Beyond Trade-offs: Unveiling Fairness-Constrained Diversity in News Recommender Systems

Célina Treuillier
celina.treuillier@loria.fr
University of Lorraine - CNRS - LORIA
Vandoeuvre-lès-Nancy, FRANCE

Sylvain Castagnos
sylvain.castagnos@loria.fr
University of Lorraine - CNRS - LORIA
Vandoeuvre-lès-Nancy, FRANCE

Özlem Özgöbek
ozlem.ozgobek@ntnu.no
Dep. of Computer Science, NTNU
Trondheim, NORWAY

Armelle Brun
armelle.brun@loria.fr
University of Lorraine - CNRS - LORIA
Vandoeuvre-lès-Nancy, FRANCE

ABSTRACT

Recommender Systems have played an important role in our daily lives for many years. However, it is only recently that their social impact has raised ethical issues and has thus been considered in the design of such systems. Particularly, News Recommender Systems (NRS) have a critical influence on individuals. NRS can provide overspecialized recommendations and enclose users into filter bubbles. Besides, NRS can influence users and make their original opinions diverge. Worse, they can orient users' opinions towards more radical views. The literature has worked on these issues by leveraging diversity and fairness in the recommendation algorithms, but generally only one of these dimensions at a time. We propose to consider both diversity and fairness simultaneously to provide recommendations that are fair, diverse, and obviously accurate. To this end, we propose a novel recommendation framework, *Accuracy-Diversity-Fairness* (ADF), which considers that fairness is not at the expense of diversity. Concretely, fairness is approached as a constraint on diversity. Experiments highlight that constraining diversity by fairness remarkably contributes to providing recommendations 5 times more diverse than models of the literature, without any loss in accuracy.

CCS CONCEPTS

• **Information systems → Recommender systems**; **Personalization**.

KEYWORDS

News Recommender Systems, Personalization, Diversity, Fairness, Calibration

ACM Reference Format:
Célina Treuillier, Sylvain Castagnos, Özlem Özgöbek, and Armelle Brun. 2024. Beyond Trade-offs: Unveiling Fairness-Constrained Diversity in News Recommender Systems. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '24), July 01–04, 2024, Cagliari, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3627043.3659571

UMAP '24, July 01–04, 2024, Cagliari, Italy

ACM ISBN 979-8-4007-0433-8/24/07
https://doi.org/10.1145/3627043.3659571

1 INTRODUCTION

Recently, the impact of RS on our societies and the ethical issues such as overspecialization and fairness have dominated the debates. Overspecialization generates poorly diversified recommendations, likely to impact both users and items providers. In News Recommender Systems (NRS), this results in filter bubbles and user political polarization [14]. Besides, fairness has been attracting attention for some years. It commonly means avoiding discrimination of certain groups, such as ethnic, gender, or demographic groups [9].

In the news domain, these challenges are of the utmost importance. Indeed, NRS should inform users about the existence of opinions that they are not familiar with (diversity), without favoring some of them (fairness). Through these considerations, a double challenge has emerged in the literature: finding a good balance (1) accuracy-diversity on the one hand, and (2) accuracy-fairness on the other hand.

As regards the first challenge, it is reasonable to seek to increase the diversity of news recommended when confronted with a lack of diversity of consumed news (*a.k.a.* selection diversity) [15]. Besides, as regards the second challenge, some RS can be biased in favor of a specific topic(s), or opinion (in NRS), so the recommendations can influence users towards an opinion that is not originally theirs [28]. By lowering the divergence between a user's profile and the recommendation list [8], this influence can be limited and fairness can thus be ensured [23].

