	District A	District B	District C
AV	2D , 6G	-	-
CC	1D	3D	1E , 1G
PAV	5G	3E	-
sPAV	5G	3E	-
Any EJR rule	5G	3E	-

Table 1: Outcomes for the PB running example (D, E and G for diamond, emerald and gold respectively).

3.1 Popular PB Voting Rules

We turn to describe several popular voting rules from the literature that we later analyze.

Approval Voting (AV) The rule selects a feasible bundle $B \subseteq P$ that maximizes the social welfare $SW(A, B)$.

Approval Chamberlin–Courant (CC)[4] The rule selects a feasible bundle $B \subseteq P$ that maximize $RP(A, B)$.

The following rule change AV such that the score achieved from a voter for projects decrease as more of his approved projects are funded. This way increasing the amount of voters that get represented. **Proportional Approval Voting (PAV)[14]** The rule selects a feasible bundle $B \subseteq P$ that maximizes the following score:

$$SC_{\text{PAV}}(A, B) = \sum_{i \in V} \sum_{k=1}^{|A(i) \cap B|} \frac{1}{k}.$$

Sequential-PAV (sPAV) Solving PAV is NP-hard [2]; Sequential PAV is an efficient heuristic that proceeds as follows. Start with an empty bundle $B_0 = \emptyset$; in iteration i select a project $p \in P$, among all projects such that $B_{i-1} \cup \{p\}$ is feasible, that maximizes $SC_{\text{PAV}}(A, B_{i-1} \cup \{p\})$. Set $B_i = B_{i-1} \cup \{p\}$. Repeat until no project p can be added.

Remark: The outcome of a voting rule may contain several optimal bundles and thus require some tie-breaking rule. We specify the approriate rule when needed.

Table 1 shows the outcome of each voting rule when applied to the running example. As shown by the table, the rules vary widely in their outcome. For example, the voting rule AV does not choose any of the projects in District B and C, while CC chooses projects from all districts.

3.2 Proportional Voting Rules

Proportional voting rules ensure that sufficiently large groups of voters that share a large set of approved projects must also receive a fair amount of projects in the outcome. The key is the notion of T-cohesive groups which are groups of voters that share a subset of projects T and are able to fund T with the proportional part of the budget. Such groups are 'entitled' to a fair representation in the outcome of the PB instance. Formally:

Definition 3.1 (T-cohesive group [16, 17]). A group of voters $S \subseteq V$ that jointly approves a set of projects $T \subseteq \cap_{i \in S} A(i)$ is T-cohesive if $\frac{L}{|V|}|S| \geq cost(T)$.

Definition 3.2 (Extended Proportionality Representation (EJR) [16, 17][1]). A bundle B for PB instance $E = (A, cost, L)$ satisfies EJR if for every $T \subseteq P$ and every T-cohesive group S, it holds that there is $i \in S$ such that $|A(i) \cap B| \geq |T|$. A voting rule R satisfies EJR, if for every PB instance E, every bundle in the outcome $R(E)$ satisfies EJR.

In our running example, district A voters are T-cohesive for any set T of 5 cheap projects (gold) in district A, and district B voters are T-cohesive for the set T of 3 cheap projects in district B. Any voting rule that satisfies EJR must include 5 projects approved by district A voters, and 3 projects approved by district B voters. As can be seen in Table 1, both PAV and sPAV satisfy EJR on this example, while AV and CC do not. In general, none of the rules in Section 3.1 are guaranteed to satisfy EJR (as shown in the running example for AV and CC, and for PAV, sPAV by Peters et al. [16]).

A well-known rule from the literature that satisfies the EJR property was suggested by Peters et al. [16], and is called Rule X (RX for short). This voting will be used later in Section 5 as a representative of the EJR voting rules. Rule X is not as simple to describe as the aforementioned rules, therefore, it will be described in detail in Appendix B.

Remark. The EJR property does not require that a voting rule exhaust the entire budget. Any voting rule that does not exhaust the budget cannot achieve an optimal social welfare, as adding any project (that at least one voter approved) with the leftover budget will increase the welfare (and possibly the representation). There are many ways to make sure the voting rule exhaust the budget, e.g. Peters et al. [16] do so by giving some very small gain to projects that voters did not approve, this way making sure that RX outcome use the entire budget.

3.3 Worst-Case Guarantees

We follow the definitions of Lackner and Skowron [12] for utilitarian and representation guarantees. Given a participatory budgeting instance E, the *utilitarian ratio* of a voting rule R for instance E is the proportion of the social welfare given by R (in this case, ties are broken according to the minimum social welfare over all bundles B in the outcome of $R(E)$) divided by the optimal social welfare over all feasible bundles, the set $S(E)$.

$$K^R_{SW}(E) := \frac{SW(A, R(E))}{\max_{B \in S(E)} SW(A, B)} \quad (1)$$

The worst-case *utilitarian guarantee* of rule R is the minimal utilitarian ratio:

$$K^R_{SW}(N, M, L, c_{min}, c_{max}) := \inf_{E \in \mathcal{E}(N,M,L,c_{min},c_{max})} K^R_{SW}(E) \quad (2)$$

In the same way $K^R_{RP}(N, M, L, c_{min}, c_{max})$ is the worst-case *representation guarantee* of rule R; this time ties are broken according to the bundle with the worse representation in $R(E)$.

