THeory and Evidence to Measure Influence in Social structures (THEMIS)

The project THeory and Evidence to Measure Influence in Social structures (THEMIS) (ANR-20-CE23-0018) is a PRC (‘*Projets de recherche collaborative*’) project has been financed by the ‘Agence nationale de la recherche’ (ANR) under the ‘AAP générique 2020’ (‘CE23 - Intelligence Artificielle’) with a total allocated financial aid of 401K nd a duration of 48 months. It officially started on March 29th 2021 and its. The THEMIS project has three institutional partners: CRIL CNRS UMR 8188, Université d’Artois, Lens, LAMSADE CNRS UMR 7243, Université Paris Dauphine, Paris and LIP6 CNRS UMR 7206, Sorbonne Université, Paris. LAMSADE is the leading partner, and members of both Pole 1 and pole 2 are involved in the project.

The THEMIS project is positioned in the core of the emerging research area on social influence analysis. Unlike existing measures of influence for social structures, this project brings insights based on different techniques from artificial intelligence including multi-agent systems, compact representation, algorithmic game theory, computational social choice, and social network analysis. Our main objective is to show that the portfolio of models proposed under the umbrella of a qualitative theory for social influence analysis is more adapted to answer important questions arising from different domains of collective decision making. For this purpose, we apply our framework to alternative research domains via a property-driven approach, keeping into account how algorithmic and strategic aspects play a role in the design of specific social ranking solutions. The originality of our project is summarized by the following list of innovative tasks:

- to conceive a novel ordinal theory of cooperative games for measuring power and influence in coalitional situations;
- to focus on a property-driven design of social ranking solutions that will be used as a top-down approach aimed at splitting the complex behaviour of groups or coalitions into more intelligible interaction situations;
- to explore natural applications of methods for compact preference representation to social ranking computation and to coalition formation;
- to formulate a portfolio of solutions accompanied with a roadmap of their properties to drive users in their practice;

- to implement most of the solutions proposed in this project as an open-source software package socialranking available on the Comprehensive R Archive Network (CRAN), which is R's central software repository.

Among the articles published so far within the framework of the THEMIS project, we mention in particular the paper [1] where the issue of manipulability for social rankings (i.e., a ranking over individuals computed keeping into account their relative positions over groups) is introduced and studied for the first time. In this context, individuals behave strategically within each group with the objective to impact groups' performance and to reach highest positions in the social ranking . So, the manipulability problem lies at the intersection of computational social choice and the algorithmic theory of power indices and perfectly fits into the perimeter of the transversal project *Games and social choice* between Pole 1 and Pole 2. In this paper, we focus in particular on classes of social rankings representing three fundamental approaches from the literature: the marginal contribution approach [2], the lexicographic approach [3] and the CP(*ceteris paribus*)-majority one [4]. In [1], we first consider some particular members of these families of social rankings analysing their resistance to a malicious behaviour of individuals. Then, we analyze the computational complexity of manipulation, and complete our theoretical results with simulations in order to analyse the manipulation frequencies and to assess the effects of manipulations.

Website of the projetc: https://www.lamsade.dauphine.fr/themis/

References

[1] Tahar Allouche, Bruno Escoffier, Stefano Moretti, and Meltem Ozturk. Social Ranking Manipulability for the CP-Majority, Banzhaf and Lexicographic Excellence Solutions. In Christian Bessiere, editor, *Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence {IJCAI-PRICAI-20}*, pages 17–23, Yokohama, Japan, January 2021. International Joint Conferences on Artificial Intelligence Organization.

[2] Hossein Khani. *Ordinal Power Indices*. Theses, Université Paris sciences et lettres, October 2020.

[3] Giulia Bernardi, Roberto Lucchetti, and Stefano Moretti. Ranking objects from a preference relation over their subsets. *Social Choice and Welfare*, 52(4):589–606, 2019.

[4] Adrian Haret, Hossein Khani, Stefano Moretti, and Meltem Ozturk. Ceteris Paribus Majority for social ranking. In *27th International Joint Conference on Artificial Intelligence (IJCAI-ECAI-18)*, pages 303–309, Stockholm, Sweden, July 2018.

Social Ranking Manipulability for the CP-Majority, Banzhaf and Lexicographic Excellence Solutions*

Tahar Allouche[1] , **Bruno Escoffier**[2,3] , **Stefano Moretti**[4] and **Meltem Öztürk**[4†]
[1]ENSTA-Paris, Institut Polytechnique de Paris, France
[2]Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
[3]Institut Universitaire de France, France
[4]Université Paris Dauphine, PSL Research University, CNRS, Lamsade, 75016 Paris, France
tahar.allouche@ensta-paris.fr, Bruno.Escoffier@lip6.fr, {stefano.moretti, meltem.ozturk}@dauphine.fr

Abstract

We investigate the issue of manipulability for social ranking rules, where the goal is to rank individuals given the ranking of coalitions formed by them and each individual prefers to reach the highest positions in the social ranking. This problem lies at the intersection of computational social choice and the algorithmic theory of power indices. Different social ranking rules have been recently proposed and studied from an axiomatic point of view. In this paper, we focus on rules representing three classical approaches in social choice theory: the marginal contribution approach, the lexicographic approach and the (*ceteris paribus*) majority one. We first consider some particular members of these families analysing their resistance to a malicious behaviour of individuals. Then, we analyze the computational complexity of manipulation, and complete our theoretical results with simulations in order to analyse the manipulation frequencies and to assess the effects of manipulations.

1 Introduction

In decision making and social choice theory, a number of studies are devoted to ranking individuals based on the performance of the coalitions formed by them. For instance, given values on coalitions of individuals, power indices map these values of coalitions on values of individuals. The seminal works of Shapley [1953] and Banzhaf III [1964] paved the way of a whole research domain and a related literature with many issues, including axiomatization [Laruelle and Valenciano, 2001; Holler and Packel, 1983], applications [Bilbao *et al.*, 2002; Moretti and Patrone, 2008], algorithmic analysis [Matsui and Matsui, 2000] and computational complexity [Deng and Papadimitriou, 1994; Bachrach and Rosenschein, 2009; Faliszewski and Hemaspaandra, 2009]. The non-manipulability (or *strategy proofness*) is another fundamental issue. In social choice, since the seminal theorems of Gibbard and Satterthwaite ([Gibbard, 1973] and [Satterthwaite, 1975]), we know that every interesting social choice function is manipulable by misrepresentation of preferences. The manipulability is also analysed for power indices. We quote in particular the literature on the paradoxical behaviour of power indices under the modification of some elements of the game, like the number of players or the size of coalitions [Felsenthal *et al.*, 1998; Laruelle and Valenciano, 2005], or the study of manipulation in weighted voting games [Aziz *et al.*, 2011; Zuckerman *et al.*, 2012]. In these models, players are analyzed from a strategic perspective to establish under which conditions they can increase their power adopting malicious behaviors like, for example, splitting or merging.