In this paper, we aim to reunify these two challenges. We argue that, beyond the multi-objective problem defined to find the optimal trade-off between accuracy-diversity and accuracy-fairness, arises the relation between diversity and fairness. Bringing opinion diversity in response to a polarized behavior, if not correctly managed, may lead to an overcompensation with recommendations from opposite and specific political ideologies [16, 20]. We thus formulate **RQ1: How to diversify news recommendations in a way that meets fairness?**

To answer this question, we propose the *Accuracy-Diversity-Fairness* (ADF) framework, designed to achieve a good balance between accuracy, diversity, and fairness. The objective of ADF is to perform a fair diversification, combined with a new re-ranking strategy to ensure accuracy. Concretely in ADF, fairness is viewed as a constraint on diversity. We thus raise a second research question **RQ2: How does fairness-constrained diversity affect accuracy?**

2 RELATED WORK

It is now obvious that RS need to address multiple objectives [13], going further the accuracy optimization.They can be addressed in a re-ranking strategy, where a trade-off between accuracy and a beyond-accuracy dimension is applied. In RS, applying re-ranking requires substantial computational costs. That is why heuristics and greedy approaches are often applied [6]. Greedy re-ranking relies on an linear combination between relevance and another dimension in the recommendation list. Among these dimensions, we find diversity and fairness.

One the one hand, diversity has been recognized as a fundamental quality of RS [7]. This has given rise to a new generation of diversity-based RS, bi-objective systems that seek to optimize an accuracy-diversity trade-off [22]. Regarding NRS, although widely discussed, the role of diversity in NRS regarding polarization decrease remains disputed [1, 12]. However, researchers agree on the fact that the diversification process should be personalized to individual users' needs [30].

One the other hand, fairness has recently gained renewed attention [9]. Common approaches includ constrained optimization [16], where fairness guides the model's optimization and can be linked to other dimensions, *e.g.* accuracy [11]. Recently, calibration has gained attention in the context of fairness in recommender systems [23]. Considering NRS and depolarization, it is crucial to ensure that users are not manipulated through recommendations. To remain fair, we propose to rely on Steck's definition [23], as recommendations should not overcompensate items that are shunned by users on the pretext of diversification [3].

NRS designers should thus pay attention to bring some diversity, without steering users towards opinions that are too far removed from their own opinion. At the opposite, NRS designers' goal is to make users aware of what exists, without manipulating them. This is where a combination between diversity and fairness becomes crucial.

To summarize, although greedy re-ranking strategies have proven their efficiency with two factors, but few studies optimize 3 factors or more. Besides, their myopic approach does not allow to finely control the nature of diversity or fairness input [21]. Finally, to the best of our knowledge, multi-objective recommender systems optimizing all accuracy, diversity, and fairness at once are lacking in the literature.

3 THE ADF FRAMEWORK

ADF is a novel news recommendation framework, intended to provide fair, diverse, and accurate recommendations. ADF is thus designed to promote users' awareness, through diversity, while not orienting users' main interests or opinions, through fairness. ADF is the first framework that explicitly addresses diversity and fairness simultaneously.

In line with [26], *diversity* of a list of news is evaluated on a set of aspects A (*i.e.* features, genres, etc. of the news) and is defined as *the extent to which the set of aspects is diversely represented in the list.* Besides, in line with [23], *fairness* is defined as *the ability to reflect the various interests of a user, according to their corresponding proportions.* ADF considers that a recommendation list is fair *w.r.t.* a user u, if its distribution on the set of aspects A is compatible with the distribution of u's interests on A.

The input of ADF is threefold. (1) The set of news each user $u \in U$ has selected. (2) The set of u's unseen news, associated with personalized relevance scores provided by a recommender system $R(u) = \{n, s(u, n)\}$, with $n \in N$. The relevance score $s(u, n)$ represents the extent to which news n fits u's interest. This input is dedicated to the accuracy dimension. ADF is agnostic to the recommendation algorithm used, so any algorithm from the literature can be used, whether it is content-based or collaborative. (3) The set of aspects A and the representation of news in the space of aspects. Concretely, ADF runs in 5 steps that are detailed in the ramaining of this section.

