Multi-Attribute Committee Selection: Diversity, Correlation, and Approximation Guarantees

Extended Abstract

Doron Heller
Technion-Israel Institute of Technology
Haifa, Israel
doron.heller@campus.technion.ac.il

Reshef Meir
Technion-Israel Institute of Technology
Haifa, Israel
reshefm@technion.ac.il

ABSTRACT

We study the problem of multi-attribute committee selection, where each candidate is described by categorical attributes and the objective is to match a given target distribution over these attributes. Building on the framework of Lang and Skowron [2], we formalize and analyze the L_p class of voting rules, including the well-studied Hamilton rule representing L_1.

We improve the additive approximation guarantees from [2], and extend our results to all L_p rules. Complementing this, we prove approximation lower bounds for the L_p class.

We then turn to analyze the effect of *correlated attributes*, which may indicate duplicates, and are likely to skew the outcome of L_p rules in a given direction. We propose a correlation-aware variant of L_2 and show it satisfies desired properties.

KEYWORDS

Computational Social Choice; Multiwinner Voting

ACM Reference Format:
Doron Heller and Reshef Meir. 2026. Multi-Attribute Committee Selection: Diversity, Correlation, and Approximation Guarantees: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/SXGS7111

1 INTRODUCTION

Many real-world scenarios require decision-makers to select a representative subset of alternatives based on a variety of attributes. For example, consider the task of forming a university admissions committee, which must reflect diversity in gender, academic background, and research fields.

These diversity criteria could be either set externally (e.g. by the university's by-laws); or decided on via voting, especially when there are many potential candidates. For example, in New York City's 2022–2023 Participatory Budgeting cycle, 46 out of 180 distinct projects were selected for funding [3]. In this case, attributes like: Domain, Modality, and Target population describe each project reasonably well. The project "Basic Necessities for Families" can be described as (Domain = Food Security, Target Population = Families, Modality = Distribution). It makes sense to have voters specify their preferences over a relatively small number of attributes rather than voting on particular projects, which also promotes higher turnout (see [1]).

To address such challenges, Lang and Skowron [2] introduced a formal framework for *multi-attribute proportional representation* (MAPR). In this model, each candidate is defined by values over a set of categorical attributes, and a vector of **target distributions** specifies the desired frequency of each attribute value in the committee.

The primary objective is to select a fixed-size committee that induces attribute distributions that are 'close' to the target distributions, according to some distance measure. Notice that every distance function induces a voting rule that minimizes said distance function.

In their work, Lang and Skowron [2] formalized the MAPR for a few different distance functions, including L_1 distance (a.k.a. Hamilton), which we discuss in this paper as well. They found that exact solutions are NP-hard and ILPs are infeasible for large instances of the problem, and focus on approximation algorithms for the different distance functions.

1.1 Our Research Problems

On the technical side, we are interested in tightening the approximation bounds on the MAPR problem. In particular we are interested in proving lower bounds, as these are absent from previous work.

On the conceptual side, each voting rule induces different soft diversity constraints, and diversity often requires guarding against the disproportionate underrepresentation of any single attribute. To address this, we consider the entire class of L_p voting rules (for finite p), which captures the whole range from equal treatment of attributes, to focusing on the worst attribute deviation.

Moreover, the outcome of any L_p voting rule may be biased towards a few similar attributes that are highly correlated, leading again to underrepresentation of the other attributes. This may happen if there are several attributes that essentially describe the same property (e.g., different ways of specifying the economic agenda of a political candidate, or the properties of a proposed PB project). These attributes are likely to have high correlation as voters are likely to vote similarly on them. Rules that only rely on the target distributions as input ignore this information, possibly allowing some latent attributes to have increased weight.

Such correlation may arise due to sloppy selection of attributes, or, more seriously, as a result of a manipulative attempt to control the agenda and the outcome. For example, a lobbyist who has some power over attribute selection can purposely select near-duplicate attributes to skew the outcome in their favor. We therefore consider

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/SXGS7111

	L_1		$L_p,\ p > 1$	
	old	new	old	new
UB	$\tilde{O}(d)$	$\tilde{O}(d/\sqrt{k})$	-	$\tilde{O}(d^{\frac{1}{p}}/\sqrt{k})$
LB	-	$\tilde{\Omega}(d^{1-\epsilon})$	-	$\tilde{\Omega}(d^{\frac{1}{p}-\epsilon})$

Table 1: Upper and Lower polynomial-time approximation bounds comparison for multi-attribute L_p problems. Parameter k represents committee size and parameter d represents the number of attributes. Bounds with ϵ mean that the bound is valid for any $\epsilon > 0$. Our new upper bounds are with high probability, whereas the old one from [2] is deterministic.