Power indices (and other indices of individual productivity based on the evaluation of revenues generated by teams [Flores-Szwagrzak and Treibich, 2020]) require a numerical evaluation of coalitions of individuals. Following classical situations in social choice where ordinal data are provided (for instance, voting theory), several articles address the question of defining ordinal notions of power indices when we only have ordinal information over coalitions. This has been formalized as the *social ranking from coalitions* (SRC) problem, where the objective is to evaluate the "influence" of individuals involved in a collective decision process like an electoral system, a parliament, a governing council, a management board, etc. ([Moretti, 2015; Moretti and Öztürk, 2017]). Basically, an SRC problem consists of a finite set N of individuals and a binary relation $\succeq$ over some subsets (hereafter called *coalitions*) of N; the binary relation $\succeq$ is called *power relation* and represents the relative power of coalitions in a decision process. A *solution* or *rule* for an SRC problem is a "suitable" method aimed to convert the information contained in a power relation $\succeq$ into a ranking over the single elements of N representing their overall individual power. Several solutions for SRC problems have been proposed in the literature. For instance, in the work by Haret *et al.* [2018] (and the one by Fayard and Öztürk [2018]) two individuals are compared using information from subsets ranking them under a *ceteris paribus* interpretation. Bernardi *et al.* [2019] axiomatically characterize a solution based on the idea that the most influential individuals are those appearing more frequently in the highest positions of the power relation. A rule based on the idea of ordinal marginal contribution has been recently introduced in the paper by Khani *et al.* [2019].

*We are grateful to anonymous reviewers for their critical and helpful comments.

†Contact author.

Following these lines of research, in this article we are interested in the analysis of the *strategic manipulation* of SRC rules, in the sense that an individual may be interested in behaving maliciously within one or more teams (weakening their group's effectiveness) in order to obtain a better position in the individual ranking. The notion of manipulability for SRC considered in this paper assumes that an individual can only weaken the ranking of teams to which she belongs. In other words, an individual i cannot affect the performances of teams not containing i and, in addition, cannot improve the position of a coalition already containing i.

Example 1. *Consider a manager who must decide how to allocate two bonuses over three employees (denoted by* $1, 2$ *and* 3*). Suppose that she can only compare the performance of teams in an ordinal way:* $\{1,2,3\} \succ \{1,3\} \succ \{1\} \succ \{1,2\} \succ \{2,3\} \succ \{2\} \succ \{3\} \sim \emptyset$[1]. *Suppose the manager wants to keep into account the attitude of employees to cooperate. So, an option is to count the number of (ordinal) positive and negative marginal contributions provided by each employee to all possible coalitions, i.e.* 1 *contributes positively to four teams (i.e.,* $\{1,2,3\} \succ \{2,3\}$, $\{1,3\} \succ \{3\}$, $\{1,2\} \succ \{2\}$ *and* $\{1\} \succ \emptyset$*),* 3 *contributes positively to three coalitions while* 2 *also contributes positively to three coalitions, but negatively to coalition* $\{1,2\}$*. Therefore the manager would end up to award players* 1 *and* 3*. Such a rule could push individual* 2 *to behave strategically and to undermine the cooperation within coalition* $\{2,3\}$*. So the new ranking being* $\{1,2,3\} \succ' \{1,3\} \succ' \{1\} \succ' \{1,2\} \succ' \{2\} \succ' \{2,3\} \succ' \{3\} \sim' \emptyset$*, individuals* 1 *and* 2 *should now get the bonus.*

To our knowledge, this article is the first one which investigates the manipulation for SRC rules. As in social choice, the manipulability is an important issue in many real-world coalitional frameworks. For instance, within a parliament, small political parties are often able to blackmail a majority coalition by threatening to withdraw from the coalition or to run a candidate of their own [Ferrara, 2004]. In international organizations, the capacity to block the proposals of a group ensures a great visibility of the blocking state's positions. In any organizational context, like a business company or an academic institution, any perceived disparity about the merits of a team's members (e.g., over-remuneration of a leader) may engender internal competition, or even the sabotage of the team by some of it's members. As a sensible example of this kind of behaviors, consider the situation following the Italian general election of March 2018, where no political party got an absolute majority. After 3 months of negotiation, a coalition government was finally formed by the two parties casting the highest number of votes, precisely, the Five Star Movement and the League. However, such a government ended one year after, in August 2019, when the League withdrew its support to the coalition government. As a consequence, the League substantially increased its position in the opinion polls done immediately after the government crisis [Chiaramonte *et al.*, 2020].

The article is organized as follows. We present basic notions and notations in Section 2. In Section 3, we introduce a formal definition of manipulability for an SRC rule and provide theoretical results on four social ranking rules: Copeland-like, Kramer-Simpson-like, Lexicographic Excellence [Bernardi *et al.*, 2019] and Ordinal Banzhaf [Khani *et al.*, 2019]. These social ranking rules display a wide variety of characteristics. The two first ones are based on *ceteris paribus* comparisons and use the majority principle in a different way. Copeland-like solution is a kind of flow analysis of majority graphs, whereas Kramer-Simpson is a minmax score. Ordinal Banzhaf rule is based on a marginal contribution principle. Lexicographic Excellence (lexcel) considers only information from the best ranked coalitions. We show that only lexcel is not manipulable. In Section 4, we analyse the computational complexity of manipulating each of the three manipulable social ranking rules, and prove that for each of them determining whether an individual can manipulate or not is an NP-hard problem. In Section 5, we present some simulation results on manipulable social ranking rules showing the manipulation frequencies and their vulnerability against the manipulation. Section 6 concludes the article.