When omitting one or more of the arguments $N, M, L, c_{min}, c_{max}$ in K_{RP} or K_{SW}, we are taking the infimum over these arguments. E.g. $K^R_{SW}(N, c_{min}) := \inf_{L,M,c_{max}} K^R_{SW}(N, M, L, c_{min}, c_{max})$.

Table 2 shows the social welfare and representation ratios for our running example. By definition, the utilitarian guarantee of AV and representation guarantee of CC equal to 1.

[1] Not to confuse with representation

	AV	CC	PAV	sPAV	Any EJR voting rule
SW	800	390	770	770	770
RP	100	200	190	190	190
$K^R_{SW}(E)$	1	0.4875	0.9625	0.9625	0.9625
$K^R_{RP}(E)$	0.5	1	0.95	0.95	0.95

Table 2: The welfare / representation scores and their ratios achieved by different voting rules for the running example (ratios are defined in Section 3.3)

4 WORST-CASE GUARANTEES OF PB VOTING RULES

In this section, we describe our first contribution of computing the worst case welfare and representation guarantees for voting rules from Section 3. We then compute the worst case guarantees for the family of rules that satisfy the EJR property. Table 3 show a summary of all our theoretical guarantees, side-by-side with the results of Lackner and Skowron [12] for multiwinner voting.

This section will feature the lower bound on social welfare guarantee of PAV and the lower and upper bounds on sPAV. Due to lack of space, the proofs of the other bounds and rules are left for the appendix. The proofs in the main body of the paper represent the spirit of how we derive lower bounds (general argumentation) and upper bounds (construction of a certain PB).

Notice that proofs from Lackner and Skowron [12] for the MW setting rely heavily on the fact that costs are uniform, which fail in the PB setting.There are a few proofs that follow the same outlines as Lackner and Skowron [12] and we shall point this out.

4.1 Common Voting Rules

We start with the guarantees for AV, CC, PAV and sPAV.

Proposition 4.1. $\forall L, c_{min} : K^{PAV}_{SW}(L, c_{min}) \geq \frac{c_{min}}{L} \log\left(\frac{L}{c_{min}}\right)$

The heart of the proof lies in the following technical lemma, whose proof is deferred to right after the proof of this claim.

Lemma 4.2. *For any feasible bundle B of projects and any approval profile A it holds* $\frac{SC_{PAV}(A,B)}{SW(A,B)} \geq \frac{c_{min}}{L} \log\left(\frac{L}{c_{min}}\right)$.

Proof of Prop. 4.1. Given a PB instance E, we denote by B_{SW} the bundle with largest SW, and B_{PAV} the one with largest PAV score. From Lemma 4.2 the following holds:

$$\begin{aligned} SW(A, B_{\text{PAV}}) &\geq SC_{\text{PAV}}(A, B_{\text{PAV}}) \geq SC_{\text{PAV}}(A, B_{SW}) \\ &\geq \frac{c_{min}}{L} \log\left(\frac{L}{c_{min}}\right) SW(A, B_{SW}), \end{aligned}$$

Which entails:

$$\frac{SW(A, B_{\text{PAV}})}{SW(A, B_{SW})} \geq \frac{\frac{c_{min}}{L} \log\left(\frac{L}{c_{min}}\right) SW(A, B_{SW})}{SW(A, B_{SW})} = \frac{c_{min}}{L} \log\left(\frac{L}{c_{min}}\right),$$

as required. □

Proof of Lemma 4.2. We will use the following notation: Let $B_i = A(i) \cap B$ be the projects in bundle B that voter i approves, $V(B)$ is the set of voters with $|B_i| > 1$ and $V1(B)$ are all voters with $|B_i| = 1$.

The harmonic sum $H(k) = 1 + \frac{1}{2} + \ldots + \frac{1}{k}$ is at least $\log(k)$ (in base e), therefore, for any bundle B and approval profile A,

$$\begin{aligned} SC_{\text{PAV}}(A, B) &= \sum_{i \in V} H(B_i) = \sum_{i \in V(B)} H(B_i) + |V1(B)| \\ &\geq \sum_{i \in V(B)} \log(|B_i|) + |V1(B)| \\ &= \log(\prod_{i \in V(B)} |B_i|) + |V1(B)| \end{aligned} \tag{3}$$

And the welfare of B is:

$$SW(A, B) = \sum_{i \in V} |B_i| = \sum_{i \in V(B)} |B_i| + |V1(B)| \tag{4}$$

From Eq. (3) and Eq. (4) we have:

$$\frac{SC_{\text{PAV}}(A, B)}{SW(A, B)} \geq \frac{\log(\prod_{i \in V(B)} |B_i|) + |V1(B)|}{\sum_{i \in V(B)} |B_i| + |V1(B)|} \geq \frac{\log(\prod_{i \in V(B)} |B_i|)}{\sum_{i \in V(B)} |B_i|} \tag{5}$$

We now want to find a lower bound on the right hand side of the last equation. Note that the lower bound does not depend on the actual B_i's but only on their size.

