributed to Armstrong [1] and reported by Lehrer and Wagner [29], is the following: a child (say a girl) may receive as a present for her birthday either a bicycle or a pony and is indifferent between the two options. If a bell is added to the bicycle, then this yields a better gift than the bicycle alone. However, the child would still be indifferent between the pony and the bicycle with a bell. Indeed, the bell is not a sufficient add-on to justify in her eyes that the bicycle should be preferred to the pony.*

This problem can be modeled by considering a set of three elements $\mathcal{E} = \{P, B, b\}$, in which P, B and b correspond to Pony, Bicycle and bell respectively, and three subsets $A_1 = \{P\}$, $A_2 = \{B\}$, and $A_3 = \{B, b\}$, defining the alternative set $\mathcal{A} = \{A_1, A_2, A_3\}$. The preferences of the child are then $A_1 \sim A_2$, $A_1 \sim A_3$ and $A_3 \succ A_2$.

Note that the preferences of Example 1 are incompatible with the linear utility model. Indeed, the first two preference statements would imply $u(A_2) = u(A_1) = u(A_3)$ and we would obtain $A_2 \sim A_3$, contradicting the last preference statement. One possibility to overcome this incompatibility is, instead of using a univariate function u associating a utility $u(A)$ to each $A \in \mathcal{A}$, to resort to a *bivariate* function $\varphi(A, B)$ reflecting the intensity of preference of A over B, with $\varphi(A, B) > 0$ (resp. $\varphi(A, B) = 0$) if $A \succ B$ (resp. $A \sim B$). It is then possible to have $\varphi(A_1, A_2) = 0$, $\varphi(A_1, A_3) = 0$ and $\varphi(A_3, A_2) > 0$, which accounts for the preferences of Example 1. More specifically, we may consider the following formula for φ:

$$\varphi(A, B) = \sum_{e \in A} w(e) - \sum_{e \in B} w(e) + \sum_{e_1 \in A} \sum_{e_2 \in B} \psi(e_1, e_2) \qquad (1)$$

where the parameters $\psi(e_1, e_2)$ reflect the intensity of preference of having e_1 (in A) "against" e_2 (in B). In line with this interpretation, the function ψ is skew-symmetric, i.e., $\psi(e_2, e_1) = -\psi(e_1, e_2)$.

Example 2. *To account for the preferences in Example 1, we can set $w(P) = w(B) = 5$, $w(b) = 1$, $\psi(b, P) = -1$ and $\psi(e_1, e_2) = 0$ otherwise. The value $\psi(b, P) = -w(b)$ reflects that the bell is not an add-on compared to a pony. We have then $\varphi(A_1, A_2) = 5 - 5 = 0$, $\varphi(A_1, A_3) = 5 - (5 + 1) + 1 = 0$ and $\varphi(A_3, A_2) = (5 + 1) - 5 = 1$.*

Nevertheless, this is not fully satisfying as, by adding $A_4 = \{P, b\}$ to $\mathcal{A}$, we obtain $\varphi(A_4, A_2) = (5+1) - 5 > 0$. Put another way, a pony with a bell would be preferred to a bike while we would want to be indifferent. Setting $\psi(b, B) = -1$ is not a solution because we would then have $\varphi(A_3, A_2) = (5 + 1) - 5 - 1 = 0$. A way to overcome this is to introduce synergy terms of the form $w(S)$, which yields the following formula for φ:

$$\varphi(A, B) = \sum_{S \subseteq A} w(S) - \sum_{S \subseteq B} w(S) + \sum_{e_1 \in A} \sum_{e_2 \in B} \psi(e_1, e_2).$$

Example 3. *Following on from Example 2, we can set $w(\{B, b\}) = 1$, $w(\{P, b\}) = 0$, and $w(b) = 0$ instead of 1 to obtain $A_1 \sim A_2$, $A_1 \sim A_3$, $A_3 \succ A_2$ and $A_4 \sim A_2$. This change in parameters[1] allows us to express that a bell is an add-on to a bike but not to a pony.*

But then A_1 is preferred to A_4, i.e., a pony alone is preferred to a pony with a bell, even though he who can do more can do less. This can be solved by a further change in the formula of φ:

$$\varphi(A, B) = \sum_{S \subseteq A} w(S) - \sum_{S \subseteq B} w(S) + \sum_{S_1 \subseteq A} \sum_{S_2 \subseteq B} \psi(S_1, S_2).$$

The parameters $\psi(S_1, S_2)$ enable us to express exactly that adding a bell *to a bike* is not enough to create a preference *over a pony*.

Example 4. *Following on from Example 3, we can set $w(P) = 5$, $w(B) = 5$, $w(\{B, b\}) = 1$, $\psi(\{B, b\}, P) = -1$, and all other parameters equal to 0 to obtain the exact set of preferences we want.*

Connection to tournament theory. The preferences expressed by function φ on a set $\mathcal{A} = \{A_1, \ldots, A_N\}$ of alternatives define a *weighted tournament*[2] $(\mathcal{A}, \Phi)$, where Φ is the skew-symmetric matrix defined by $\Phi_{ij} = \varphi(A_i, A_j)$ if $i \neq j$, else 0. This tournament can be represented by a valued digraph $G = (V, E, v)$ where $V = \mathcal{A}$, $E = \{(A_i, A_j) : \Phi_{ij} > 0\}$, and $v(A_i, A_j) = \Phi_{ij}$ $\forall (A_i, A_j) \in E$. In other words, there is one vertex per alternative and an edge from A_i to A_j iff $A_i \succ A_j$, valued by the preference intensity of A_i over A_j.

Example 5. *Back to the example: if we set $\mathcal{A} = 2^{\mathcal{E}} \setminus \{\emptyset\}$, and we define φ by setting $w(B) = w(P) = 5$, $w(\{B, b\}) = 1$, $\psi(P, \{B, b\}) = 1$, and $\psi(\{P, B, b\}, P) = 1$, we obtain the tournament in Figure 1.*

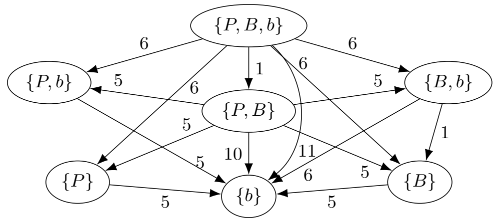

Figure 1. The tournament obtained in Example 5.

As seen in Examples 1 to 5, many model parameters remain at 0. Hence, to keep the representation *sparse*, we circumscribe the set of parameters of the model to a subset $\theta = \{S_1, \ldots, S_t\} \subseteq 2^{\mathcal{E}} \setminus \{\emptyset\}$, to finally obtain the *SSB model* (with synergies) on sets we propose to consider in this paper. In the following, we assume that the parameters $\psi(S_i, S_j)$ (with i and j in $\{1, \ldots, t\}$) are stored in a matrix Ψ defined by $\Psi_{ij} = \psi(S_i, S_j)$ if $i \neq j$ and 0 otherwise. Furthermore, we abbreviate $w(S_i)$ by ω_i and we denote by $I_\theta(A)$ the indices of subsets $S_i \in \theta$ included in A, i.e., $I_\theta(A) = \{i \in \{1, \ldots, t\} : S_i \subseteq A\}$.

Definition 1. *Given a set $\mathcal{E}$ of elements, a set $\theta = \{S_1, \ldots, S_t\} \subseteq 2^{\mathcal{E}} \setminus \{\emptyset\}$, a vector ω encoding a weighting function $w : \theta \to \mathbb{R}$ and a matrix Ψ encoding a skew-symmetric bivariate function $\psi : \theta^2 \to \mathbb{R}$, the preference intensity of A over B according to the SSB model is*

$$\varphi_{\theta,\omega,\Psi}(A, B) = \sum_{i \in I_\theta(A)} \omega_i - \sum_{i \in I_\theta(B)} \omega_i + \sum_{i \in I_\theta(A)} \sum_{j \in I_\theta(B)} \Psi_{ij}.$$

In the next section, given a skew-symmetric matrix M obtained from a set of known preference intensities between sets in $\mathcal{A}$, we investigate how to learn the parameters of an SSB model accounting for the weighted tournament $(\mathcal{A}, M)$.

