D-RDW: Diversity-Driven Random Walks for News Recommender Systems

Runze Li
runze.li@uzh.ch
Department of Informatics, University of Zurich
Zurich, Switzerland

Lucien Heitz
heitz@ifi.uzh.ch
Department of Informatics, University of Zurich
Zurich, Switzerland

Oana Inel
inel@ifi.uzh.ch
Department of Informatics, University of Zurich
Zurich, Switzerland

Abraham Bernstein
bernstein@ifi.uzh.ch
Department of Informatics, University of Zurich
Zurich, Switzerland

Abstract

This paper introduces **D**iversity-Driven **R**an**d**om **W**alks (D-RDW), a lightweight algorithm and re-ranking technique that generates diverse news recommendations. D-RDW is a societal recommender, which combines the diversification capabilities of the traditional random walk algorithms with customizable target distributions of news article properties. In doing so, our model provides a transparent approach for editors to incorporate norms and values into the recommendation process. D-RDW shows enhanced performance across key diversity metrics that consider the articles' sentiment and political party mentions when compared to state-of-the-art neural models. Furthermore, D-RDW proves to be more computationally efficient than existing approaches.

CCS Concepts

• **Information systems → Information retrieval diversity**; **Recommender systems**; • **Theory of computation → Random walks and Markov chains**.

ACM Reference Format:
Runze Li, Lucien Heitz, Oana Inel, and Abraham Bernstein. 2025. D-RDW: Diversity-Driven Random Walks for News Recommender Systems. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25), September 22–26, 2025, Prague, Czech Republic.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3705328.3748016

1 Introduction

In news recommender systems (NRSs), diversity is one of the most important criteria for assessing the societal impact of these systems [6, 34, 40]. This is particularly true for recommending political articles, as a diverse selection of news can impact opinion formation and social deliberation [17–19]. When designing NRSs, however, diversity optimization is predominantly achieved in a post-processing step [30, 38]. As such, it plays only a secondary role; it is limited to re-ranking items from the candidate list of the model, instead of being able to access the entire item pool [14, 43]. To address the limitations of re-ranking approaches, this paper introduces ***D**iversity-Driven **R**an**d**om **W**alks* (D-RDW) that provides diversity optimization at the model stage.

RecSys '25, Prague, Czech Republic

ACM ISBN 979-8-4007-1364-4/2025/09
https://doi.org/10.1145/3705328.3748016

D-RDW is an extension of the random walk algorithms RP^3_β [9, 28]. In the past, RP^3_β was shown to outperform neural models across a wide range of accuracy and diversity metrics in the movie and e-commerce domain [9, 13]. While state-of-the-art neural models can account for diversity through re-ranking, the benefits of using random walks for news recommendations are that they allow for 1) an explainable recommendation approach (i.e., they are not a black box model), 2) an easily extendable approach to include heuristics of editors and journalists, and 3) a lightweight and cost-efficient approach that is highly scalable and parallelizable. D-RDW achieves this by introducing a new step in the random walk pipeline that enforces a so-called normative target distribution (NTD) of article properties in the recommendation list.

As a secondary contribution, we demonstrate the effectiveness of enforcing NTD to achieve higher diversity by creating separate re-ranking strategies for existing neural models. When speaking of normativity in the context NRS, we follow the definition of Vrijenhoek et al. [42], which refers to the process of operationalizing and including societal values in the optimization and evaluation of NRS (e.g., optimizing news feeds for source or political viewpoint diversity and providing equal exposure for minority and majority parties). In doing so, we offer a first diversity-optimizing model and re-ranker for a *deliberative recommender* according to the normative NRSs framework of Helberger [19].

2 Related Work

Diversity has increasingly been recognized as a crucial aspect of assessing the quality of news recommendations [5], by looking at aspects such as topics, viewpoints, and political mentions, among others [2, 3, 6, 22, 32]. NRSs also take into account other forms of diversity, including those related to sources, people, events, semantics, sentiments, authors, and temporal aspects [8, 47]. In recent years, however, the research community has begun to recognize the critical role that NRSs play in democratic societies and has proposed four different models from the normative framework of democracy to consider exposure diversity as a societal goal [19, 20]. These normative frameworks, however, have not yet been implemented into any recommendation model.