Intuitively, we want to show that the lowest value is obtained when the sets sizes' are most unbalanced—essentially when there is only one nonempty set.

For this, we solve the following optimization problem instead, which is a relaxation of the above problem. Let $T = |V(B)|$ and $Q = \sum_{i \in V(B)} |B_i|$. Define the convex set

$$C = \{(q_1, \ldots, q_T) \text{ s.t. } \forall i\, q_i \geq 2 \text{ and } \sum_{i=1}^{T} q_i = Q\}.$$

Using the notation, for any bundle B, the right hand side in Eq. (5) is lower bounded by

$$\inf_C \frac{\log(\prod_{i \in T} q_i)}{\sum_{i \in T} q_i} = \inf_C \frac{\log(\prod_{i \in T} q_i)}{Q}. \tag{6}$$

The product of the q_i's is minimal when the distribution of q_i's is the most unbalanced. By setting the minimal value $q_i = 2$ for all $i > 1$, and $q_1 = Q - 2(T-1)$, we get

$$(6) \geq \inf_T \frac{\log(q_1 2^{T-1})}{Q} = \inf_T \frac{(T-1)\log 2 + \log(Q - 2(T-1))}{Q}.$$

Taking the derivative w.r.t. T, we get that this is a convex function with a maximum at $T = \frac{Q}{2} + 1 - \frac{1}{\log 2} > \frac{Q}{2} - 1$. However $Q \geq q_1 - 2(T-1) \geq 2 - 2(T-1) = 2T$, i.e. $T \leq \frac{Q}{2}$ so the only possible integer solutions are $T = 1$ and $T = \frac{Q}{2}$, which map to $\frac{\log Q}{Q}$ and $\frac{\log 2}{2}$, respectively.

Hence the minimum is obtained at $T = 1$, which means that $q_1 = Q$ (the solution of the relaxed problem is also a valid solution

Rule	Participatory budgeting Lower	Participatory budgeting Upper	Multi-winner Lower	Multi-winner Upper
AV	1	1	1	1
CC	$\Omega\left(\frac{1}{L}\right)$ (Prop. A.3)	$O\left(\frac{1}{L}\right)$ (Prop. A.4)	$\Omega\left(\frac{1}{L}\right)$	$O\left(\frac{1}{L}\right)$
PAV	$\Omega\left(\frac{\log(L)}{L}\right)$ (Prop. 4.1)	$O\left(\frac{\log(L)}{L}\right)$ (Prop. A.5)	$\Omega\left(\frac{1}{\sqrt{L}}\right)$	$O\left(\frac{1}{\sqrt{L}}\right)$
sPAV	$\Omega\left(\frac{1}{NL}\right)$ (Prop. 4.3)	$O\left(\frac{1}{N}\right)$ (Prop. 4.3)	$\Omega\left(\frac{1}{\sqrt{L}}\right)$	$O\left(\frac{1}{\sqrt{L}}\right)$
EJR rules	$\Omega\left(\frac{1}{Nc_{max}}\right)$ (Prop. 4.4)	$O\left(\frac{1}{\sqrt{Nc_{max}}}\right)$ (Prop. A.9)	$\Omega\left(\frac{1}{N}\right)$(Prop. 4.6)	$O\left(\frac{1}{N-L}\right)$ (Prop. A.10)

Rule	Participatory budgeting Lower	Participatory budgeting Upper	Multi-winner Lower	Multi-winner Upper
AV	$\Omega\left(\frac{1}{Lc_{max}}\right)$ (Prop. A.1)	$O\left(\frac{1}{L}\right)$ (Prop. A.2)	$\Omega\left(\frac{1}{L}\right)$	$O\left(\frac{1}{L}\right)$
CC	1	1	1	1
PAV	$\Omega\left(\frac{1}{\log(L)}\right)$ (Prop. A.6)	$O\left(\frac{1}{\log(L)}\right)$ (Prop. A.7)	$\frac{1}{2}$	$\frac{1}{2}+O\left(\frac{1}{L}\right)$
sPAV	$\Omega\left(\frac{1}{N}\right)$ (Prop. 4.3)	$O\left(\frac{1}{N}\right)$ (Prop. 4.3)	$\Omega\left(\frac{1}{\log(L)}\right)$	$\frac{1}{2}+O\left(\frac{1}{L}\right)$
EJR rules	$\Omega(\frac{1}{N})$ (Prop. A.8)	$O(\frac{1}{N})$ (Prop. A.8)	$\Omega\left(\frac{1}{N}\right)$(Prop. 4.6)	$O\left(\frac{1}{N-L}\right)$ (Prop. A.10)

Table 3: (top) Utilitarian and (bottom) representation guarantees for PB and multi-winner for rules studied in the paper, as a function of the budget (L), the number of voters (N) and the highest project cost c_{max}. We assume w.l.o.g. that the cost of the cheapest project is 1. The multi-winner guarantees are taken from Lackner and Skowron [12].