2 Preliminaries

Let $N = \{1, \dots, n\}$ be a finite set of elements called *individuals* and let $R \subseteq N \times N$ be a *binary relation* on N. A *preorder* is a reflexive and transitive binary relation. A preorder that is total is called *total preorder*. An antisymmetric[2] total preorder is called *linear order*. We denote by $\mathcal{T}(N)$ the set of all total preorders on N and by 2^N the powerset of N, i.e. the set of all subsets (also called, *coalitions*) of N. Let $\mathcal{P} \subseteq 2^N$ be a non-empty collection of subsets of N. A *power relation* on $\mathcal{P}$ is a total preorder $\succeq \subseteq \mathcal{P} \times \mathcal{P}$. We denote by $\mathcal{T}(\mathcal{P})$ the family of all power relations on every non-empty collection $\mathcal{P} \subseteq 2^N$. Given a *power relation* $\succeq \in \mathcal{T}(\mathcal{P})$ on $\mathcal{P} \subseteq 2^N$, we denote by $\sim$ its symmetric part (i.e. $S \sim T$ if $S \succeq T$ and $T \succeq S$) and by $\succ$ its asymmetric part (i.e. $S \succeq T$ and not $T \succeq S$). So, for each pair of subsets $S, T \in \mathcal{P}$, $S \succ T$ means that S is *strictly stronger* than T, whereas $S \sim T$ means that S and T are *indifferent*.

Let $\succeq \in \mathcal{T}(\mathcal{P})$ be of the form $S_1 \succeq S_2 \succeq S_3 \succeq \cdots \succeq S_{|\mathcal{P}|}$. The *quotient order* of $\succeq$ is denoted as $\Sigma_1 \succ \Sigma_2 \succ \Sigma_3 \succ \cdots \succ \Sigma_m$ in which the subsets S_j are grouped in the *equivalence classes* Σ_k generated by the symmetric part of $\succeq$. This means that all the sets in Σ_1 are indifferent to S_1 and are strictly better than the sets in Σ_2 and so on. So, $\Sigma_i = S_i$ for any $i = 1, \dots, |\mathcal{P}|$ if and only if $\succeq$ is a linear order.

A *social ranking solution* or *solution on* N, is a function $R : \mathcal{T}(\mathcal{P}) \longrightarrow \mathcal{T}(N)$ associating to each power relation $\succeq \in \mathcal{T}(\mathcal{P})$ a total preorder $R(\succeq)$ (or $R^{\succeq}$) over the elements of N. By this definition, the notion $iR^{\succeq}j$ means that applying the social ranking solution to the power relation $\succeq$ gives the result that i is ranked higher than or equal to j. We denote by $I^{\succeq}$ the symmetric part of $R^{\succeq}$, and by $P^{\succeq}$ its asymmetric part. The *social score* $p_i(R^{\succeq})$ of individual $i \in N$ in $R^{\succeq}$ is defined as the number of individuals in $N \setminus \{i\}$ that are ranked lower than i minus the number of individuals in $N \setminus \{i\}$ that are ranked higher than i, that is

$$p_i(R^{\succeq}) = |\{j \in N \setminus \{i\} : iR^{\succeq}j\}| - |\{j \in N \setminus \{i\} : jR^{\succeq}i\}|.$$

[1] $\succ$: strict preference, $\sim$: indifference.

[2] $\forall i, j \in N$, iRj and $jRi \Rightarrow i = j$.

3 Manipulability

In this paper we focus on a particular notion of manipulation, intended as the "unlimited" capacity of individuals to undermine the position of coalitions to which they belong in a power relation $\succeq$ on $\mathcal{P} \subseteq 2^N$.

Definition 1. Let $\succeq \in \mathcal{T}(\mathcal{P})$ on $\mathcal{P}$ be a power relation with the associated quotient order $\succ$:

$$\Sigma_1 \succ \Sigma_2 \succ \cdots \succ \Sigma_j \succ \cdots \succ \Sigma_m. \tag{1}$$

Let i be an individual, and $\mathcal{C} \subseteq \mathcal{P}$ be a collection of coalitions in $\mathcal{P}$ all containing i. For all $S \in \mathcal{C}$, let $j(S) \in \{1, \ldots, m\}$ be such that $S \in \Sigma_{j(S)}$ [3].

A manipulation of $\succeq$ by individual i via collection $\mathcal{C}$ is another power relation $\succeq^{\mathcal{C}}$ on $\mathcal{P}$ with $\succeq^{\mathcal{C}} \neq \succeq$ and with the associated quotient order $\succ^{\mathcal{C}}$ such that the following two conditions hold:

i) $\Sigma_1 \backslash \mathcal{C} \succ^{\mathcal{C}} \Sigma_2 \backslash \mathcal{C} \succ^{\mathcal{C}} \cdots \succ^{\mathcal{C}} \Sigma_j \backslash \mathcal{C} \succ^{\mathcal{C}} \cdots \succ^{\mathcal{C}} \Sigma_m \backslash \mathcal{C}$;

ii) $T \succ^{\mathcal{C}} S$ for all $S \in \mathcal{C}$ and $T \in \bigcup_{i=1,\ldots,j(S)} \Sigma_i \setminus \mathcal{C}$.

A social ranking $\mathcal{R}$ is *manipulable by* i on a power relation $\succeq$ on $\mathcal{P}$ if there exists a collection of coalitions $\mathcal{C} \subseteq \mathcal{P}$ containing i, a manipulation $\succeq^{\mathcal{C}}$ of $\succeq$ by i via $\mathcal{C}$ such that

$$p_i(R^{\succeq^{\mathcal{C}}}) > p_i(R^{\succeq}).$$

[recall that $p_i(R^{\succeq})) = |\{j, iR^{\succeq}j\}| - |\{j, jR^{\succeq}i\}|$.]

A social ranking solution $\mathcal{R}$ is *manipulable* on a power relation $\succeq$ if it is manipulable by some individual i.

Condition (ii) says that $\succeq^{\mathcal{C}}$ is obtained from $\succeq$ moving each coalition $S \in \mathcal{C}$ from the equivalence class to which it belongs in $\succeq$, to a strictly lower equivalence class (that can also be a new singleton equivalence class containing only S in $\succeq^{\mathcal{C}}$), while the relation among all the other coalitions not in $\mathcal{C}$ is maintained as in $\succeq$ (condition(i)). The family of all manipulations of $\succeq$ via collection $\mathcal{C}$ is denoted by $\mathcal{M}_{\mathcal{C}}(\succeq)$.