[1] To alleviate the notations, the braces are omitted for singletons.

[2] Strictly speaking, we consider *weak* tournaments, as we may have $A_i \neq A_j$ for which $\varphi(A_i, A_j) = 0$. For brevity, we omit the term "weak".

3 Learning the Parameters of an SSB Model

We first give a constructive way to find an SSB model to account for a set of observed preference intensities over subsets. Indeed, while we have seen that the SSB model induces a weighted tournament on the alternative set $\mathcal{A}$, we now show conversely that, given a skew-symmetric matrix M defining a weighted tournament $(\mathcal{A}, M)$, there always exists an SSB model φ such that $(\mathcal{A}, \Phi) = (\mathcal{A}, M)$.

Theorem 1. *Let $\mathcal{E}$ be a set of n elements, $\mathcal{A} = \{A_1, \dots, A_N\} \subseteq 2^{\mathcal{E}}$ be an alternative set, and $(\mathcal{A}, M)$ be a weighted tournament on $\mathcal{A}$. Without loss of generality, assume that $A_N = \emptyset$ if $\emptyset \in \mathcal{A}$.*

Let $N' = 2^n - 1$. By setting $\theta = 2^{\mathcal{E}} \setminus \{\emptyset\} = \{S_1, \dots, S_{N'}\}$ with $S_i = A_i$ $\forall i \in \{1, \dots, N - \mathbb{1}_{\mathcal{A}}(\emptyset)\}$ (where $\mathbb{1}_{\mathcal{A}}(\emptyset) = 1$ if $\emptyset \in \mathcal{A}$, else 0), and for $i \in \{1, \dots, N'\}$ and $j \in \{1, \dots, N'\}$

$$\omega_i = \begin{cases} \sum_{k \in I_{\mathcal{A}}(S_i)} (-1)^{|S_i \setminus A_k|} M_{kN} & \text{if } \emptyset \in \mathcal{A}, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

$$\Psi_{ij} = \sum_{k \in I_{\mathcal{A}}(S_i)} \sum_{l \in I_{\mathcal{A}}(S_j)} (-1)^{|S_i \setminus A_k| + |S_j \setminus A_l|} M_{kl}, \tag{3}$$

where $I_{\mathcal{A}}(S) = \{k \in \{1, \dots, N\} : A_k \subseteq S\}$, we obtain that:

$$\forall A_p, A_q \in \mathcal{A}, \quad \varphi_{\theta,\omega,\Psi}(A_p, A_q) = M_{pq}.$$

Proof. This is a generalization of the Möbius transform of a capacity function [see e.g., 23] to the case of a binary function. Let us first assume that $\emptyset \notin \mathcal{A}$. We consider the extended tournament $(\mathcal{A}', M')$ where $\mathcal{A}' = \{A_1, \dots, A_N, A_{N+1}, \dots, A_{N'}\} = 2^{\mathcal{E}} \setminus \{\emptyset\}$ with $A_i = S_i$ $\forall i \in \{N+1, \dots, N'\}$, $M'_{ij} = M_{ij}$ for $i, j \in \{1, \dots, N\}$, and $M'_{ij} = 0$ otherwise. As $\omega_i = 0$ for $i \in \{1, \dots, N'\}$, we have by Definition 1:

$$\begin{aligned}
\varphi_{\theta,\omega,\Psi}(A_p, A_q) &= \sum_{i \in I_\theta(A_p)} \sum_{j \in I_\theta(A_q)} \Psi_{ij} \\
&= \sum_{i \in I_\theta(A_p)} \sum_{j \in I_\theta(A_q)} \sum_{k \in I_{\mathcal{A}}(S_i)} \sum_{l \in I_{\mathcal{A}}(S_j)} (-1)^{|S_i \setminus A_k| + |S_j \setminus A_l|} M_{kl} \\
&= \sum_{i \in I_{\mathcal{A}'}(A_p)} \sum_{j \in I_{\mathcal{A}'}(A_q)} \sum_{k \in I_{\mathcal{A}'}(S_i)} \sum_{l \in I_{\mathcal{A}'}(S_j)} (-1)^{|S_i \setminus A_k| + |S_j \setminus A_l|} M'_{kl}
\end{aligned}$$

where we have used Equation 3 on the second line, and the fact that $\mathcal{A}' = \theta$ and $M'_{kl} = 0$ when k or l does not belong to $\{1, \dots, N\}$ on the third line. As $[k \in I_{\mathcal{A}'}(S_i)$ and $i \in I_{\mathcal{A}'}(A_p) \Leftrightarrow A_k \subseteq S_i = A_i \subseteq A_p]$, and $[l \in I_{\mathcal{A}'}(S_j)$ and $j \in I_{\mathcal{A}'}(A_q) \Leftrightarrow A_l \subseteq S_j = A_j \subseteq A_q]$, the coefficient of each term M'_{kl} in this formula can be written:

$$\begin{aligned}
&\sum_{S_i} \sum_{S_j} (-1)^{|S_i \setminus A_k| + |S_j \setminus A_l|} \text{ for } A_k \subseteq S_i \subseteq A_p \text{ and } A_l \subseteq S_j \subseteq A_q \\
&= \sum_{S_i : A_k \subseteq S_i \subseteq A_p} (-1)^{|S_i \setminus A_k|} \Big(\sum_{S_j : A_l \subseteq S_j \subseteq A_q} (-1)^{|S_j \setminus A_l|} \Big).
\end{aligned}$$

The binomial theorem states that $(x+y)^n = \sum_{t=0}^{n} \binom{n}{t} x^{n-t} y^t$. Taking $x = 1$ and $y = -1$, we know that $\sum_{t=0}^{n} \binom{n}{t} (-1)^t = (1-1)^n = 0$ for $n \in \mathbb{N}^*$. Noting that the expression within the parentheses in the above formula is equal to $\sum_{t=0}^{|A_q \setminus A_l|} \binom{|A_q \setminus A_l|}{t} (-1)^t = 0$ (by the previous argument with $n = |A_q \setminus A_l|$), we deduce that the coefficient of M'_{kl} is 0 for $l \neq q$ (so that $A_l \neq A_q$). Similarly, we can show that the coefficient of M'_{kl} is also 0 for $k \neq p$. The only exception is the case $(k, l) = (p, q)$, for which the coefficient of M'_{pq} is $(-1)^0 = 1$. Thus $\varphi_{\theta,\omega,\Psi}(A_p, A_q) = M'_{pq} = M_{pq}$ for $p, q \in \{1, \dots, N\}$, which concludes the proof in the case where $\emptyset \notin \mathcal{A}$.

When $\emptyset \in \mathcal{A}$, we apply the same proof with $\theta' = \theta \cup \{\emptyset\}$ and the following formula:

$$\varphi_{\theta,\omega,\Psi}(A_p, A_q) = \sum_{i \in I_{\theta'}(A_p)} \sum_{j \in I_{\theta'}(A_q)} \Psi_{ij}.$$

Once the parameters Ψ_{ij} that satisfy this formula are determined for $i, j \in \{1, \dots, N\}$, the expression of $\varphi_{\theta,\omega,\Psi}(A_p, A_q)$ as stated in Definition 1 can indeed be recovered by setting $\omega_i = \Psi_{iN}$ for $i \in \{1, \dots, N'\}$, and keeping unchanged the values Ψ_{ij} for $i, j \in \{1, \dots, N'\} \setminus \{N\}$ because $\varphi_{\theta,\omega,\Psi}(A_p, A_q)$ can be rewritten as:

$$\sum_{i \in I_\theta(A_p)} \Psi_{iN} + \sum_{i \in I_\theta(A_q)} \Psi_{Ni} + \sum_{i \in I_\theta(A_p)} \sum_{j \in I_\theta(A_q)} \Psi_{ij}.$$

The parameter $\omega_i = \Psi_{iN}$ is equal to:

$$\sum_{k \in I_{\mathcal{A}}(S_i)} \sum_{l \in I_{\mathcal{A}}(\emptyset)} (-1)^{|S_i \setminus A_k| + |\emptyset \setminus A_l|} M_{kl} = \sum_{k \in I_{\mathcal{A}}(S_i)} (-1)^{|S_i \setminus A_k|} M_{kN}$$

Hence, we obtain formulas for parameters ω_i and Ψ_{ij} that correspond to Equations 2 and 3 in Theorem 1. □

Theorem 1 shows how to construct an SSB model to account for a set of observed preferences over subsets. Sadly, it uses $\theta = 2^{\mathcal{E}} \setminus \{\emptyset\}$ and thus yields a model with a possibly prohibitive number of parameters. We now investigate how to learn a sparser representation.

