Equality of Voice: Towards Fair Representation in Crowdsourced Top-K Recommendations

Abhijnan Chakraborty, Gourab K Patro, Niloy Ganguly, Krishna Gummadi, Patrick Loiseau

Equality of Voice: Towards Fair Representation in Crowdsourced Top-K Recommendations

Abhijnan Chakraborty
MPI for Software Systems, Germany
IIT Kharagpur, India

Gourab K Patro
IIT Kharagpur, India

Niloy Ganguly
IIT Kharagpur, India

Krishna P. Gummadi
MPI for Software Systems, Germany

Patrick Loiseau
Univ. Grenoble Alpes, Inria, CNRS
Grenoble INP, LIG & MPI SWS

ABSTRACT

To help their users to discover important items at a particular time, major websites like Twitter, Yelp, TripAdvisor or NYTimes provide Top-K recommendations (e.g., 10 Trending Topics, Top 5 Hotels in Paris or 10 Most Viewed News Stories), which rely on crowdsourced popularity signals to select the items. However, different sections of a crowd may have different preferences, and there is a large *silent majority* who do not explicitly express their opinion. Also, the crowd often consists of actors like bots, spammers, or people running orchestrated campaigns. Recommendation algorithms today largely do not consider such nuances, hence are vulnerable to strategic manipulation by small but hyper-active user groups.

To *fairly aggregate the preferences of all users* while recommending top-K items, we borrow ideas from prior research on social choice theory, and identify a voting mechanism called Single Transferable Vote (STV) as having many of the fairness properties we desire in top-K item (s)elections. We develop an innovative mechanism to attribute preferences of silent majority which also make STV completely operational. We show the generalizability of our approach by implementing it on two different real-world datasets. Through extensive experimentation and comparison with state-of-the-art techniques, we show that our proposed approach provides maximum user satisfaction, and cuts down drastically on items disliked by most but hyper-actively promoted by a few users.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Human-centered computing → Social media**;

KEYWORDS

Top-K Recommendation; Fair Representation; Twitter Trends; Most Popular News; Fairness in Recommendation

ACM Reference format:
Abhijnan Chakraborty, Gourab K Patro, Niloy Ganguly, Krishna P. Gummadi, and Patrick Loiseau. 2019. Equality of Voice: Towards Fair Representation in Crowdsourced Top-K Recommendations. In *Proceedings of FAT* '19: Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, January 29–31, 2019 (FAT* '19),* 10 pages.
https://doi.org/10.1145/3287560.3287570

FAT '19, January 29–31, 2019, Atlanta, GA, USA*
2019. ACM ISBN 978-1-4503-6125-5/19/01…$15.00
https://doi.org/10.1145/3287560.3287570

1 INTRODUCTION

Many websites today are deploying top-K recommendations to help their users find important items. For instance, social media sites like Twitter recommend 10 'Trending Topics' to let users know about breaking news stories. Review aggregators like Yelp or TripAdvisor show top 10 restaurants or hotels in a particular city. News websites like CNN or NYTimes show 10 most viewed or most shared stories. While some of these recommendations are *personalized*, i.e., tailored to a particular user, others are *non-personalized* and the same items are recommended to all users (at least in a geographical area).

Such recommendations implicitly rely on crowd-sourced popularity signals to select the items. Recently, concerns have been raised about the potential for *bias* in such crowdsourced recommendation algorithms [2]. For instance, Google's search query autocomplete feature has been criticized for favoring certain political parties [38]. In another work [9], we showed that the majority of Twitter trends are promoted by crowds whose demographics differ significantly from Twitter's overall user population, and certain demographic groups (e.g., middle-aged black female) are severely under-represented in the process.

In this paper, we propose to reimagine top-K non-personalized crowdsourced recommendations (e.g., trending topics or most viewed news articles) as the *outcomes of a multi-winner election* that is periodically repeated. We show that the observed biases in top-K recommendations can be attributed to the unfairness in the electoral system. More specifically in Twitter, we observe that during any single election cycle (5 to 15 minutes), (a) only a tiny fraction (< 0.1%) of the overall user population express candidate (topics or hashtag) preferences, i.e., *a vast majority of voters are silent*, (b) some people vote multiple times, i.e., there is *no one person, one vote principle*, and (c) voters choose from several thousands of potential candidates (topics or hashtags), splitting their votes over several moderate and reasonable topics, and thereby, allowing extreme topics (representing highly biased view points) to be selected. Today's trending topic (s)election algorithms are vulnerable to electing such fringe trends with as low as 0.001% of the electorate support.

To address the unfairness in item selections, we borrow ideas from extensive prior research on social choice theory. We focus on electoral mechanisms that attempt to ensure two types of fairness criteria: *proportional representation* that requires the divisions in (the topical interests of) the electorate to be reflected proportionally in the elected body (i.e., selected items) and *anti-plurality*, where an extremist candidate (item) highly disliked by a vast majority of voters has little chance of getting (s)elected. We survey existing

literature and identify a voting mechanism *Single Transferable Vote (STV)* as having the properties we desire in top-K item (s)elections.

However, implementing STV-based item selection poses a technical challenge: to deter strategic manipulation, STV requires every user to provide a preference ranking over all candidates. Requiring the website users to rank thousands of candidate items makes the scheme impractical. We solve this challenge by proposing to *automatically infer* the preference rankings for users. Fortunately, we can leverage the rich existing literature on personalized recommendations to rank items according to individual personal preferences of users. In fact, sites like Facebook and Twitter already use personal preferences to order topics in users' newsfeeds [30]. Additionally, our approach enables us to account for (i.e., automatically infer the ranking choices for) the large fraction of the electorate that is otherwise silent and inactive during any election.

We demonstrate the practicality and effectiveness of our ideas by conducting a comparative analysis of different mechanisms for top-K recommendations using real-world data from social media site Twitter and news media site Adressa. Over the course of a month, we collected trending topics recommended by Twitter itself, and computed in parallel the topics that would be recommended by four different election mechanisms including plurality voting (where the candidates with most first place votes win) and STV. At a high-level, our findings demonstrate that trending topics elected by STV are significantly less demographically biased than those selected by both plurality-based voting schemes and Twitter itself. At a lower-level, our analysis reveals how the improvement in STV selected topics arise from STV's fairness criteria of proportional representation (which selects topics such that most users have at least one of their highly preferred topics included in the elected set) and anti-plurality (which rejects highly biased topics disliked by a majority of users). We further evaluate the mechanisms for recommending most popular Adressa news stories every day throughout a two-months period, and make similar observations.

In summary, we make the following contributions in this paper: (a) by mapping crowdsourced recommendations to multi-winner elections, we show how the bias in recommendation can be traced back to the unfairness in the electoral process; (b) we establish the fairness properties desired in crowdsourced recommendations, and identify an electoral method, STV, which ensures fair representation in such contexts; (c) we implement STV by devising a mechanism to provide equality of voice even to the users who are otherwise silent during the election cycle. To the best of our knowledge, ours is the first attempt to introduce fairness in crowdsourced recommendations, and we hope that the work will be an important addition to the growing literature on fairness, bias and transparency of algorithmic decision making systems.