Looking at models, several approaches in the literature focused on generating such diversified recommendations by using random walk-based models, i.e., models that simulate user navigation in an item-user interaction graph as a probabilistic transition [4, 11, 12, 27, 44, 49]. Liu et al. [26] introduced a collaborative filtering method based on directed random walks that amplifies the influence of users with small degrees to reach a more diverse set of recommendations in the movie domain.

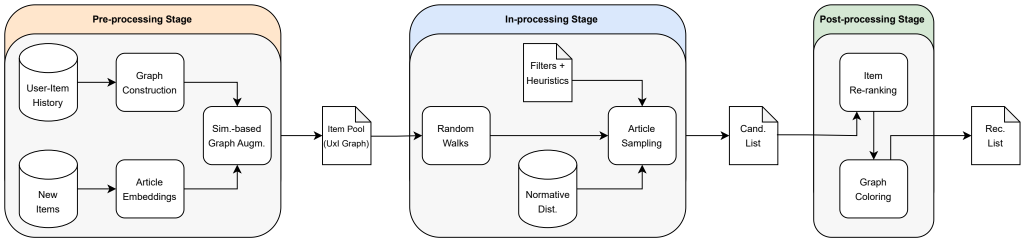

Figure 1: The D-RDW pipeline consists of three stages: 1) a pre-processing stage for graph construction, 2) an in-processing stage for generating a candidate list, and 3) a post-processing step for creating the recommendation list.

Christoffel et al. [9, 29] created an improved variant, RP^3_β, that showed the capability of random walks to optimize for both accuracy as well as diversity. Paudel and Bernstein [28] then proposed Random Walk with Erasure (RWE-D), where they successfully leveraged random walks to diversify recommendations within a left-right political spectrum. These approaches, however, do not offer any direct control to editors or journalists when it comes to specifying the precise dimension of diversity (be that, e.g., source or viewpoint diversity). The underlying algorithm only looks at user interactions and has no information on the article content. D-RDW addresses this issue with NTD-based filtering by accessing annotated news articles, which we will outline in the next section.

Algorithm 1: Diversity-Driven Random Walks

```
Input  : graph, articleAttributes, targetDimensions
         targetDistribution, targetSize, filterCriteria
         samplingObjective, maxHops
Output : recommendations
1  recommendations = []
2  currentHop = 3
3  while True do
4      candidates = newHop(graph, currentHop)
5      candidates = filterHeuristics(candidates,
6          articleAttributes, filteringCriteria)
7      recommendations = sampling(candidates,
8          articleAttributes, targetDimensions,
9          targetDistribution, targetSize,
10         samplingObjective)
11     if length(recommendations) >= targetSize ∨
12         currentHop >= maxHops then
13         break
14     currentHop = currentHop + 2
15 if length(recommendations) < targetSize then
16     recommendations = addRandomArticles(candidates,
17         articleAttributes, targetDimensions,
18         targetDistribution, targetSize,
19         samplingObjective)
20 recommendations = rankArticles(recommendations,
21     targetSize, rankingObjectives)
22 return recommendations
```

3 Diversity-Driven Random Walks

D-RDW follows the recommendation pipeline shown in Figure 1. It consists of three main steps: 1) a pre-processing stage for graph construction based on user-item interactions, 2) an in-processing stage for running random walks and applying filter criteria to enforce a target normative item distribution, and 3) a post-processing stage for re-ranking the candidate list to achieve a homogeneous distribution of articles across the news feed.[1]

Stage 1: Graph Construction and Augmentation. The random walk algorithm uses a bipartite graph with user and item nodes [9, 28], with edges representing user-item interactions (i.e., clicks). To address the cold start problem, unconnected nodes (i.e., articles not yet read by any user) are integrated by 1) selecting the most similar items already in the graph[2] and 2) creating edges between the new item and *all users* connected to the 3 most similar articles that are part of the graph.

