Participatory Budgeting: Data, Tools, and Analysis

Piotr Faliszewski[1] , **Jarosław Flis**[2] , **Dominik Peters**[3] , **Grzegorz Pierczyński**[4] ,
Piotr Skowron[4] , **Dariusz Stolicki**[2,5] , **Stanisław Szufa**[1] and **Nimrod Talmon**[6]

[1]AGH University
[2]Jagiellonian University in Krakow
[3]CNRS, LAMSADE Université Paris Dauphine - PSL
[4]University of Warsaw
[5]Center for Quantitative Political Science
[6]Ben-Gurion University

Abstract

We provide a library of participatory budgeting data (Pabulib) and open source tools (Pabutools and Pabustats) for analysing this data. We analyse how the results of participatory budgeting elections would change if a different selection rule was applied. We provide evidence that the outcomes of the Method of Equal Shares would be considerably fairer than those of the Utilitarian Greedy rule that is currently in use. We also show that the division of the projects into districts and/or categories can in many cases be avoided when using proportional rules. We find that this would increase the overall utility of the voters.

1 Introduction

Participatory budgeting (PB) [Cabannes, 2004] is a form of public consultation in which residents decide how to spend a part of the municipal budget. First, a number of projects are submitted, and after the initial formal evaluation, some of these projects are admitted to voting. Next, each citizen can participate in an election, and express their preferences over the projects. Finally, given the voters' ballots, a decision is made as to which of the projects to fund.

Each step in this process must be carefully designed in order to ensure that the selected projects match the voters' preferences to the highest possible degree. For example, some cities regulate the submission process by putting upper bounds on the costs of the projects (e.g., to avoid situations where a single expensive project consumes the whole budget, leaving a large fraction of the voters unsatisfied) or by assigning each project to one of a few predefined categories (e.g., to ensure that projects from the less popular categories also receive some funding). Choosing the right format of the ballots for the election also is a significant and challenging issue [Benade *et al.*, 2021]. For example, some cities use approval ballots, where the voters simply indicate which projects they support (sometimes with additional constraints, such as regarding the number of projects a voter can approve), while others turn to (forms of) score ballots, which allow the voters to indicate the degree of support for the respective projects. Further, some

cities allow the voters to vote only on the projects from the district where they live, while others do not put such restrictions [Hershkowitz *et al.*, 2021].

Finally, the voting rule used for aggregating the ballots and selecting the projects is of utmost importance for the whole process. For example, if the rule is capable of representing the voters proportionally, then it may not be necessary to partition the projects into categories or to put constraints on their costs, whereas a majority-driven rule may indeed require such interventions. Consequently, there is a growing interest in the design and analysis of voting rules for participatory budgeting [Peters *et al.*, 2021; Goel *et al.*, 2019; Aziz and Shah, 2020; Talmon and Faliszewski, 2019; Aziz *et al.*, 2018; Fain *et al.*, 2018; Jiang *et al.*, 2020; Munagala *et al.*, 2022; Skowron *et al.*, 2020]. However, with a notable exception of [Benade *et al.*, 2023] (who performed lab experiments assessing how humans perceive different ballot formats), all this research is focused on theoretical analysis.

In this paper we take a step towards understanding how various selection rules for participatory budgeting operate in practice. We do this by releasing and analysing data from over 650 PB elections, mainly conducted in Poland.[1] Our contribution is the following:

PABULIB. This is a library of participatory budgeting data (PArticipatory BUdgeting LIBrary), and can be accessed via the following URL: http://pabulib.org.

PABUTOOLS. This is a Python library providing a parser of Pabulib files and implementations of selected rules for participatory budgeting. The library is accessible via PyPI (https://pypi.org/project/pabutools/).

PABUSTATS. This is a web application for comparing various PB rules based on the data from Pabulib. It also offers the possibility to run simulations on files uploaded by the

---

[1]Poland is a good source of PB instances because the law requires every major city to spend at least 0.5% of its annual budget through PB. In 2021, over 42% of Polish cities with populations above 5 000 organised PB elections, spending 627.5 million PLN in total.

users. The application is accessible via the following URL: http://pabulib.org/pabustats.

We apply our tools on the collected data and perform an extensive analysis that compares different voting rules for PB.

We defer the detailed description of the Pabulib data format as well as additional figures with the results of our analysis to the full version of the paper [Faliszewski *et al.*, 2023].

2 Preliminaries

An *election* is a tuple $E = (P, N, b, \text{cost})$, where $P = \{p_1, \ldots, p_m\}$ is a set of *projects*, $N = \{1, 2, \ldots n\}$ is a set of *voters*, $b \in \mathbb{N}$ is the *budget*, and $\text{cost} \colon P \to \mathbb{N}$ is a function that associates each project with its *cost*. The cost function naturally extends to sets: for each $W \subseteq P$, we let $\text{cost}(W) = \sum_{p \in W} \text{cost}(p)$. The voters express their preferences by casting *ballots*: each voter $i \in N$ assigns to each project $p \in P$ a score $s_i(p) \in \mathbb{N}$ that reflects her level of support for p. If $s_i(p) \in \{0, 1\}$ for all $i \in N$ and $p \in P$, then we have an *approval election*. Otherwise, we say it is a *cardinal election*. Intuitively, in an approval election each voter simply indicates the projects that she supports; in cardinal elections the voters provide more fine-grained information on how much they supports particular projects.