To put it formally, given a set $\theta = \{S_1, \dots, S_t\} \subseteq 2^{\mathcal{E}} \setminus \{\emptyset\}$, we outline a method to derive an SSB model (i.e., learn ω and Ψ) from a set of m weighted preferences $R = \{(A_k, B_k, p_k) : 1 \leq k \leq m\}$, where $(A, B, p) \in R$ means that we have observed that A is preferred to B with intensity p. We assume that R is *noise free*, in the sense that a possible incompatibility of the model with R comes from the parameter set θ used and not from errors in the observed preferences.

As θ is fixed beforehand, and since we have no guarantee that a given θ yields a model $\varphi_{\theta,\omega,\Psi}$ that can account perfectly for preferences in R, we introduce a gap variable $\varepsilon_{A,B}$ for each preference (A, B, p) in R [see e.g., 11]. We obtain the following constraints:

$$\sum_{i \in I_\theta(A)} \omega_i - \sum_{i \in I_\theta(B)} \omega_i + \sum_{i \in I_\theta(A)} \sum_{j \in I_\theta(B)} \Psi_{ij} = p + \varepsilon_{A,B} \tag{4}$$

for all $(A, B, p) \in R$. We add the following set of constraints to ensure that Ψ is skew-symmetric:

$$\Psi_{ij} = -\Psi_{ji} \text{ for all } (i, j) \in \{1, \dots, t\}^2. \tag{5}$$

Let $\mathcal{P}_\theta^R$ be the polyhedron defined by Constraints 4–5 on $[\omega, \Psi, \epsilon]$.

Learning a compact model. To compute a specific model, following the approach proposed by [22] to implement Occam's razor principle, we define objectives that are lexicographically minimized over $\mathcal{P}_\theta^R$. We first aim to minimize the sum of gap variables $\varepsilon_{A,B}$:

$$\mathcal{L}_\varepsilon = \sum_{(A,B) \in R} \varepsilon_{A,B}$$

Compactness involves minimizing non-zero parameters, requiring loss functions for both additive (ω) and bilinear (Ψ) parameters. We opt for the L_1 loss function, defined for ω (left) and Ψ (right) by:

$$\mathcal{L}_\omega = \sum_{i=1}^{t} |\omega_i| \qquad \mathcal{L}_\Psi = \sum_{i=1}^{t} \sum_{j=1}^{t} |\Psi_{ij}|$$

Moreover, prioritizing the use of additive coefficients over bilinear ones, given their expressive yet less versatile nature, demands separate minimization objectives in a lexicographic order, starting with the loss on gap variables. This yields the following procedure:

1. Compute $\mathcal{L}_\varepsilon^* = \min_{[\omega,\Psi,\varepsilon] \in \mathcal{P}_\theta^R} \mathcal{L}_\varepsilon$.
2. Compute $\mathcal{L}_\omega^* = \min_{[\omega,\Psi,\varepsilon] \in \mathcal{P}_\theta^R} \mathcal{L}_\omega$ s.t. $\mathcal{L}_\varepsilon = \mathcal{L}_\varepsilon^*$.
3. Compute $[\omega^*, \Psi^*] = \operatorname{argmin}_{[\omega,\Psi,\varepsilon] \in \mathcal{P}_\theta^p} \mathcal{L}_\Psi$ s.t. $\mathcal{L}_\varepsilon = \mathcal{L}_\varepsilon^*$ and $\mathcal{L}_\omega = \mathcal{L}_\omega^*$.

We now present the results of numerical tests on synthetic and real-world data. The code is available online [33].

Numerical tests on synthetic data. The tests were conducted with $n = 6$. We consider two types of synthetic instances, the *Gaussian instances* and the *SSB instances*. Let $[\mathcal{E}]^k = \{A : A \subseteq \mathcal{E}, 1 \leq |A| \leq k\}$. The set of alternatives is $\mathcal{A} = [\mathcal{E}]^k$ (for a given k) in Gaussian instances, and $\mathcal{A} = 2^{\mathcal{E}}$ in SSB instances. For both types of instances, to define a weighted tournament on $\mathcal{A}$, a weighted undirected graph of vertex set $\mathcal{A}$ is first generated using the Erdös-Rényi model [17], where each edge is included in the graph with probability p. The difference lies in the generation of edge weights. In Gaussian instances, the weight M_{ij} of each edge $\{A_i, A_j\}$ $(i < j)$ is randomly drawn using a normal distribution: $M_{ij} \sim \mathcal{N}(0, 1)$. In SSB instances, the weight is the value of a randomly drawn SSB function (common to all edges): $M_{ij} = \varphi_{\theta,\omega,\Psi}(A_i, A_j)$, where $\theta = [\mathcal{E}]^k$, $\omega_i \sim \mathcal{N}(0, 1)$ and $\Psi_{ij} \sim \mathcal{N}(0, 1)$. We call k the *degree* of the SSB model. In both cases (Gaussian and SSB), edges are directed according to the sign of M_{ij} (from A_i to A_j if $M_{ij} > 0$, otherwise the converse).

Our experiments on synthetic data aim to measure how well an SSB model of degree d can replicate the arc weights for both types of instances. For this purpose, we use a variant of the r^2 score, namely the score $1 - (\sum_{i<j} |M_{ij} - \hat{\Phi}_{ij}|) / \sum_{i<j} |M_{ij}|$ (the closer to 1, the better the fit), where $\hat{\Phi}$ is the learned SSB model. This score is averaged over 10 randomly drawn instances for each instance type and tuple (p, k, d), with $p \in \{0.1, 0.6\}$, $k \in \{2, \ldots, 5\}$ and $d \in \{1, \ldots, k\}$.

The results are presented in Figure 2. The top (resp. bottom) diagrams are obtained for Gaussian (resp. SSB) instances with $p = 0.1$ (left) and $p = 0.6$ (right). As expected, increasing the density of the graph (with p) complicates the representation task, as witnessed by the Gaussian instances with $p = 0.1$ and $d \geq 3$, for which the scores do not exceed 0.1. Nevertheless, as soon as the weights M_{ij} are functions of A_i and A_j (SSB instances), even low-degree SSB model can replicate them very satisfactorily, as witnessed by the scores that are all greater than 0.8 for $d \geq 2$ in the bottom diagrams, even if $p = 0.6$ (dense graph). Overall, the SSB model demonstrates good replication power for the synthetic data considered, as can be seen by considering the rows $d = 3$ in the different diagrams.

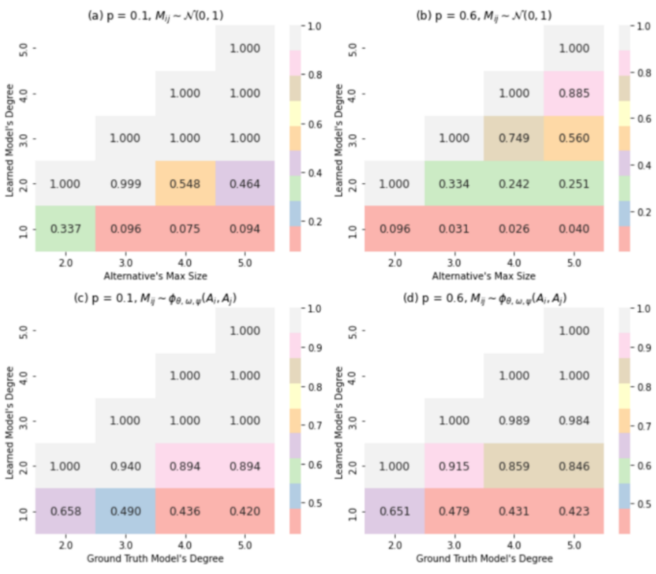

Figure 2. Mean scores on synthetic data with $n = 6$.