Example 2. *Consider the power relation $\succeq$ such that*[4] $23 \succ (123 \sim 12) \succ 13$ *(hence, $\Sigma_1 = \{23\}, \Sigma_2 = \{123, 12\}, \Sigma_3 = \{13\}$). Imagine that individual 1 wants to manipulate by deteriorating the positions of coalitions in $\mathcal{C} = \{12, 13\}$. Condition i) of Definition 1 imposes to maintain $23 \succ^{\mathcal{C}} 123$, condition ii) imposes $23 \succ^{\mathcal{C}} 12$, $23 \succ^{\mathcal{C}} 13$, $123 \succ^{\mathcal{C}} 12$ and $123 \succ^{\mathcal{C}} 13$. Hence, the family of all possible manipulations of $\succeq$ by 1 via collection $\mathcal{C}$ is $\mathcal{M}_{\mathcal{C}}(\succeq) = \{\succeq^{\mathcal{C}}_a, \succeq^{\mathcal{C}}_b, \succeq^{\mathcal{C}}_c\}$, with $23 \succ^{\mathcal{C}}_a 123 \succ^{\mathcal{C}}_a 12 \succ^{\mathcal{C}}_a 13$, $23 \succ^{\mathcal{C}}_b 123 \succ^{\mathcal{C}}_b 13 \succ^{\mathcal{C}}_b 12$ and $23 \succ^{\mathcal{C}}_c 123 \succ^{\mathcal{C}}_c 12 \sim^{\mathcal{C}}_c 13$.*

We will now analyse the manipulability of different social ranking rules.

3.1 Copeland-like and Kramer-Simpson-like Rules

Copeland-like and Kramer-Simpson (KS)-like rules are both based on *Ceteris Paribus*-majority relation, where individuals i and j are ranked according to their relative success over comparisons of coalitions of the type $S \cup i$ *vs.* $S \cup j$ (CP-comparisons), more precisely:

[3] $j(S)$ represents the rank of the equivalence class to which S belongs in the initial power relation $\succeq$.

[4] To avoid cumbersome notations later, sets will be written for short without commas and parentheses, e.g., 123 instead of $\{1, 2, 3\}$, and $S \cup i$ instead of $S \cup \{i\}$.

Definition 2 (CP-Majority [Haret *et al.*, 2018])**.** Let $\succeq \in \mathcal{T}(\mathcal{P})$. The *Ceteris Paribus (CP-) majority* relation is the binary relation $R^{\succeq}_{CP} \subseteq N \times N$ such that for all $i, j \in N$:

$$iR^{\succeq}_{CP}j \Leftrightarrow d_{ij}(\succeq) \geq d_{ji}(\succeq),$$

where $d_{ij}(\succeq)$ represents the cardinality of the set $D_{ij}(\succeq)$, the set of all coalitions $S \in 2^{N-\{i,j\}}$ for which $S \cup i \succeq S \cup j$[5].

Example 3. *Consider:* $(123 \sim 12 \sim 3 \sim 1) \succ (2 \sim 23) \succ 13$. *Then, we obtain* $1I^{\succeq}_{CP}2, 1P^{\succeq}_{CP}3, 2I^{\succeq}_{CP}3$. *For instance,* $1I^{\succeq}_{CP}2$ *:* $D_{12}(\succeq) = \{\emptyset\}$ *(*$1 \succ 2$*);* $D_{21}(\succeq) = \{3\}$ *(*$23 \succ 13$*).*

The CP-Majority relation has a major drawback: it can generate cycles within the individual ranking (except under some particular domain restrictions, as suggested in [Haret *et al.*, 2018]). For this reason, we investigate the manipulability of two transitive solutions derived from the CP-Majority, which are inspired, respectively, by the Copeland [Copeland, 1951] and Kramer-Simpson [Simpson, 1969] [Kramer, 1977] voting schemes. These two rules are known to be Condorcet coherent, meaning that when a Condorcet winner (a candidate beating all the other candidates by the majority rule) exists, it is chosen by them. Interestingly, while it can be easily proved that CP-majority relation is not manipulable, Copeland like and KS like solutions are manipulable.

Copeland-like Method

Strongly inspired by the Copeland score of social choice theory, we define Copeland-like solution based on the net flow of CP-majority graph. According to the Copeland solution, individuals are ordered according to the number of pairwise winning comparisons, minus the one of pairwise losing comparisons, over the set of all CP-comparisons.

Definition 3 (Copeland-like solution)**.** Let $\succeq \in \mathcal{T}(\mathcal{P})$. The *Copeland-like* relation is the binary relation $R^{\succeq}_{Cop} \subseteq N \times N$ such that for all $i, j \in N$:

$$iR^{\succeq}_{Cop}j \Leftrightarrow Score^{\succeq}_{cop}(i) \geq Score^{\succeq}_{cop}(j).$$

where $Score^{\succeq}_{cop}(i) = p_i(R^{\succeq}_{CP}) =$

$$|\{j \in N \setminus \{i\} : iR^{\succeq}_{CP}j\}| - |\{j \in N \setminus \{i\} : jR^{\succeq}_{CP}i\}|.$$

Theorem 1. *The Copeland-like solution is manipulable.*

Proof. See Example 4 for an instance of manipulation. □

Example 4. *Consider $\succeq$ of Example 3. Then,*

$$Score^{\succeq}_{cop}(1) = 1, Score^{\succeq}_{cop}(2) = 0, Score^{\succeq}_{cop}(3) = -1.$$

Hence, the Copeland-like relation is: $1P^{\succeq}_{Cop}2P^{\succeq}_{Cop}3$. *Now imagine that 3 deteriorates the performance of 23 ($\mathcal{C} = \{23\}$)*

$$(123 \sim^{\mathcal{C}} 12 \sim^{\mathcal{C}} 3 \sim^{\mathcal{C}} 1) \succ^{\mathcal{C}} 2 \succ^{\mathcal{C}} 13 \succ^{\mathcal{C}} 23.$$

Now we have: $Score^{\succeq^{\mathcal{C}}}_{cop}(1) = 2$, $Score^{\succeq^{\mathcal{C}}}_{cop}(2) = -1$, $Score^{\succeq^{\mathcal{C}}}_{cop}(3) = -1$. *So, now, 3 shares the second position with 2.*

[5] Note that S can be $\emptyset$.

Kramer-Simpson-like Method

Strongly inspired by the Kramer-Simpson method of social choice theory (*Minmax*), individuals are ranked inversely to their greatest pairwise defeat over all possible CP-comparisons.