2.1 Utility Models

Given the ballot of voter $i \in N$, there are two natural ways to define i's utility from a given set of projects $W \subseteq P$. The first approach is to assume that the utility does not depend on the costs of the projects; in this case we speak of *score utilities*, defined as $u_i^{\text{sc}}(W) = \sum_{p \in W} s_i(p)$. For approval elections, this is the number of projects in W that the voter supports.

An alternative approach is to assume that expensive projects carry more value to the voters; then, we speak of *cost utilities*, $u_i^{\text{cost}}(W) = \sum_{p \in W} s_i(p)\text{cost}(p)$. For approval elections, cost utilities can be interpreted as the amount of funds spent on the projects supported by a given voter.

Sometimes it will be clear from the context whether we refer to score or to cost utilities. In such cases, we will often omit the superscripts and write $u_i(W)$.

2.2 Voting Rules

A *voting rule* is a function that takes an election as input and returns a subset of projects W, called an *outcome*, such that $\text{cost}(W) \leqslant b$. We say that outcome W is *exhaustive* if for each project $c \in C \setminus W$ we have that $\text{cost}(W) + \text{cost}(c) > b$ (i.e., no additional project can be funded without violating the budget constraint). A voting rule f is exhaustive if it always returns an exhaustive outcome.

Among proposed voting rules [Rey and Maly, 2023], we focus on two: Utilitarian Greedy and Method of Equal Shares. Each has two variants, depending on the type of utilities.

Utilitarian Greedy (UG).   We start with an empty outcome $W = \emptyset$, and repeatedly select a project p maximising the

ratio $\sum_{i \in N} u_i(p)/\text{cost}(p)$. If $\text{cost}(W) + \text{cost}(p) \leqslant b$ then we add project p to W; otherwise, we remove the project from consideration and repeat, until no more projects remain.

This rule aims at maximising the total utility of the voters, $\sum_{i \in N} u_i(W)$. Indeed, the Utilitarian Greedy rule is *optimal up to one project* for this objective [Dantzig, 1957], i.e., for each outcome W returned by UG there exists $p \notin W$ s.t.:

$$\sum_{i \in N} u_i(W \cup \{p\}) \geqslant \max_{W' \colon \text{cost}(W') \leqslant b} \sum_{i \in N} u_i(W').$$

Note that if we use this rule with cost utilities, then the projects will be selected in descending order of their total scores (where the total score of project p is $\sum_{i \in N} s_i(p)$). This voting rule is currently used by the vast majority of cities that do PB. We could view this choice of voting rule as a "revealed preference" of the city for interpreting ballots based on cost utility. If we use Utilitarian Greedy with score utilities, the rule selects projects in order of descending "value-for-money", which for project p is equal to $\sum_{i \in N} s_i(p)/\text{cost}(p)$.

Method of Equal Shares.   This is a recent method, introduced by [Peters and Skowron, 2020] and [Peters *et al.*, 2021]. In the first step, we divide the budget equally among the voters: for each voter $i \in N$, we set $b_i \leftarrow b/n$. We say that a project p is *α-affordable* for $\alpha \in \mathbb{R}$ if the following equality holds:

$$\sum_{i \in N} \pi_i(p) = \text{cost}(p), \text{ where } \pi_i(p) = \min\left(b_i, \alpha u_i(p)\right).$$

Here, $\pi_i(p)$ is the amount that voter i needs to pay if project p is selected. The condition says that the cost of project p can be covered while (1) no voter exceeds their budget share of b_i, and (2) each voter pays at most α per unit of utility. If such an α does not exist (which happens if and only if the total budget shares b_i of the voters who assign a positive score to p is lower than $\text{cost}(p)$), then the project is *not affordable*.

The method starts with $W = \emptyset$. It then repeatedly computes α-affordability for all not-selected projects, chooses a project p that is α-affordable for minimal α, adds p to the outcome, and updates the voters' individual budgets: for each $i \in N$, $b_i \leftarrow b_i - \pi_i(p)$. It stops when no project is affordable.

As in the case of Utilitarian Greedy, the rule can work both with cost utilities (in which case in the first iteration it chooses an affordable project with the highest total score) and with score utilities (in which case in the first iteration it chooses an affordable project with the highest value-for-money). In the following iterations the rule continues to focus on the total scores and the value-for-money, respectively, but it also takes into account how much the voters have already spent.

The Method of Equal Shares satisfies strong proportionality properties [Peters and Skowron, 2020; Peters *et al.*, 2021], but is not exhaustive (typically spending 60-80% of the budget on real instances). We consider three ways to make it exhaustive (see also [Rey and Maly, 2023]):

1. Completion by the Utilitarian Greedy algorithm (U): We first select an outcome W_{ES} using Equal Shares; next, we select additional candidates from $C \setminus W_{\text{ES}}$ using the Utilitarian Greedy rule with the budget set to $b - \text{cost}(W_{\text{ES}})$. We return W_{ES} together with the additional candidates.