Numerical tests on real-world data. The tests were conducted on Kaggle data [9] from 81,272 online games of "League of Legends" (Season 8). The data were gathered using Riot's public API [35]. In this video game, each player selects a character from one of five distinct classes and engages in combat as two teams of five individuals. Each team must comprise a character from each class, ensuring a balanced composition. The meta-game, a term that players use to describe the current strategic trends and power balances among characters, is often quantified by the "win rate"—a metric that reflects the probability of winning a game when selecting a particular character.

Yet, only considering win rates between individual characters has its limits. It fails to account for synergies between the characters and overlooks the nuanced strength dynamics that emerge when certain characters face off against one another. Our tests aim precisely to assess the benefits of using an SSB model to explain win rates between character *combinations* and to compare the relative impact of interaction parameters $w(S)$ (with $|S| \geq 2$) and bilinear parameters $\psi(e_i, e_j)$ in the explanatory power of the SSB model.

For this purpose, we leverage data from the game to build a weighted tournament $T = (\mathcal{A}, M)$ on characters and pairs of characters, where M_{ij} reflects the win rate of a team comprising A_i against a team comprising A_j. The set $\mathcal{E}$ corresponds to the characters. Note that not all subsets of $\mathcal{E}$ of size 1 or 2 are feasible due to the distinct classes of characters. Formally, we have $\mathcal{A} = [\mathcal{E}]^1 \cup (\cup_{i \neq j}\{\{e, e'\} : e \in X_i, e' \in X_j\})$, where X_i is the set of characters of class i. In the definition of the tournament, we restrict ourselves to the 5 most played characters for each class i ($|X_i| = 5$ for $i \in \{1, \ldots, 5\}$). The Kaggle dataset synthesizes the results of numerous games that were played online between various teams of 5 characters, from which we derive the weights M_{ij} as follows. Let W_{ij} be the win rate computed from the data by examining the outcomes of games between teams comprising A_i against teams comprising A_j, viewed as an estimate of the binomial probability p_{ij} that the former wins against the latter. We use the Wilson score interval [37] to produce a 90% confidence interval $[\alpha_{ij}, \beta_{ij}]$ for p_{ij}. An arc from A_i to A_j (resp. A_j to A_i) of weight $M_{ij} = W_{ij} - 0.5$ (resp. $M_{ji} = 0.5 - W_{ij}$) is included in T if $\alpha_{ij} > 0.5$ (resp. $\beta_{ij} < 0.5$), otherwise there is no arc between them.

From this tournament, we learn 1) an SSB model where $\theta = [\mathcal{E}]^1$ (that we call 1-SSB model), 2) a linear additive model that corresponds to a degenerate SSB model where $\theta = [\mathcal{E}]^1$ and Ψ is the null matrix (1-ADD model), and 3) an SSB model with $\theta = [\mathcal{E}]^2$ and a null Ψ matrix (2-ADD model). The 1-ADD model involves n parameters, while the 2-ADD and 1-SSB models involve $n(n+1)/2$ parameters. Players and game designers widely use the 1-ADD model. However, it does not account for the synergies and antagonisms between characters. The 1-SSB (resp. 2-ADD) model only accounts for antagonisms (resp. pairwise synergies).

The obtained scores (with the same formula as for synthetic data) are given in Table 1. We can see that model 1-ADD performs poorly, which may indicate that its failure to account for synergies and intransitivities significantly restricts its ability to fit the data. In contrast, model 2-ADD achieves a good score. Nonetheless, its inability to achieve a score of 1 shows that it cannot perfectly fit the tournament data without incorporating intransitivities via bilinear coefficients (which would be sufficient because $\mathcal{A}$ involves only singletons and pairs). Model 1-SSB shows the best performance. In the remainder of the paper, devoted to studying the complexity of determining the winning sets in a tournament induced by an SSB model, we show that it also provides some computational complexity advantages.

Learned Model	1-SSB	1-ADD	2-ADD
Scores	0.94	0.16	0.80

Table 1. Scores on real-world data from League of Legends.

4 Negative Complexity Results

This section studies the problem of finding an "optimal" alternative in $\mathcal{A}$ given an SSB model. As the SSB model induces a tournament on $\mathcal{A}$, which is combinatorial in nature, determining an optimal subset raises the question of the complexity of implementing a weighted tournament solution on a tournament (implicitly) defined by the SSB model. A weighted tournament solution [for a review, see e.g., 6] is a function γ that associates to a weighted tournament (X, M) a non-empty subset of alternatives $\gamma(X, M) \subseteq X$, called the *choice set*. We call TOURNAMENT-ALL/ONE-γ (resp. SSB-ALL/ONE-γ) the problem of computing all/one alternative(s) of $\gamma(X, M)$ in a weighted tournament (resp. a tournament defined by an SSB model).[3]

> TOURNAMENT-ALL/ONE-γ
> **INPUT:** A weighted tournament (X, M), where X is a set of alternatives and M is a skew-symmetric matrix.
> **OUTPUT:** All/one alternative(s) of $\gamma(X, M)$.

> SSB-ALL/ONE-γ
> **INPUT:** An element set $\mathcal{E}$, an alternative set $\mathcal{A} \subseteq 2^{\mathcal{E}}$, a parameter set $\theta \subseteq 2^{\mathcal{E}} \setminus \{\emptyset\}$, a weight vector $\omega \in \mathbb{Q}^{|\theta|}$, and a bilinear coefficient matrix $\Psi \in \mathbb{Q}^{|\theta| \times |\theta|}$ that yield a tournament $T = (\mathcal{A}, \Phi)$.
> **OUTPUT:** All/one alternative(s) of $\gamma(\mathcal{A}, \Phi)$.

Several comments need to be made to clarify the complexity analyses in the sequel. Note that in the SSB-ALL/ONE-γ problem, the size of the input depends on the way the alternative set $\mathcal{A}$ is represented and on the size of θ. If $\mathcal{A}$ is represented explicitly, the size of the input may be as large as 2^n while it can be polynomial in n when represented implicitly (typically, by linear constraints, e.g., $A \in \mathcal{A}$ iff $|A| = k$ for parameter k). In the latter case, the number of alternatives in $\gamma(\mathcal{A}, \Phi)$ may hence be non-polynomial in the input size. However, this may not preclude the polynomiality of the SSB-ALL-γ problem as a compact representation of $\gamma(\mathcal{A}, \Phi)$ may be achievable.

We start with the following result showing that, even when $\mathcal{A}$ is represented explicitly, the problem of computing a solution in a weighted tournament defined by an SSB model is at least as hard as its version in a regular weighted tournament.

Theorem 2. *If no specific assumption is made on θ, ω and Ψ, then problem* SSB-α-γ *is at least as hard as problem* TOURNAMENT-α-γ *for* $\alpha \in \{$ALL,ONE$\}$.