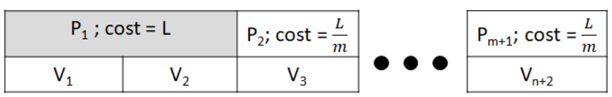

Figure 2: PB instance for Prop. 4.3 showing projects $p_1, p_2, \ldots p_{m+1}$ and voters $v_1, v_2, \ldots, V_{n+1}$. Each voter appears below its approved projects. In this example the sPAV selects project p_1.

of our original problem). Back to bundle problem, $T = 1$ entails $|V(B)| = 1$ and $|B_1| = Q$. Plugging this back into Eq. (5),

$$\arg\min_B \frac{SC_{\text{PAV}}(A,B)}{SW(A,B)} \geq \frac{\log(|B_1|)}{|B_1|} \geq \frac{\log(|B|)}{|B|} \tag{7}$$

Since the size of any feasible bundle is at most $\lfloor \frac{L}{c_{min}} \rfloor \leq \frac{L}{c_{min}}$, we get from Eq. (7):

$$\arg\min_B \frac{SC_{\text{PAV}}(A,B)}{SW(A,B)} \geq \frac{c_{min}}{L} \log(\frac{L}{c_{min}}) \tag{8}$$

This completes the proof of Lemma 4.2. □

Proposition 4.3.

$$\forall N, L, c_{min}: \frac{c_{min}}{NL} \leq K_{SW}^{sPAV}(N, L, c_{min}) \leq \frac{2}{N}$$
$$\forall N: \frac{1}{N} \leq K_{RP}^{sPAV}(N) \leq \frac{2}{N}$$

Proof. Consider the PB instance presented in Figure 2 with a budget of L. There are $N = n+2$ voters $\{v_1, \ldots, v_{n+2}\}$ and $M = m+$ projects $\{m_1, \ldots, m_{m+1}\}$.

The first 2 voters approve project p_1 that costs L and the rest of the voters approve one project each, at a cost of $\frac{L}{m}$.

At its first iteration, sPAV chooses p_1, adding 2 to the score, while the addition of any other project adds 1. Therefore, sPAV will fund $B_{\text{sPAV}} = \{p_1\}$ and stop, having an outcome with welfare of 2. The bundle $B_{SW} = B_{RP} = \{p_2, \ldots, p_{m+1}\}$ maximizes both welfare and representation with value n. Putting things together we get that

$$\frac{SW(A, B_{\text{sPAV}})}{SW(A, B_{SW})} = \frac{RP(A, B_{\text{sPAV}})}{RP(A, B_{\text{CC}})} = \frac{2}{n} = \frac{2}{(N-2)} \geq \frac{2}{N} \tag{9}$$

As for the lower bound, any voting rule will fund at least one project and for any instance there can be at most $\lfloor \frac{L}{c_{min}} \rfloor$ projects funded that all voters want, therefore:

$$\frac{SW(A, B_{\text{sPAV}})}{SW(A, B_{SW})} \geq \frac{1}{N\lfloor \frac{L}{c_{min}} \rfloor} \geq \frac{c_{min}}{NL} \tag{10}$$

$$\frac{RP(A, B_{\text{sPAV}})}{RP(A, B_{\text{CC}})} \geq \frac{1}{N} \tag{11}$$

□

The rest of the results for AV, CC, PAV and sPAV can be seen in Table 3. Their proofs are left for the appendix, where the upper guarantees follow the same idea as in Proposition 4.3, and the rest take into advantage the voting rule properties as done for Proposition 4.1 (Propositions A.2, A.1, A.4 follow the same outlines as Lackner and Skowron [12]).

4.2 EJR Voting Rule Guarantees

In this section we will present the utilitarian and representation guarantees for the family of EJR voting rules.

Proposition 4.4. *Let R be a voting rule that satisfies the EJR property. Then the utilitarian guarantee satisfies*

$$\forall N, L, c_{min} : K^R_{SW}(N, L, c_{min}) \geq \frac{c_{min}}{NL} \lfloor \frac{L}{c_{max}} \rfloor.$$

Proof. To avoid trivialities, we consider rules that exhaust the entire budget (EJR does not require that). First let us lower bound the SW of an EJR rule R with respect to some PB instance E. Let $T \subseteq P$ be the largest set of projects that is T-cohesive with respect to E. Let B be a bundle in the outcome of R and $B \subseteq B'$ its extension to consume the remaining budget. From Definitions 3.1 and 3.2 it readily follows that any bundle B in the outcome of R satisfies $SW(A, B) \geq \lceil |T| \frac{N}{L} \rceil$.

In the worse case $T = \emptyset$, namely R is EJR in an empty way. Since we assume that all the budget is consumed, then B' (perhaps even B) contains at least $\lfloor \frac{L}{c_{max}} \rfloor$ projects (otherwise the budget is not consumed). Therefore $SW(A, B') \geq \lfloor \frac{L}{c_{max}} \rfloor$.

As for the bundle B_{SW} that maximizes the social welfare, there are at most $\lfloor \frac{L}{c_{min}} \rfloor$ projects possible to fund, each one of them is supported by at most all N voters. This means that

$$SW(\text{AV}) \leq N \lfloor \frac{L}{c_{min}} \rfloor$$

Putting everything together we get that

$$K^R_{SW} \geq \frac{SW(A, B')}{SW(A, B_{SW})} \geq \frac{\lfloor \frac{L}{c_{max}} \rfloor}{N \lfloor \frac{L}{c_{min}} \rfloor} \geq \frac{c_{min}}{NL} \lfloor \frac{L}{c_{max}} \rfloor \quad \square$$

The rest of the results for EJR voting rules in PB can be seen in Table 3. The proofs for those results are in Prop. A.8 and Prop. A.9.