2 BACKGROUND AND MOTIVATION

As mentioned earlier, *non-personalized* top-K recommendations in different websites rely on crowdsourced popularity signals to select the contents. For example, Twitter recommends hashtags and key-phrases as trending when their popularity among the crowds exhibit a sudden spike [42]. Many news websites like NYTimes (`nytimes.com`) or BBC (`bbc.com/news`) recommend stories that are most read or most shared by their audience. Multiple recent works have highlighted the potential for bias in such recommendations.

2.1 Biases in crowdsourced recommendations

Google's search query autocomplete feature has been criticized as favoring certain political parties [38], while concerns about political biases in Facebook's trending topic selection have led a fierce debate about the need for human editorial oversight of the recommended trends [31]. Interestingly, after Facebook removed the human editors who used to oversee the topics (popular among the crowds) before they were recommended to the users [29], it was accused of featuring fake news as trending [32]. In our earlier work [9], we showed that the demographics of promoters of Twitter trends differ significantly from Twitter's overall user population, and certain demographic groups are under-represented in the process. Similarly, Baker *et al.* [3] found that the gender and racial stereotypes get perpetuated in Google search auto complete suggestions.

Going beyond demographic bias, different types of actors (such as spammers, trend hijackers or automated bots) disguise under the umbrella term 'crowd'. As crowdsourced algorithms are driven by data generated by them, their outputs will reflect the biases in the composition of the crowds. A recent investigation by Politico [27] revealed that Twitter bots were largely responsible for the trend #ReleaseTheMemo. Multiple works have also investigated the roles of spammers and trend hijackers around Twitter trends [40, 45].

We hypothesize that one of the main reasons behind the bias in crowdsourced recommendations is the **lack of fair representation** of various segments among the crowd considered in the algorithms. Using the datasets described next, we attempt to specifically identify the root causes behind the bias in the recommendations.

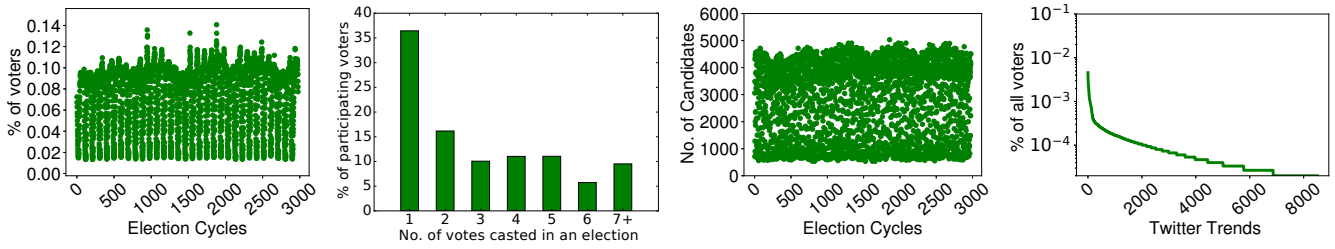
2.2 Datasets gathered

In this work, we consider two different recommendations: recommendation of (i) trending topics, and (ii) most popular news stories.

(i) Trending Topics: Social media sites like Twitter recommend a set of trending topics to help their users find happening events. We gathered extensive data from Twitter during February to July, 2017. Throughout this 6 months period, we collected 1% sample of all tweets posted in the US by applying the appropriate location filters in the Twitter Streaming API [43]. In total, we collected $230M+$ tweets posted by around 15 million US-based Twitter users throughout this period. Simultaneously, by querying the Twitter REST API [44] every 15-minutes during the month of July 2017, we collected all topics which became trending in the US. During this month, $10,877$ topics became trending, out of which $4,367$ were *hashtags* and the rest were multi-word phrases. For simplicity, we restrict our focus on trending hashtags in this paper.

(ii) Most Popular News: All major news websites recommend a set of stories which are most popular (e.g., most read, most shared) among the crowd. To consider such recommendations, we use the 'Adressa News Dataset' [18] which consists of the news reading behavior of around 3.1 million users on the Norwegian news website Adresseavisen[1] during the 3 months period from January, 2017 to March, 2017. The dataset not only provides the information about the stories read by every user (out of total $48,486$ news stories), but also includes how much time a reader spent in each story. Using these reading times as popularity signal, we simulate the recommendation of 10 most read news stories every day.

[1] https://www.adressa.no, henceforth referred as 'Adressa'



(a) (b) (c) (d)

Figure 1: (a) Percentage of all voters participating during different election cycles in Twitter. (b) Average number of votes casted by different voters during an election. (c) Number of potential candidates for becoming trending during election cycles. (d) Percentage of overall population needed to make different topics trending.

2.3 Reimagining Top-K recommendation as a multi-winner election

In this paper, *we propose to see crowdsourced recommendation as the result of an election*, where the users vote for an item (e.g., a hashtag, a news story) by tweeting, reading, or sharing it. We can think of every x time interval as an election cycle (where x can be any duration: 5 minutes, 1 hour, or 1 day), the topics or stories tweeted (or read) in an interval as the candidates, and user activities during this interval serving as the ballots. The recommendation algorithm can then be viewed as an election method which considers these ballots and selects the winner items for recommendation. If only one item is to be selected (i.e., $K = 1$), then it is a single winner election. For $K > 1$, the corresponding election is multi-winner.

2.4 Unfair election biases selection of items

The mapping between top-K recommendation and multi-winner election allows us to hypothesize that the bias in the recommended items originates from a lack of *fair representation* in the underlying election mechanism. More specifically, we identify a number of potential root causes, as discussed next.

2.4.1 **Not everyone votes in every election cycle.** Out of the thousands (or millions) of visitors to many websites, only a small fraction of them actively participate during any particular election cycle. For example, Figure 1(a) shows the percentage of Twitter users in our dataset who participated in the trending topic selection during different cycles throughout July, 2017. Although there are around 15 Million Twitter users in our dataset (all of whom are eligible voters), we can observe from Figure 1(a) that on average, only 0.052% of them influence the trending topic selection. Similarly, on average, only 4.54% of the Adressa readers read any news on a given day. Therefore, we can conclude that there is a large majority of website users who are *silent* during an election.

2.4.2 **One person can cast multiple votes in an election.** Different voters participating in an election may have different activity levels. For example, Figure 1(b) shows the average percentage of participating voters who cast different number of votes during trending topic election in Twitter. We can see that only 35% of the voters vote once (i.e., use a single hashtag), and rest of the voters either vote for different candidates (by using multiple hashtags) or vote for same candidate multiple times. Although, we do not know for sure whether Twitter's Trending Topic selection algorithm considers multiple votes from the same person, here we highlight that it may be vulnerable to such multi-voting. We see similar trends among Adressa readers, where there is a huge variation in individual users' reading activities.

2.4.3 **Too many candidates to choose from.** In different websites today, the number of potential candidates for recommendations is much more than a user can possibly notice. News websites are producing hundreds of news stories everyday, and news readers have very limited time and attention. The problem is more acute for social media – the amount of information generated is a lot more, and a user will encounter only those coming from her neighborhood, thus triggering a natural bias.

Figure 1(c) shows the number of candidate hashtags in Twitter during any election cycle. On average, at least 3,000 candidates compete to become trending in an election. Similarly, around 2,000 stories compete to become daily most popular news in Adressa.