Proof. We show that there is a polynomial-time reduction from TOURNAMENT-α-γ to SSB-α-γ for $\alpha \in \{$ALL,ONE$\}$. Given a weighted tournament (X, M), let $n \in \mathbb{N}^*$ be such that $2^n \geq |X| > 2^{n-1}$. We define a set $\mathcal{E}$ of n elements and an alternative set $\mathcal{A} \subseteq 2^{\mathcal{E}}$ such that $|\mathcal{A}| = |X|$, and consider the weighted tournament $(\mathcal{A}, M)$. Note that, in this way, there is a one-to-one correspondence between alternatives in X and alternatives in $\mathcal{A}$. Theorem 1 showed that, by setting $\theta = 2^{\mathcal{E}} \setminus \{\emptyset\}$, we can find ω and Ψ such that:

$$\forall A_i, A_j \in \mathcal{A}, \quad \varphi_{\theta,\omega,\Psi}(A_i, A_j) = M_{ij}.$$

Computing ω_i for all $S_i \in \theta$ and Ψ_{ij} for all $S_i, S_j \in \theta$ can be done in polynomial time in $|A|$ by using Equations 2 and 3. Indeed, there are $O(2^n) \equiv O(|X|)$ values ω_i (resp. $O(2^n \cdot 2^n) \equiv O(|X|^2)$ values Ψ_{ij}) to compute, each of which is a weighted sum of $O(|X|)$ operands (resp. $O(|X|^2)$ operands) M_{kl}. The weight of each summand is 1 or -1 according to the parity of the exponent, which is computed in $O(n)$ because the exponent is upper bounded by n. □

In the previous result, the SSB-tournament did not necessarily provide a compact formulation of the tournament. By *compact formulation*, we mean that the size of the representation of $\mathcal{A}$ and the size of θ are polynomial in the number n of elements in $\mathcal{E}$.

Regarding the representation of $\mathcal{A}$, we focus on the two following cases, that we call the *complete* and *partition* cases:

- The case $\mathcal{A} = 2^{\mathcal{E}}$, i.e., each subset of $\mathcal{E}$ is a valid alternative (as in the SSB instances of synthetic data), for its simple definition.
- The case $\mathcal{A} = \{\{e_1, \ldots, e_p\} : (e_1, \ldots, e_p) \in X_1 \times \ldots \times X_p\}$, where $X_1, \ldots, X_p$ is a partition of $\mathcal{E}$, i.e., $X_i \cap X_j = \emptyset$ and $\bigcup_{i=1}^p X_i = \mathcal{E}$ (as in our real-world data), which occurs for instance in the configuration of complex products [4, 7].

Regarding θ, a particular case in which its size is polynomial in n occurs when θ only contains sets of cardinalities upper bounded by a constant. This is likely to happen in practice as we expect synergies to appear between small groups of elements. Let $[\mathcal{E}]^k = \{S : S \subseteq \mathcal{E}, 1 \leq |S| \leq k\}$ contain all non-empty sets of elements of size lower than or equal to k. When k is a small constant, the parameters θ, ω, Ψ characterizing the SSB model provide a compact representation of the tournament matrix Φ.

Unfortunately, we show that computing an alternative in the choice set for most tournament solutions is hard in the complete case for $\theta = [\mathcal{E}]^k$ with a constant $k \geq 2$, namely tournament solutions which refine *Top Cycle*. The Top Cycle (TC) is a well-known tournament solution that returns the unique smallest dominant set of alternatives, i.e., a set $Y \subseteq X$ of alternatives such that each alternative in Y is preferred to all alternatives in $X \setminus Y$. TC is a weak tournament solution in the sense that most tournament solutions γ are such that $\gamma(X, M) \subseteq TC(X, M)$ [see 6, p. 76], abbreviated by $\gamma \subseteq TC$ in the sequel. We have the following negative result, which is a direct consequence of a result by Fishburn and LaValle [20]:

Theorem 3. *For $k \geq 2$ a fixed constant, $\theta = [\mathcal{E}]^k$, and $\gamma \subseteq TC$, there is no polynomial-time algorithm to solve* SSB-ONE-γ *in the complete case, unless $P = NP$.*

Proof. Fishburn and Lavalle [20, p.189] have proven that, given a weighting function $w : [\mathcal{E}]^2 \to \mathbb{Q}$, determining $A \subseteq \mathcal{E}$ that maximizes $\sum_{S \in [\mathcal{E}]^2} w(S) I_A(S)$, where $I_A(S) = 1$ if $S \subseteq A$ and 0 otherwise, is an NP-hard problem. Indeed, we can easily make a reduction from the NP-complete maximum independent set problem. This corresponds to the special case of SSB-ONE-γ where $\mathcal{A} = 2^{\mathcal{E}}$, $\theta = [\mathcal{E}]^2$, ω encodes w, and Ψ is the null matrix. Note that in this case the alternatives in $TC(\mathcal{A}, \Phi)$ are exactly the ones maximizing $\sum_{S \in \theta} w(S) I_A(S)$. As k is a constant, it takes a polynomial time in n to set $\theta = [\mathcal{E}]^k$ and $w(S) = 0$ for $|S| > 2$ (Ψ is the null matrix). Thus the reduction is polynomial-time. □

Due to Theorem 3, to establish positive complexity results, we restrict our attention to the case in which $\theta = [\mathcal{E}]^1$. Hence, φ may have bilinear coefficients Ψ_{ij} but no synergy terms and writes as Equation 1. Considering the complete and partition cases to represent $\mathcal{A}$ will make it possible to give tractability results for two well-known weighted tournament solutions: the Borda set and the essential set.[4]

5 Computing the Borda Set

The first specific weighted tournament solution we consider is the Borda set, which adapts the Copeland set to the weighted case [12].

[3] For complexity reasons, the parameters ω and Ψ are from now on assumed to be rational numbers.

[4] Our intuition is that, for many other tournament rules, even in these restricted settings, the choice set (winners of the tournament) is hard to determine because a concise representation of it is required.

Given a weighted tournament $T=(X,M)$, this weighted tournament solution assigns to each alternative $A_i \in X$ a score $\texttt{Borda}_T(A_i) = \sum_j M_{ij}$ (the Borda score) and returns the set $\texttt{BordaSet}(T) = \arg\max_{A\in X} \texttt{Borda}_T(A)$ of alternatives of maximal score.

For a tournament $T=(\mathcal{A},\Phi)$, the Borda score of $A\in\mathcal{A}$ is:

$$\texttt{Borda}_T(A) = \sum_{B\in\mathcal{A}} \varphi_{\theta,\omega,\Psi}(A,B).$$

Using the structure underlying function $\varphi_{\theta,\omega,\Psi}$, we now show that problem SSB-ALL-BORDA can be solved in polynomial-time in the complete and partition cases when $\theta=[\mathcal{E}]^1$. Indeed, we can compute in polynomial-time, not only one solution of the Borda set, but a compact representation of the Borda set itself.

Theorem 4. SSB-ALL-BORDA *can be solved in time*[5] $O^*(n^2)$ *in the complete and partition cases when* $\theta = [\mathcal{E}]^1$. *The Borda set is characterized by values* $C(i)$ $\forall i\in\mathcal{E}$ *defined by:*

- *Complete case:* $C(i) = 2\omega_i + \sum_{j\in\mathcal{E}} \Psi_{ij}$.
- *Partition case:* $C(i) = (\prod_{t=1}^{p} |X_t|)\omega_i + \sum_{j\in\mathcal{E}} (\prod_{t\in[p]:j\notin X_t} |X_t|)\Psi_{ij}$.

This yields an implicit *compact formulation of the Borda set:*

- *Complete case:* $\texttt{BordaSet}(T)=\{A : C_{>0} \subseteq A \subseteq C_{\geq 0}\}$, *where* $C_{>0}=\{i\in\mathcal{E} : C(i)>0\}$ *and* $C_{\geq 0}=\{i\in\mathcal{E} : C(i)\geq 0\}$.
- *Partition case:* $\texttt{BordaSet}(T)=\prod_{t=1}^{p} \arg\max_{i\in X_t} C(i)$.