Step 1: Building u's profile . This first step is dedicated to user profiling, *i.e.* the representation of u's interests over the set of aspects A. Interests are evaluated from u's interactions on news. Concretely, u's interest for a given aspect $a \in A$ is formulated as a probability $p(a|u)$. A user profile is thus a probability distribution $P(u)$ over the set of aspects, with $\sum_{a \in A} p(a|u) = 1$. It is termed the *selection distribution.* In ADF, u's *selection distribution* is not only used for fairness purposes, as proposed by the literature, but also for diversity purposes.

Step 2: Evaluating u's selection diversity. Given u's *selection distribution* $P(u)$ (from Step 1), the goal here is to evaluate the diversity of u's interests $diversity(P(u))$, *i.e.* the diversity associated with this distribution. Several measures from the literature can be used to instantiate this diversity: polarization score [2], entropy [24], etc. The framework supposes that, as for any diversity measure, $\forall u \in U,\ diversity(P(u)) \in [0; 1]$.

Step 3: Estimating u's personalized target diversity. The goal of this step is to estimate this personalized level of diversity, termed the *target diversity.* We propose to promote u's personalized *target diversity* by exploiting u's *selection diversity.* We thus define the *target diversity* as a function f of the *selection diversity* $f(diversity(P(u)))$.

As we aim to provide diversified recommendations to promote u's awareness of the set of aspects A, personalized *target diversity* should be at least equal, but more importantly higher than *selection diversity.* Considering $f()$, we thus expect two main characteristics: (1) the *target diversity* of a user can not be lower than her *selection diversity.* Concretely, this means that users will not receive recommendations with a diversity lower than their own *selection diversity.* Thus, $\forall x, x \in [0; 1], f(x) >= x$, (2) if a user u_1 has a *selection diversity* higher than a user u_2, the *target diversity* of u_1 will be higher or equal to the *target diversity* of u_2. Concretely, f has to be increasing. Thus, $diversity(P(u_1)) > diversity(P(u_2)) \Rightarrow f(diversity(P(u_1))) >= f(diversity(P(u_2)))$. We propose the following general formulation for f, that fits both characteristic: $f(x) = \beta + (1 - \beta)x^{1-\alpha}$, with $0 <= \alpha, \beta <= 1$. β represents the minimal *target diversity* and α represents the steepness and the slope of the curve. The higher α, the higher the *target diversity*. When $\alpha = 0$, $f()$ is linear, and the increase in diversity is constant. In this case, if $\beta = 0$, it comes down to $f(x) = x$. At the opposite, when $\alpha = 1$, the *target diversity* is maximal (equal to 1), and is independent of the *selection diversity*. In this case, the value of β has no impact on the *target diversity*.

Step 4: Determining u's fair target distribution. Given u's *target diversity* (Step 3), the goal of this step is to determine u's associated *target distribution*, $P^*(u) = \{p^*(a|u)\}, a \in A$, *i.e.* whose diversity is equal to this *target diversity*. The solution to this question is not unique. Indeed, different probability distributions can have the same diversity, whatever the way diversity is evaluated. However, this *target distribution* has to be compatible with u's *selection distribution* $P(u)$, to make the *target distribution* fair. This constrains the *target distribution* and makes it unique.

The originality of our objective is twofold. First, it lies in the problem definition. As previously mentioned, to the best of our knowledge, no work has explicitly focused on managing both diversity and fairness simultaneously. Second, it lies in the way the trade-off between both dimensions is defined. Recall that the literature generally defines the trade-off (accuracy-fairness and accuracy-diversity) as a simple bi-objective function [23], that does not guarantee that fairness is met by the resulting distribution. By contrast, we define this trade-off by considering the *selection distribution* as a constraint on the *target distribution*. Concretely, this is a compatibility constraint, used to guarantee the fairness of the *target distribution*.