2. Completion by tweaking voters' utilities (Eps): It is known that the outcome of Equal Shares is exhaustive if every voter's utility for every project is strictly positive [Peters *et al.*, 2021]. The most straightforward way to use this idea is to override each $s_i(p) = 0$ with $s_i(p) = \varepsilon$. We use an implementation which in fact does not depend on the ε.[2]

3. Completion by increasing the initial endowments (Add1): Observe that Equal Shares can be run with the initial voter endowments b_i set to a different value than b/n. We start with the endowments set to b/n. If the outcome is not exhaustive, we increase the initial endowment b_i by one unit, and run Equal Shares from scratch. We repeat this procedure until the outcome is exhaustive or until the moment when the next increase of endowments would result in exceeding the original budget b.

Add1 does not necessarily return an exhaustive outcome, but the amount of unspent funds is typically small (as we will see). To get an exhaustive rule, we can combine Add1 with the Utilitarian Greedy completion; we call it Add1U.

3 Participatory Budgeting Library (PABULIB)

PABULIB is an open library of participatory budgeting instances that we collected in this project (but we invite and encourage interested researchers to submit their data). In the full version of this paper [Faliszewski *et al.*, 2023] we define the *.pb* format, which we recommend for representing PB instances, and which is used in our library.

The aim of PABULIB is to gather participatory budgeting data from as many cities and as many countries as possible, but currently most of the instances come from several large cities in Poland (in particular, from Warsaw, with a population of 1.7 million people; from Krakow, Wroclaw, and Gdansk, with populations between 500 000 and 1 million; and from Czestochowa, Zabrze, and Katowice, with populations between 150 000 and 300 000). As these cities use different voting formats, the library accepts several different types of ballots. For example, in Warsaw every voter is asked to approve up to 15 local and up to 10 citywide projects; in Krakow each voter must assign three different scores (3 points, 2 points, and 1 point) to three different projects; and in Czestochowa each voter distributes a total score of 10 points between the projects. In Wroclaw and Zabrze, a voter can only vote for a single local and a single citywide project.

In Figure 1, we illustrate the distribution of project costs and the total numbers of approvals per project for two sample cities, Warsaw and Wroclaw. (The distributions for other cities and the correlation between the projects costs and their support are provided in the full version of the paper [Faliszewski *et al.*, 2023]). We observe a sharp difference between the cities, depending on the type of ballots they use. The data suggest that

(a) Warsaw 2020-2023   (b) Wroclaw 2019-2021

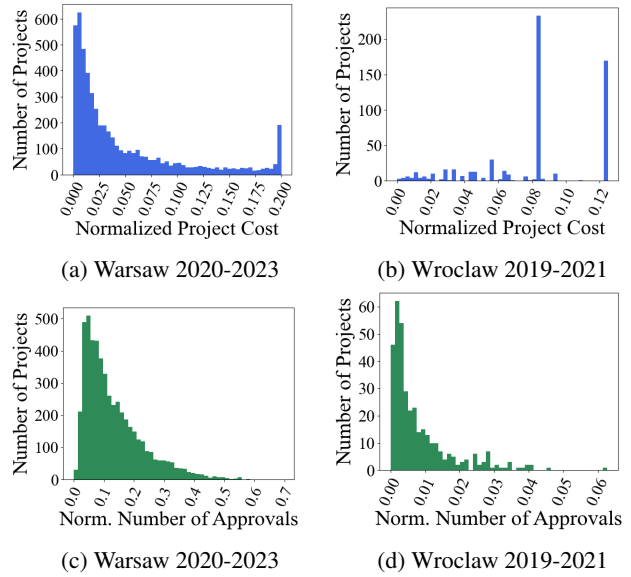(c) Warsaw 2020-2023   (d) Wroclaw 2019-2021

Figure 1: The distribution of costs and the total number of approvals per project for two sample cities, aggregated over different years. For each PB instance, we normalize the costs by dividing them by the total budget, and the number of approvals by dividing it by the largest number of approvals obtained by any project.

forcing the voters to vote only for a single project typically discourages submitting smaller and cheaper projects. We consider this effect undesirable and so this is an argument against such voting formats.

4 Basic Metrics of Fairness and Efficiency

In this section we consider several basic metrics for comparing outcomes returned by different voting rules.

The *average utility*, which for outcome W is defined as $1/n \sum_{i \in N} u_i(W)$, is perhaps the most natural metric of efficiency. We are also interested in the degree to which a given voting rule respects the diversity of voters' opinions, which is less straightforward to measure. One possible way would be to apply known measures of statistical dispersion (such as the Gini index) to the vector of voters' utilities: a smaller value of the dispersion would suggest that the utility is more evenly distributed, hence the opinions of a greater part of the population are taken into account. However, this approach is problematic in the context of cost utilities. Indeed, some residents choose to vote only for local, relatively cheap projects, while others support mainly large and expensive initiatives. Thus, it is expected that the cost utilities of different voters can vary substantially. When performing the comparative analysis of two rules, R_1 and R_2, we find it more informative to compute the *dominance margin* of R_1 over R_2, i.e., the fraction of voters who enjoy a strictly higher utility from the outcome of R_1 than from the one of R_2. A related metric is the *improvement margin* of R_1 over R_2 which is the dominance margin of R_1 over R_2 minus the dominance margin of R_2 over R_1. We also measure the *exclusion ratio*—the fraction of voters who support none of the selected projects.

