Recommendation Fairness in e-Participation: Listening to Minority, Vulnerable and NIMBY Citizens

Marina Alonso-Cortés,
Iván Cantador[0000−0001−6663−4231], and Alejandro Bellogín[0000−0001−6368−2510]

Escuela Politécnica Superior, Universidad Autónoma de Madrid, 28049 Madrid, Spain
marina.alonso-cortes@estudiante.uam.es,
ivan.cantador@uam.es, alejandro.bellogin@uam.es

Abstract. E-participation refers to the use of digital technologies and online platforms to engage citizens and other stakeholders in democratic and government decision-making processes. Recent research work has explored the application of recommender systems to e-participation, focusing on the development of algorithmic solutions to be effective in terms of personalized content retrieval accuracy, but ignoring underlying societal issues, such as biases, fairness, privacy and transparency. Motivated by this research gap, on a public e-participatory budgeting dataset, we measure and analyze recommendation fairness metrics oriented to several minority, vulnerable and NIMBY (Not In My Back Yard) groups of citizens. Our empirical results show that there is a strong popularity bias (especially for the minority groups) due to how content is presented and accessed in a reference e-participation platform; and that hybrid algorithms exploiting user geolocation information in a collaborative filtering fashion are good candidates to satisfy the proposed fairness conceptualization for the above underrepresented citizen collectives.

Keywords: Recommender system · Citizen participation · Fairness · Bias · Minority group · Vulnerable group · NIMBY.

1 Introduction

Citizen participation (a.k.a. public participation) refers to the active involvement of individuals in government decision-making processes that affect their lives and communities [20]. It entails the citizens' engagement in various forms, such as taking part in consultations, attending public meetings, and collaborating with governments, organizations or other stakeholders to solve citizenry problems and to shape public decisions. It thus has the potential to strengthen democracy by creating more inclusive, transparent and informed problem-solving, empowering citizens to have a more active role in public initiatives and policies, and increasing their trust in public institutions [8].

Citizen participation is progressively being conducted on the internet, through the so-called electronic participation (or *e-participation*) platforms [19]. Participedia[1] is an open-source, collaborative project aimed to document, promote and analyze public

[1] https://participedia.net

participation and democratic innovations around the world. It serves as a global resource with information on developed citizen participation initiatives, allowing individuals and organizations to share their experiences, case studies, and best practices. As of October 2023, with more than 2,400 cases recorded, it shows that 30% (111 out of 362) of the considered participatory method types are online. Representative examples of these methods are e-deliberation, e-voting, and e-participatory budgeting.

By harnessing the power of ICT, e-participation can overcome physical barriers to participation and amplify the voices of individuals who may be marginalized or excluded from the government decision-making [35]. It thus can offer greater accessibility, participation convenience (flexibility) and efficiency, and broader reach and inclusivity. Nonetheless, as stated by the OECD [24], e-participation has difficulties, such as the digital divide (i.e., required digital literacy) and technological barriers (e.g., access to the internet). Moreover, it has been shown to entail certain limitations, such as unequal (biased) citizen representation, and non-in-depth deliberation and polarization [5].

These problems are originated or intensified by the overload of information and the lack of personalization in the used platforms [4], among other causes. For e-participatory processes in large cities or at regional or national levels, the number and size of citizen proposals, discussions and debates may be overwhelming for a user. However, in general, current platforms do not filter or rank the content provided to the user according to her profile and preferences. It is for this situation that researchers have begun to explore the use of recommender systems in e-participation [28].

Among other applications, in the research literature, recommender systems for online citizen participation have been proposed for presenting political candidates with similar ideological positions [12,30], and suggesting relevant citizen proposals based on personal preferences [4,6]. Previous work has mainly focused on the development of algorithmic solutions to be effective in terms of personalized content retrieval accuracy. These solutions, by contrast, have not been studied considering underlying societal issues, such as biases, fairness, privacy and transparency.

Motivated by this research gap, in this work, we aim to investigate whether traditional recommendation algorithms applied on a representative e-participation platform entail bias and unfairness effects for certain citizens. Hence, our study targets the following two research questions:

- **RQ1**. How can recommendation fairness be conceptualized and formalized in e-participation taking societal issues into account?
- **RQ2**. How do standard recommendation algorithms behave on considered social fairness dimensions in a real e-participation case?

To address these questions, we evaluate diverse recommendation algorithms according to both accuracy and fairness metrics, on a public dataset of the *Decide Madrid*[2] e-participatory budgeting platform, and for several general, underrepresented citizen collectives. More specifically, in the context of making personalized recommendations of citizen proposals publicly available and discussed online in Decide

[2] `https://decide.madrid.es`

Madrid, we carefully identify a number of minority, vulnerable and NIMBY[3] (Not In My Back Yard) groups of city residents, and assign to them the proposals that deal with their concerns, needs and problems. Then, we propose a novel social fairness conceptualization, and measure and analyze associated fairness metrics based on *Generalized Cross Entropy* [10], with which, for the above groups, we study existing biases of recommendations generated by content-based, collaborative filtering and hybrid algorithms. We release our data and code as open access[4].

2 Related Work

2.1 Recommender Systems for Online Citizen Participation

In the context of cities, recommender systems have been proposed for a large variety of problems and tasks. In [28], the authors survey the state of the art, characterizing and categorizing existing recommendation approaches in terms of several city dimensions, namely *smart economy*, *environment*, *mobility*, *governance*, *living*, and *people* dimensions.