This compatibility constraint can not be simply defined as the equivalence between both distributions, as it would not allow to estimate a distribution with a higher *target diversity*. Compatibility has to allow differences between both distributions. We propose to instantiate this compatibility by exploiting the order relationships of the distributions. Concretely, we will consider that two probability distributions are compatible if their order relationships are equal, *i.e.* the order between elements is the same between both distributions.

For now, let us put aside the *target diversity*, and focus on how to transform a distribution, $P(u)$, into a compatible distribution $P^*(u)$, *i.e.* that remains fair. Let $smooth()$ be this transformation function, presented in Equation (1).

$$smooth(u, \delta) = (1 - \delta)P(u) + \delta\frac{1}{|A|} \tag{1}$$

where $\frac{1}{|A|}$ is the uniform distribution and $0 \leq \delta \leq 1$ represents its weight (in other words, the strength of the smoothing). This function is applied for each $a \in A$, where $P(u)$ is instantiated by $p(a|u)$.
Concretely, $smooth()$ computes a trade-off between the selection probability distribution ($P(u)$) and the uniform probability distribution ($\frac{1}{|A|}$). As the same trade-off is applied to every element of the probability distribution, the distribution is simply smoothed and the resulting distribution is thus fair. If $\delta = 0$, no smoothing is performed. At the opposite, if $\delta = 1$ the resulting distribution is the uniform probability distribution.

With the $smooth()$ function in mind, let us now go back to the *target diversity* estimated in Step 3. The objective here is to determine, for a given user u, the δ value that smooths the *selection distribution* $P(u)$ so that the diversity associated with the smoothed distribution approximates the *target diversity*. This problem is thus a mono-objective optimization problem, presented in Equation (2). As a result, with δ^* the optimal δ value, $smooth(P(u), \delta^*)$ corresponds to the *target distribution*, *i.e.* $P^*(u)$.

$$\delta^* = \underset{\delta}{\operatorname{argmin}}\ |f(diversity(P(u))) - diversity(smooth(P(u), \delta))| \tag{2}$$

Step 5: Re-ranking u's recommendations. Given the *target distribution* $P^*(u)$, re-ranking is performed to maximize the accuracy of recommendations. It consists in re-ordering $R(u)$, *i.e.* the input list of news associated with u's personalized relevance score, into a final recommendation list $R^*_k(u)$ of length k. Re-ranking strategies from the literature are mainly based on a greedy approach [27, 33]. In our view, such a greedy approach can not guarantee that fairness remains fulfilled, which conflicts with our objectives. We thus re-rank the recommendations, in a way that ensures that the resulting recommendation list $R^*_k(u)$ complies with the fairness constraint, while maximizing accuracy. Fairness acts as a constraint and the distribution over aspects of $R^*_k(u)$ is intended to be equal to the *target distribution* $P^*(u)$.

The re-ranking strategy proposed here has three inputs: (1) $R(u)$ the list of news with relevance score, (2) $P^*(u)$ the personalized *target distribution*, (3) k the size of the expected recommendation list.
The proposed re-ranking first computes the expected number of news per aspect $a \in A$ that the final recommendation list R^*_k has to contain so that it fits the *target distribution* ($P^*(u)$). Given an aspect a, this number is $l_a = round(p^*(a|u) \cdot k)$. Once l_a is defined, to maximize accuracy, re-ranking selects the associated top-l_a news in R, that are labeled with aspect a. The output of this re-ranking strategy is thus the final recommendation list R^*_k, made up of top-l_a news for each aspect a, and whose distribution is as close as possible to the personalized *target distribution* $P^*(u)$.

4 DATA AND EXPERIMENTAL SETUP

4.1 Data

We evaluate ADF on the real world benchmark MIND dataset [29]. MIND gathers 24M click behaviors from about 1M users, interacting with 160k English news from the Microsoft News website between October 12 and November 22, 2019. The news is categorized into 20 categories, we select a subset of news that deals with the "news" category which is the closest category to the political domain, useful when focusing on political polarization. To rule out extreme behaviors, we selected 10k users having between 20 and 200 interactions, resulting in 18,186 distinct news.