Definition 4 (Kramer-Simpson-like solution)**.** Let $\succeq \in \mathcal{T}(\mathcal{P})$. The KS-like relation (KS relation) is the binary relation $R^{\succeq}_{KS} \subseteq N \times N$ such that for all $i, j \in N$:

$$iR^{\succeq}_{KS}j \Leftrightarrow Score^{\succeq}_{KS}(i) \leq Score^{\succeq}_{KS}(j),$$

where $Score^{\succeq}_{KS}(i) = \max_{j \in N}(d_{ji}(\succeq))$

Theorem 2. *The Kramer-Simpson (KS)-like solution is manipulable.*

Proof. See Example 5 for an instance of manipulation. □

Example 5. *Consider* $\succeq : 2 \succ (1 \sim 3) \succ 12 \succ (13 \sim 23) \succ \emptyset \succ 123$. *Then,* $Score^{\succeq}_{KS}(1) = 1, Score^{\succeq}_{KS}(2) = 0, Score^{\succeq}_{KS}(3) = 2$. *Hence, KS-like solution is:* $2P^{\succeq}_{KS}1P^{\succeq}_{KS}3$. *Now consider the following manipulation operated by 1 on* $\mathcal{C} = \{12\}$*:*

$$2 \succ^{\mathcal{C}} (1 \sim^{\mathcal{C}} 3) \succ^{\mathcal{C}} (13 \sim^{\mathcal{C}} 23) \succ^{\mathcal{C}} 12 \succ^{\mathcal{C}} \emptyset \succ^{\mathcal{C}} 123.$$

The new scores are: $Score^{\succeq^{\mathcal{C}}}_{KS}(1) = 1, Score^{\succeq^{\mathcal{C}}}_{KS}(2) = 1, Score^{\succeq^{\mathcal{C}}}_{KS}(3) = 1$. *So, individual 1 now gets the first position.*

Remark 1. *For* $n = 2$*, the Copeland-like solution and the KS-like solution coincide with the CP-majority relation, hence these solutions are not manipulable for* $n = 2$.

3.2 Ordinal Banzhaf

In the same spirit of the Banzhaf index [Banzhaf III, 1964], the ordinal Banzhaf solution is based on counting the number of positive and negative ordinal marginal contributions.

Definition 5 (Ordinal marginal contribution [Khani *et al.*, 2019])**.** Let $\succeq \in \mathcal{T}(\mathcal{P})$. The *ordinal marginal contribution* $m^S_i(\succeq)$ of player i w.r.t. coalition S, $i \notin S$, in power relation $\succeq$ is defined as:

$$m^S_i(\succeq) = \begin{cases} 1 & \text{if} \quad S \cup \{i\} \succ S, \\ -1 & \text{if} \quad S \succ S \cup \{i\}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Definition 6 (Ordinal Banzhaf relation)**.** Let $\succeq \in \mathcal{T}(\mathcal{P})$. The *ordinal Banzhaf relation* is the binary relation $R^{\succeq}_{Banz}$ such that for all $i, j \in N$:

$$iR^{\succeq}_{Banz}j \Leftrightarrow Score^{\succeq}_{Banz}(i) \geq Score^{\succeq}_{Banz}(j),$$

where $Score^{\succeq}_{Banz}(i) = u^{+,\succeq}_i - u^{-,\succeq}_i$ and $u^{+,\succeq}_i$ ($u^{-,\succeq}_i$) is defined as the number of coalitions S with $i \notin S$ such that $m^S_i(\succeq) = 1$ ($m^S_i(\succeq) = -1$).

Theorem 3. *The Ordinal-Banzhaf solution is manipulable.*

Proof. See Example 6 for an instance of manipulation. □

Example 6. *Consider* $\succeq : 13 \succ 1 \succ 12 \succ 23 \succ 2 \succ 3 \succ 123 \succ \emptyset$. *Then,* $1I^{\succeq}_{Banz}3P^{\succeq}_{Banz}2$ *since* $Score^{\succeq}_{Banz}(1) = 2, Score^{\succeq}_{Banz}(2) = 0, Score^{\succeq}_{Banz}(3) = 2$ *. However, if* 2 *undermines the cooperation with* 1 *and* 3 *(*$\mathcal{C} = \{12, 23\}$*):* $13 \succ^{\mathcal{C}} 1 \succ^{\mathcal{C}} 2 \succ^{\mathcal{C}} 23 \succ^{\mathcal{C}} 3 \succ^{\mathcal{C}} 12 \succ^{\mathcal{C}} 123 \succ^{\mathcal{C}} \emptyset$. *Then the three individuals would have a null Banzhaf score and would be ranked equally.*

3.3 Lexicographic Excellence Solution

The idea of lexicographic excellence is based on the lexicographic comparison of the frequency of individuals within equivalence classes, and taking care to reward individuals within the most excellent ones. Given the power relation $\succeq$ and its associated quotient ranking $\Sigma_1 \succ \Sigma_2 \succ \Sigma_3 \succ \cdots \succ \Sigma_m$, we denote by i_k the number of sets in Σ_k containing i:

$$i_k = |\{S \in \Sigma_k : i \in S\}|$$

for $k = 1, \ldots, l$. Now, let $\theta^{\succeq}(i)$ be the l-dimensional vector $\theta^{\succeq}(i) = (i_1, \ldots, i_l)$ associated to $\succeq$. Consider the lexicographic order $\geq_L$ among vectors **i** and **j**: $\mathbf{i} \geq_L \mathbf{j}$ if either $\mathbf{i} = \mathbf{j}$ or there exists $t : i_r = j_r, r = 1, \ldots, t-1,$ and $i_t > j_t$.

Definition 7 (Lexicographic-excellence solution [Bernardi *et al.*, 2019])**.** Let $\succeq \in \mathcal{T}(\mathcal{P})$. The *lexicographic excellence (lexcel) relation* is the binary relation $R^{\succeq}_{lexcel}$ such that for all $i, j \in N$:

$$iR^{\succeq}_{lexcel}j \Leftrightarrow \theta^{\succeq}(i) \quad \geq_L \quad \theta^{\succeq}(j).$$

Example 7. *Consider the power relation of Example 2. We have* $\theta^{\succeq}(1) = (0, 2, 1)$ *(since 1 is twice in* Σ_2 *and once in* Σ_3*),* $\theta^{\succeq}(2) = (1, 2, 0), \theta^{\succeq}(3) = (1, 1, 1)$*, which yields the following lexcel ranking:* $2P^{\succeq}_{lexcel}3P^{\succeq}_{lexcel}1$.

Theorem 4. *The lexcel solution is not manipulable.*

Proof. Let $\succeq \in \mathcal{T}(\mathcal{P})$ be a power relation on $\mathcal{P} \subseteq 2^N$ with the associated quotient order $\succ$:

$$\Sigma_1 \succ \Sigma_2 \succ \cdots \succ \Sigma_j \succ \cdots \succ \Sigma_k \cdots \succ \Sigma_m, \tag{3}$$

and let $\mathcal{C} = \{S_1, S_2, \ldots, S_l\} \subseteq \mathcal{P}$ and $i \in N$ be such that $i \in \bigcap_{S \in \mathcal{C}} S$ (wlog, assume $S_1 \succeq S_2 \succeq \ldots \succeq S_l$).