Next, we consider a metric recently introduced by [Lackner *et al.*, 2021] which, informally speaking, measures the amount of spending that different voters had an influence on [Maly *et al.*, 2023]. The measure assumes that the supporters of a selected project contributed to the decision on spending money on this project proportionally to the score that they assigned to it. Consequently, given an outcome W, voter i's share is:

$$\text{share}_i(W) = \sum_{p \in W} \frac{s_i(p)}{\sum_{j \in N} s_j(p)} \cdot \text{cost}(p).$$

Note that the shares of the voters sum up to the total cost of the selected projects. In an ideally fair solution we would like all these shares to be equal. This suggests the following metric, which we call *power inequality*:

$$^{1}/_{n} \cdot \sum_{i \in N} |\text{share}_i(W) - {}^{b}/_{n}| \cdot {}^{n}/_{b}.$$

We used the mentioned metrics to compare the outcomes of various election rules. We took data from PB elections carried out in seven major Polish cities: Czestochowa (2020), Gdansk (2020), Katowice (2020–2021), Krakow (2018–2022), Warsaw (2017–2023), Wroclaw (2015–2021), and Zabrze (2020–2021). In the plots we show only cities for which we have data for at least 3 years (Warsaw, Krakow, and Wroclaw). The analysis of the remaining cities led to similar conclusions.

To deal with city districts, for each city and each year we tested the rules using two different schemes:

1. The *citywide* (C) scheme, where we put all the projects from different districts and categories in the same pool, and we kept the original voters' ballots. Thus, for a fixed city and year we have a single election.

2. The *districtwise* (D) scheme which corresponds to how the cities currently organize their elections: a separate election is run in each city district; typically there is also one additional election involving the same set of voters but concerning citywide projects (these are projects that are potentially interesting to voters from multiple districts). The outcome for a given city and year is obtained by adding together the outcomes of these smaller elections. The metrics are computed for this combined outcome, with respect to all the voters in the city, just like in the citywide scheme.

When it does not lead to confusion, we will often speak of citywide and districtwise "elections" rather than schemes.

We present results for each city averaged over all years, with figures showing error bars corresponding to standard deviations over the years. While we consider averages, the conclusions of our analysis also hold for each year separately. The results for separate editions can be checked through our web application: http://pabulib.org/pabustats.

In our first set of simulations, we compare the three different approaches to making the Method of Equal Shares exhaustive. We observe that the completion strategy plays a critical role. For example, in citywide and districtwise elections, Equal Shares without completion uses, on average, only 32% and 50% of the available funds, respectively. The Add1 completion uses on average 98% and 88%, and Add1U (an exhaustive variant) uses 99.9% and 95% of the funds, respectively.

In Figure 2, we compare our metrics for the three completion strategies, the two types of utilities, and citywide and
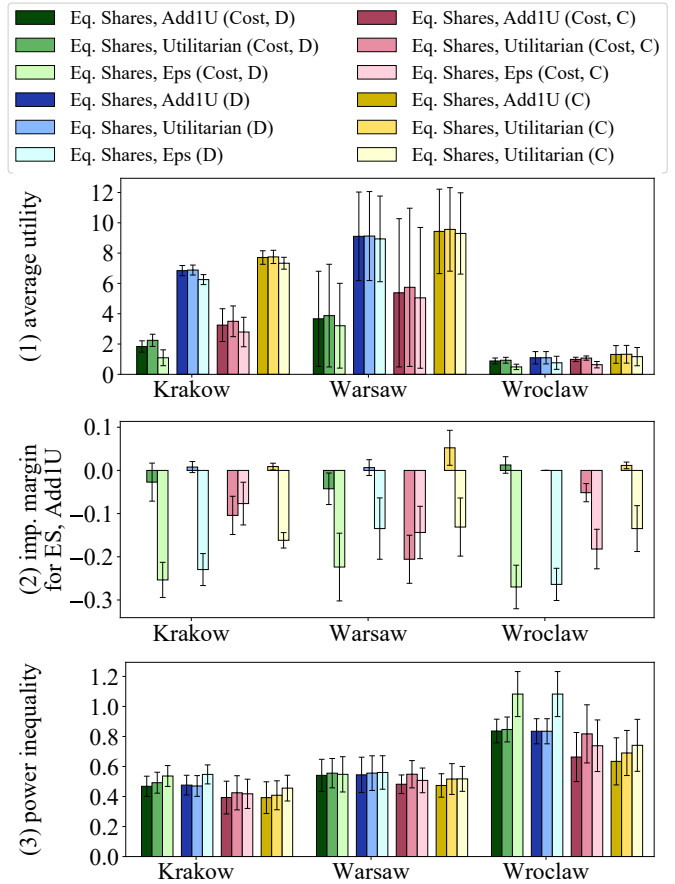

Figure 2: Comparison of different completion rules for the Method of Equal Shares. The label "Cost" means that we are referring to the cost-utility variant of the method; otherwise we are referring to its score-utility variant. The symbols "D" and "C" stand for the districtwise and citywide schemes, respectively. We compare (1) voters' average utility (for the score-utility variants of the methods we give the average score utility; for the cost-utility variants we give the average cost utility in millions), (2) the improvement margin over Equal Shares with the Add1U completion (the improvement margin is with respect to cost utilities for the cost-utility variants of the methods and with respect to score utilities for the remaining variants), and (3) the power inequality.