Proof. Let $\mathcal{E} = \{1,\dots,n\}$. Consider an SSB induced tournament $T=(\mathcal{A},\Phi)$, with $\theta=[\mathcal{E}]^1=\{S_1,\dots,S_n\}$ with $S_i=\{i\}$. Denoting by $\mathcal{A}_j$ the set $\{A\in\mathcal{A} : j\in A\}$, we rewrite the Borda score as follows:

$$\begin{aligned}\texttt{Borda}_T(A) &= \sum_{B\in\mathcal{A}} \varphi_{\theta,\omega,\Psi}(A,B)\\ &= \sum_{B\in\mathcal{A}} (\sum_{i\in A}\omega_i - \sum_{j\in B}\omega_j + \sum_{i\in A}\sum_{j\in B}\Psi_{ij})\\ &= \sum_{B\in\mathcal{A}}\sum_{i\in A}\omega_i - \sum_{B\in\mathcal{A}}\sum_{j\in B}(\omega_j - \sum_{i\in A}\Psi_{ij})\\ &= \sum_{i\in A}|\mathcal{A}|\omega_i - \sum_{j\in\mathcal{E}}|\mathcal{A}_j|(\omega_j - \sum_{i\in A}\Psi_{ij})\\ &= \sum_{i\in A}|\mathcal{A}|\omega_i + \sum_{i\in A}\sum_{j\in\mathcal{E}}|\mathcal{A}_j|\Psi_{ij} - \sum_{j\in\mathcal{E}}|\mathcal{A}_j|\omega_j\\ &= \sum_{i\in A}\underbrace{(|\mathcal{A}|\omega_i + \sum_{j\in\mathcal{E}}|\mathcal{A}_j|\Psi_{ij})}_{C(i)} - \underbrace{\sum_{j\in\mathcal{E}}|\mathcal{A}_j|\omega_j}_{cst}\end{aligned}$$

where $C(i)$ denotes the contribution of element $i\in\mathcal{E}$ to the Borda Score of alternative $A\in\mathcal{A}$, and with the last summand being a constant cst independent of A. This yields the formula for $C(i)$ given in Theorem 4, with the exception that a 2^{n-1} multiplicative factor is omitted for the complete case. Hence, finding an alternative A maximizing $\texttt{Borda}_t(A)$ amounts to solving the problem $\max_{A\in\mathcal{A}}\sum_{i\in A} C(i)$. The mathematical expressions of the sets of alternatives maximizing this sum are given in the statement of the theorem. These expressions provide a compact representation of the Borda set in the complete and partition cases. The values $C(i)$ can all be computed in time $O^*(n^2)$. Indeed, note that in the partition case, the term $\prod_{t\in[p]}|X_t|$ can be computed in time $O^*(n)$, which then makes it possible to compute values $\prod_{t\in[p]\setminus\{l\}}|X_t|$ for $l\in[p]$ in $O^*(n)$ time. These precomputations can be used to compute each $C(i)$ in $O^*(n)$ time. Moreover note that, in both cases, one element of the Borda set can then be computed in time $O^*(n)$. □

[5] We use the symbol $*$ in O^* to indicate that we omit the logarithmic terms induced by the arithmetic operations in the complexity formula.

6 Computing the Essential Set

The *essential set* [16] generalizes the *bipartisan set* solution concept [27] to weak tournaments. It is defined as the unique largest possible support of a Nash equilibrium in mixed strategies [12] when viewing the matrix M of a (weighted) tournament $T=(X,M)$ as a two-player symmetric zero-sum game. In the following, we denote by $\texttt{EssentialSet}(T)$ the essential set of a tournament T.

In a mixed strategy π, each alternative A_i is played with a probability π_i, where $\pi_i\geq 0$ for $i\in\{1,\dots,N\}$ and $\sum_{i=1}^N \pi_i = 1$. Let $\texttt{sp}(\pi)$ be the indices of the alternatives in the support of π, i.e., the set $\{i\in\{1,\dots,N\} : \pi_i>0\}$. From a complexity viewpoint, the size of π is $O(|\texttt{sp}(\pi)|)$. Let $\Delta(X)$ be the set of mixed strategies over X. Given $\pi\in\Delta(X)$, a best response to π is a mixed strategy $\rho\in\Delta(X)$ that maximizes $\sum_{j\in\texttt{sp}(\rho)}\sum_{i\in\texttt{sp}(\pi)}\rho_j\pi_i M_{ji}$ (i.e., the expected value obtained when ρ is played against π). It is well known that a best response can always be found as a pure strategy, i.e., an alternative $A\in X$; hence we focus here on best responses in pure strategies. Moreover, in a two-player symmetric zero-sum game, a mixed strategy π is part of a Nash equilibrium iff its best response yields a value of 0. To summarize, the essential set corresponds to the maximal support of a mixed strategy $\pi\in\Delta(X)$ for which:

$$\forall A_j\in X,\ \sum_{i\in\texttt{sp}(\pi)}\pi_i M_{ji}\leq 0.$$

In our SSB setting, this corresponds to the maximal support of a mixed strategy $\pi\in\Delta(\mathcal{A})$ for which:

$$\forall A_j\in\mathcal{A},\ \sum_{i\in\texttt{sp}(\pi)}\pi_i\varphi_{\theta,\omega,\Psi}(A_j,A_i)\leq 0.$$

This mixed strategy π induces the essential set defined as:

$$\texttt{EssentialSet}(\mathcal{A},\Phi) = \{A_i\in\mathcal{A} : i\in\texttt{sp}(\pi)\}.$$

As a warm-up, we first study the simpler issue of computing the set $\texttt{BR}_T(\pi)$ of best responses in pure strategies to a given mixed strategy π in a weighted tournament $T=(\mathcal{A},\Phi)$ when $\theta=[\mathcal{E}]^1$.

SSB-ALL/ONE-BESTRESPONSE
INPUT: A set of elements $\mathcal{E}$, a set of alternatives $\mathcal{A}\subseteq 2^{\mathcal{E}}$, a parameter set $\theta\subseteq 2^{\mathcal{E}}\setminus\{\emptyset\}$, a weight vector $\omega\in\mathbb{Q}^{|\theta|}$, a bilinear coefficient matrix $\Psi\in\mathbb{Q}^{|\theta|\times|\theta|}$ that yield a tournament $T=(\mathcal{A},\Phi)$, and a mixed strategy π defined on $\mathcal{A}$.
OUTPUT: All/one alternative(s) of $\texttt{BR}_T(\pi)$.

Theorem 5. SSB-ALL-BESTRESPONSE *can be solved in time* $O^*(n^2|\texttt{sp}(\pi)|)$ *in the complete and partition cases when* $\theta=[\mathcal{E}]^1$. *The set* $\texttt{BR}_T(\pi)$ *is characterized by values* $C(i,\pi)$ *of having element* $i\in\mathcal{E}$ *in the alternative in response to* π*:*

$$\forall i\in\mathcal{E},\ C(i,\pi)=\omega_i+\sum_{t\in\texttt{sp}(\pi)}\sum_{j\in A_t}\pi_t\Psi_{ij}.$$

This yields an implicit *compact formulation of the set* $\texttt{BR}_T(\pi)$*:*

- *Complete case:* $\texttt{BR}_T(\pi)=\{A : C_{>0}(\pi)\subseteq A\subseteq C_{\geq 0}(\pi)\}$, *where* $C_{>0}(\pi)=\{i\in\mathcal{E} : C(i,\pi)>0\}$ *and* $C_{\geq 0}(\pi)=\{i\in\mathcal{E} : C(i,\pi)\geq 0\}$.
- *Partition case:* $\texttt{BR}_T(\pi)=\prod_{t=1}^{p}\arg\max_{i\in X_t} C(i,\pi)$.

Proof. Let $\mathcal{E}=\{1,\dots,n\}$. Consider an SSB induced tournament $T=(\mathcal{A},\Phi)$, with $\theta=[\mathcal{E}]^1=\{S_1,\dots,S_n\}$ with $S_i=\{i\}$. Consider a mixed strategy $\pi\in\Delta(\mathcal{A})$. Solving the problem is equivalent to finding an alternative in $\arg\max_{A\in\mathcal{A}}\varphi_{\theta,\omega,\Psi}(A,\pi)$, where the definition of the SSB function is extended by linearity to mixed strategies: $\varphi_{\theta,\omega,\Psi}(A,\pi)=\sum_{t\in\texttt{sp}(\pi)}\pi_t\varphi_{\theta,\omega,\Psi}(A,A_t)$.