2.4.4 **% of voters needed to get an item selected is too low.** As only a small fraction of voters participate in any election and their votes can get split across a large number of candidates, effectively a tiny fraction of overall user population can make an item get (s)elected. For example, Figure 1(d) shows that most of the Twitter trends enjoy the support of less than 0.001% of the overall Twitter population. This makes the elections vulnerable to biased and manipulated trends.

3 FAIRNESS CRITERIA FOR TOP-K RECOMMENDATIONS

In this work, we express the process in which top-K recommendations are chosen through crowdsourcing as an election mechanism. Extensive prior research in social choice theory have identified several fairness criteria (properties) desired from the electoral systems [14, 37]. However, all fairness criteria are not applicable in a given context. In Section 2.4, we identified the potential unfairness in the election mechanism that leads to bias in the crowdsourced recommendations. In this section, we propose three fairness properties an election mechanism should satisfy to make the recommendations fairly representative: (a) Equality of Voice, (b) Proportional Representation, and (c) Anti-Plurality.

3.1 Equality of Voice

Most of the top-K recommendations in use today (e.g., 'Most Viewed Stories' in news websites like `nytimes.com`) can be intuitively categorized as a particular type of electoral systems – 'Weighted Voting' [41] (also known as 'Plural Voting' [28]), where a single voter can vote for any number of candidates and that too multiple times. The candidates getting maximum votes are selected for recommendation (e.g., the stories getting maximum views regardless of users reading multiple stories or reading the same story multiple times). We saw earlier that there is a large variety in the activity levels of different users. Thus, effectively a hyper-active user can influence the election much more than a lesser-active user.

To avoid this issue, we propose that *the item (s)election algorithm should treat all website users (i.e., all voters) similarly, where no user is more privileged or discriminated to determine the set of winners.* In social choice, this property is known as *Anonymity Criterion* [37], and intuitively termed as 'one person one vote'. One way to achieve this is to require the voters to specify their preferences over a set of candidates. In the recommendation context, we can compute these preference rankings based on the activities of the users (e.g., a user may post her highly preferred topic more than a lesser preferred topic). In addition, to give equal voice to the *silent* (i.e., less active) users, we also need to *infer* their ranked choices over candidate items. Fortunately, we can utilize the long lines of works in personalized recommendations for inferring different user's *personalized preferences* towards different items (detailed in the next section).

Let $\sigma_i(j)$ denote the preference rank user i gives to item j (where $\sigma_i(j) = 1$ denotes that j is the most preferred item to i). $\beta_i = \{\sigma_i(j)\ \forall j \in C\}$ denotes the preference ranking (i.e., ranked ballot) of the user i, where $C = \{c_1, c_2, ..., c_m\}$ is the set of candidate items. Then, the top-K recommendation can be formally expressed as the 4-tuple (C, P, f, K), where $P = \{\beta_1, \beta_2, ..., \beta_n\}$ is the preference rankings from all users, and the selection algorithm is a function $f : C, P \rightarrow W$ which selects the set of K winner items W for recommendation from the candidate set C (i.e., $W \subseteq C$) using the preference rankings P.

3.2 Proportional Representation

Even after considering everyone's preferences, due to the presence of too many candidate items, users' choices get split across many *irrelevant alternatives* [1]. Furthermore, some alternatives may be very similar to each other (e.g., two hashtags or news stories referring to the same event), and there the vote splitting can be sharper. Consequently, items preferred by only a small number of users may end up being selected despite being disliked by the majority.

To illustrate this point, let us consider a toy example in Table 1, depicting a 2-winner election with 5 candidate items and 100 voters who rank them according to their choices. Assume that items 1 and 2 belong to a category C1, and items 3 and 4 belong to another category C2. Further assume that there is another item 5 representing an extreme opinion. We can see that although 39% of the voters are interested in each of C1 and C2, due to splitting of votes (especially between items 3 and 4), the extreme item 5 can win in a *plurality vote*[2], along with item 1. This is despite item 5 being disliked by 78% voters who have put it as their last choice. To alleviate the problems illustrated above, we consider two fairness criteria: (i) proportionality for solid coalitions, and (ii) Anti-plurality.

[2]where the candidates getting most first choice votes are (s)elected.

	1st	2nd	3rd	4th	5th	Category
Item 1	30	20	30	17	3	C1
Item 2	9	30	15	38	8	C1
Item 3	20	20	25	30	5	C2
Item 4	19	30	30	15	6	C2
Item 5	22	0	0	0	78	Extreme

Table 1: How 5 candidate items are present across different ranked choices of 100 users.

Proportionality for solid coalitions

We propose that *in a fair top-K recommendation, the diversity of opinions in the user population should be proportionally represented in the recommended items.* To formalize proportional representation, we consider the criterion of *proportionality for solid coalitions* [13]. In social choice, a solid coalition for a set of candidates $C' \subseteq C$ is defined as a set of voters V' who all rank every candidate in C' higher than any candidate outside of C'. The proportionality for solid coalitions criterion requires that if V' has at least $q \cdot \frac{n}{K+1}$ voters, then the set of winning candidates W should contain at least q candidates from C' (where $q \in \mathbb{N}$, $n = |P|$ and $K = |W|$).

In the context of crowdsourced recommendations, different groups of users may prefer a group of items (e.g., hashtags or news) more than other items. Then, the proportional representation criteria means that if a set of items is preferred by a group of users who represent a fraction $q \cdot \frac{1}{K+1}$ of the population, then at least q items from this set should be selected. The quantity $\lfloor \frac{n}{K+1} \rfloor + 1$ is known as *Droop Quota* [12]. In Table 1, the Droop Quota is 34 and hence, to satisfy proportional representation, a recommendation algorithm should select one item each from the categories C1 and C2.

3.3 Anti-plurality

In social choice theory, the *Majority Loser Criterion* [46] was proposed to evaluate single-winner elections, which requires that if a majority of voters prefer every other candidate over a given candidate, then that candidate should not be elected. We extend this criterion to top-K recommendations (and implicitly to multi-winner elections), where the *anti-plurality* property intuitively mandates that no item disliked by a majority of the users should be recommended. We can formalize this criterion by requiring that no candidate item among the bottom x percentile of the ranked choices for majority of the voters should be selected, where x is a parameter of the definition. For example, any recommendation algorithm satisfying anti-plurality will not select Item 5 in Table 1 because it is the last choice for 78% of the users.

4 FAIR TOP-K RECOMMENDATION WITH EQUALITY OF VOICE

Several electoral mechanisms have been proposed for multi-winner elections, which include *Plurality Voting, k-Borda, Chamberlain-Courant, Monroe or Approval Voting* [15]. Subsequent research works have investigated different fairness criteria that these mechanisms satisfy [14, 37]. In this paper, we consider a particular electoral mechanism, *Single Transferable Vote (STV)*, that satisfies two fairness criteria we described in Section 3 – proportional representation and

Input : Candidate list C, Preference rankings P, K
Output: Set of K winners (W)
W = Φ ; ▷ Start with an empty winner set
$dq = \lfloor \frac{|P|}{K+1} \rfloor + 1$; ▷ Droop quota
while $|W| < K$ **do**
  using P, assign votes to the first choice candidates ;
  if *a candidate j has votes* $\geq dq$ **then**
    W = W $\cup \{j\}$; ▷ Add j to the winner set
    remove dq voters from P who rank j first ;
    transfer j's surplus votes to the next preference of the corresponding voters ;
    remove j from all voters' preference rankings ;
  else
    eliminate a candidate c with the smallest tally ;
    redistribute c's votes to its voters' next preferences ;
  end
end

Algorithm 1: Single Transferable Vote (STV)

anti-plurality[3], and apply it in the context of crowdsourced top-K recommendations.