districtwise elections. This figure is quite involved, so let us provide some guidance on how to read it. First, we have a separate plot for each of the metrics, i.e., for (1) average utility, (2) the improvement margin over Equal Shares with the Add1U method, and (3) power inequality. Within each of the plots, different shades of each color correspond to different completion types (darkest for Add1U, middle for U, and lightest for Eps). The green and blue shades correspond to the districtwise elections, and red and yellow shades to citywide elections. Cost utilities correspond to green and red, whereas score utilities correspond to blue and yellow. We conclude the following (the exclusion ratios are given in the full version):

1. The completion by tweaking voters' utilities (Eps, lightest shade for each color) gives the worst results in terms of the average utility, power inequality, exclusion ratio, and
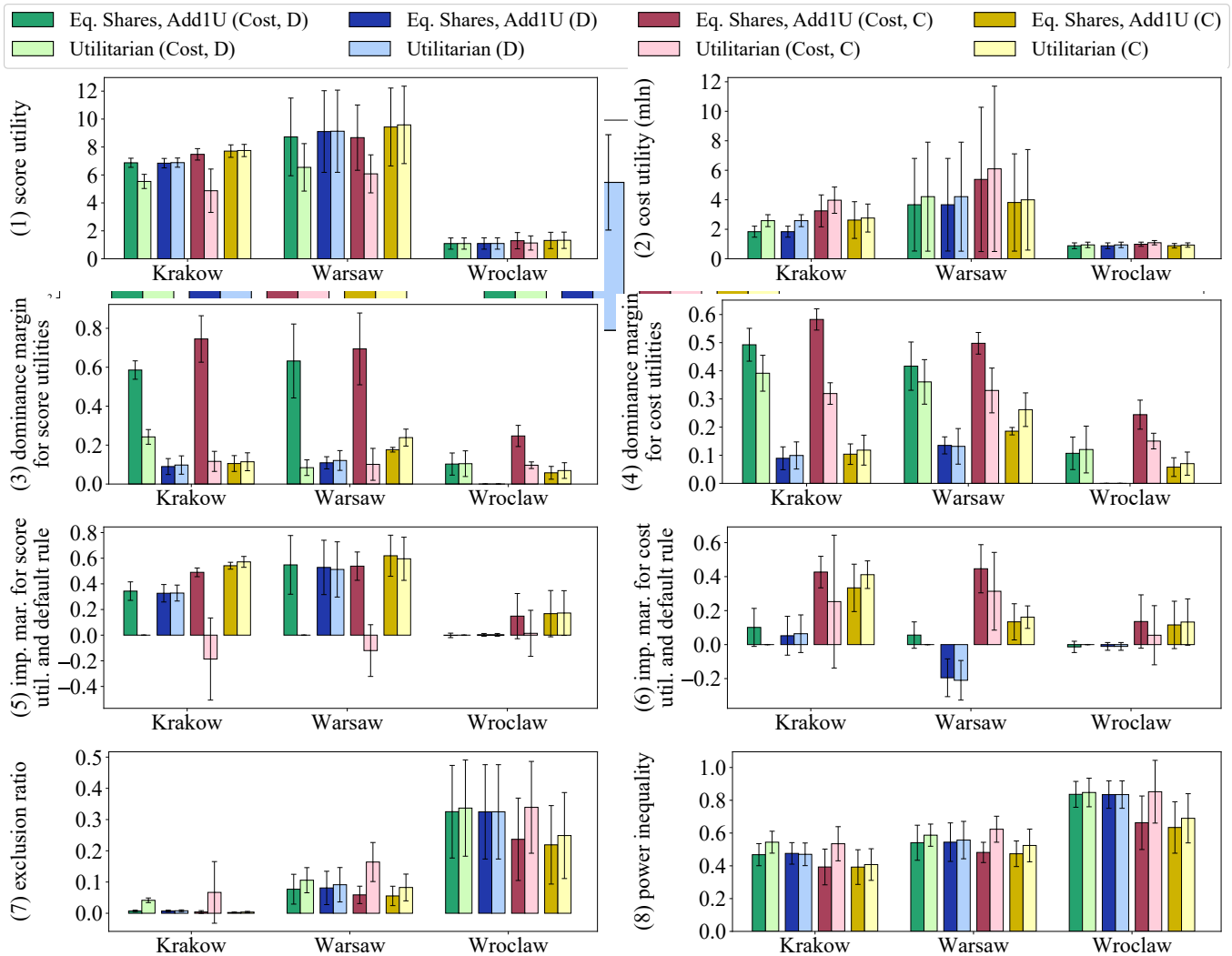
Figure 3: Comparison of Equal Shares and Utilitarian Greedy (in different variants), with respect to (1) average score utility, (2) average cost utility, (3) dominance margin with respect to score utility, (4) dominance margin with respect to cost utility, (5) improvement margin over the currently used rule (Utilitarian, Cost, D) with respect to score utilities, (6) improvement margin over the currently used rule with respect to cost utilities (7) exclusion ratio, and (8) power inequality. The label "Cost" means that we are referring to the cost-utility variant of the method; otherwise we are referring to its score-utility variant. The symbols "D" and "C" stand for districtwise and citywide schemes, respectively.

improvement over Add1U.

2. The utilitarian completion (U, middle shades) gives a bit higher average utility than Add1U (darkest shades), but also a worse power inequality and exclusion ratio.