In order to have complete comparison of the PB results to multi-winner, we will also find the guarantees for EJR voting rules in multi-winner. Notice that in multi-winner context, $L \in \mathbb{N}^+$ and tell how many projects should be selected.

Proposition 4.5. *Let R be a voting rule that satisfies the EJR property. Then the utilitarian and representation guarantees satisfies*

$$\forall N, L, \ s.t.\ N \geq 2L : K^R_{SW}(N, L, c_{min} = 1, c_{max} = 1) \leq \frac{1}{N - L}$$

$$\forall N, L, \ s.t.\ N \geq 2L : K^R_{RP}(N, L, c_{min} = 1, c_{max} = 1) \leq \frac{1}{N - L}$$

.

It is worth noting that while the bound in Prop. 4.5 (proof in Prop. A.10) can slightly improve on our bound for general PB problems (Prop. A.9), asymptotically Prop. A.9 provides a tighter bound, that also shows some PB instances are worse (in terms of welfare) than any MW instance.

Proposition 4.6. *Let R be a voting rule that satisfies the EJR property. Then the representation guarantee satisfies*

$$\forall N : K^R_{RP}(N, c_{min} = 1, c_{max} = 1) \geq \frac{1}{N}$$

This is the trivial guarantee, as the optimal outcome represents at most all voters, and the voting rule outcome represents at least a single voter.

Proposition 4.7. *Let R be a voting rule that satisfies the EJR property. Then the utilitarian guarantee satisfies*

$$\forall N : K^R_{SW}(N, c_{min} = 1, c_{max} = 1) \geq \frac{1}{N}.$$

Proposition 4.7 is proven the same way as Proposition 4.4.

We end this section with three conclusions that we draw from Table 3:

- The guarantees for CC and AV are the same for the multi-winner and PB settings (up to a c_{max} factor in the representation lower bound of AV). The case for PAV and PAV is very different. The PB guarantees are an order of magnitude lower, and for PAV they also depend on N; no multi-winner guarantee depends on N. This caused because they "ignore" the projects cost, however, less significant for PAV, as it take the cost into account indirectly when solving the optimization problem.
- Our results induce a nearly-strict order over the voting rules:

  $$\text{welfare: AV} \gg \text{PAV} \gg \text{CC} \gg EJR \gg \text{PAV}$$

  $$\text{representation: CC} \gg \text{PAV} \gg \text{AV} \gg \text{PAV}, EJR$$

  These two rankings are similar to the ones obtained in the multi-winner setting, except for PAV which drops to the bottom when introducing costs (the PB setting).
- Most voting rules' guarantees depend on the budget and projects cost, while EJR voting rules guarantees depend on the number of voters and projects cost. This means that as the number of voters grows, the "cost of proportionality" may be rising as well.

5 EXPERIMENTAL EVALUATION

In this section we examine the performance of the rules in practice on real world and synthetic data, beyond the worst-case scenario.

For every dataset and every voting rule we calculate the utilitarian ratio and representation ratio for all the instances in that dataset. We report the average and standard error. Tables 4,5 and 6 report the results for the three datasets.

Poland A dataset that was taken from Pabulib.org [22], a library of PB instances available to the research community. We looked at 130 instances that took place in different districts of Warsaw, Poland, in the years 2017–2021. Each instance included between 50-10,000 voters (2,982 on average) and between 20-100 projects (36 on average).

Euclidean This dataset consists of 1,000 synthetic PB instances, each containing 1,000 voters, 100 projects and a budget of $L = 10^5$. In a city, most of the population lives near the city center, and so the location of a project is more likely to be there as well. To this end, for every instance, the locations of the voters and projects are generated randomly according to a 2-dimensional euclidean model [5, 21, 24]. Each voter v and project p are given some location ℓ_v, ℓ_p in the unit square $[0, 1] \times [0, 1]$, according to the normal distribution

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \Sigma = \begin{pmatrix} 0.2^2 & 0 \\ 0 & 0.2^2 \end{pmatrix}$$

The costs of the projects are parameterized with two values $c_{min} \in [100, 500]$, $c_{avg} \in [10^4, 2 \cdot 10^4]$, the minimum project cost and the average one, are both chosen uniformly at random.

The project costs are chosen according to the following procedure. For every project p_i choose c_i from the exponential distribution with $\lambda = c_{avg} - c_{min}$ and the cost of p_i will be $c_{min} + c_i$. This simulates a scenario with many cheap projects, and a few expensive ones.

To create the approval profile for each voter i, a number a_i is chosen from the normal distribution with $\mu = 10, \sigma = 3$, and the set of projects $A(i)$ approved by voter i consists of the $\max(a_i, 1)$ closest projects to the location of voter i.