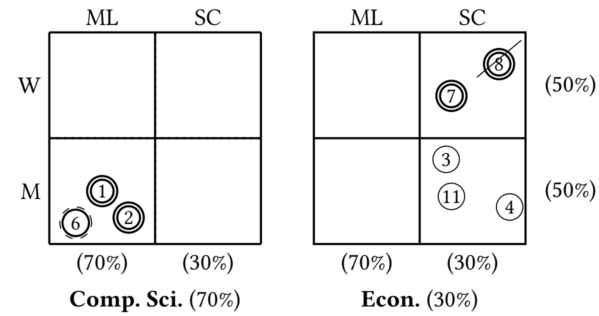

Figure 1: A three-attribute selection problem.

how duplicate and near-duplicate attributes should be treated to avoid such excessive bias.

2 APPROXIMATION RESULTS

In this section, we provide the first approximation hardness result for the MAPR model, specifically on the class of L_p rules. We also improve the approximation upper bound for L_1 and provide new upper bounds for the entire L_p class.

Since distributions induced by committees can have a distance of zero from target distributions and finding the best committee is NP-hard, there can be no interesting multiplicative approximation results. Therefore, the analysis in this model is focused on additive approximations.

Table 1 summarizes the main results in this section.

2.1 Approximation algorithm

On the algorithmic side, we formulate a natural integer program whose objective is an L_p norm over per-attribute deviations. Allowing for fractional committee members yields a *conic program* (linear when $p = 1$). We then round a fractional optimum using dependent rounding, preserving cardinality exactly and preserving marginal selection probabilities.

Because the algorithm independently draws committees from a distribution, we can repeat the randomized rounding multiple times and choose the best committee to amplify the probability of achieving the bound as much as we like.

Note that this is the first approximation result that shows larger committees allow better polynomial-time approximations.

2.2 Hardness of additive approximation

We provide the first strong additive inapproximability for multi-attribute L_p rules via a gap-preserving reduction from the maximum 3-dimensional matching problem.

Theorem 1. *For any constant $p \geq 1$ and any $\beta < \frac{1}{p}$, there exists a constant $\gamma > 0$ such that no polynomial-time algorithm can guarantee an additive $\gamma \cdot d^{\beta}$ approximation for multi-attribute L_p unless* P = NP.

This lower bound is particularly relevant where the number of attributes grows relative to committee size, and it quantifies a fundamental trade-off: richer attribute descriptions can make the approximation worse.

At first glance, it may seem the upper and lower bound conflict with each other. However, in the instances we generate in the lower-bound proof, it holds that $1 \leq k \leq d^{\epsilon}$. We conjecture that the best possible approximation is $\frac{d^{\frac{1}{p}}}{k}$.

3 CORRELATED ATTRIBUTES

Consider the problem of selecting new members to the board of the AAMAS conference in Figure 1. In this example, we have three attributes: Gender, Research Area, and Faculty. In this context, it is logical to assume that candidates researching Machine Learning (ML) are from the Computer Science faculty, and those researching Social Choice (SC) are from the Economics faculty. It is also safe to assume that voters would vote the same way on both attributes, resulting in a high correlation.

under the assumptions above, if we use the L_2 rule, the winning committee is $\{1, 2, 6, 7\}$ (or $\{1, 2, 6, 8\}$). However, if we remove either one of the 'Research Area' or 'Faculty' attributes, the winning committee would be $\{1, 2, \underline{7}, \underline{8}\}$. This problem in L_2 allows a lobbyist with some control over selecting the attributes to change the outcome in support of certain candidates.

Specifically, even though no information is lost from the removal of either attribute, the outcome of L_2 changed. To solve this problem, we aim to find a voting rule that satisfies the follwing properties:

(1) **Coincides with L_p.** When attributes are uncorrelated, the voting rule coincides with an L_p voting rule.
(2) **Eliminates duplicates** If two attributes are complete duplicates, the winning committee is the same as the winning committee if only one of them exists.

These two properties alone are trivial to satisfy, e.g. by manually removing all duplicate attributes and then use some L_p. However, this would be missing the point, as 'near-duplicate' attributes could be almost as bad, and any hard criterion for near-duplicates would be arbitrary. We therefore also require *continuity* of the voting rule. Intuitively this would mean that highly correlated attributes will be discounted more.

In the full version of the paper, we present the uncorrelated L_2 (UL_2) voting rule, which satisfies a formal variant of these properties. UL_2 coincides with L_2 when attributes are uncorrelated, and the problem of finding uncorrelated voting rules that coincide with other L_p rules is open.

REFERENCES

[1] Alan S. Gerber, Gregory A. Huber, and Seth J. Hill. 2013. Identifying the Effect of All-Mail Elections on Turnout: Staggered Reform in the Evergreen State. *Political Science Research and Methods* 1, 1 (June 2013), 91–116. https://doi.org/10.1017/psrm.2013.5

[2] Jérôme Lang and Piotr Skowron. 2018. Multi-attribute proportional representation. *Artificial Intelligence* 263 (2018), 74–106.

[3] New York City Civic Engagement Commission. 2023. Civic Engagement Commission Announces Implementing Partners for 46 Projects Funded for a Total of $5M in the First Cycle of "The People's Money" Participatory Budgeting Process. https://www.nyc.gov/site/civicengagement/news/cec-announces-46-projects-funded.page