Suppose there exists a manipulation $\succeq^{\mathcal{C}}$ of $\succeq$ by i such that $p_i(R^{\succeq^{\mathcal{C}}}) \geq p_i(R^{\succeq})$. Then there must be some $k \in N \setminus \{i\}$ such that

$$kR^{\succeq}_{lexcel}i \quad \text{and} \quad iR^{\succeq^{\mathcal{C}}}_{lexcel}k. \tag{4}$$

First, notice that there exists some coalition $S \in \mathcal{C}$ such that $k \notin S$ (otherwise, if $i, k \in \bigcap_{S \in \mathcal{C}} S$, the manipulation would have no impact on the relative comparison of i and k, since in this case $\theta^{\succeq}(i) \geq_L \theta^{\succeq}(k) \Leftrightarrow \theta^{\succeq^{\mathcal{C}}}(i) \geq_L \theta^{\succeq^{\mathcal{C}}}(k)$).

Now let $S^* \in \mathcal{C}$ be a coalition not containing k with the smallest index in $\mathcal{C}$, and let $j(S^*) \in \{1, \ldots, m\}$ be such that $S^* \in \Sigma_{j(S^*)}$. Since $kR^{\succeq}_{lexcel}i$, we distinguish two cases:

i) $kI^{\succeq}_{lexcel}i$ (k and i are indifferent in $\succeq$ according to the lexcel relation). Then, by Definition 7, $\theta^{\succeq^{\mathcal{C}}}_v(k) = \theta^{\succeq}_v(k) = \theta^{\succeq}_v(i) = \theta^{\succeq^{\mathcal{C}}}_v(i)$ for all $v < j(S^*)$, while $\theta^{\succeq^{\mathcal{C}}}_{j(S^*)}(k) > \theta^{\succeq^{\mathcal{C}}}_{j(S^*)}(i)$. So, $kP^{\succeq^{\mathcal{C}}}_{lexcel}i$;

ii) $kP^{\succeq}_{lexcel}i$ (k is strictly stronger than i in $\succeq$ according to the lexcel relation). So, let t be the smallest index such that $\theta^{\succeq}_{t}(k) > \theta^{\succeq}_{t}(i)$. Moreover, let $q = \min(t, j(S^*))$. By Definition 7, $\theta^{\succeq^{\mathcal{C}}}_{v}(k) = \theta^{\succeq}_{v}(k) = \theta^{\succeq}_{v}(i) = \theta^{\succeq^{\mathcal{C}}}_{v}(i)$ for all $v < q$ and $\theta^{\succeq^{\mathcal{C}}}_{q}(k) > \theta^{\succeq^{\mathcal{C}}}_{q}(i)$. So, again, $kP^{\succeq^{\mathcal{C}}}_{lexcel}i$.

In both cases we get a contradiction with the fact that $iR^{\succeq^{\mathcal{C}}}_{lexcel}k$ by relation (4). Since $\mathcal{C}$ is an arbitrary collection of coalitions in $\mathcal{P}$, we have proved that the lexcel solution is not manipulable. □

A drawback of the lexcel solution is that it makes use of a limited amount of information, giving a quasi-dictatorial power to the best ranked coalitions. On the other hand, Copeland-like and KS-like rules are based on the intuitive idea of *ceteris paribus* comparisons and have the advantage of choosing the Condorcet winner (with respect to the CP-majority relation) when it exists. The Ordinal Banzhaf rule, based on the intuitive notion of ordinal marginal contribution of individuals, also offers another interesting interpretation in terms of fairness. In order to further study the effective impact of manipulation on these three appealing solutions, in Section 4 we analyse the computational complexity of manipulation and in Section 5 we introduce some simulations on several numerical instances.

4 Computational Complexity of Manipulation

As we have seen in the previous section, the Copeland-like solution, the KS-like solution, and the Ordinal Banzhaf solution are manipulable. We strengthen these results in this section by showing that, for each of these social ranking solutions, determining whether an individual can manipulate or not is an NP-hard problem. Let us state the problem precisely. As an instance, we have a set $N = \{1, \dots, n\}$ of individuals with a manipulator $t \in N$, a set $\mathcal{P} \subseteq 2^N$, and a power relation $\succeq$ on $\mathcal{P}$. The question is to determine whether a given solution is manipulable by t on $\mathcal{P}$, as defined in Definition 1.

Theorem 5. *For the Copeland-like solution, the KS-like solution, and the Ordinal Banzhaf solution, the manipulation problem is NP-hard.*

Proof. Due to lack of space, we only present the proof for the Ordinal Banzhaf solution. We build the following instance of the manipulation problem under the Ordinal Banzhaf solution. First, let us consider the following individuals:

- we associate to each edge $e_i \in E$ an individual that we call e_i as well (for convenience), and to each vertex v an individual that we call v as well (for convenience);
- two other individuals: t (the manipulator), and α.

For each vertex v, let us call $\mathcal{P}_v$ the set containing the subsets of individuals $\{v, \alpha, t\}$, $\{v, t\}$ and all subsets $\{v, e_i, \alpha, t\}$ for each edge e_i incident to v. $\mathcal{P}_v$ is ordered as follows in $\succeq$: $\{v, \alpha, t\}$ is the first one (strictly preferred to any other sets in $\mathcal{P}_v$), $\{v, t\}$ is the last one, and all $\{v, e_i, \alpha, t\}$ are equivalent, ranked between $\{v, \alpha, t\}$ and $\{v, t\}$.

Note that each set in $\mathcal{P}_v$ contains v, so for the scores the relative positions of 2 sets in $\mathcal{P}_v$ and $\mathcal{P}_{v'}$ do not matter; we do not specify it.

The contribution of these sets $\mathcal{P}_v$ to the scores are: $+|V|$ for α (due to $\{v, \alpha, t\} \succ \{v, t\}$ for each vertex/individual v), -2 for each object e_i (due to $\{v, \alpha, t\} \succ \{v, e_i, \alpha, t\}$, for each of the two extremities of edge e_i), and 0 for t (each set contains t).

The idea of the reduction is that, in order to manipulate, t has to become first (defeating α). To do this she shall put $\{v, \alpha, t\}$ below $\{v, t\}$ in some $\mathcal{P}_v$. But doing this, the score of the edges incident to v increases. t cannot do this for the two extremities of an edge, otherwise e_i defeats him.