Party-list This is also a synthetic dataset containing 1,000 party-list [12] PB instances that satisfy the following condition: every pair of voters i, j, either approve the same list of projects, $A(i) = A(j)$, or don't approve any mutual one, $A(i) \cap A(j) = \emptyset$. Each instance includes 200 voters which are split uniformly at random into groups of sizes 5 to 20. Each group of voters approves uniformly at random between 10 to 30 different projects. The cost of the projects is linear in the group size, such that the more voters a group has, the higher the cost of the projects they approve.

Due to the criteria above, PB instances in the party-list dataset contain large parties that tend to approve expensive projects. Funding these projects will contribute significantly to the overall SW, but will consume large part of the budget, risking that small parties will not be represented, thus violating the EJR property.

5.1 Voting Rules

To apply the voting rules described in Section 3.1 to the PB datasets, we extended the python framework used by Lackner et al. [11], originally designed to find committees in the multi-winner setting. The AV, CC and PAV were solved using linear programming with the Gurobi solver [9].

For the sake of efficiency, instead of breaking ties for the worst-case ratio, we broke them at random. This allowed to test the voting rules on larger instances in reasonable time. On several random instances that we sampled, we did break ties for the worst-case and also for the best-case, and noted no significant change in the final score compared to the random policy.

One exception is CC in the party-list dataset. In this dataset, it is possible to get 100% representation with only a small portion of the budget spent. This leads to a variety of CC-optimal bundles, achieving a wide range of welfare scores. The Gurobi solver selected, for unclear reasons, only solutions with high social welfare. To compensate for that, in this case only, we took the worst-case SW solution, and reported this result in Table 6 (we added a penalty for SW in the objective function and then ran Gurobi).

As mentioned in Section 3.1, none of the voting rules satisfies the EJR property. We use Rule X [16] (RX) as a representative from the EJR family. We refer the reader to the original paper for a description of this rule [16]. One caveat is that RX does not necessarily exhaust the entire budget, in contrast to the other rules we consider. To allow a fair comparison with the other rules we will also consider two extension to Rule-X. The first variant, RX-ϵ, is described in Peters et al. [16]. The second variant, RX-PAV, applies RX to the PB instance, and runs PAV on the remaining budget over the unfunded projects. The outcome of RX-PAV is defined as the union of the outcomes of the RX and PAV rules.

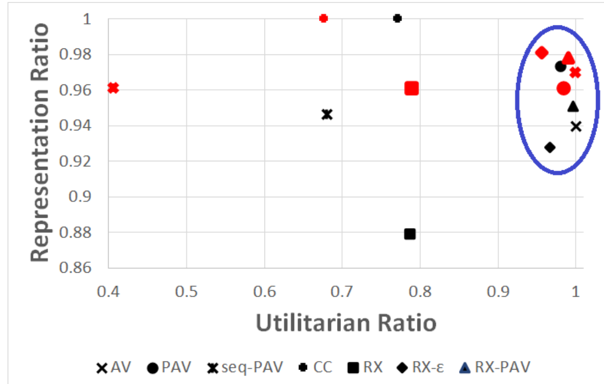

Figure 3: Utilitarian ratio vs. Representation ratio results for the Poland (black) and Euclidean (bigger red).

5.2 Results

Figure 3 shows the trade-off between welfare and representation when applying the voting rules on Poland and Euclidean datasets (the party-list dataset was omitted in this figure as all voting rules except AV gave an outcome with 100% representation). We see a cluster of voting rules (marked with blue circle) that includes AV, PAV, RX-ϵ and RX-PAV, achieving the best trade-off between welfare and representation. On the other, the outlier sPAV achieves low welfare ratio in the Poland dataset and a very low welfare ratio in the Euclidean dataset.

Tables 4,5 and 6 provide a finer level of granularity of the results, by displaying the average ratios and percentages (with standard error) of PB instances that satisfy EJR for the Poland, Euclidean and Party-list datasets, respectively. As can be seen in the three tables, the ratios are similar across datasets. Specifically, both PAV and RX-PAV succeed in achieving high utilitarian and representation ratios for all datasets. Another noticeable result, is the fact that sPAV achieves quite poor results in both ratios, which is in-line with the ranking that we presented at the end of Section 4. In addition, sPAV exhibits large variance in all three datasets (compared to the other voting rules), which further emphasizes that the sPAV rule is unstable.

The results from the Poland dataset in Table 4 demonstrate the disadvantage of using an EJR voting rule that does not guarantee to exhaust the budget. For this dataset, RX achieved poorer results for both welfare and representation compared to all other voting rules. In contrast, RX-ϵ and RX-PAV, which are similar to RX, but make sure to exhaust the entire budget, succeed in gaining a significant improvement in both measurements. This result emphasizes the benefit of using the entire budget, even if the EJR requirement is satisfied before exhausting the budget.