3. For score utilities (blue and yellow), the utilitarian completion (U, middle shade) and Add1U (darkest shade) are comparably good. For cost utilities (green and red), the Add1U strategy (darkest shade) has a large advantage, especially for improvement margins (the other rules have negative values in almost all settings).

Based on these observations, we suggest using the Add1U completion as the default option.

In our second set of simulations we compare the Add1U variant of Equal Shares with the Utilitarian Greedy rule. The results are depicted in Figure 3. In this figure, each scenario

corresponds to a color. The darker shades represent Equal Shares and the lighter ones represent Utilitarian Greedy. Our findings can be summarised as follows:

1. For rules based on score utilities (blue and yellow), the results of the Method of Equal Shares and of Utilitarian Greedy are comparable. Equal Shares selects outcomes with slightly lower exclusion ratio as well as lower power inequality, at the cost of lower average utility. However, these differences are relatively small.

2. For rules based on cost utilities (green and red), we see a significant difference between the two rules. Unsurprisingly, the outcomes of Utilitarian Greedy have better average cost utility, but the difference is relatively small (e.g., it is the largest in Warsaw for citywide elections, 13%). For all other metrics, Equal Shares outperforms Utilitarian

City	Add1U, C	Util. G, D	Util. G, C
Czestochowa	0.23	0.28	0.39
Gdansk	0.27	0.33	0.46
Katowice	0.19	0.26	0.51
Krakow	0.08	0.24	0.23
Warsaw	0.20	0.41	0.41
Wroclaw	0.15	0.26	0.22
Zabrze	0.38	1.24	0.41

Table 1: Average dispersion of the budget allocation. We compare cost-utility variants of Equal Shares (with the Add1U completion) and Utilitarian Greedy (on the districtwise and the citywide schemes).

Greedy by a large margin. For example, for citywide elections in Warsaw, the average score utility of the outcomes of Equal Shares is 43% higher, and using Equal Shares would result in a drop of the exclusion ratio from 16% to 6%. The improvement margin over UG in terms of score and cost utility is on average, respectively, 59% and 17%.

3. We observe a significant difference between the citywide (red and yellow) and districtwise (blue and green) schemes. Global elections (that do not divide projects *a priori* into districts) result in a much higher average utility and much lower exclusion ratio, for Equal Shares.[3]

Finally, it is interesting to see how fairly the budgets were distributed among districts when using the citywide scheme. Let $\mathcal{D} = \{D_1, D_2, \ldots, D_t\}$ be the set of districts, which is formally a partition of N. Ideally, the voters from a district $D \in \mathcal{D}$ should get a share of the budget that is proportional to the size of D.[4] The *dispersion of the budget allocation* is:

$$\frac{1}{|\mathcal{D}|} \cdot \sum_{D \in \mathcal{D}} \frac{\left| \sum_{i \in D} \text{share}_i(W) - |D|/n \cdot b \right|}{|D|/n \cdot b}.$$

This metric captures the average relative difference between how much money the district got and how much we would expect it to get. Table 1 shows average dispersion values, which are lower for Equal Shares than for Utilitarian Greedy.

---

[3]The large difference between districtwise and citywide elections arises because in some districts no popular projects are submitted, and their residents would prefer to fund citywide projects instead.

As an example, in 2021, Warsaw had 2 projects about the same street: (A) new plants for Modlinska street (citywide project), and (B) new pavement for Modlinska street (district project in Bialoleka). Project A received six times as many votes as B (12 463 vs 1 932). Even among the voters of Bialoleka, A was twice as popular (4 365 vs 1 932). Project A was cheaper (435k vs 630k PLN). Even though A was more popular and cheaper than B, it was not selected (being a citywide project) while B was selected.

[4]This assumes that the budget should be divided proportionally to the number of voters and not to the number of residents of a district. If turnout varies between districts, the difference matters. Being proportional to the number of voters promotes participation, incentivizes districts to encourage their residents to vote, and follows the "one person, one vote" principle. If the city prefers being proportional to the number of residents, the citywide scheme can be adapted by giving voters from districts with lower turnout a larger initial endowment.

City	Add1U, C	Util. G, D	Util. G, C
Czestochowa	0.80	0.35	0.39
Gdansk	0.87	0.26	0.39
Katowice	0.83	0.56	0.42
Krakow	0.78	0.52	0.41

Table 2: Robustness ratio for different voting rules. We compare cost-utility variants of Equal Shares (with the Add1U completion) and Utilitarian Greedy (on the districtwise and the citywide schemes).
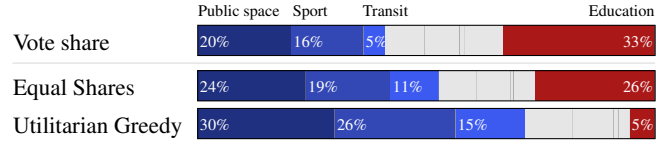
Figure 4: The vote share and the spending share of different tags in district Bielany, Warsaw 2020. The picture is similar for 2021–23.

5 Robustness to Changing the Type of Ballots

Recently, [Benade *et al.*, 2023] performed lab experiments about different input formats for PB elections. One of their findings was that the Method of Equal Shares is robust to changing the type of ballots. In this section we reinforce their conclusions by analyzing data from real PB elections.