4.1 Single Transferable Vote (STV)

STV considers the ranked choices of all voters, and then executes a series of iterations, until it finds K winners. Algorithm 1 presents the pseudocode of the STV procedure. Consider the example in Table 1 with $K = 2$. Here, Droop Quota (dq) is 34; hence there is no winner in the first iteration. Item 2 gets eliminated transferring all 9 votes to Item 1 (assuming Item 1 to be those voters' second choices). In the second iteration, Item 1 wins and transfers excess 5 votes to Item 3 or Item 4 (lets assume Item 4). In the third iteration, Item 3 gets eliminated, transferring all its votes to Item 4. Finally, in the fourth iteration, Item 4 wins resulting in $W = \{$Item 1, Item 4$\}$. The worst case time complexity of STV is $O(n \cdot m \cdot (m-K))$ where there are n voters and m candidates. However, some performance speedup is possible over the vanilla algorithm presented in Algorithm 1.

By transferring votes to the preferred candidates beyond first choice, STV achieves proportional representation, where every selected candidate gets about $\frac{1}{K+1}$ fraction of electorate support [13]. Similarly, for candidates disliked by a majority of the users, unless they are preferred by at least $\frac{1}{K+1}$ fraction of all users, STV will not include them in the winner list, thus satisfying anti-plurality. More importantly, STV has been proved to be resistant to *strategic voting*, where determining a preference that will elect a favored candidate is NP-complete [4]. Thus, STV would make it much harder for malicious actors to manipulate the selection of items for recommendation. For example, consider the case of trending topic selection. Essentially, it would require at least $\frac{n}{K+1}$ compromised or bot accounts (which is a very large number considering n = number of all Twitter users and $K = 10$) to make one topic trending.

[3]For brevity, we skip the comparison of STV with all other electoral methods. Interested readers are referred to [14] to see how STV compares with other methods across different fairness properties.

Although, STV satisfies the fairness properties required in crowdsourced item selection, it considers the ranked choice over all candidates for every user[4], which gives rise to the following two problems that hinder the applicability of STV in recommending items:
(i) A large majority of the users do not vote during an election, and
(ii) Even for the users who participate, it is not possible to get the ranked choice over all candidate items (because they may vote for only a few candidates during an election).
Next, we propose approaches to circumvent these two issues, enabling us to *guarantee equality of voice to everyone* (including silent users) and apply STV for selecting items in top-K recommendations.

4.2 Getting preference rankings of all users

Intuitively, we can think of getting the ranked choices of a user u as determining how interested u is in different candidate items. Then, the problem gets mapped to *inferring user interests in personalized item recommendations*, and there is a large body of works on the same, which can be categorized into two broad classes: content based methods [24] and collaborative filtering [22]. We first attempt to get the personalized ranked choices of Adressa readers, by applying a collaborative filtering approach based on Non-negative Matrix Factorization (NMF), as described next.

Inferring preferences for Adressa readers
As mentioned in Section 2.2, the Adressa dataset contains information about the time different readers spent on different news stories. We first convert this *implicit* feedback to *explicit* ratings by normalizing with respect to both users' reading habits and the length of different articles. If a user u spent $v_{u,i}$ time reading news story i, then we compute the normalized view duration $nv_{u,i}$ as

$$nv_{u,i} = \frac{v_{u,i}}{\mu_i} \tag{1}$$

where μ_i is the average time spent by all users reading story i. Note that this normalization removes the bias of having possibly longer view duration for lengthier articles.

Once $nv_{u,i}$ values are computed for different stories for the user u, we divide them into 5 quantiles and convert them into ratings. For example, the top 20 percentile $nv_{u,i}$ values are converted to rating 5, the next 20 percentile values to rating 4, and so on. We apply this mapping to every user-news interaction and end up with a user-news rating matrix R where $R_{u,i}$ denotes the rating of news story i computed for user u.

Matrix factorization approaches map both users and news stories into a joint latent feature space of Z dimensions such that the interaction between the users and the news stories are modeled as inner products in that space. For example, if the vectors x_u and y_i denote the latent feature vectors for user u and news story i, then the estimated rating $\hat{r}_{u,i}$ for a given user u and a story i is given by the scalar product:

$$\hat{r}_{u,i} = x_u^T \cdot y_i \tag{2}$$

The challenge then is to find the latent feature vectors by observing the existing ratings. Multiple approaches have been proposed to efficiently optimize the following objective [22]

$$min \sum_{(u,i)\in\kappa} (r_{u,i} - \hat{r}_{u,i})^2 + \lambda(||x_u||_2 + ||y_i||_2) \tag{3}$$

[4]Though it is technically possible to apply STV with incomplete preference rankings, STV guarantees strategyproofness only when the preferences are complete [4].

where κ is set of user-news pairs for which the ratings are known.

In this paper, we apply the Non-negative Matrix Factorization approach proposed by Luo *et al.* [25] which solves Equation (3) by using stochastic gradient descent with non-negativity constraints on the feature values. Once we get the feature vectors for different users and news stories, then the ratings can be predicted even for the unread stories. Thereafter, we compute preference ranking for the users based on the predicted and actual ratings (with actual ratings getting precedence and ties being broken randomly).

Inferring preferences for Twitter users

To infer Twitter users' preferences, we considered both content based recommendation and collaborative filtering:
(i) Compute content based similarity between a user u and hashtag h by considering the set of all tweets posted by u and the set of tweets containing h. However, we found that most users do not post enough tweets, and thus we can not accurately compute the content based similarity between a user and the candidate hashtags.
(ii) As there is no explicit rating available, we tried to apply a collaborative filtering based approach to compute personalized ranking using implicit feedback like favoring or retweeting [35]. However, two independence assumptions in such approaches – items are independent of each other and the users act independently – do not hold in the context of Twitter. Hashtags are often related [36], and Twitter users often influence other users. Further, the amount of implicit feedback is very low (in our dataset, only 21% tweets get any retweets, or likes or favorites) and the set of hashtags are constantly changing. Hence, the collaborative filtering approaches could not be applied in this context.

To circumvent these difficulties, we utilize prior works involving topical experts on Twitter [5, 17, 48]. Using the 'List' feature in Twitter, users can create named groups of people they follow. By giving meaningful names to the lists created, they implicitly describe the members of such groups. Ghosh *et al.* [17] gathered these list information from a large number of Twitter users, and identified thousands of topical experts on Twitter, where the topics are very fine-grained. Then, both Bhattacharya *et al.* [5] and Zafar *et al.* [48] utilized these topical experts to infer interest of a particular user as well as topic of a particular hashtag. The basic intuition of [5] is that if a user is following multiple experts in some area, then he is likely to be interested in that area. Similarly, if multiple topical experts are posting some hashtag, then the probability that the hashtag belongs to that topic is very high [48].