Lastly, looking at percentage of instances where the chosen bundle satisfied EJR, we can see that for both the Poland and Euclidean datasets (Tables 4,5), almost all rules succeed in getting an outcome which satisfies EJR. This result is interesting, in that even though a voting rule is not guaranteed to always produce a solution that satisfies EJR, this is often the case. This phenomenon may be explained by the fact that there are T-cohesive groups only for small sets of projects T in those datasets, in other words voters are entitled to only a few projects. This makes the EJR requirement easier

	Utilitarian ratio	Representation ratio	EJR%
AV	1 ± 0	0.9401 ± 0.054	100
PAV	0.9802 ± 0.025	0.9735 ± 0.026	100
sPAV	0.6801 ± 0.18	0.9468 ± 0.073	94.6
CC	0.7713 ± 0.12	1 ± 0	100
RX	0.7868 ± 0.065	0.8793 ± 0.08	100
RX-ϵ	0.9666 ± 0.025	0.9278 ± 0.063	100
RX-PAV	0.9966 ± 0.006	0.9508 ± 0.044	100

Table 4: Ratios and percentage of instances which satisfy EJR (Poland dataset)

	Utilitarian ratio	Representation ratio	EJR%
AV	1 ± 0	0.9696 ± 0.021	100
PAV	0.9845 ± 0.011	0.9866 ± 0.008	100
sPAV	0.4063 ± 0.095	0.8423 ± 0.089	64.7
CC	0.6768 ± 0.075	1 ± 0	99.9
RX	0.7890 ± 0.037	0.9609 ± 0.024	100
RX-ϵ	0.9569 ± 0.017	0.9808 ± 0.012	100
RX-PAV	0.9905 ± 0.007	0.9780 ± 0.014	100

Table 5: Ratios and percentage of instances which satisfy EJR for euclidean dataset

	Utilitarian ratio	Representation ratio	EJR%
AV	1 ± 0	0.6628 ± 0.044	0
PAV	0.8459 ± 0.023	1 ± 0	79.2
sPAV	0.7348 ± 0.053	1 ± 0	0
CC	0.6558 ± 0.056	1 ± 0	0
RX	0.8125 ± 0.024	1 ± 0	100
RX-ϵ	0.8579 ± 0.023	1 ± 0	100
RX-PAV	0.8536 ± 0.022	1 ± 0	100

Table 6: Ratios and percentage of instances which satisfy EJR for party-list dataset

to satisfy. This property is satisfied for example in the Euclidean dataset where all projects have roughly the same, small, number of users that approve them.

In contrast to the above, in the party-list dataset (Table 6) most outcomes do not satisfy EJR (unless of course when the rule is part of the EJR family). AV, sPAV and CC did not satisfy EJR in any instance. PAV satisfies EJR for about 80% of the instances and provides a good utilitarian and representation ratios. However, the RX-variants, RX-ϵ and RX-PAV, achieve the same ratios, but, also satisfy the EJR property, making them more attractive than PAV.

The poor EJR percentages of AV, sPAV and CC for the party-list dataset can be explained by observing the following. First, this dataset forces large cohesive voter groups, making it more difficult for voting rules to satisfy EJR. Second, the PB instances contain projects that give high welfare or representation, but are also more expensive. The voting rules AV, sPAV, PAV and CC ignore cost, and thus by choosing such expensive projects, the budget is eaten fast, and small groups, with cheap projects, are not funded, violating the EJR property.

Finally, we consider the relationship between run-time and the guarantees. The rules PAV, RX and RX-ϵ run in polynomial time in the number of projects and voters, while the rest of the voting rules take exponential time. While RX-ϵ provides a good trade-off between all measurements, in practice the run-time of this voting rule is an order of magnitude more time consuming than *all* of the other voting rules, across all three datasets. Therefore, using the exponential-time PAV or RX-PAV might be preferred in relatively small instances, where in practice we observed fast termination.

If we were to prepare a recommendation list for which rule to use when, taking into consideration the performance of the rules with respect to both guarantees and the run-time, then the PAV rule offers a good compromise across the board but it is computationally feasible only on small instances. The RX-PAV rule achieves similar results to PAV in addition to satisfying EJR, and it exhibits shorter run-time, since the exponential part of RX-PAV (the PAV part) is applied to the remaining budget and unfunded projects (which is a much smaller instance). Hence RX-PAV is suitable both for small and medium instances. Lastly, the rule RX-ϵ provides lower utilitarian and representation guarantees compared to PAV and RX-PAV in addition to satisfying EJR, but it runs in polynomial time, making it the rule of choice for large instances.

6 CONCLUSIONS AND FUTURE WORK

We presented a theoretical and empirical investigation of the trade-off between welfare and representation for different voting rules for participatory budgeting. From the theoretical perspective, we analyzed the worst-case guarantees of common voting rules from the literature. We show that when introducing costs to projects, these guarantees do not generalize to the PB setting, with some rules (e.g., sPAV) exhibiting significantly lower guarantees than the multi-winner setting. From the empirical perspective, we show that some proportional voting rules (namely RX-PAV and RX-ϵ) are able to achieve high social welfare and representation on real PB instances, in contrast to their theoretical guarantees.

Taking into consideration the trade-off between welfare and representation, we concluded the PAV rule to be the clear winner from both theory and practice perspectives, however, it is not proportional, and exhibits worst case exponential running time. This led us to analyze two variants of RX (RX-PAV and RX-ϵ) that exhibit proportionality and provide similar results to PAV in practice, despite their lower theoretical guarantees (expressing the cost of proportionality). Specifically, we claimed that RX-PAV is suitable for solving medium sized PB instances, while RX-ϵ is suitable for large PB instances as a result of their run-time. Our results provide a deeper understanding of the trade-offs between welfare and representation for voting rules in Participatory Budgeting and can lead to more efficient outcomes that will satisfy the citizens.