To make this true, we need to add dummy individuals to adjust the initial scores of α and t. For the score of α, we add $\lambda = 2k - |V|$ individuals $b_1, \dots, b_\lambda$. For each b_i, we order $\{b_i, \alpha\} \succ \{b_i\}$. This gives an extra score of $\lambda = 2k - |V|$ to α, while b_i has score 0. Finally, we add an object γ, and order $\{\gamma, t\} \succ \{\gamma\}$, giving an extra score of 1 to the manipulator t.

Note that as previously we do not need to further specify $\succeq$, since the relative positions of sets containing different b_i, and/or γ, and/or in different $\mathcal{P}_v$, does not matter with respect to the scores (there is no other set inclusion).

To sum up, we have $|E| + |V| + 2 + (2k - |V|) + 1$ individuals: each individual e_i has score -2, individual t has score 1, and α has score $|V| + 2k - |V| = 2k$. All dummy objects and objects v have score 0. Note that the size of $\mathcal{P}$ is polynomial in the size of G.

We claim that t can manipulate if and only if there is an independent set of size k in G.

Suppose that there is an independent set S of size k in G. Then consider the manipulation where, for each v in S, t puts $\{v, e_i^j, \alpha, t\}$ down to the last position in $\mathcal{P}_v$. Then the score of α decreases by $2k$ and becomes 0. The score of e_i is modified in at most one $\mathcal{P}_v$, since S is an independent set, so it is at most 0. The score of t is still 1, and t manipulated the election.

Conversely, suppose that t can manipulate. Note that t cannot increase her own score, so she must make the score of α at most 1. This means that she has to put $\{v, \alpha, t\}$ in the last position in at least k sets $\mathcal{P}_v$. Let S be the corresponding set of vertices. If S contains both extremities of one edge e_i, then the score of e_i becomes $+2$, and t is not better of. So, in order to manipulate, the set S must be an independent set, and it is of size at least k. □

5 Simulations

Inspired by previous works in voting theory [Chamberlin, 1985], we study to what extend the three rules are manipulable. In other terms, based on computer simulations of various power relations, we estimate the probability that a manipulation occurs and we analyse the vulnerability of each solution to manipulation. We only consider situations with a single manipulator and power relations on the whole set 2^N. In order to perform our simulations we need to find a manipulation strategy for each rule under the assumption that the manipulator has a complete knowledge about the power relation.

To find a manipulation, we set up an integer linear programming (ILP) formulation of the problem (not detailed here due to lack of space), the variables of which represent the ranking after manipulation. This ILP is efficient enough

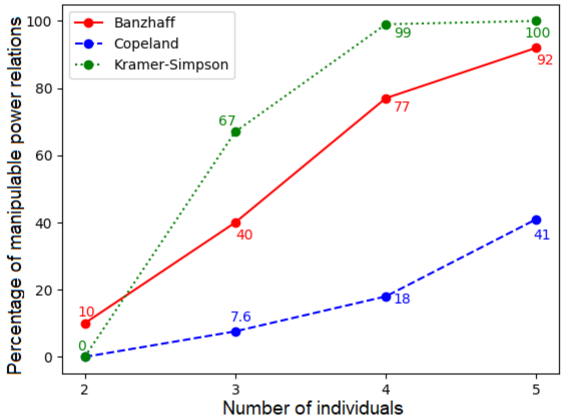

Figure 1: Percentage of manipulable cases as a function of $n = |N|$.

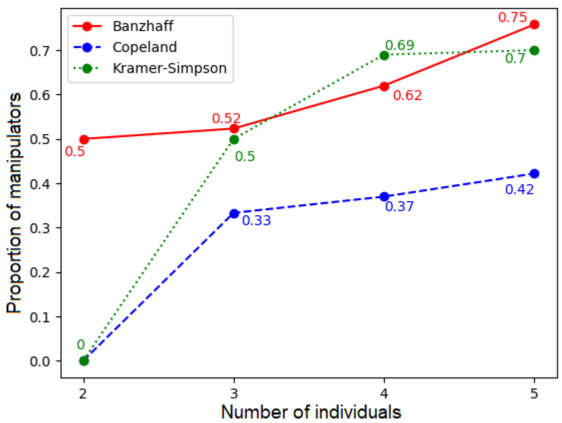

Figure 2: Proportion of manipulators as a function of $n = |N|$.

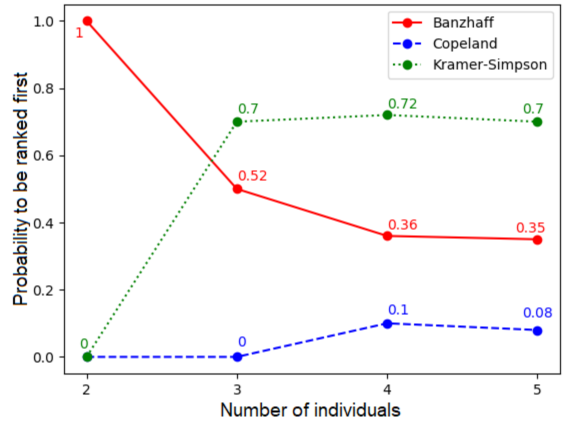

Figure 3: Probability to be ranked first as a function of $n = |N|$.

manipulable by	$N = 3$	$N = 4$	$N = 5$
Banz. +Cop. +KS	7.4	16.4	33.1
Banz. +Cop.	0	0	0
Banz. +KS	29.6	61.2	56.2
Banz.	1	0	0
Cop. +KS	0.2	0.4	4.7
Cop.	0	0	0
KS	28.7	22.0	6.0
none	33.1	0	0

Table 1: Percentage of manipulable power relations.

for small values of n, and we perform our simulation on total power relations over 2^N up to $n = 5$, i.e., power relations on up to $2^5 = 32$ coalitions. The data generation is done using Monte Carlo methods, following uniform (*impartial culture*) model, which assumes that all power relations over coalitions are equally likely to occur. For each number of individuals n, we generated 1000 random total power relations.

The proportion of manipulable cases. A power relation is manipulable if there exists at least one individual who can manipulate it. The probability of having a manipulable power relation increases rapidly with the number of individuals for the three solutions, especially for the KS relation for which it reaches 99.6% for $n = 4$ and 100% for $n = 5$. However, for Copeland it reaches 41% and 92% for Banzhaf (see Figure 1).

Number of possible manipulators. We look at the number of possible manipulators for each manipulable case. The results are shown in Figure 2. The proportion of manipulators grows with the number of individuals. We note also that for Copeland solution, there are on average less possible manipulators for each power relation, and thus has a lower probability of being actually manipulated by one of them.