4.2 Recommendation

Any recommendation algorithm can be used for the recommendation process upstream of the ADF framework. We adopt a content-based approach, which is the most widely used by NRS [19]. We choose the recent ClayRS library [17]. As a recommendation algorithm, we select the *CentroidVector* algorithm, which provides satisfactory performance, is easily interpretable, scalable to large datasets, and is computationally efficient [31]. To process the content of each news, we retrieve the full content of news by processing provided URLs. Then we apply the popular Latent Dirichlet Allocation (LDA), an unsupervised approach to represent news [4] For the recommendation process, the dataset is temporarily split per user with 75% used for the training set, and the remaining 25% used for the test set.

4.3 Aspects definition

Working on a news dataset, we choose to define the set of aspects A as news topics. Thus from now on, we will refer to aspects as topics. The news in the MIND dataset are binary-categorized and sub-categorized. To have a refined identification of news topics, especially to evaluate the degree of belonging of a news item to a specific topic, we use automatic topic modeling. We thus apply a dimensionality reduction algorithm on LDA embeddings using the UMAP algorithm [18]. Second, a clustering step is performed with the traditional density-based HDBSCAN algorithm [5].

4.4 Building u's profile

Given a user u, ADF first builds her profile, *i.e.* the distribution of u's interests over the set of topics A (Section 3, Step 1). In line with [10], we choose to define u's interest on topic a as the ratio between the interest of u on this topic and the total interest of u over A, as presented in Equation (3).

$$p(a|u) = \frac{\sum_{n \in N_u} s(u,n) q(n|a)}{\sum_{a \in A} \sum_{n \in N_u} s(u,n) q(n|a)} \tag{3}$$

where $s(u,n)$ is the relevance score provided as input of ADF. N_u is the set of news u interacted with. $q(n|a)$ is the strength of belonging of news n to a topic a, provided by the HDBSCAN algorithm. This probability allows to have a finer evaluation of u's interest in a specific topic. Recall that u's profile (*selection distribution*) $P(u) = \{p(a|u) \forall a \in A\}$.

4.5 Evaluating u's selection diversity

u's *selection diversity*, based on her *selection distribution* $P(u)$, is evaluated by the normalized Shannon entropy, as proposed in [25]. The resulting value ranges in $[0,1]$. Higher values indicate that u has homogeneously selected all aspects $a \in A$, while lower values indicate that u has strong preferences for specific aspects.

4.6 Evaluation metrics

We fix the size of the recommendation list to $k = 20$. Note that this value can have an impact on evaluation metrics since k can be higher than the number of accessed news in the test set. Nevertheless, this impact will be similar between models, which does not bias analysis.

To evaluate recommendation performance, we rely on accuracy, diversity and fairness metrics.These will contribute to evaluate the three dimensions at the core of ADF. First, to evaluate accuracy, we use *Precision@k*, which evaluates the ability of the NRS to select news that is relevant to u. Second, to evaluate diversity, we compute *Intra-List Diversity (ILD)* [33], which measures the pair-wise distance of the news in the recommendation list, and *S-Recall@k* metric [32], that evaluates the ratio of topics covered in the recommendation list. Finally, for the fairness dimension, we evaluate if the topic distribution in the recommendation list ($P(R^*_k(u))$) is compatible with a smoothed *selection distribution*. To do this, we inspire from the calibration metric proposed by Steck [23], with the Hellinger distance. For the statistical analysis, as our data does not follow a normal distribution (Shapiro-Wilk test, $p_{value} < 0.01$), we apply the Wilcoxon test. We fix the significance level at 0.01. Data and code are available: https://github.com/Celina-07/ADF_framework.