In all these dimensions, we can find purposes and goals for recommender systems where citizens are active participants, such as fostering healthy lifestyle [7] and promoting cultural city heritage [2] in *smart living*, and optimizing parking space usage [34] and supporting evacuation management [18] in *smart mobility*.

For the *smart governance* dimension –aimed to increase efficiency in municipal management, and promote citizen participation and inclusion–, as presented in [9], recommender systems can be further categorized according to whether they are applied to government-to-citizen (G2C), government-to-business (G2B), or government-to-government (G2G) services.

According to [9], G2C recommenders are predominant in the research literature, and mainly focus on providing citizens with personalized government e-notifications and e-services [1]; and keeping the government informed about the citizens' problems, concerns and opinions expressed in social media, and e-consultation and e-participation platforms [30].

Besides, recommender systems for online citizen participation have been proposed for assisting voters by presenting political candidates with similar ideological positions [12,30], retrieving comments from individuals who hold similar and dissimilar opinions [22], providing local news relevant to citizen discussions [17], supporting processes of public participation in urban planning [21], finding relevant citizen proposals based on personal preferences [4,6], and assisting citizens in tagging personal e-petitions [33].

As [4,6], in this paper, we consider the task of providing personalized suggestions of citizen proposals existing in an e-participation platform. However, differently to previous work, we evaluate recommendation algorithms by analyzing to what extent they

[3] NIMBY phenomenon: residents' opposition to certain development projects, facilities or infrastructures that they believe could have a negative impact on their immediate surroundings.

[4] `https://github.com/malonsocortes/fairness-eparticipation-recsys`

underrepresent proposals about minority and vulnerable collectives, and are biased towards proposals that affect citizen majorities. This approach could be explored in other recommendation tasks for online citizen participation, such as those mentioned above.

2.2 Fairness in Recommender Systems

The evaluation of recommender systems has been mainly conducted from the point of view of the accuracy and effectiveness of the generated recommendations. In recent years, nonetheless, there has been growing interest and concern for issues related to fairness [32,11], since recommenders may have algorithmic biases or may be influenced by biases existing in input data –understanding bias as a systematic deviation of the results that benefits some users or items against others.

The notion of fairness in the recommender systems field is acknowledged to be multi-faceted, entailing several factors, such as the target stakeholders, the type of benefit expected from recommendations, and the context of the recommendation application, among others [11]. Besides, traditional fairness definitions have been associated with a homogeneous distribution of benefits among the different groups of users or items involved [13]. For this, it is usually necessary to define protected groups (e.g., female or black candidates in job recommendations) or sensitive attributes, either from users or items (e.g., the user's race or gender, or a job's leadership requirements).

Recent work, by contrast, challenges this idea, and shows that, depending on the application context, definitions of fairness related to non-homogeneous distributions might be needed [10], in part because not all the stakeholders understand fairness under the same perspective. For instance, a streaming service company with free and paid user accounts could consider it fair that the error of a not relevant recommendation would be smaller for a free account than for a paid account.

This consideration was modeled in [10] through a probabilistic framework based on an adaptation of the Generalized Cross Entropy (GCE) metric, which is used to quantify the difference between two probability distributions. Specifically, in the context of recommendation fairness, one probability distribution would be estimated according to the output provided by the recommendation algorithm, whereas the other one would be the ideal or target distribution, either uniform (as the standard *equality* fairness perspective) or non-uniform. Moreover, depending on the probability space, the metric could be used to assess user-, item- or even context-oriented fairness.

The flexibility granted by GCE makes it suitable for our work, as we shall test it with different definitions of fairness according to item attributes. In particular, being a novel contribution to the field, we will use GCE to measure recommendation fairness based on sensitive item distributions, going beyond item popularity, as done in [10].

Nonetheless, while our study advances the understanding of fairness in recommender systems, it does not address their potential role in promoting polarization, confirmation bias, and echo chamber effects. These phenomena undermine (political) deliberation and diversity of thought, as highlighted in Pariser's work on filter bubbles [25] and Nguyen et al.'s work on echo chambers [23]. Future research should focus on identifying and mitigating these effects, ensuring exposure to diverse viewpoints in recommender systems.

3 Case Study

3.1 Decide Madrid: A Participatory Budgeting e-Platform

As a specific citizen participation method, participatory budgeting is a democratic process of deliberation and decision-making in which residents choose how to spend part of a municipal budget. In this process, participants raise awareness of problems and issues related to their city within a wide range of topics –e.g., urban planning and housing, environment, education, health, transport and security–, and propose, debate and support public investment in solutions and initiatives for such problems and issues.

Participatory budgeting was originated in 1988 in Porto Alegre, Brazil [29], and since then it has gained popularity spreading to over 7,000 cities around the world[5], especially after the adoption of ICT and the digitization of the process through the so-called e-participatory budgeting. In this context, ad hoc e-participation platforms have entailed increasing transparency and saving time for citizens [26].