For $A \in \mathcal{A}$:

$$\varphi_{\theta,\omega,\Psi}(A,\pi) = \sum_{j\in A}\omega_j - \sum_{t\in \mathtt{sp}(\pi)}\sum_{i\in A_t}\pi_t\omega_i + \sum_{j\in A}\sum_{t\in \mathtt{sp}(\pi)}\sum_{i\in A_t}\pi_t\Psi_{ji}$$
$$= \sum_{j\in A}\underbrace{(\omega_j + \sum_{t\in \mathtt{sp}(\pi)}\sum_{i\in A_t}\pi_t\Psi_{ji})}_{C(j,\pi)} - \underbrace{\sum_{t\in \mathtt{sp}(\pi)}\sum_{i\in A_t}\pi_t\omega_i}_{cst} \tag{6}$$

where the second summand is a constant cst which only depends on π. As a result:

$$\mathtt{BR}_T(\pi) = \arg\max_{A\in\mathcal{A}}\sum_{j\in A}C(j,\pi).$$

This yields the compact representation of $\mathtt{BR}_T(\pi)$ given in the statement of the theorem for the complete and partition cases. Note that each value $C(j,\pi)$ for $j\in\mathcal{E}$ can be computed in $O^*(n|\mathtt{sp}(\pi)|)$. Once computed the values $C(j,\pi)$ for $j\in\mathcal{E}$, in both cases, an alternative in $\mathtt{BR}_T(\pi)$ can be computed in time $O^*(n)$. □

We now return to the problem SSB-ALL-ESSENTIAL.

Theorem 6. *A concise representation of the essential set in problem* SSB-ALL-ESSENTIAL *can be computed in polynomial time in the complete and partition cases when* $\theta = [\mathcal{E}]^1$.

Proof. Let π^* denote a mixed Nash equilibrium of maximal support for the game defined by matrix Φ. The maximal support $\mathtt{sp}(\pi^*)$ is unique by convexity of the set of mixed Nash equilibria: it is the union of their supports. The essential set of $T=(\mathcal{A},\Phi)$ is:

$$\mathtt{EssentialSet}(\mathcal{A},\Phi) = \{A_t\in\mathcal{A} : t\in\mathtt{sp}(\pi^*)\},$$

whose size may be exponential in $n=\mathcal{E}$. Let $\mathcal{S}=\cup_{t\in\mathtt{sp}(\pi^*)}A_t\subseteq\mathcal{E}$, let $\mathcal{T}=\cap_{t\in\mathtt{sp}(\pi^*)}A_t\subseteq\mathcal{E}$ and $p_i=\sum_{t\in\mathtt{sp}(\pi):i\in A_t}\pi_t$ denote the probability that i is part of a set sampled by a mixed strategy π. The idea of the proof is to solve a sequence of at most $2n$ Linear Programs (LP) in variables p_i to determine the elements in $\mathcal{S}$ and $\mathcal{T}$. Each LP can be solved in polynomial time by using an interior point method [32], as there are only $n=|\mathcal{E}|$ variables p_i (while there would have been 2^n variables if the LP had been expressed in variables π_t). The essential set then corresponds to $\{A : \mathcal{T}\subseteq A\subseteq\mathcal{S}\}$, thus is concisely expressed by $\mathcal{S}$ and $\mathcal{T}$. For space reasons, we only give the proof in the complete case (the proof in the partition case is similar).

We first need to define the LP in variables p_i allowing us to determine a mixed Nash equilibrium. Given $A\in\mathcal{A}$, from Equation 6, we can express $\varphi_{\theta,\omega,\Psi}(A,\pi)$ in function of probabilities $p_1,\ldots,p_n$:

$$\sum_{j\in A}(\omega_j + \sum_{t\in\mathtt{sp}(\pi)}\sum_{i\in A_t}\pi_t\Psi_{ji}) - \sum_{t\in\mathtt{sp}(\pi)}\sum_{i\in A_t}\pi_t\omega_i$$
$$= \sum_{j\in A}(\omega_j + \sum_{i\in\mathcal{E}}\sum_{t\in\mathtt{sp}(\pi):i\in A_t}\pi_t\Psi_{ji}) - \sum_{i\in\mathcal{E}}\sum_{t\in\mathtt{sp}(\pi):i\in A_t}\pi_t\omega_i$$
$$= \sum_{j\in\mathcal{E}}\underbrace{(\omega_j + \sum_{i\in\mathcal{E}}p_i\Psi_{ji})}_{C(j,\pi)}\mathbb{1}_A(j) - \sum_{j\in\mathcal{E}}p_j\omega_j$$

If $A\in\mathtt{BR}_T(\pi)$, then the contribution of $j\in\mathcal{E}$ in $\varphi_{\theta,\omega,\Psi}(A,\pi)$ is $\max\{0,C(j,\pi)\}$, i.e., 0 if $j\notin A$ and $\omega_j+\sum_{i\in\mathcal{E}}\Psi_{ji}p_i>0$ if $j\in A$. The value $\varphi_{\theta,\omega,\Psi}(A,\pi)$ thus corresponds to the optimum of the LP:

$$\min_y \sum_{j\in\mathcal{E}}(y_j - p_j\omega_j)$$
$$\omega_j + \sum_{i\in\mathcal{E}}\Psi_{ji}p_i \le y_j \quad \forall j\in\mathcal{E},$$
$$y_j \ge 0 \quad \forall j\in\mathcal{E}.$$

because $y_j=\max\{0,C(j,\pi)\}$ at the optimum. As the value of any symmetric zero-sum game is 0, a set of values $p_i\in[0,1]$ defining a mixed Nash equilibrium satisfy the following set C of constraints:

$$(C)\quad\begin{cases}\sum_{i\in\mathcal{E}}(y_i - p_i\omega_i)\le 0 \\ \omega_j + \sum_{i\in\mathcal{E}}\Psi_{ji}p_i \le y_j \quad \forall j\in\mathcal{E}, \\ y_j\ge 0 \quad \forall j\in\mathcal{E}, \\ p_i\in[0,1] \quad \forall i\in\mathcal{E}.\end{cases} \tag{7}$$

Constraint 7 ensures that the values of variables p_i define a mixed Nash equilibrium. Indeed, from any mixed Nash equilibrium π we can induce values $p_i=\sum_{t\in\mathtt{sp}(\pi):i\in A_t}\pi_t$ and $y_j=\max\{0,\omega_j+\sum_{i\in\mathcal{E}}\Psi_{ji}p_i\}$ that are feasible for C. The converse is also true, as from any feasible solution (p,y), we can build a mixed Nash equilibrium π using the following greedy procedure:

1. Sort the elements $i\in\mathcal{E}$ in increasing order of probabilities p_i, to obtain a permutation $(\sigma(1),\sigma(2),\ldots,\sigma(n))$ of $\mathcal{E}$ s.t. $p_{\sigma(i)}\le p_{\sigma(i+1)}$. Let $\hat{p}=(p_{\sigma(1)},p_{\sigma(2)},\ldots,p_{\sigma(n)})$;
2. If $\hat{p}$ has no strictly positive element then return π, complementing with the $\emptyset$ to obtain a valid mixed strategy with $\sum_{t=1}^N\pi_t=1$.
3. Let i be the first index such that $\hat{p}_i>0$; let l be the index such that $A_l=\{\sigma(j): i\le j\le n\}$. Add A_l to π with $\pi_l=\hat{p}_i$; update $\hat{p}$ by subtracting π_l to the elements of indices i to n. Come back to 2.

At the end of this greedy algorithm (after at most n iterations), we obtain a mixed strategy π such that $|\mathtt{sp}(\pi)|\le n+1$, $\sum_t\pi_t=1$, and $p_i=\sum_{t:i\in A_t}\pi_t$. This last point is due to an invariant argument: at each iteration of the method we have $p_i=\hat{p}_{\sigma^{-1}(i)}+\sum_{t:i\in A_t}\pi_t$, and at termination $\hat{p}$ is a null vector.

Computing $\mathcal{S}$ (resp. $\mathcal{T}$) amounts to find all $i\in\mathcal{E}$ such that $p_i>0$ (resp. $p_i=1$) in at least one (resp. all) feasible solution (p,y) for C: $\mathcal{S}=\{i:\max p_i>0 \text{ for } (p,y)\in C\}$ and $\mathcal{T}=\{i:\min p_i=1 \text{ for } (p,y)\in C\}$. Thus, finding $\mathcal{S}$ and $\mathcal{T}$ requires to solve $2n$ LPs.