Implementing the approaches proposed in [5] and [48], for a user u, we infer an interest vector I_u considering the experts u follows, and similarly, we compute a topic vector T_h for a hashtag h by taking into account the experts tweeting h. Then, for every user u, we normalize the interest topics in I_u such that every entry in I_u lies between 0 and 1, and all entries sum to 1. Similarly, for every hashtag h, we calculate the tf-idf scores over the topics in T_h. We repeat this process for every user and every candidate hashtag during an election. Finally, we compute the preference scores between all users and all candidate hashtags as

$$A = U \times T \times H^T \tag{4}$$

where $A_{n\times m}$ is the User-Hashtag Affinity Matrix with $A_{v,h}$ denoting affinity between user u and hashtag h; $U_{n\times t}$ is the User-Interest Matrix with $U_{u,j}$ representing normalized interest of u in some interest topic j; $T_{t\times t}$ is the Interest-Topic Similarity Matrix, $T_{i,j}$ representing the similarity between two topics i and j (we compute $T_{i,j}$ as the Jaccard Similarity between the set of experts in topic i and j respectively). Finally, $H_{m\times t}$ is the Hashtag-Topic Matrix where $H_{h,j}$ denotes tf-idf of topic j in hashtag h.

Using A computed above, we can get the preference ranking of any user over the candidate hashtags. If a user u participates in an election and votes for tag h, then h is considered as the top choice in u's preference ranking and other ranked positions are shifted accordingly. If a user votes for k hashtags, top k positions are assigned to these k candidates according to their usage frequency.

Accuracy of the preference inference

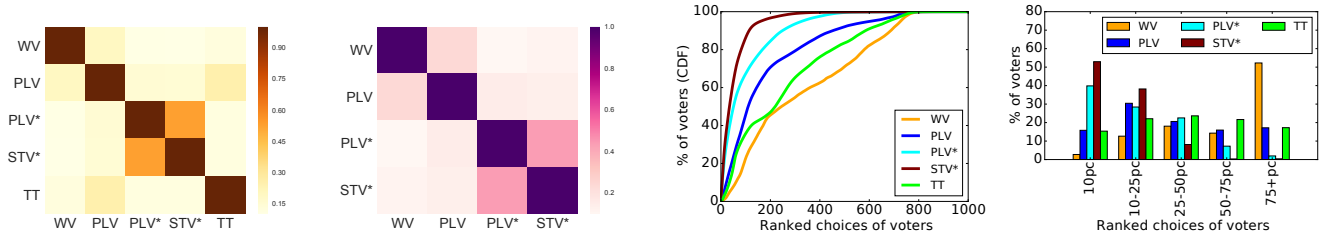
For inferring the preferences of Adressa readers, we attempted another technique based on Singular Value Decomposition (SVD) [33]. Comparing the Root Mean Squared Error (RMSE) between the actual ratings and ratings inferred by both SVD and NMF based approaches, we found that the NMF based approach (RMSE: 0.87) works better than the SVD based approach (RMSE: 0.97).

In Twitter, there is no ground truth rating or ranking. Hence, to check the accuracy of the inference of Twitter users' preference rankings, we asked 10 volunteers (who are active Twitter users) to rank 10 hashtags during 15 election cycles. Then, we compute their preference ranking using our approach and checked Kendall's rank correlation coefficient τ [20] between the inferred and actual rankings for every volunteer. We find the average τ value to be 0.702 (compared to 0.317 for random ordering), which suggests that our method can infer the ranked choices of users reasonably well.

5 EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of our proposed approaches in selecting items for recommendation. For that, in Twitter, we consider every 15 minute intervals throughout the month of July, 2017 as election cycles. During any election, from the large set of available hashtags, we select 1,000 candidate hashtags which experience highest jump in usage during that election cycle (compared to their usage in the previous cycle). While computing the preference rankings of Twitter users, due to Twitter API rate limits, it is not possible to infer ranked choice for everyone. We take 2% random sample from the 15 million Twitter users in our dataset (resulting in a large sample of $300K$ users)[5], and gather the ranked choices of all of them (and of no other) over the 1,000 candidates. For the Adressa dataset, we consider every day during February and March, 2017 as election cycles. We select as candidates top 1,000 stories based on the number of users clicking on them. Then, we compute the preference rankings of all users in our dataset over the candidates. After getting the preference rankings of the users, we apply two methods:
(i) Consider the preference rankings, and select K items which are the first choice for most users. We denote this method as PLV*, because this is an extension of Plurality Voting described next.
(ii) Run STV using the preference rankings and select the winners

[5]The idea of 'random voting' is not new. Getting everyone to vote in an election is often impractical or too costly. Dating back to ancient Athenian democracy, philosophers including Aristotle argued for selecting a large random sample of voters and then mandating them to vote during an election [19]. More recently, Chaum [10] proposed a technique to hold random elections. Fishkin *et al.* [16] proposed an alternate 'Deliberative Polling', where the idea is to select a random sample of voters, give them proper information, ask them to discuss issues and then consider only their votes.



(a) (b) (c) (d)

Figure 2: Heatmaps depicting the Jaccard Coefficients between different methods for selecting (a) trending topics in Twitter, and (b) most popular news in Adressa. (c) Average ranked choices and (d) Percentile choices for the trending topics selected by different methods throughout July, 2017.

of the election. Following the convention used for PLV*, in this section, we denote STV as STV* to reflect the fact that the ranked choices of everyone have been considered, not only the active users. Next, we describe the baselines we compare PLV* and STV* against.

5.1 Baseline approaches

In addition to the preference rankings, we also gather the votes given by the users participating in an election cycle. Then, using the data, we apply the following approaches to select the winners.

i. Weighted Voting (WV) : Here, K candidates getting maximum votes win the election regardless of who voted for them and how many times one user voted. Hence, it is vulnerable to manipulation by hyper-active users.

ii. Plurality Voting (PLV) : Plurality Voting[6], or Single Non-Transferable Vote (SNTV), considers only one vote from a participating user. So, if a particular user voted multiple times, we count only one vote for the candidate she voted the most (with randomly breaking ties). Then, K candidates with maximum votes win.

iii. Twitter Trending Topics (TT) : We are not privy to the exact algorithm Twitter employs to select the set of trending topics during an election cycle. Therefore, we consider the algorithm as black-box and compare the hashtags selected by the methods with the hashtags declared as trending by Twitter.

5.2 Quantifying pairwise overlaps

We first investigate whether these different methods pick very different items or they end up selecting same items during an election. To check that, we gather all the items (i.e., hashtags and news stories) selected by each of the methods, and then compute pairwise overlaps between them. Figure 2(a) shows the heatmap of Jaccard coefficient between different methods, where Jaccard coefficient between methods i and j is measured as $\frac{|S_i \cap S_j|}{|S_i \cup S_j|}$, where S_i is the set of items selected by method i throughout all election cycles.