Decide Madrid is a representative example of this type of platforms, and has supported the annual participatory budgeting campaigns of Madrid, Spain, since 2016. It is a website that allows the city residents to propose, discuss and vote for proposals which the City Council commits to implement if they satisfy certain feasibility and citizen support requirements. The tool is built upon the CONSUL framework[6], which is accessible as open source, and, as of October 2013, has been used by at least 135 institutions of 35 countries and 90 million citizens around the world. Its architecture and user interface are analogous to those of other popular platforms, such as the Stanford Participatory Budgeting[7] and EU Open Budgets[8] platforms, and are based on traditional *web forums* with tree structures of conversation threads (i.e., nested comments), which are commonly used by other e-participation tools.

In this work, we focus our attention on the citizen proposals posted in Decide Madrid, and the comments that the proposals received from registered users. Due to the large number of proposals (around four thousand per year) in the platform, the need for personalized proposal recommendations is justified. In fact, there already exist scientific papers on the topic [4,6]. However, to the best of our knowledge, no previous research has addressed the potential bias and unfairness effects that recommender systems for e-participation may have.

3.2 The Decide Madrid Dataset

Among hundreds of data collections, the open data portal[9] of the Madrid City Council gathers and makes publicly available citizen-generated content of Decide Madrid; specifically, its citizen proposals with their metadata and comments. From this open data collection, in [4,5], the authors built a dataset for the proposals made in four participatory budgeting campaigns (2015-2019), which we extended as explained in Section 4,

[5] `https://www.participatorybudgeting.org/about-pb/#what-is-pb`
[6] `https://consuldemocracy.org`
[7] `https://pbstanford.org`
[8] `https://openbudgets.eu`
[9] `https://datos.madrid.es`

Table 1. Statistics of the Decide Madrid dataset.

Campaign	Users (U)	Proposals (P)	Comments (C)	Sparsity	C/U	C/P
C1	8,009	9,677	34,149	99.96%	4.26	3.52
C2	2,374	3,805	7,190	99.94%	3.02	1.88
C3	1,275	2,814	4,075	99.91%	3.19	1.44
C4	1,210	2,746	4,829	99.89%	3.99	1.75
Total	**12,868**	**19,042**	**50,243**			

and used for the work presented herein. In the dataset, the comments that users made on the proposals are considered as feedback of interest, and are assumed as implicit, unary ratings. Besides, processing the textual metadata in Spanish of the proposals (i.e., titles, tags, summaries and descriptions), the authors assigned categories, topics and locations (districts) to the proposals. Table 1 shows some statistics of the dataset for the covered campaigns. The rating sparsity levels are extremely high (99.9%), being greater than those of popular datasets in the recommender systems research field.

4 Proposed Recommendation Fairness for e-Participation

4.1 Fairness Conceptualization

As mentioned in Section 2.2, to address **RQ1**, we propose using the Generalized Cross Entropy (GCE) metric for measuring recommendation fairness. This metric allows considering distinct conceptualizations of fairness, and is based on the difference between the distribution of recommendations generated by an algorithm and an ideal or target distribution (*perspective*) of recommendations with respect to a certain user/item/context variable (*attribute*).

For e-participation, we propose to instantiate and analyze the GCE metric by taking as item **attribute** of interest the belonging of a citizen proposal to a minority, vulnerable or NIMBY citizen group, potentially discriminated by a recommendation algorithm. From now on, we will refer to this attribute as *group*. Its possible values correspond to three broad categories: `Minority` (for minority and vulnerable citizen groups), `NIMBY` (for NIMBY groups), and `Other` (for other groups, presumably non-underrepresented). Additionally, we measure several versions of the GCE metric from the following **perspectives**: uniform (p_u), test proportion (p_t), and biased towards discriminated groups, namely minority/vulnerable (p_m), NIMBY (p_n), and both minority/vulnerable and NIMBY (p_{m+n}). Table 2 gathers all the above attribute and perspective values.

For establishing the minority, vulnerable and NIMBY groups to study, we proceeded in a two-fold process, counting on the collaboration of a political science professor expert in citizen participation. First, we generated a list of potential groups from a compilation of technical reports and scientific papers (e.g., [14,27]). Next, we checked if the Decide Madrid dataset had citizen proposals belonging to any of the identified groups. For such purpose, using the Apache Lucene library[10], we created a

[10] `https://lucene.apache.org`

Table 2. Possible values of the proposed fairness attribute and perspective.

Attribute	`Minority`
	`NIMBY`
	`Other`

Perspective	Uniform (p_u) Test proportion (p_t) Biased towards discriminated groups: minority (p_m), NIMBY (p_n), both (p_{m+n})

search index storing all the citizen proposals of the dataset. The index allowed us to launch regular expression-based queries against the proposals' titles and summaries[11].

In an iterative fashion, we collaboratively defined a vocabulary of terms (i.e., keywords and regular expressions) for each group. For instance, the 'People with disabilities' minority group was represented with terms such as `disabilit*`, `handicap*`, `impairment*`, `accessibility`, `reduced mobility`, and `architectural barrier*`, where the asterisk `*` stands for none or several word letters. Each term was carefully selected so that it did not generate ambiguity in finding its corresponding group. We provide the groups and their vocabularies in our public repository.

Then, in the index, for each term, we searched for proposals whose title or summary had the term. Boosting in the queries the term matches occurred on the titles, we created aggregated scores for every retrieved proposal and group, by counting the frequencies of a group's terms in the proposal's title and summary. Finally, each proposal was assigned to its matched group with the highest score. Table 3 shows the minority/vulnerable and NIMBY groups, respectively listed within the broad categories `Minority` and `NIMBY`. The groups are sorted by decreasing number of proposals in the dataset, showing which citizen collectives are less represented in Decide Madrid.