We now show that the essential set corresponds to $\{A_t:\mathcal{T}\subseteq A_t\subseteq\mathcal{S}\}$, i.e., if $\mathcal{T}\subseteq A_t\subseteq\mathcal{S}$ then there exists a mixed Nash equilibrium π such that $t\in\mathtt{sp}(\pi)$. The previous $2n$ LPs result in a sequence of $2n$ optimal solutions $(\overline{y}^i,\overline{p}^i)$ (when maximizing p_i) and $(\underline{y}^i,\underline{p}^i)$ (when minimizing p_i). By convexity of the feasible set for C, we can consider a linear combinations of these solutions (e.g., with weights $1/2n$) to obtain a feasible solution (y,p) for C in which $\{i:p_i>0\}=\mathcal{S}$, and $\{i:p_i=1\}=\mathcal{T}$. Then, we can adapt the previous greedy procedure to generate from p a mixed Nash equilibrium π s.t. $t\in\mathtt{sp}(\pi)$. It consists in adding the following preliminary step:

0. Add A_t with probability $\pi_t=\min(\{p_i:i\in A_t\}\cup\{1-p_j:j\notin A_t\})$ to π, subtract π_t to each value p_i, with $i\in A_t$. □

7 Future Work

This work paves the way for different research questions. First, it would be interesting to have a more complete picture of the complexity of finding the best alternatives. This could be obtained by studying more tournament solutions or other domain restrictions. Moreover, more work on the learning procedure should be undergone, notably to better account for noise in the preference data. For instance, one could design a Bayesian approach for this purpose, where the prior parameters would favor simpler models (e.g., with few interactions, and as additive as possible). Last, finding additional real-world datasets on which our model could yield an increased descriptive power is a worthwhile research direction, e.g., by looking at sports competitions or participatory budgeting.

Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR20-CE23-0018 (THEMIS). We also thank Nassim Hemdane for insightful discussions about the game League of Legends, as part of the design of numerical tests.

References

[1] W. E. Armstrong. The determinateness of the utility function. *The Economic Journal*, 49(195):453–467, 1939.

[2] D. Bigot, H. Fargier, J. Mengin, and B. Zanuttini. Using and learning GAI-decompositions for representing ordinal rankings. In *Proceedings of the ECAI'2012 workshop on Preference Learning, Montpellier, France, August 28, 2012*, pages 5–10, 2012.

[3] M. Binshtok, R. I. Brafman, S. E. Shimony, A. Martin, and C. Boutilier. Computing optimal subsets. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*, pages 1231–1236, 2007.

[4] C. Boutilier, R. Patrascu, P. Poupart, and D. Schuurmans. Constraint-based optimization and utility elicitation using the minimax decision criterion. *Artificial Intelligence*, 170(8-9):686–713, 2006.

[5] R. I. Brafman, C. Domshlak, S. E. Shimony, and Y. Silver. Preferences over sets. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence, July 16-20, 2006, Boston, Massachusetts, USA*, pages 1101–1106, 2006.

[6] F. Brandt, M. Brill, and P. Harrenstein. Tournament solutions. In *Handbook of Computational Social Choice*, page 57–84. Cambridge University Press, 2016.

[7] D. Braziunas and C. Boutilier. Minimax regret based elicitation of generalized additive utilities. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI 2007), July 19-22, 2007, Vancouver, BC, Canada*, pages 25–32, 2007.

[8] R. Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12:331–370, 2002.

[9] P. Campanelli. League of Legends Ranked Matches — kaggle.com. kaggle.com/datasets/paololol/league-of-legends-ranked-matches, 2017. [Accessed 18-07-2024].

[10] W. Chu and Z. Ghahramani. Preference learning with Gaussian processes. In *Proceedings of the 22nd International Conference on Machine Learning (ICML 2005), August 7-11, 2005, Bonn, Germany*, pages 137–144, 2005.

[11] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[12] P. De Donder, M. Le Breton, and M. Truchon. Choosing from a weighted tournament. *Mathematical Social Sciences*, 40(1):85–109, 2000.

[13] M. Desjardins and K. Wagstaff. DD-PREF: A language for expressing preferences over sets. In *Proceedings of the 20th AAAI Conference on Artificial Intelligence, July 9-13, 2005, Pittsburgh, Pennsylvania, USA*, pages 620–626, 2005.

[14] M. Desjardins, E. Eaton, and K. L. Wagstaff. Learning user preferences for sets of objects. In *Proceedings of the 23rd International Conference on Machine Learning (ICML 2006), June 25-29, 2006, Pittsburgh, Pennsylvania, USA*, pages 273–280, 2006.

[15] C. Domshlak and T. Joachims. Unstructuring user preferences: efficient non-parametric utility revelation. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI 2005), July 26-29, 2005, Edinburgh, Scotland*, pages 169–177, 2005.

[16] B. Dutta and J.-F. Laslier. Comparison functions and choice correspondences. *Social Choice and Welfare*, 16(4):513–532, 1999.

[17] P. Erdös and A. Rényi. On random graphs I. *Publ. math. debrecen*, 6 (290-297):18, 1959.

[18] P. C. Fishburn. Intransitive indifference in preference theory: a survey. *Operations Research*, 18(2):207–228, 1970.

[19] P. C. Fishburn. Nontransitive measurable utility. *Journal of Mathematical Psychology*, 26(1):31–67, 1982.

[20] P. C. Fishburn and I. H. LaValle. Binary interactions and subset choice. *European Journal of Operational Research*, 92(1):182–192, 1996.

[21] W. V. Gehrlein. Condorcet's paradox. *Theory and Decision*, 15(2): 161–197, 1983.

[22] H. Gilbert, M. Ouaguenouni, M. Öztürk, and O. Spanjaard. A hybrid approach to preference learning with interaction terms. In *Proceedings of the 26th European Conference on Artificial Intelligence (ECAI 2023), September 30 - October 4, 2023, Kraków, Poland*, pages 835–842, 2023.

[23] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*, volume 127. Cambridge University Press, 2009.

[24] S. Greco, V. Mousseau, and R. Słowiński. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, 191(2):416–436, 2008.

[25] Y. Guo and C. P. Gomes. Learning optimal subsets with implicit user preferences. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009), July 11-17, 2009, Pasadena, California, USA*, pages 1052–1057, 2009.

[26] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2):273–324, 1997.

[27] G. Laffond, J.-F. Laslier, and M. Le Breton. The bipartisan set of a tournament game. *Games and Economic Behavior*, 5(1):182–201, 1993.

[28] R. Lahdelma, J. Hokkanen, and P. Salminen. SMAA-Stochastic Multiobjective Acceptability Analysis. *European Journal of Operational Research*, 106(1):137–143, 1998.

[29] K. Lehrer and C. Wagner. Intransitive indifference: The semi-order problem. *Synthese*, pages 249–256, 1985.

[30] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu. Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6):1–45, 2017.

[31] R. D. Luce. Semiorders and a theory of utility discrimination. *Econometrica, Journal of the Econometric Society*, pages 178–191, 1956.

[32] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *Siam studies in applied mathematics*. SIAM, 1994.

[33] M. Ouaguenouni. Code for "Learning and Optimizing with an SSB Representation of Intransitive Preferences on Sets". Available at https://github.com/ouaguenouni/Eunomia, 2024.

[34] A. Parameswaran, P. Venetis, and H. Garcia-Molina. Recommendation systems with complex constraints: A course recommendation perspective. *ACM Transactions on Information Systems (TOIS)*, 29(4):1–33, 2011.

[35] Riot Games. Riot Developer Portal — developer.riotgames.com. https://developer.riotgames.com. [Accessed 18-07-2024].

[36] B. Scholkopf and A. J. Smola. *Learning with kernels: Support Vector Machines, regularization, optimization, and beyond*. MIT press, 2018.

[37] E. B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158): 209–212, 1927.