5 EXPERIMENTS

We compare the performance of ADF to those of three baseline models: (1) The **A-baseline** that manages the **A**ccuracy dimension. The recommendation list is $R_k(u)$, the list of k news with top relevance scores $s(u,n)$ in $R(u)$; (2) The **AF-baseline** that manages **A**ccuracy and **F**airness, and provides a fully calibrated recommendation list, *i.e.* whose topic distribution equals u's *selection distribution*; (3) the **AD-baseline** that provides an **A**ccurate and **D**iversified recommendation list [33] by applying a greedy re-ranking.

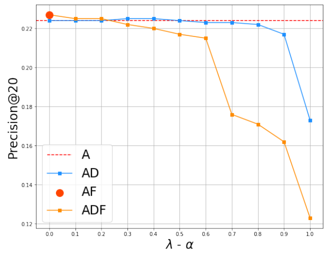
To evaluate the impact of constraining diversity, we run ADF with values of $\alpha \in [0,1]$. As the average *selection diversity* in MIND is quite high (0.653) and no user is fully polarized ($min(diversity(P(u)) = 0.11$), we set $\beta = 0$.

We first compare the performance between a full calibration ($\alpha = 0$) and a full diversity ($\alpha = 1$). When $\alpha = 0$ it corresponds to the AF-baseline, that results in an increase in both accuracy (5%) and diversity (3%). Besides, when α=1, the impact on accuracy is massive, with a drop of $Precision@20$ by 45% (from 0.224 to 0.123) (Figure 1a).

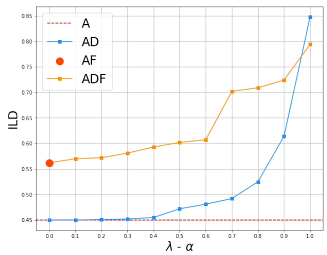
Let us now take a closer look at intermediate values of α. Up to $\alpha = 0.2$, accuracy is not decreased, it is even significantly higher (+0.4%) than the AD-baseline (from 0.224 to 0.225). In this case, not only *ILD* increases: 27%, going from 0.450 to 0.572, but diversity of topics, *S-Recall@20*, also increases: 46% (from 0.383 to 0.559). This higher increase in *S-Recall@20* was expected as ADF manages topic diversity. The calibration metric C_H equals 0.030 (Figure 1d), confirming that the recommendation list is fair, which was not the case for AD-baseline. Let us compare this configuration and the one of the AD-baseline with $\lambda = 0.4$, where in both cases $Precision@20 = 0.225$. The ADF framework leads to a content diversity 26% higher than the AD-baseline (0.450 and 0.572) and a topic diversity 45% higher (0.386 and 0.559).

When α increases (smoother *target distribution*) and up to $\alpha = 0.6$, the drop in $Precision@20$ is limited: 4% (0.215). Comparing with the AD-baseline, when $\alpha <= 0.6$ and $\lambda <= 0.6$, $Precision@20$ values are close (Figure 1a). However, both content and topic diversities are further increased with ADF. Regarding the fairness dimension, as expected, ADF allows to maintain fairness, and C_H values remain low on a stable basis for all values of α, which is not the case at all for the AD-baseline. When $\alpha >= 0.6$, the accuracy is significantly reduced, as for AD-baseline.

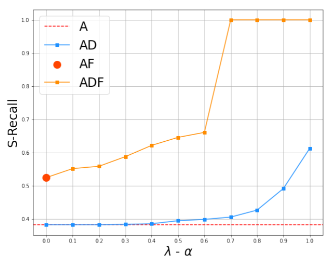
To summarize, ADF provides significantly more diverse recommendations while being fair, with no impact on $Precision@20$, which answers **RQ2**. This allows us to confirm that the myopic nature of the greedy approach does not allow to manage the nature of diversity, limiting the quality of the trade-off [21]. On the contrary,



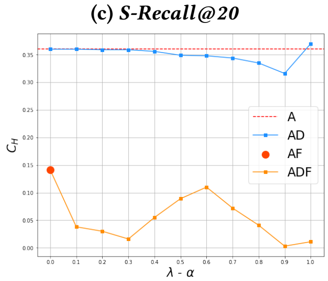
(a) *Precision*@20



(b) *ILD*



(c) ***S-Recall@20***



(d) C_H

Figure 1: Metrics values for A-baseline (red dashed line), AD-baseline (blue line), AF-baseline (red dot), and ADF (orange line).

by constraining fairness, the diversity input is finely controlled and fostered.