We note that regardless of its category, any of the sets of proposals corresponds to an unrepresented collective of citizens which is not limited to Madrid, but could exist in any (large) city of the world. We thus believe that the built group list and vocabularies, as well as the sets of citizen proposals, may be of interest for the research community.

4.2 Fairness Formalization

In [10], the GCE metric is defined as follows:

$$GCE_\beta(A,R;p_f) = \frac{1}{\beta(1-\beta)}\left[\sum_{a\in A} p_f^\beta(a)\cdot p_R^{(1-\beta)}(a) - 1\right] \quad (1)$$

where A is the attribute space upon which probability distributions are defined, R is the recommendation algorithm whose fairness is assessed, and p_f is the ideal or target fairness distribution, against which GCE will compare the estimated p_R distribution from R –in particular, if $p_R = p_f$ then $GCE = 0$, i.e., R is considered a perfectly fair model. By definition, GCE outputs negative values, so the closer they are to 0, the

[11] We discarded using the proposals' descriptions since they entailed information noise and ambiguities on the proposals' main topics.

Table 3. Considered minority/vulnerable and NIMBY groups with their respective numbers of citizen proposals in the dataset.

Minority and vulnerable collectives		**NIMBY issues**	
People with mental disorders	9	Shantytowns	6
Evicted and rehoused	10	Ethnicities and xenophobia	11
LGTBIQ+ collective	11	Squatting	12
Poor and people in social exclusion	14	Gambling and betting houses	14
People with dependency or special needs	16	Landfills, incinerators and crematoriums	28
Parents [family conciliation]	18	Antennas and electrical towers	34
Retirees (pensioners)	33	Drugs [on the streets]	35
Women [equality and gender violence]	33	Prostitution	44
Immigrants and refugees	33	Urban planning [at city level]	69
Cyclists and motorcyclists [no bike lanes issues]	57	Alcohol [on the streets]	81
Homeless (vagrants and destitute people)	62	Terraces	86
Unemployed	65	Burials	107
Young people (youth)	75	Resident parking	121
Elderly (Third Age)	142	Noise [on the streets]	173
People with disabilities	262	Feces and urine [on the streets]	207
Children (childhood)	709	Garbage [no recycling issues]	235
	1,549		**1,263**

closer the two distributions are and, hence, p_R might be considered fairer with respect to p_f. The choice of β is critical and entails a particular divergence metric depending on its value. As in [10], we shall use a value of $\beta = 2$, which corresponds to Pearson's χ^2, since the metric becomes more robust to outliers.

To estimate p_R, we first obtain the mapping $\text{att}(i)$ of each recommended item i to the **attribute space** A. As specified in the previous section, A allows encoding whether a citizen proposal (item) refers to a minority, NIMBY, or other group. Hence, $|A| = 3$. Based on this, we estimate the value p_R of each element $a \in A$ by considering how often the items of the group appear in a given recommendation list:

$$p_R(a) = \frac{1}{Z} \sum_{i \in \mathcal{I}:\text{att}(i)=a} rg_R(i) \tag{2}$$

$$rg_R(i) = \sum_{u \in \mathcal{U}} \phi(i, Rec_u^K) \cdot \text{gain}(u, i, r) \tag{3}$$

where $Z = \sum_{i \in \mathcal{I}} rg_R(i)$ is the normalization factor so that $\sum p_R(a_j) = 1$, Rec_u^K is the set of top-K items recommended by model R to user u, and $\phi(i, Rec_u^K) = 1$ if $i \in Rec_u^K$ and 0 otherwise. The function $\text{gain}(u,i,r)$ encodes the recommendation gain of item i for user u in position $r = \text{rank}(i, Rec_u^K)$. In this work, we define the gain as the normalized Discounted Cumulative Gain (nDCG) metric. We refer the reader to [10] for further details and possible configurations of GCE.

Afterwards, to compute the metric, we have to specify the *fairness perspective* to assess. By doing this, we define what is considered a fair recommendation. As explained before, in GCE, this is equivalent to setting a target distribution probability p_f, with the constraint that such distribution needs to live in the same attribute space A as p_R. We shall consider three possibilities, as summarized in Table 2.

By considering a **uniform perspective**, we would assume that fairness is equivalent to *equality*; that is, a fair recommendation algorithm is such that it equally

Table 4. Percentage of test citizen proposals that belong to each group by campaign.

Campaign	`Minority`	`NIMBY`	`Other`
C1	6.7%	10.3%	83.0%
C2	6.2%	7.9%	85.9%
C3	4.4%	7.0%	88.6%
C4	7.7%	7.3%	85.0%

suggests items of the three categories (`Minority`, `NIMBY`, `Other`): $p_u = \left[\frac{1}{3},\frac{1}{3},\frac{1}{3}\right]$. While this perspective puts some stress on the minority, vulnerable and NIMBY groups (since they do not frequently appear in the dataset), it still assumes a compromise between the groups must be met. However, in the long term, it may continue reinforcing a sufficiently large number of recommendations from the `Other` category, as they are the most popular ones. To better understand this rationale, Table 4 shows the number of proposals of each category in the test sets of the four campaigns. It is clear that in the dataset, the `Other` category is the largest one, followed by the `NIMBY` and `Minority` categories, in most of the campaigns.

