

# Manipulabilité des solutions de classement social

Tahar Allouche<sup>1</sup>, Bruno Escoffier<sup>2</sup>, Stefano Moretti<sup>3</sup>, Meltem Öztürk<sup>4</sup>

<sup>1</sup> ENSTA-Paris, France

`tahar.allouche@ensta-paris.fr`

<sup>2</sup> LIP6, Sorbonne Université, et Institut Universitaire de France

`Bruno.Escoffier@lip6.fr`

<sup>3</sup> Université Paris Dauphine, PSL Research University, CNRS, Lamsade, France

`Stefano.moretti@dauphine.fr`

<sup>4</sup> Université Paris Dauphine, PSL Research University, CNRS, Lamsade, France

`meltem.ozturk@dauphine.fr`

**Mots-clés :** *Classement social, Choix social, Manipulabilité, Compléxité, Simulations*

## 1 Introduction et contexte

Le problème de classement social (Social ranking) consiste en l'inférence d'un classement sur des individus à partir d'un classement sur les coalitions qu'ils forment. Par exemple, considérez un gestionnaire des ressources humaines qui doit promouvoir un des 3 employés  $N = \{1, 2, 3\}$ . Au cours du temps, on a suivi les performances des différents groupes de travail formés par les différentes combinaisons possibles des individus  $N$ , et on les a classées comme suit :

$$13 \succ 1 \succ 12 \succ 23 \succ 2 \succ 3 \succ 123 \succ \emptyset^1$$

L'objectif du manager est bien entendu celui de trouver la manière la plus équitable de classer les employés pour en choisir celui qui bénéficiera de la promotion.

Formellement, étant donné un ensemble d'individu  $N$ , une solution de classement social est une fonction  $R : C \subseteq \mathcal{B}(2^N) \rightarrow \mathcal{T}(N)$  qui à tout pré-ordre total  $\succeq$  sur un sous-ensemble de coalitions de  $2^N$  associe un pré-ordre total  $R(\succeq)$  sur  $N$ .

Plusieurs méthodes ont été récemment proposées et étudiées d'un point de vue axiomatique (des caractérisations ont été trouvées).

- La solution de CP-Majorité [5] compare les individus deux-à-deux en se basant sur des comparaisons *ceteris paribus* entre les coalitions qui les contiennent.
- La solution ordinale de Banzhaf [6] se base sur le principe des contributions marginales des individus.
- La solution d'excellence lexicographique [1] classe les individus en ne considérant que les informations sur les coalitions les mieux classées.

Par analogie au problème de l'élection en choix social, où des candidats sont classés selon les résultats des votes, et pour lequel chaque fonction non-triviale d'agrégation des choix est manipulable [4], on cherche à étudier la résistance des méthodes de classement social aux comportements individuels pouvant biaiser le classement final. Notons que la question de manipulabilité est une question centrale pour l'acceptabilité des systèmes électoraux et elle a engendré plusieurs questions théoriques et appliquées dans la littérature.

## 2 Contribution

En résumé, on a fourni une formalisation du problème de manipulation dans le cadre du classement social. Dans ce cadre, on a considéré un type particulier de comportement malicieux

---

1.  $13 \succ 1$  traduit le fait que  $\{1, 3\}$  est strictement préféré à  $\{1\}$

des individus, celui de choisir stratégiquement de paresser avec quelques groupes et de donner à fond avec d'autres pour se garantir un rang individuel supérieur.

On a étudié la manipulabilité des trois méthodes axiomatisées et montré que la solution ordinale de Banzhaf est manipulable alors que la solution d'excellence lexicographique et la solution de CP-Majorité ne le sont pas. Puisque cette dernière présente l'inconvénient de pouvoir générer des cycles paradoxaux du type Condorcet [3], on a proposé deux nouvelles variantes transitives inspirées des solutions de Kramer-Simpson [7] et de Copeland [2], et on a montré qu'elles sont également manipulables.

De plus, on a étudié la complexité de la manipulation des différentes solutions, et on a montré qu'il s'agit d'un problème NP-difficile (réduction au problème de stable dans un graphe). Pour compléter les résultats théoriques, on a introduit un programme linéaire en nombres entiers permettant de manipuler, si possible, un classement donné, en vue d'effectuer des simulations (voir figure 1) qui nous a permis de comparer les solutions de classement social selon la fréquence des manipulations, le nombre d'individus ayant la possibilité de manipuler simultanément et l'impact que leurs comportements ont sur le classement individuel final.

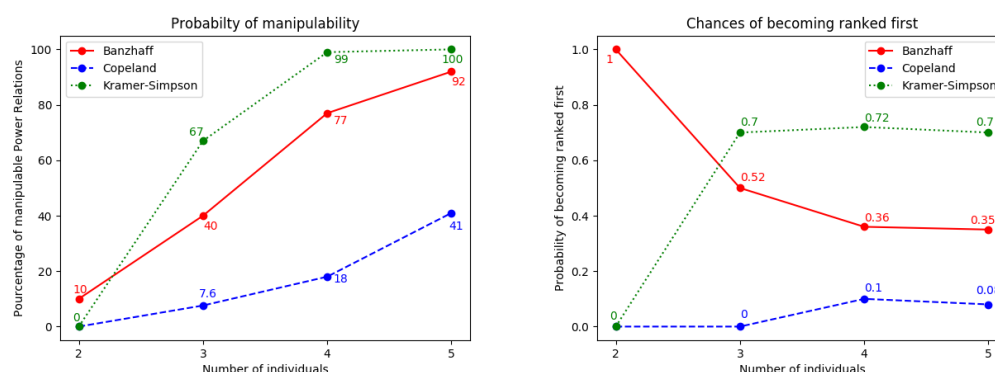


FIG. 1 – Quelques résultats des simulations

## Références

- [1] G. Bernardi, R. Lucchetti, and S. Moretti. Ranking objects from a preference relation over their subsets. *Social Choice and Welfare*, 52(4) :589–606, April 2019.
- [2] A.H Copeland. A reasonable social welfare function. Technical report, mimeo, 1951. University of Michigan, 1951.
- [3] N. Fayard and M. Öztürk. Ordinal Social ranking : simulation for CP-majority rule. In *Multiple Criteria Decision Aid to Preference Learning, DA2PL 2018*, Poznan, Poland, November 2018.
- [4] A. Gibbard. Manipulation of voting schemes : A general result. *Econometrica*, 41(4) :587–601, 1973.
- [5] A. Haret, H. Khani, S. Moretti, and M. Öztürk. Ceteris paribus majority for social ranking. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 303–309, 2018.
- [6] H. Khani, S. Moretti, and M. Öztürk. An ordinal banzhaf index for social ranking. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI 2019*, 2019.
- [7] Gerald H. Kramer. A dynamical model of political equilibrium. *Journal of Economic Theory*, 16(2) :310–334, December 1977.