From a NRS and polarization perspective, thanks to ADF users can be exposed to a wider range of news topics, while getting accurate recommendations. Besides, ADF guarantees that users' opinions are not artificially steered towards topics or news that are too far removed from their own interests.

6 CONCLUSION AND PERSPECTIVES

In this work, we have proposed a novel recommendation framework, ADF, designed to answer **RQ1**. ADF is the first framework that simultaneously optimizes three traditional recommendation dimensions generally optimized by pair: accuracy, diversity, and fairness. In ADF, fairness is not viewed as a simple dimension that has to be maximized, as for accuracy and diversity. Fairness is viewed as a constraint on the distribution of the recommendation list, ADF thus goes beyond traditional trade-offs. We have shown that contrary to what is usually thought, when fairness is managed as a constraint, it does not hinder accuracy of recommendations, which answers **RQ2**. Importantly, we have shown that the traditional greedy approach used to identify trade-offs limits the improvement in diversity, which is exceeded by ADF. The fairness of the recommendations, associated with personalized diversity make that ADF can be used repeatedly to increase user awareness step-by-step, without orienting her towards a specific topic.

This work leaves many directions for future work. ADF has been designed to be deployed in any recommender system. The recommendation algorithm used in this work is content-based and we wonder to what extent similar conclusions will be drawn when a collaborative filtering (CF) algorithm will be used. Especially, as CF allows to provide diverse recommendations, what is the impact of ADF in such a configuration? Besides, ADF has been designed to be used in NRS, but the issues it tackles hold in other domains. It would be interesting to evaluate ADF on these domains. To make the evaluation of ADF more complete, a user study is unavoidable, it will provide interesting feedback, especially about recommendation acceptance and users' perception of the fairness-constrained diversity.

REFERENCES

[1] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221.

[2] Carolina Becatti, Guido Caldarelli, Renaud Lambiotte, and Fabio Saracco. 2019. Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections. *Palgrave Communications* 5, 1 (2019), 1–16.

[3] Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 405–414.

[4] David M Blei, Andrew Y Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3, 4-5 (2003), 993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9

[5] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 160–172.

[6] Pablo Castells, Neil Hurley, and Saul Vargas. 2021. Novelty and diversity in recommender systems. In *Recommender systems handbook*. Springer, 603–646.

[7] Pablo Castells, Saúl Vargas, Jun Wang, et al. 2011. Novelty and diversity metrics for recommender systems: choice, discovery and relevance. In *International Workshop on Diversity in Document Retrieval (DDR 2011) at the 33rd European Conference on Information Retrieval (ECIR 2011)*. Citeseer, 29–36.

[8] Diego Corrêa da Silva, Marcelo Garcia Manzato, and Frederico Araújo Durão. 2021. Exploiting personalized calibration and metrics for fairness recommendation. *Expert Systems with Applications* 181 (2021), 115112.

[9] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. 2023. Fairness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction* (2023), 1–50.

[10] Farzad Eskandanian, Bamshad Mobasher, and Robin Burke. 2017. A clustering approach for personalizing diversity in collaborative recommender systems. In

Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. 280–284.

[11] Ruoyuan Gao and Chirag Shah. 2019. How fair can we go: Detecting the boundaries of fairness optimization in information retrieval. In *Proceedings of the 2019 ACM SIGIR international conference on theory of information retrieval*. 229–236.