We see from Figure 2(a) that there is 50% overlap between the trending hashtags selected by PLV* and STV*. TT has around 35% overlap with PLV. There is little overlap between hashtags selected by other methods. Similarly, for Adressa dataset (Figure 2(b)), we see around 45% overlap between the news stories selected by PLV* and STV*. The rest of the methods do not have much overlap.

Takeaway: Different methods select mostly different items during election cycles. We only see some common items being selected by our two proposed approaches: PLV* and STV*, possibly because both consider the top preferences of all users. Interestingly, actual Twitter Trending Topics (TT) has the highest overlap with the tags selected by Plurality Voting (PLV). Thus, the Twitter algorithm can be conceptualized as running plurality-based elections.

[6]Not to be confused with 'Plural Voting', which is a variant of 'Weighted Voting'.

5.3 Comparing ranked choices of users

Different users have different interests, and thus their ranked choices over different candidate items can vary considerably. We now investigate how different election methods capture the choices of these users. Figure 2(c) shows, on average, how the hashtags selected by different methods represent different ranked choices of Twitter users. Figure 2(d) presents the user choices in different percentile bins. We can observe in both Figure 2(c) and Figure 2(d) that the STV* selected tags correspond to top 10 percentile choices for a majority of users. While PLV* captures users' top choices to some extent, both PLV and TT appeal very differently to different voters. Finally, WV tends to pick tags which represent top choices of only a few voters and bottom choices for a majority of the voters.

Takeaway: STV* consistently selects tags that are the top preferences for all voters; whereas other methods capture both top and bottom preferences, with WV performing the worst. Few actual trends selected by Twitter are least preferred by a lot of voters. We see similar result for the Adressa news data as well.

5.4 Comparing desirable fairness properties

We now compare different methods along the desirable fairness properties identified in Section 3. WV does not satisfy 'Equality of Voice' because effectively a voter voting multiple times exerts more power than a voter voting once during an election. PLV considers one vote per participating voter; however, it does not consider votes of the silent users. Our proposed PLV* and STV* both guarantee voice equality by giving all users an equal chance to participate in the item selection process. Regarding the other two properties, we empirically observe to what extent the methods satisfy them.

5.4.1 User Satisfaction Index.

The proportional representation criterion requires that if a candidate is preferred by $\frac{1}{K+1}$ fraction of the users, it should be selected, and only STV* theoretically satisfies this criterion. An alternate way to consider representation is from users' perspective. We propose a user satisfaction criterion which requires that every user should have at least one elected candidate from her top choices. Formally, we consider a user to be satisfied if at least one of its top 10 choices is selected by a method during an election. Then, *User*

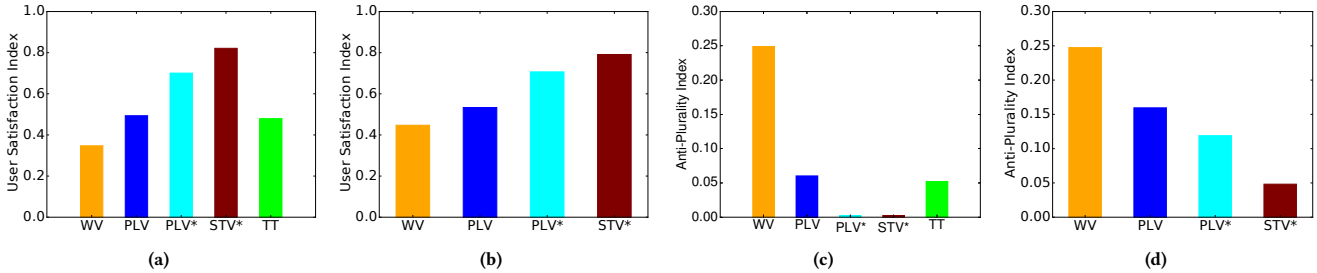


Figure 3: User Satisfaction Index of different methods for computing (a) Twitter Trends, and (b) Most Popular Adressa News. Anti-Plurality Index of different methods for computing (c) Twitter Trends, and (d) Most Popular Adressa News.

Satisfaction Index is measured as the fraction of users who are satisfied by a method. Figure 3(a) shows the average User Satisfaction Index for different methods to compute Twitter trends, and we see that both PLV* and STV* are able to satisfy more than 70% of the users; whereas, the other methods cannot satisfy even 50% users. We see similar results for Adressa news dataset as well (Figure 3(b)).

5.4.2 **Anti-plurality Index.**
The notion of anti-plurality captures whether a method selects items that are disliked by most of the users. We consider a item i to be disliked by a user u if t appears among v's bottom 10 percentile choices. Then for every such i, we compute what percentage of users dislike i and aggregate this over all the items selected by different methods. Figure 3(c) and Figure 3(d) shows the average Anti-plurality Index for all methods of selecting Twitter Trends and Most Popular Adressa News. We can see in Figure 3(c) that both STV* and PLV* select almost no tags which are disliked by any of the users. On the other hand, WV picks tags which, on average, are disliked by 25% users. For both PLV and TT, the selected tags are disliked by around 5% of all users. Similarly, we can see in Figure 3(d) that STV* has the lowest anti-plurality value (less than 5%) while stories selected by WV are disliked by 25% of users.

Specific to Twitter, we observe that there were some extremist tags (e.g., #exterminate_syrians, #IslamIsTheProblem), spammy tags (e.g., #houstonfollowtrain, #InternationalEscorts) or politically biased tags (e.g., #fakeNewsCNN, #IdiotTrump) which were disliked by more than 90% users, yet got selected by WV or PLV due to the presence of some hyper-active user groups. However, STV* and PLV* did not select any of such hashtags.

5.5 Demographic bias and under-representation in selected topics

In our earlier work [9], we found that most of the Twitter trends are promoted by users whose demographics vary significantly from Twitter's overall population. Next, we check whether the voting methods considered in this paper amplify or reduce these demographic biases. We use the demographic information of Twitter users as obtained in [9]. Then, demographic bias of tag i is computed as the euclidean distance between the demographics d_i of the people tweeting on i and the reference demographics d_r of the Twitter population in the US [9]: $Bias_i = ||d_i - d_r||$. The higher the score $Bias_i$, more biased are the users using the tag i.

Figure 4(a) shows the average bias across the tags selected by different methods throughout all election cycles. We see in Figure 4(a) that the tags selected by WV are most gender, racially and age biased. On the other hand, STV* selects tags that are least biased. We further observe that considering the preferences of the silent users helps reducing the bias as the average bias of tags selected by PLV* is lower than the average bias of PLV selected tags.