For each election where voters used cardinal ballots, we construct a corresponding approval election by letting a voter approve all the projects to which she assigned a positive score. Then, we compare the outcomes of different rules for these two elections. Let W_{sc} and W_{appr} be the outcomes of a given voting rule for the original and the approval elections, respectively. We define the *robustness ratio* as $\text{cost}(W_{\text{appr}} \cap W_{\text{sc}})/\text{cost}(W_{\text{sc}})$. Table 2 summarises the results of our analysis. We can see that the outcomes of Equal Shares change much less after switching to approval compared to Utilitarian Greedy. For users of Equal Shares, this provides an argument in favor of approval ballots, which is the voting format that [Benade *et al.*, 2023] find to be the easiest to use.

6 Budget Distribution among Categories

Cities often organize projects by topics (such as public space, environment, education) to make browsing the list of projects easier. Warsaw, for example, categorizes projects using 10 different tags (where projects can get multiple tags). This allows us to ask whether voter preferences across categories are well-reflected by the spending of the rules.

We focus on Warsaw district elections (2020–23), which use approval voting. Denote by A_i the set of projects approved by $i \in N$. For each project p, denote by $\text{tags}(p) \subseteq \{\text{public space}, \ldots, \text{education}\}$ the tags assigned to p. For each tag t, we can then compute its *vote share*:

$$\frac{1}{n} \sum_{i \in N} \sum_{p \in A_i : t \in \text{tags}(p)} \frac{1}{|A_i| \cdot |\text{tags}(p)|}$$

This intuitively counts the fraction of the votes that went to projects with tag t, in a way that each voter contributes the

Mokotow, 2020   Ochota, 2021
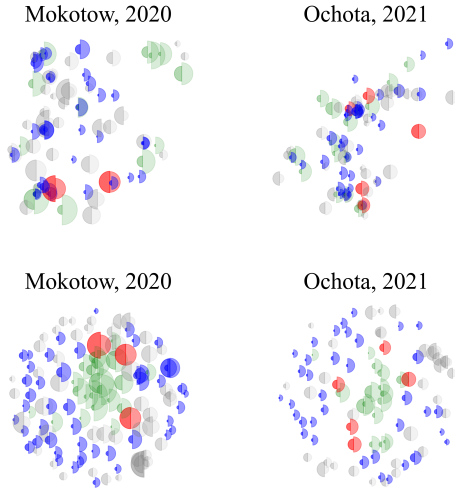
Mokotow, 2020   Ochota, 2021

Figure 5: Visualisation of projects in PB elections using GPS data (upper row) and using the Jaccard distance (lower row). Each project is represented by two glued-together half-discs. The size of the left half is proportional to the project's cost, whereas the size of the right half is proportional to the total number of votes the project received. The figures compare the outcomes of the cost-utility variant of Equal Shares with Add1U completion, with the outcomes of the Utilitarian Greedy rule. Specifically, gray projects were not selected by either of the rules, green projects were selected by both, blue projects were selected only by Equal Shares, and red projects were selected only by Utilitarian Greedy.

same amount to the vote share, and for projects with multiple tags, splitting their contribution equally between them. Note that the vote shares of all the tags sum to 1. For an outcome W, we can similarly define the spending share of the tag: $\frac{1}{\text{cost}(W)} \sum_{p \in W : t \in \text{tags}(p)} \text{cost}(p)/|\text{tags}(p)|$.

We can now compute the ℓ_2 distance between the vector of vote shares and the spending shares of all the tags, to see how well they align. While it is not necessarily desirable for the two vectors to perfectly coincide, a large distance indicates that an outcome neglects certain categories.

We find that for 93% of districts, Utilitarian Greedy gives outcomes with a larger distance to the vote shares than the Equal Shares outcome (cost utilities, Add1U); see the full version of the paper [Faliszewski *et al.*, 2023]. In some cases, the distance is much larger, like in the district Bielany, where in each year, Utilitarian Greedy spends much less on education projects than suggested by the vote shares. For example, in 2020, when education had a vote share of 33%, Utilitarian Greedy spent only 5% of the budget on these projects (Equal Shares spent 26%), see Figure 4.

7 Maps of Participatory Budgeting Elections

We provide easy-to-interpret visualisations of the outcomes of different voting rules. For elections that were carried out in Warsaw between 2020 and 2023, most of the projects (but not all) were associated with their GPS locations. Thus, we can depict those submitted projects that have GPS data in such a way that their relative locations correspond to their

physical locations in the respective districts. We present such visualisations in Figure 5 (additional figures are given in the full version of the paper [Faliszewski *et al.*, 2023]).

A different approach is to create a map that illustrates voters' preferences rather than geographic locations of the projects. Here, for a given approval PB election we first compute the Jaccard distances between all pairs of projects. Recall that for two projects, p_1 and p_2, their Jaccard distance is $|N(p_1) \triangle N(p_2)|/|N(p_1) \cup N(p_2)|$, where $N(p)$ denotes the set of voters who support project p (in other words, we assume that two projects are similar if similar groups of voters voted for them). Next, based on these distances, we create a two-dimensional embedding, using the Multidimensional Scaling Algorithm MDS[5] [Kruskal, 1964; de Leeuw, 2005]. Eventually, we obtain a plot where the closer two projects are, the larger is the fraction of their common supporters (however we do mention that MDS is only a heuristic and, more importantly, a perfect embedding may not exist, so these plots provide intuitions, but should by interpreted with care).