Under a different perspective, lying at the other extreme of the spectrum, we define fairness as the situation where recommendations match the observed items in the system – i.e., based on *exposure* [3]. This conceptualization assumes that a recommender imposes the user a citizen proposal she would not have seen by herself. Hence, it aims to maintain as much as possible the status of previous user preferences, even if they are biased towards non-underrepresented proposals. This estimation could be measured based on the entire dataset, only the training set, or only the test set. We decided towards the latter. Hence, as shown in Table 2, we refer to it as the **test proportion perspective**. In particular, according to the data split of our experiments (Table 4), the ideal distribution according to this perspective for C1 would be $p_t = [0.067, 0.103, 0.830]$.

Finally, we consider a perspective to specifically account for biases or *discriminations* on the considered sensitive citizen categories (and consequently, on their underlying groups), by increasing the weight of each category. We thus would be imposing (and measuring the extent of) positive discrimination or affirmative actions for the groups. Based on this rationale, we define three ideal distributions: a first one biased only towards the `Minority` category (groups) ($p_m = [0.8, 0.1, 0.1]$), following a **minority perspective**; a second one biased towards the `NIMBY` category ($p_n = [0.1, 0.8, 0.1]$), following a **NIMBY perspective**; and a third one where all the weight is shared across the two sensitive categories ($p_{m+n} = [0.45, 0.45, 0.1]$), following an **underrepresented perspective**.

5 Experiments

5.1 Data Processing

On the dataset described in Section 3.2, we removed those proposals without location, category and topic, as this metadata is needed by some of the evaluated recommendation algorithms. We also removed the users with no interactions (i.e., comments) in the dataset, avoiding cold-start cases.

As mentioned in Section 3.2, the dataset had citizen-generated content of the four e-participatory budgeting campaigns dating from 2015 to 2019. Some citizen proposals had comments from several campaigns. To avoid this situation, we discarded those comments that did not belong to their proposal's campaign. Consequently, at the campaign level, we then repeated the removal of possible users with no comments.

With all the above, we built four datasets, each of them associated to an isolated campaign. Due to lack of space, in the subsequent sections, we only report and analyze average empirical results from experiments on the four campaigns. Nonetheless, we did not find significant differences in the results across campaigns.

5.2 Recommendation Algorithms

We experimented with five families of recommendation algorithms. In all cases, as done in [4,6], we considered the comment a user makes on a citizen proposal as a (unary) rating, indicating implicit feedback of interest, regardless the comment was positive or negative, i.e., in favor or against the proposal or a previous comment.

Specifically, we evaluated algorithms that generate recommendations randomly (*rand*) and based on popularity according to the number of users commenting a proposal (pop_u) and the number of comments of a proposal (pop_c); collaborative filtering algorithms, both heuristic based on items (*ib*) and users (*ub*), and model-based via matrix factorization (*mf*) and Bayesian personalized ranking (*bpr*); content-based algorithms exploiting user and item profiles with category (cb_{cat}), topic (cb_{top}), or location (cb_{loc}) information; and hybrid algorithms using either the *ib* or *ub* heuristic with content-based similarities, i.e., $cbib_{cat}$, $cbib_{top}$, $cbib_{loc}$. $cbub_{cat}$, $cbub_{top}$, and $cbub_{loc}$.

For some of these algorithms, we used the implementations given in the *Implicit* library[12]. Those algorithms that are not included in the library were implemented on top of it, and their source code was made publicly available.

5.3 Accuracy of Recommendation Algorithms

To evaluate the accuracy of the recommenders, we computed the following ranking-based metrics: Precision, Recall, MAP, nDCG, MRR, and F1, which are described in [15]. As mentioned before, we report and analyze the average of these metrics over the four campaigns of the dataset. To ensure stable results, the algorithms were evaluated through a 5-fold cross-validation process, and tuned with respect to nDCG@100.

Table 5 shows the achieved accuracy values at cutoff 50. From it, we first highlight that the popularity algorithms pop_u and pop_c reached the best results for all ranking metrics. This evidences a strong popularity bias existing in the Decide Madrid platform, which likely also appears in (many) other similar e-participation tools, and certainly in other domains and applications [13]. A possible explanation for this bias is the way the content is presented and accessed in Decide Madrid. When a proposal is voted in the platform, it gets higher relevance for being shown at the top positions of the platform's interface, which consequently increases the probability of being accessed, voted and commented (i.e., rated in the database). It is also interesting to note that pop_u, which is based on the number of users who have commented the proposal, is

[12] `https://github.com/benfred/implicit`

Table 5. Average ranking-based accuracy values achieved by the recommendation algorithms evaluated on the four campaigns of the dataset. Best values per column in darker colors.