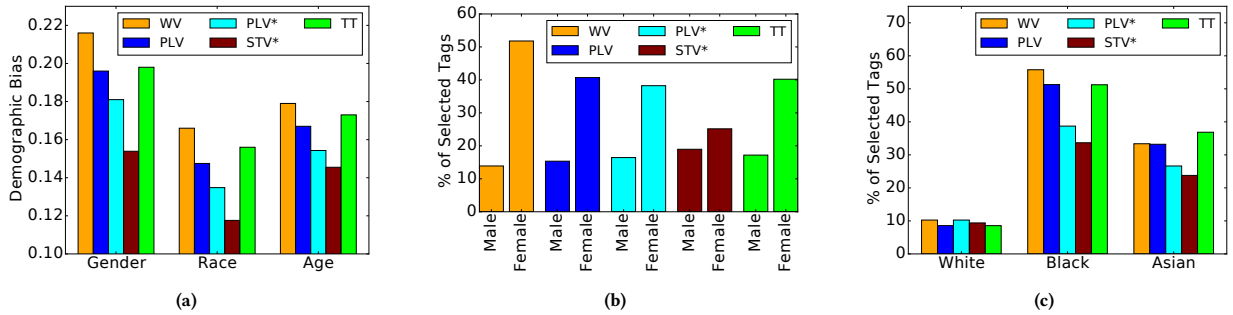
We next consider the under-representation of different socially salient groups among the users of the tags selected by different methods (where we consider a group i to be under-represented if the fraction of i among the trend users is $< 80\%$ of the fraction of i in the overall population [9]). Figure 4(b) shows the under-representation of men and women. In almost all the methods, women are under-represented for over 40% of the selected tags; whereas, men are under-represented for only around 15% of the tags. However, in the tags selected by STV*, although under-representation of men slightly increases, under-representation of women greatly reduces, having almost equal under-representation of both gender groups.

Figure 4(c) shows the under-representation of different racial groups: Whites, Blacks and Asians. Even though none of the methods achieve similar under-representation of all three racial groups, STV* reduces the under-representation of Blacks and Asians considerably, while keeping the under-representation of Whites similar to other methods. We observe similar trends for age groups where under-representation of Mid-Aged and Adolescents decrease in the tags selected by STV*. The detailed result is omitted for brevity.

How does considering preference rankings reduce demographic bias?
The reduction in demographic bias and under-representation of different social groups among STV* selected tags is surprising because the method has not explicitly taken into account the preference rankings of voters belonging to different demographic groups. We investigate the reason by considering the 100 most and 100 least biased tags along all three demographic dimensions – gender, race and age, and then by checking how they rank in different voters' preference rankings. Figure 5 clearly shows that highly biased tags rank low in most of the voter choices. On the other hand, tags with low bias tend to be ranked higher by most of the voters. This interesting observation explains why methods like PLV* or STV* which relies on preference rankings of all the voters tend to select tags with low bias as compared to other methods like WV or TT which only consider votes by the active users.

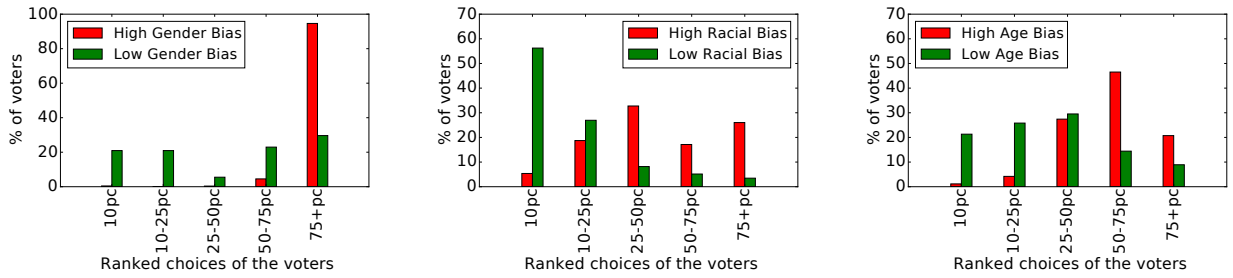
6 RELATED WORKS

In this section, we briefly review the related works along two dimensions: top-K item recommendations, and fairness in algorithmic decision making systems.



(a) (b) (c)

Figure 4: (a) Demographic bias, (b) Gender and (c) Racial under-representation in tags selected by different methods.



(a) (b) (c)

Figure 5: Ranked choices of voters for 100 hashtags most biased and least biased along (a) gender, (b) race and (c) age.

Top-K recommendations: Top-K item recommendations is traditionally associated with personalized recommendation which attempts to find K items a particular user would be mostly interested in [11]. In content-based recommendations, a user profile is generated based on what she likes or dislikes, and then similar content is identified depending on her past likes [34]. In collaborative filtering, the preference of a particular user can be inferred based on their similarity to other users [22]. However, the recommendation scenario we are considering here is *non-personalized*, where the same K items are recommended to everyone. In fact, the problem we are focusing on is how to *fairly aggregate personalized preferences of all users* of a website.

Bringing fairness in algorithmic decisions: Multiple recent works have focused on biases and unfairness in algorithmic decision making [7, 39, 49]. Yao *et al.* [47] proposed a few fairness notions for personalized recommendations. Zehlike *et al.* [50] introduced fairness in top-k ranking problem through utility based multi-objective formulation. Burke [6] and Chakraborty *et al.* [8] argued for preserving fairness of consumers (users) as well as suppliers (item producers) in two-sided matching markets. Complementary to earlier efforts, in this paper, we present the notions of fairness in crowdsourced non-personalized recommendations, and utilize electoral mechanisms to satisfy them in practice.

7 CONCLUSION AND FUTURE DIRECTIONS

Recently, there has been a lot of debate and concerns regarding the bias in algorithms operating over big crowd-sourced data. In this paper, by conceptualizing crowdsourced recommendation as a multi-winner election, we showed that the bias originates from the unfairness in the electoral process. Then, utilizing long lines of works in social choice theory, we established the fairness properties desired in crowdsourced selections, and identified a particular mechanism STV which satisfies most of these properties. As a result, extensive evaluation over two real-world datasets shows that STV can reduce unfairness and bias in crowdsourced recommendations. Moreover, STV can also resist strategic manipulation by requiring a lot of user support behind potential candidates for recommendation, thereby making it difficult for spammers, bots, or trend hijackers to influence the recommendation process.

There are multiple research directions we want to explore in future. First, our proposed approach can potentially be applied in *personalized news recommendation* scenario which combine both user choices and the news trends among the crowds (e.g., Google News [23]). In such context, at the first level, the candidate stories for recommendation can be selected by standard personalized recommendation algorithms which consider a particular user's interest. Then, an election method like STV can be applied to take into account the crowd choices for electing news stories to recommend to the user. Second, in this work, we conceptualized item (s)election to happen at every fixed intervals; however, there is a streaming component in recommendations like Trending Topics [26] (with occasional burstiness in user activities [21]). Regular election methods are not designed to tackle such scenarios, and we plan to develop mechanisms to handle continuous elections, while simultaneously satisfying the desired fairness properties.

Acknowledgments: This research was supported in part by a European Research Council (ERC) Advanced Grant for the project

"Foundations for Fair Social Computing", funded under the European Union's Horizon 2020 Framework Programme (grant agreement no. 789373). P. Loiseau was supported by the French National Research Agency through the "Investissements d'avenir" program (ANR-15-IDEX-02) and the Alexander von Humboldt Foundation. A. Chakraborty was a recipient of Google India PhD Fellowship and Prime Minister's Fellowship Scheme for Doctoral Research, a public-private partnership between Science & Engineering Research Board (SERB), Department of Science & Technology, Government of India and Confederation of Indian Industry (CII).