There are several directions that are interesting to explore in future work. First, extending our analysis to consider additional voting rules from the literature, and considering more families of rules (e.g. voting rules with constrain on minority representation). Second, finding sufficient conditions on PB instances such that a voting rule would satisfy EJR. This could be a way to reconcile proportionality with other requirements, albeit for a restricted class of participatory budgeting problems.

7 ACKNOWLEDGEMENTS

This work was supported in part by Israeli Science Foundation (ISF) Grant No. 773/16.

REFERENCES

[1] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. 2017. Justified representation in approval-based committee voting. *Social Choice and Welfare* 48, 2 (2017), 461–485.

[2] Haris Aziz, Serge Gaspers, Joachim Gudmundsson, Simon Mackenzie, Nicholas Mattei, and Toby Walsh. 2014. Computational aspects of multi-winner approval voting. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

[3] Haris Aziz, Barton E Lee, and Nimrod Talmon. 2018. Proportionally Representative Participatory Budgeting: Axioms and Algorithms. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 23–31.

[4] John R Chamberlin and Paul N Courant. 1983. Representative deliberations and representative decisions: Proportional representation and the Borda rule. *American Political Science Review* 77, 3 (1983), 718–733.

[5] Edith Elkind, Piotr Faliszewski, Jean-François Laslier, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. 2017. What do multiwinner voting rules do? An experiment over the two-dimensional euclidean domain. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.

[6] Brandon Fain, Ashish Goel, and Kamesh Munagala. 2016. The core of the participatory budgeting problem. In *International Conference on Web Information Engineering (ICWIE)*. Springer, 384–399.

[7] Brandon Fain, Kamesh Munagala, and Nisarg Shah. 2018. Fair allocation of indivisible public goods. In *Proceedings of the 2018 ACM Conference on Economics and Computation (ACM-EC-2018)*. 575–592.

[8] Ashish Goel, Anilesh K Krishnaswamy, Sukolsak Sakshuwong, and Tanja Aitamurto. 2019. Knapsack voting for participatory budgeting. *ACM Transactions on Economics and Computation (TEAC)* 7, 2 (2019), 1–27.

[9] Gurobi Optimization, LLC. 2021. Gurobi Optimizer Reference Manual. https://www.gurobi.com

[10] Pallavi Jain, Krzysztof Sornat, and Nimrod Talmon. 2020. Participatory budgeting with project interactions. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*. 386–392.

[11] Martin Lackner, Peter Regner, Benjamin Krenn, and Stefan Schlomo Forster. 2021. abcvoting: A Python library of approval-based committee voting rules. https://doi.org/10.5281/zenodo.3904466 Current version: https://github.com/martinlackner/abcvoting.

[12] Martin Lackner and Piotr Skowron. 2020. Utilitarian welfare and representation guarantees of approval-based multiwinner rules. *Artificial Intelligence* 288 (2020), 103366.

[13] Marcin Michorzewski, Dominik Peters, and Piotr Skowron. 2020. Price of fairness in budget division and probabilistic social choice. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 2184–2191.

[14] Hans Christian Orsted, Georg Forchhammer, and Johannes Japetus Sm Steenstrup. 1894. *Oversigt over det Kongelige Danske videnskabernes selskabs forhandlinger...*

[15] Madeleine Pape and Josh Lerner. 2016. Budgeting for equity: How can participatory budgeting advance equity in the United States? *Journal of Public Deliberation* 12, 2 (2016).

[16] Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. 2020. Proportional Participatory Budgeting with Cardinal Utilities. *arXiv preprint arXiv:2008.13276* (2020).

[17] Dominik Peters and Piotr Skowron. 2020. Proportionality and the limits of welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation (ACM-EC-2021)*. 793–794.

[18] Luis Sánchez-Fernández, Edith Elkind, Martin Lackner, Norberto Fernández, Jesús Fisteus, Pablo Basanta Val, and Piotr Skowron. 2017. Proportional justified representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.

[19] Yves Sintomer, Carsten Herzberg, and Anja Röcke. 2008. From Porto Alegre to Europe: potentials and limitations of participatory budgeting. *International Journal of Urban and Regional Research* 32, 1 (2008), 164–178.

[20] Piotr Skowron. 2021. Proportionality degree of multiwinner rules. In *Proceedings of the 22nd ACM Conference on Economics and Computation*. 820–840.

[21] Piotr Skowron, Arkadii Slinko, Stanisław Szufa, and Nimrod Talmon. 2020. Participatory Budgeting with Cumulative Votes. *arXiv preprint arXiv:2009.02690* (2020).

[22] Dariusz Stolicki, Stanisław Szufa, and Nimrod Talmon. 2020. Pabulib: A Participatory Budgeting Library. *arXiv preprint arXiv:2012.06539* (2020).

[23] Celina Su. 2017. From Porto Alegre to New York City: Participatory Budgeting and Democracy. *New Political Science* 39, 1 (2017), 67–75.

[24] Nimrod Talmon and Piotr Faliszewski. 2019. A framework for approval-based budgeting methods. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 2181–2188.