These two types of maps for four sample elections are depicted in Figure 5. We observe that Equal Shares selects more diverse and more representative sets of projects both in terms of their geographic locations and in terms of their supporters. Further, Utilitarian Greedy mainly selects large and expensive projects, whereas Equal Shares selects a mixture of projects, including some large and some small ones.

8 Conclusions

Our results lead to several conclusions that should be taken into account by election designers. First, the types of ballots to be used affect the process of preparing project proposals. For example, voting for single projects discourages submitting small and medium proposals. Second, we see that the cost-utility variants of the rules select fewer but larger projects compared to their score-utility counterparts. Third, the proportional methods such as the cost-utility variants of Equal Shares result in much fairer solutions, where the different voters' opinions are better represented; the difference is substantial. Fourth, running citywide elections, without dividing the budget and the projects *a priori* between districts and/or categories results in a significant improvement in terms of the average utility as well as the number of voters whose opinions are taken into account. Fifth, Equal Shares is more robust to changing the ballot types than the utilitarian rule.

Acknowledgments

---

[5]All the distances d lie between 0 and 1, but almost all are above 0.5, so we rescale them by setting $d' = \max(0, d - 0.5)$.

References

[Aziz and Shah, 2020] H. Aziz and N. Shah. Participatory budgeting: Models and approaches. In Tamás Rudas and Gábor Péli, editors, *Pathways Between Social Science and Computational Social Science: Theories, Methods, and Interpretations*. Springer, 2020.

[Aziz *et al.*, 2018] H. Aziz, B. Lee, and N. Talmon. Proportionally representative participatory budgeting: Axioms and algorithms. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2017)*, pages 23–31, 2018.

[Benade *et al.*, 2021] G. Benade, S. Nath, A. D. Procaccia, and N. Shah. Preference elicitation for participatory budgeting. *Management Science*, 67(5):2813–2827, 2021.

[Benade *et al.*, 2023] G. Benade, R. Fairstein, and Y. Gal. Participatory budgeting design for the real world. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI-2023)*, 2023. To appear.

[Cabannes, 2004] Y. Cabannes. Participatory budgeting: a significant contribution to participatory democracy. *Environment and Urbanization*, 16(1):27–46, 2004.

[Dantzig, 1957] G. B. Dantzig. Discrete-variable extremum problems. *Operations Research*, 5(2):266–288, 1957.

[de Leeuw, 2005] J. de Leeuw. Modern multidimensional scaling: Theory and applications. *Journal of Statistical Software*, 14:1–2, 2005.

[Fain *et al.*, 2018] B. Fain, K. Munagala, and N. Shah. Fair allocation of indivisible public goods. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 575–592, 2018. Extended version arXiv:1805.03164.

[Faliszewski *et al.*, 2023] P. Faliszewski, J. Flis, D. Peters, G. Pierczyński, P. Skowron, D. Stolicki, S. Szufa, and N. Talmon. Participatory budgeting: Data, tools, and analysis. *arXiv:2305.11035*, 2023.

[Goel *et al.*, 2019] A. Goel, A. K. Krishnaswamy, S. Sakshuwong, and T. Aitamurto. Knapsack voting for participatory budgeting. *ACM Transactions on Economics and Computation (TEAC)*, 7(2):1–27, 2019.

[Hershkowitz *et al.*, 2021] D. Hershkowitz, A. Kahng, D. Peters, and A. D. Procaccia. District-fair participatory budgeting. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI-2021)*, pages 5464–5471, 2021.

[Jiang *et al.*, 2020] Z. Jiang, K. Munagala, and K. Wang. Approximately stable committee selection. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC-2020)*, pages 463–472, 2020.

[Kruskal, 1964] J. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[Lackner *et al.*, 2021] M. Lackner, J. Maly, and S. Rey. Fairness in long-term participatory budgeting. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI-2021)*, pages 299–305, 2021.

[Maly *et al.*, 2023] J. Maly, S. Rey, U. Endriss, and M. Lackner. Fairness in participatory budgeting via equality of resources. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2023)*, 2023.

[Munagala *et al.*, 2022] K. Munagala, Y. Shen, K. Wang, and Z. Wang. Approximate core for committee selection via multilinear extension and market clearing. In *Proceedings of the 33rd ACM-SIAM Symposium on Discrete Algorithms (SODA-2022)*, pages 2229–2252, 2022.

[Peters and Skowron, 2020] D. Peters and P. Skowron. Proportionality and the limits of welfarism. In *Proceedings of the 2020 ACM Conference on Economics and Computation (ACM EC)*, pages 793–794, 2020. Extended version arXiv:1911.11747.

[Peters *et al.*, 2021] D. Peters, G. Pierczyński, and P. Skowron. Proportional participatory budgeting with additive utilities. *Advances in Neural Information Processing Systems*, 34:12726–12737, 2021.

[Rey and Maly, 2023] S. Rey and J. Maly. The (computational) social choice take on indivisible participatory budgeting. *arXiv:2303.00621*, 2023.

[Skowron *et al.*, 2020] P. Skowron, A. Slinko, S. Szufa, and N. Talmon. Participatory budgeting with cumulative votes. *arXiv:2009.02690*, 2020.

[Talmon and Faliszewski, 2019] N. Talmon and P. Faliszewski. A framework for approval-based budgeting methods. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI-2019)*, pages 2181–2188, 2019.