REFERENCES

[1] Kenneth J Arrow. 1950. A difficulty in the concept of social welfare. *Journal of political economy* 58, 4 (1950).
[2] Ricardo Baeza-Yates. 2016. Data and algorithmic bias in the web. In *WebScience*.
[3] Paul Baker and Amanda Potts. 2013. 'Why do white people have thin lips?' Google and the perpetuation of stereotypes via auto-complete search forms. *Critical Discourse Studies* 10, 2 (2013).
[4] John J Bartholdi and James B Orlin. 1991. Single transferable vote resists strategic voting. *Social Choice and Welfare* 8, 4 (1991).
[5] Parantapa Bhattacharya, Muhammad Bilal Zafar, Niloy Ganguly, Saptarshi Ghosh, and Krishna P Gummadi. 2014. Inferring user interests in the twitter social network. In *ACM RecSys*.
[6] Robin Burke. 2017. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093* (2017).
[7] Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2015. Can trending news stories create coverage bias? on the impact of high content churn in online news media. In *Computation and Journalism Symposium*.
[8] Abhijnan Chakraborty, Aniko Hannak, Asia J Biega, and Krishna P Gummadi. 2017. Fair Sharing for Sharing Economy Platforms. (2017).
[9] Abhijnan Chakraborty, Johnnatan Messias, Fabricio Benevenuto, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2017. Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations. In *AAAI ICWSM*.
[10] David Chaum. 2016. Random-sample voting. (2016).
[11] Mukund Deshpande and George Karypis. 2004. Item-based top-n recommendation algorithms. *ACM TOIS* 22, 1 (2004).
[12] Henry Richmond Droop. 1881. On methods of electing representatives. *Journal of the Statistical Society of London* 44, 2 (1881).
[13] Michael Dummett. 1984. *Voting procedures*.
[14] Edith Elkind, Piotr Faliszewski, Piotr Skowron, and Arkadii Slinko. 2017. Properties of multiwinner voting rules. *Social Choice and Welfare* 48, 3 (2017).
[15] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. 2017. Multiwinner voting: A new challenge for social choice theory. *Trends in Computational Social Choice* (2017).
[16] James S Fishkin, Max Senges, Eileen Donahoe, Larry Diamond, and Alice Siu. 2017. Deliberative polling for multistakeholder internet governance: considered judgments on access for the next billion. *Information, Comm. & Society* (2017).
[17] Saptarshi Ghosh, Naveen Sharma, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi. 2012. Cognos: crowdsourcing search for topic experts in microblogs. In *ACM SIGIR*.
[18] Jon Atle Gulla, Lemei Zhang, Peng Liu, Özlem Özgöbek, and Xiaomeng Su. 2017. The Adressa dataset for news recommendation. In *ACM WI*.
[19] Mogens Herman Hansen. 1991. *The Athenian democracy in the age of Demosthenes: structure, principles, and ideology*.
[20] Maurice G Kendall. 1955. Rank correlation methods. (1955).
[21] Jon Kleinberg. 2003. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery* 7, 4 (2003).
[22] Yehuda Koren and Robert Bell. 2015. Advances in collaborative filtering. In *Recommender systems handbook*. Springer.
[23] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized news recommendation based on click behavior. In *ACM IUI*.
[24] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2011. Content-based recommender systems: State of the art and trends. In *Rec. Sys. handbook*.
[25] Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu. 2014. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics* 10, 2 (2014).
[26] Michael Mathioudakis and Nick Koudas. 2010. Twittermonitor: trend detection over the twitter stream. In *ACM SIGMOD*. ACM.
[27] Molly K. Mckew. 2018. How Twitter Bots and Trump Fans Made #ReleaseTheMemo Go Viral. https://www.politico.com/magazine/story/2018/02/04/trump-twitter-russians-release-the-memo-216935. (2018).
[28] J Joseph Miller. 2003. JS Mill on plural voting, competence and participation. *History of political thought* 24, 4 (2003), 647–667.
[29] Facebook Newsroom. 2016. Search FYI: An Update to Trending. https://newsroom.fb.com/news/2016/08/search-fyi-an-update-to-trending/. (2016).
[30] Facebook Newsroom. 2018. News Feed FYI. https://newsroom.fb.com/news/category/news-feed-fyi. (2018).
[31] Michael Nunez. 2016. Former Facebook Workers: We Routinely Suppressed Conservative News. https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006. (2016).
[32] Abby Ohlheiser. 2016. Three days after removing human editors, Facebook is already trending fake news. https://www.washingtonpost.com/news/the-intersect/wp/2016/08/29/a-fake-headline-about-megyn-kelly-was-trending-on-facebook. (2016).
[33] Arkadiusz Paterek. 2007. Improving regularized singular value decomposition for collaborative filtering. In *ACM SIGKDD*.
[34] Michael J Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The adaptive web*. Springer.
[35] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *IUAI*.
[36] Bidisha Samanta, Abir De, Abhijnan Chakraborty, and Niloy Ganguly. 2017. LMPP: a large margin point process combining reinforcement and competition for modeling hashtag popularity. In *IJCAI*.
[37] Piotr Skowron, Piotr Faliszewski, and Arkadii Slinko. 2016. Axiomatic characterization of committee scoring rules. *arXiv preprint arXiv:1604.01529* (2016).
[38] Olivia Solon and Sam Levin. 2016. How Google's search algorithm spreads false information with a rightwing bias. theguardian.com/technology/2016/dec/16/google-autocomplete-rightwing-bias-algorithm-political-propaganda. (2016).
[39] Till Speicher, Hoda Heidari, Nina Grgic-Hlaca, Krishna P Gummadi, Adish Singla, Adrian Weller, and Muhammad Bilal Zafar. 2018. A Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual & Group Unfairness via Inequality Indices. In *ACM SIGKDD*.
[40] Grant Stafford and Louis Lei Yu. 2013. An evaluation of the effect of spam on twitter trending topics. In *IEEE SocialComm*.
[41] Alan Taylor and William Zwicker. 1992. A characterization of weighted voting. *Proc. of the American mathematical society* 115, 4 (1992).
[42] Twitter. 2010. To Trend or Not to Trend. https://blog.twitter.com/official/en_us/a/2010/to-trend-or-not-to-trend.html. (2010).
[43] Twitter. 2018. Filtering Tweets by location. https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location. (2018).
[44] Twitter. 2018. Get trends near a location. https://developer.twitter.com/en/docs/trends/trends-for-location/api-reference/get-trends-place. (2018).
[45] Courtland VanDam and Pang-Ning Tan. 2016. Detecting hashtag hijacking from twitter. In *ACM WebScience*.
[46] Gerhard J Woeginger. 2003. A note on scoring rules that respect majority in choice and elimination. *Mathematical Social Sciences* 46, 3 (2003).
[47] Sirui Yao and Bert Huang. 2017. Beyond parity: Fairness objectives for collaborative filtering. In *NIPS*.
[48] Muhammad Bilal Zafar, Parantapa Bhattacharya, Niloy Ganguly, Saptarshi Ghosh, and Krishna P Gummadi. 2016. On the wisdom of experts vs. crowds: discovering trustworthy topical news in microblogs. In *ACM CSCW*.
[49] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. 2017. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *WWW*.
[50] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. Fa* ir: A fair top-k ranking algorithm. In *ACM CIKM*.