	Precision	**Recall**	**MAP**	**nDCG**	**MRR**	**F1**
rand	0.001	0.014	0.002	0.004	0.002	0.001
pop_u	0.006	0.217	0.062	0.096	0.067	0.011
pop_c	0.005	0.198	0.044	0.078	0.051	0.010
ib	0.002	0.041	0.013	0.020	0.017	0.003
ub	0.003	0.063	0.018	0.030	0.027	0.005
mf	0.003	0.120	0.052	0.068	0.057	0.006
bpr	0.003	0.101	0.017	0.035	0.021	0.005
cb_{cat}	0.001	0.023	0.003	0.008	0.005	0.002
cb_{top}	0.002	0.033	0.006	0.013	0.011	0.003
cb_{loc}	0.001	0.033	0.004	0.010	0.006	0.003
$cbib_{cat}$	0.001	0.020	0.004	0.008	0.006	0.002
$cbib_{top}$	0.002	0.046	0.010	0.020	0.015	0.004
$cbib_{loc}$	0.002	0.038	0.010	0.017	0.013	0.003
$cbub_{cat}$	0.002	0.066	0.015	0.028	0.022	0.004
$cbub_{top}$	0.002	0.066	0.026	0.037	0.032	0.004
$cbub_{loc}$	0.003	0.070	0.013	0.027	0.019	0.004

Table 6. Average GCE values (the closer to 0, the better) achieved by the recommendation algorithms on the four campaigns of the dataset. Each column refers to a fairness perspective. nDCG values are included for completeness. Best values per column in darker colors.

	nDCG	p_u	p_t	p_m	p_n	p_{m+n}
rand	0.004	-1.264	-0.007	-3.806	-5.151	-2.604
pop_u	0.096	-1.932	-0.031	-10.243	-2.605	-3.817
pop_c	0.078	-1.005	-0.011	-4.716	-2.702	-2.125
ib	0.020	-1.034	-0.004	-3.974	-3.626	-2.181
ub	0.030	-0.954	-0.005	-3.888	-3.237	-2.034
mf	0.068	-2.071	-0.018	-7.292	-6.393	-4.077
bpr	0.035	-1.270	-0.028	-4.482	-4.496	-2.611
cb_{cat}	0.008	-1.600	-0.007	-6.388	-4.543	-3.219
cb_{top}	0.013	-1.757	-0.010	-6.399	-5.450	-3.505
cb_{loc}	0.010	-1.157	-0.006	-3.549	-4.774	-2.407
$cbib_{cat}$	0.008	-1.633	-0.008	-6.617	-4.509	-3.280
$cbib_{top}$	0.020	-1.864	-0.012	-6.605	-5.873	-3.701
$cbib_{loc}$	0.017	-1.225	-0.020	-2.989	-5.726	-2.529
$cbub_{cat}$	0.028	-1.352	-0.007	-6.228	-3.242	-2.764
$cbub_{top}$	0.037	-1.319	-0.007	-6.035	-3.237	-2.702
$cbub_{loc}$	0.027	-1.387	-0.018	-6.780	-2.882	-2.824

more accurate than pop_c, which considers popularity as the number of comments a proposal has. The more users have commented (rated) a proposal, the more likely the proposal is in the test, since it has a greater number of ratings from distinct users.

The next best performing algorithms are the *mf*, *bpr* and *ub* collaborative filtering recommenders, followed by the user-based hybrid recommenders. This reinforces the idea that the preferences of users in Decide Madrid can be related to each other for personalized recommendation purposes.

5.4 Fairness Impact of Recommendation Algorithms

Addressing **RQ2**, we report (Table 6) and analyze the GCE values achieved by the evaluated recommendation algorithms according to the fairness perspectives presented in Section 4. Recall that *fairness as equality* is represented by the **uniform perspective** (p_u). In this case, *ub* and pop_c achieved the best (highest, closer to 0) results, which means that their recommendations are uniform for the considered fairness attribute values: minority/vulnerable, NIMBY or any other citizen collective. On the other extreme, we observe that pop_u and *mf* as the most biased ones.

According to the **test perspective**, that is, how close the generated recommendations are to the item distribution observed in test, the *ib* and *ub* heuristic collaborative filtering algorithms, those based on content (in particular, cb_{loc}), and the hybrid algorithms achieve the best results. This means that these algorithms are good approaches to recommend items of each citizen collective in the same proportion as they interest the users, i.e., as they actually appear in the test set. However, considering a tradeoff between fairness and accuracy, *ib* performs worse, since its nDCG value is lower. In this context, the success of exploiting location information in a collaborative filtering fashion could indicate that location is a good indicator of fair preferences, reflecting that users, in addition to popular proposals, tend to explore and comment on proposals about their surroundings (districts or neighborhoods).

Finally, regarding the **perspective biased towards discriminated groups**, the algorithms exploiting location information –in particular, $cbib_{loc}$ for p_m and $cbub_{loc}$ for p_n– and the heuristic collaborative filtering algorithms stand out. This could be related to the idea that users in the same surrounding (district or neighborhood) tend to be interested in the same proposals, and in this way, proposals from the minority and NIMBY groups that affect a certain environment are relevant and consequently should be recommended to users in that environment. It is worth noting the negative bias that the pop_u algorithm has on the `Minority` category is accentuated, moving far away from the idea of fairness defined by p_m. This is probably linked to the fact that the minority proposals are not popular on the platform. For the `NYMBY` category, popularity algorithms achieve better results, which could be explained by the fact that, by their nature, the NIMBY proposals are more controversial, and thus tend to have more comments and more users commenting on them [5].

6 Conclusions

The integration of recommender systems in e-participation platforms has been envisioned as a solution to reduce the information overload problem of the platforms. However, for such purpose, it is essential to evaluate the generated recommendations not only in terms of personalized content accuracy, but also according to social fairness dimensions, with the aim of avoiding or mitigating biases that could exist or be amplified by how information is presented and accessed in the platforms.