[12] Lucien Heitz, Juliane A Lischka, Alena Birrer, Bibek Paudel, Suzanne Tolmeijer, Laura Laugwitz, and Abraham Bernstein. 2022. Benefits of Diverse News Recommendations for Democracy: A User Study. *Digital Journalism* (2022), 1–21.

[13] Dietmar Jannach and Himan Abdollahpouri. 2023. A survey on multi-objective recommender systems. *Frontiers in big Data* 6 (2023), 1157899.

[14] Emily Kubin and Christian von Sikorski. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 45, 3 (2021), 188–206.

[15] Matev Kunaver and Toma Porl. 2017. Diversity in Recommender Systems A Survey. 123, C (2017), 154––162.

[16] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Juntao Tan, Shuchang Liu, and Yongfeng Zhang. 2023. Fairness in Recommendation: Foundations, Methods, and Applications. *ACM Trans. Intell. Syst. Technol.* 14, 5 (2023), 48 pages.

[17] Pasquale Lops, Marco Polignano, Cataldo Musto, Antonio Silletti, and Giovanni Semeraro. 2023. ClayRS: An end-to-end framework for reproducible knowledge-aware recommender systems. *Information Systems* 119 (2023), 102273.

[18] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[19] Shaina Raza and Chen Ding. 2022. News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review* (2022), 1–52.

[20] Laura Schelenz. 2021. Diversity-aware recommendations for social justice? exploring user diversity and fairness in recommender systems. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 404–410.

[21] Sinan Seymen, Himan Abdollahpouri, and Edward C Malthouse. 2021. A constrained optimization approach for calibrated recommendations. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 607–612.

[22] Barry Smyth and Paul McClave. 2001. Similarity vs. Diversity. In *Case-Based Reasoning Research and Development*, David W. Aha and Ian Watson (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 347–361.

[23] Harald Steck. 2018. Calibrated recommendations. In *Proceedings of the 12th ACM conference on recommender systems*. 154–162.

[24] Celina Treuillier, Sylvain Castagnos, and Armelle Brun. 2023. A Multi-Factorial Analysis of Polarization on Social Media. In *UMAP'23*. Limassol, Cyprus. https://hal.science/hal-04108988

[25] Celina Treuillier, Sylvain Castagnos, and Armelle Brun. 2023. All Polarized but Still Different: a Multi-factorial Metric to Discriminate between Polarization Behaviors on Social Media. *arXiv preprint arXiv:2312.04603* (2023).

[26] Saúl Vargas. 2015. *Novelty and diversity enhancement and evaluation in Recommender Systems*. Ph. D. Dissertation. Universidad Autónoma de Madrid.

[27] Saúl Vargas and Pablo Castells. 2013. Exploiting the diversity of user preferences for recommendation. In *Proceedings of the 10th conference on open research areas in information retrieval*. Citeseer, 129–136.

[28] Chuhan Wu, Fangzhao Wu, Xiting Wang, Yongfeng Huang, and Xing Xie. 2021. Fairness-aware news recommendation with decomposed adversarial learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4462–4469.

[29] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, et al. 2020. Mind: A large-scale dataset for news recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. ACL, Online, 3597–3606.

[30] Wen Wu, Li Chen, and Yu Zhao. 2018. Personalizing recommendation diversity based on user personality. *User Modeling and User-Adapted Interaction* 28, 3 (2018), 237–276.

[31] Sobia Zahra, Mustansar Ali Ghazanfar, Asra Khalid, Muhammad Awais Azam, Usman Naeem, and Adam Prugel-Bennett. 2015. Novel centroid selection approaches for KMeans-clustering based recommender systems. *Information sciences* 320 (2015), 156–189.

[32] ChengXiang Zhai, William W Cohen, and John Lafferty. 2015. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Acm sigir forum*, Vol. 49. ACM New York, NY, USA, 2–9.

[33] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. 22–32.