Manipulating to be the best ranked. We analyse in the following the probability of becoming the best ranked one (ties are possible) thanks to a manipulation (see Figure 3).

Cross-simulation. We end our analysis by a cross simulation where for a given power relation we analyze the manipulability with respect to each social ranking rule (Table 1). Table 1 is coherent with our previous results (see Figure 1). Based on the results of simulations, it seems that if a power relation is manipulable by the Copeland-like solution, it is also by the KS-like solution. The most common case is to be manipulable by ordinal the Ordinal Banzhaf solution and the KS-like solution. For these reasons, our conjecture (suggested by the experimental results) is that the Copeland-like solution is not manipulable alone.

6 Conclusion

We have studied the problem of manipulating social ranking solutions. We have shown that lexcel is not manipulable and the manipulation of three other rules is NP-hard. Using simulation, we have remarked that Copeland-like is more resistant to manipulation than the Ordinal Banzhaf solution and the KS-like solution. Our study opens the way for many future works. We quote some of them : An axiomatic characterization of SRC rules taking into account strategy-proofness (like the one of Gibbard and Satterhwaite ([Gibbard, 1973] and [Satterthwaite, 1975])), analysis of the impact on the manipulability of some domain restrictions, study of coalitional manipulation or of simultaneous manipulation (game theoretical issues).

References

[Aziz *et al.*, 2011] Haris Aziz, Yoram Bachrach, Edith Elkind, and Mike Paterson. False-name manipulations in weighted voting games. *J. Artif. Intell. Res.*, 40:57–93, 2011.

[Bachrach and Rosenschein, 2009] Yoram Bachrach and Jeffrey S Rosenschein. Power in threshold network flow games. *Autonomous Agents and Multi-Agent Systems*, 18(1):106, 2009.

[Banzhaf III, 1964] John F Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev.*, 19:317, 1964.

[Bernardi *et al.*, 2019] Giulia Bernardi, Roberto Lucchetti, and Stefano Moretti. Ranking objects from a preference relation over their subsets. *Social Choice and Welfare*, 52(4):589–606, 2019.

[Bilbao *et al.*, 2002] Jesús Mario Bilbao, Julio R Fernandez, Nieves Jiménez, and Jorge J Lopez. Voting power in the european union enlargement. *European Journal of Operational Research*, 143(1):181–196, 2002.

[Chamberlin, 1985] John R. Chamberlin. An investigation into the relative manipulability of four voting systems. *Behavioral Science*, 30(4):195–203, 1985.

[Chiaramonte *et al.*, 2020] Alessandro Chiaramonte, Lorenzo De Sio, and Vincenzo Emanuele. Salvini's success and the collapse of the five-star movement: The european elections of 2019. *Contemporary Italian Politics*, pages 1–15, 2020.

[Copeland, 1951] Arthur H. Copeland. A reasonable social welfare function. Technical report, mimeo, 1951. University of Michigan, 1951.

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

[Faliszewski and Hemaspaandra, 2009] Piotr Faliszewski and Lane Hemaspaandra. The complexity of power-index comparison. *Theoretical Computer Science*, 410(1):101–107, 2009.

[Fayard and Öztürk, 2018] Nicolas Fayard and Meltem Öztürk. Ordinal Social ranking: simulation for CP-majority rule. In *Multiple Criteria Decision Aid to Preference Learning, DA2PL 2018*, Poznan, Poland, November 2018.

[Felsenthal *et al.*, 1998] Dan S Felsenthal, Moshé Machover, et al. The measurement of voting power. *Books*, 1998.

[Ferrara, 2004] Federico Ferrara. Electoral coordination and the strategic desertion of strong parties in compensatory mixed systems with negative vote transfers. *Electoral studies*, 23(3):391–413, 2004.

[Flores-Szwagrzak and Treibich, 2020] Karol Flores-Szwagrzak and Rafael Treibich. Teamwork and individual productivity. *Management Science*, 2020.

[Gibbard, 1973] Allan Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–601, 1973.

[Haret *et al.*, 2018] Adrian Haret, Hossein Khani, Stefano Moretti, and Meltem Öztürk. Ceteris paribus majority for social ranking. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 303–309, 2018.

[Holler and Packel, 1983] Manfred J Holler and Edward W Packel. Power, luck and the right index. *Zeitschrift für Nationalökonomie*, 43(1):21–29, 1983.

[Khani *et al.*, 2019] Hossein Khani, Stefano Moretti, and Meltem Öztürk. An ordinal banzhaf index for social ranking. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI 2019*, 2019.

[Kramer, 1977] Gerald H. Kramer. A dynamical model of political equilibrium. *Journal of Economic Theory*, 16(2):310–334, December 1977.

[Laruelle and Valenciano, 2001] Annick Laruelle and Federico Valenciano. Shapley-shubik and banzhaf indices revisited. *Mathematics of operations research*, 26(1):89–104, 2001.

[Laruelle and Valenciano, 2005] Annick Laruelle and Federico Valenciano. A critical reappraisal of some voting power paradoxes. *Public Choice*, 125(1-2):17–41, 2005.

[Matsui and Matsui, 2000] Tomomi Matsui and Yasuko Matsui. A survey of algorithms for calculating power indices of weighted majority games. *Journal of the Operations Research Society of Japan*, 43(1):71–86, 2000.

[Moretti and Öztürk, 2017] Stefano Moretti and Meltem Öztürk. Some axiomatic and algorithmic perspectives on the social ranking problem. In *Algorithmic Decision Theory - 5th International Conference, ADT 2017, Luxembourg, Luxembourg, October 25-27, 2017, Proceedings*, pages 166–181, 2017.

[Moretti and Patrone, 2008] Stefano Moretti and Fioravante Patrone. Transversality of the shapley value. *Top*, 16(1):1, 2008.

[Moretti, 2015] Stefano Moretti. An axiomatic approach to social ranking under coalitional power relations. *Homo Oeconomicus*, 32(2):183–208, 2015.

[Satterthwaite, 1975] Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[Shapley, 1953] Lloyd S Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317, 1953.

[Simpson, 1969] Paul B. Simpson. On Defining Areas of Voter Choice: Professor Tullock on Stable Voting. *The Quarterly Journal of Economics*, 83(3):478–490, 08 1969.

[Zuckerman *et al.*, 2012] Michael Zuckerman, Piotr Faliszewski, Yoram Bachrach, and Edith Elkind. Manipulating the quota in weighted voting games. *Artif. Intell.*, 180-181:1–19, 2012.