The work reported in this paper represents a seminal research in that direction. In particular, we have proposed a conceptualization and metrics of recommendation fairness oriented to minority, vulnerable and NIMBY groups of citizens, and have experimented with measuring such metrics for heterogeneous recommender systems on a real e-participation dataset. Our empirical results have been revealing. First, they have confirmed initial suspicions about the fact that there is a strong popularity bias on the platform data that affects the recommendation algorithms. Second, they have shown that those recommendations generated by exploiting the users' geolocation information in a collaborative filtering fashion are less affected by such bias for the target underrepresented citizen collectives.

Nonetheless, we are aware of limitations in our study. Most importantly, we should conduct more exhaustive experiments to further corroborate and generalize our findings and conclusions. In this sense, we plan to thoroughly test further hyperparameter settings and recommendation algorithms, and use additional datasets, such as those published in [6] from the e-participatory budgets of New York City, Miami, and Cambridge, which would allow us to explore citizen-generated content in English and distinct types of participatory budgeting processes. Moreover, we note we have focused on the identification and measuring of unfair recommendations. Ad hoc, fairness-aware recommendation algorithms and mitigation techniques should be investigated. We believe diversification [31] could be an effective approach for such purpose.

A second priority line of future work is extending the analysis of e-participation recommendation fairness to additional attributes of users, items and contexts. On the one hand, we plan to run our experiments for particular groups of citizens (rather than considering all together in the broad `Minority` and `NIMBY` categories), aiming to better understand the nature of the identified biases. On the other hand, looking at the behavior of the algorithms according to geolocation information, we could analyze the fairness with respect to different areas of a city, and study whether biases correlate with certain city or citizen participation features. In fact, similarly to [5], achieved results could be contrasted with demographic, socioeconomic and ideological data of the city's districts and neighborhoods.

Another issue that has arisen in our work is the sensitivity of the GCE metric to the differences in the proportion of test ratings for the different groups. In this sense, a possible line of future research would be incorporating into the approximation of the target recommendation distributions a normalization factor based on the number of ratings of each group. The objective would be to define a metric that is more robust to these differences in testing, and therefore more versatile when applied to other cases.

Regardless of these issues, we must highlight the relevance of the targeted citizen groups and proposals, which address pressing concerns within the community: cases of poverty and hunger, deficiencies in educational and healthcare systems, social inequalities and environmental, urban planning, and economic growth problems. Notably, these concerns are related to the Sustainable Development Goals (SDGs) established by the European Union [16]. Therefore, we believe that our work can be an inspiration for future redesigns of e-participation platforms, in such a way that, through novel, fairness-aware methods of information access, retrieval and recommendation, they would allow policymakers to be more aware of existing (city) problems and act (at local level) more effectively on some of the SDGs.

Acknowledgements

This work was supported by Grant PID2019-108965GB-I00 of the Spanish Ministry of Science and Innovation, and Grant PID2022-139131NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe."

References

1. Al-Hassan, M., Lu, H., Lu, J.: A semantic enhanced hybrid recommendation approach: a case study of e-government tourism service recommendation system. Decision Support Systems **72**, 97–109 (2015)
2. Barile, F., Calandra, D.M., Caso, A., D'Auria, D., Di Mauro, D., Cutugno, F., Rossi, S.: ICT solutions for the OR.C.HE.STRA project: from personalized selection to enhanced fruition of cultural heritage data. In: Proceedings of the 10th International Conference on Signal-Image Technology and Internet-Based Systems. pp. 501–507. IEEE (2014)
3. Boratto, L., Fenu, G., Marras, M.: Interplay between upsampling and regularization for provider fairness in recommender systems. User Modeling and User-Adapted Interaction **31**(3), 421–455 (2021)
4. Cantador, I., Bellogín, A., Cortés-Cediel, M.E., Gil, O.: Personalized recommendations in e-participation: offline experiments for the 'Decide Madrid' platform. In: Proceedings of the International Workshop on Recommender Systems for Citizens. pp. 1–6. ACM (2017)
5. Cantador, I., Cortés-Cediel, M.E., Fernández, M.: Exploiting open data to analyze discussion and controversy in online citizen participation. Information Processing & Management **57**(5), 102301 (2020)
6. Cantador, I., Cortés-Cediel, M.E., Fernández, M., Alani, H.: What's going on in my city? Recommender systems and electronic participatory budgeting. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 219–223. ACM (2018)
7. Casino, F., Batista, E., Patsakis, C., Solanas, A.: Context-aware recommender for smart health. In: Proceedings of the 1st Int. Smart Cities Conference. pp. 1–2. IEEE (2015)
8. Church, J., Saunders, D., Wanke, M., Pong, R., Spooner, C., Dorgan, M.: Citizen participation in health decision-making: past experience and future prospects. Journal of Public Health Policy **23**, 12–32 (2002)
9. Cortés-Cediel, M.E., Cantador, I., Gil, O.: Recommender systems for e-governance in smart cities: state of the art and research opportunities. In: Proceedings of the International Workshop on Recommender Systems for Citizens. pp. 1–6. ACM (2017)
10. Deldjoo, Y., Anelli, V.W., Zamani, H., Bellogín, A., Di Noia, T.: A flexible framework for evaluating user and item fairness in recommender systems. User Modeling and User-Adapted Interaction **31**, 457–511 (2021)
11. Deldjoo, Y., Jannach, D., Bellogín, A., Difonzo, A., Zanzonelli, D.: Fairness in recommender systems: research landscape and future directions. User Modeling and User-Adapted Interaction (in press) (2023)
12. Dyczkowski, K., Stachowiak, A.: A recommender system with uncertainty on the example of political elections. In: Proceedings of the 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. pp. 441–449. Springer (2012)
13. Ekstrand, M.D., Tian, M., Azpiazu, I.M., Ekstrand, J.D., Anuyah, O., McNeill, D., Pera, M.S.: All the cool kids, how do they fit in?: popularity and demographic biases in recommender evaluation and effectiveness. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency. vol. 81, pp. 172–186. PMLR (2018)
14. Eranti, V.: Re-visiting NIMBY: from conflicting interests to conflicting valuations. The Sociological Review **65**(2), 285–301 (2017)
15. Gunawardana, A., Shani, G., Yogev, S.: Evaluating recommender systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 547–601. Springer (2022)
16. Hák, T., Janoušková, S., Moldan, B.: Sustainable Development Goals: a need for relevant indicators. Ecological Indicators **60**, 565–573 (2016)

17. Kavanaugh, A., Ahuja, A., Pérez-Quiñones, M., Tedesco, J., Madondo, K.: Encouraging civic participation through local news aggregation. In: Proceedings of the 14th International Conference on Digital Government Research. pp. 172–179. ACM (2013)
18. Lujak, M., Billhardt, H., Dunkel, J., Fernández, A., Hermoso, R., Ossowski, S.: A distributed architecture for real-time evacuation guidance in large smart buildings. Computer Science and Information Systems **14**(1), 257–282 (2017)
19. Macintosh, A.: Characterizing e-participation in policy-making. In: Proceedings of the 37th Hawaii International Conference on System Sciences. pp. 10–pp. IEEE (2004)
20. Macintosh, A.: E-democracy and e-participation research in Europe. In: Digital Government: e-Government Research, Case Studies, and Implementation, pp. 85–102. Springer (2008)
21. Marsal-Llacuna, M.L., de la Rosa-Esteva, J.L.: The representation for all model: an agent-based collaborative method for more meaningful citizen participation in urban planning. In: Proceedings of the 13th International Conference on Computational Science and its Applications (Part III). pp. 324–339. Springer (2013)
22. Nelimarkka, M., Nonnecke, B., Krishnan, S., Aitamurto, T., Catterson, D., Crittenden, C., Garland, C., Gregory, C., Huang, C.c.A., Newsom, G., et al.: Comparing three online civic engagement platforms using the "spectrum of public participation" framework. In: Oxford Internet, Policy & Politics Conference. pp. 1–22 (2014)
23. Nguyen, T.T., Hui, P.M., Harper, F.M., Terveen, L., Konstan, J.A.: Exploring the filter bubble: the effect of using recommender systems on content diversity. In: Proceedings of the 23rd International Conference on World Wide Web. pp. 677–686. ACM (2014)
24. Organisation for Economic Co-operation and Development: OECD guidelines for citizen participation processes. OECD Publishing (2022)
25. Pariser, E.: The filter bubble: what the internet is Hiding from you. Penguin (2011)
26. Peixoto, T.: Beyond theory: e-participatory budgeting and its promises for eParticipation. European Journal of ePractice **7**(5), 1–9 (2009)
27. Peroni, L., Timmer, A.: Vulnerable groups: the promise of an emerging concept in European Human Rights Convention law. International Journal of Constitutional Law **11**(4), 1056–1085 (2013)
28. Quijano-Sánchez, L., Cantador, I., Cortés-Cediel, M.E., Gil, O.: Recommender systems for smart cities. Information Systems **92**, 101545 (2020)
29. de Sousa Santos, B.: Participatory budgeting in Porto Alegre: toward a redistributive democracy. Politics & Society **26**(4), 461–510 (1998)
30. Terán, L., Meier, A.: A fuzzy recommender system for eElections. In: Proceedings of the 1st International Conference on Electronic Government and the Information Systems Perspective. pp. 62–76. Springer (2010)
31. Vargas, S., Castells, P.: Rank and relevance in novelty and diversity metrics for recommender systems. In: Proceedings of the 5th ACM Conference on Recommender Systems. pp. 109–116. ACM (2011)
32. Wang, Y., Ma, W., Zhang, M., Liu, Y., Ma, S.: A survey on the fairness of recommender systems. ACM Transactions on Information Systems **41**(3), 1–43 (2023)
33. Yang, Z., Feng, J.: Explainable multi-task convolutional neural network framework for electronic petition tag recommendation. Electronic Commerce Research and Applications **59**, 101263 (2023)
34. Yavari, A., Jayaraman, P.P., Georgakopoulos, D.: Contextualised service delivery in the internet of things: parking recommender for smart cities. In: Proceedings of the 3rd World Forum on Internet of Things. pp. 454–459. IEEE (2016)
35. Zheng, Y., Schachter, H.L.: Explaining citizens' e-participation use: the role of perceived advantages. Public Organization Review **17**, 409–428 (2017)