Stage 2: Random Walks and Normative Distributions. Recommendations are generated by following four steps: 1) running random walks on the graph with 3 hops, 2) a heuristic rule for removing items included in the history of the target user, 3) applying a normative target distribution (NTD),[3] and 4) sampling random items from test impressions to fill open positions in case the candidate list is smaller than the target feed. The novel element of D-RDW is the article sampling in Algorithm 1 on Line 7 to solve the constraint satisfaction problem of the NTD. For example, let I be an all-ones vector in a matrix form $I^\top X = 20$, where X is the target size of the news feed of 20 items. We then define a binary indicator matrix for each category. Take P, where $P_i = 1$ represents the political item i and let $S_i = 1$ be an item i with a positive sentiment. In this example, our NTD includes 15 political articles and 10 articles with a positive sentiment. By combining these linear equations into their matrix form, we arrive at Equation 1:

$$\begin{pmatrix} I^\top \\ P^\top \\ S^\top \end{pmatrix} X = \begin{pmatrix} 20 \\ 15 \\ 10 \end{pmatrix} \tag{1}$$

[1] A reference implementation of our algorithm is available online in our repository: https://github.com/Informfully/Experiments

[2] Items are matched based on the semantic similarity of the article text. Similarity is calculated with an off-the-shelf model: https://huggingface.co/sentence-transformers/all-mpnet-base-v2

[3] We provide a norm-aware deliberative recommender [19] by focusing on article sentiment and political party. However, this is one possible example, as deliberative recommenders can be operationalized in different ways (see Section 4 for details).

The variable X is restricted to binary values (0 or 1), making it a binary integer programming problem. As multiple solutions for X may satisfy these linear equations, we use recency as an objective function. Let C represent a vector of item attributes that the sampling algorithm aims to maximize in the final item selection. When C is defined as the item recency vector, the sampling goal is to select items that satisfy the linear constraints while maximizing overall recency, formalized as $max \sum_i c_i \cdot x_i$.

Stage 3: Graph Coloring and Item Re-ranking. D-RDW sorts the candidate list by the prediction score in descending order and applies a graph coloring algorithm [23] to the candidate list (where the article categories represent colors). This ensures a homogeneous category distribution across the entire recommendation list. Line 15 performs random sampling from the most recent news items to complete the targeted size of the news feed. We note that the graph coloring algorithm is not interfering with the diversity of the recommended items and the offline experiments presented here. Instead, it is a necessary step for the visualization of the recommended items. We address this matter as future work.

Normativity and Explainability. With D-RDW, our goal is to provide an algorithm for NRSs that is both norm-aware and easy to explain. In terms of operationalizing norms, we base our algorithm on the work of Helberger [19], where four types of democratic recommenders are proposed. They are each defined by the different distributions of, e.g., topics, viewpoints, and emotions in the news feed/recommendation list. In our approach, we use article category as a proxy for modeling the distribution of topics, party mentions as a proxy for viewpoints, and sentiment as a proxy for emotion. Based on this, D-RDW implements what is referred to as the *deliberative* approach, focusing on an equal representation of viewpoints (party mentions) and diversity of emotions (article sentiment).[4]

By using a customizable NTD for item filtering, D-RDW provides a way to tweak the output of random walks that takes into account article properties. As such, it goes beyond existing algorithms whose parameters only allowed for changes to the output based on the graph (e.g., tweaking hops or using node degree [9, 29]). Our parameters, in contrast, are distributions based on article categories, party mentions, and sentiments are easy to quantify and explain. They present non-technical terms that editors in the newsrooms are familiar with.

4 Experiment

We evaluate D-RDW on the RecSys 2024 Challenge benchmarking Ekstra Bladet News Recommendation Dataset (EB-NeRD), specifically EB-NeRD small [24]. Recommendations are calculated using the test set with $15,339$ users. In our analysis, we compare the performance of D-RDW with the neural models LSTUR [1], NPA [45], and NRMS [46]. To optimize for diversity, we re-ranked the output of these models using three customized approaches: greedy KL (G-KL) [35], PM-2 [10], and MMR [7].[5] Additionally, RP^3_β, random walks with erasure (RWE-D), and a randomized list (Random) are added as baseline algorithms.[6]

[4]Please see Section 4 for the precise numbers used in the underlying NTD.
[5]G-KL and PM-2 use the same NTD sampler as D-RDW.
[6]We used 3 hops and for augmenting the graphs of the random walk models, we connected each cold item with the users of their 3 most similar items in the graph.

In preparing our submission, we experimented with different numbers of epochs and ratios for the sampling strategy of negative and positive items for our baselines. However, similar to previous experiments with EB-NeRD that used a subset of our baselines [24], we were unable to significantly improve AUC beyond 0.6 and used the original hyperparameters instead.[7]

For our NTD, we consider sentiment and political party distributions in this experiment. Sentiment ranges from negative (-1.0) to positive (1.0).[8] The articles are put into 4 discretized buckets with value ranges of $[-1, -0.5)$ (20%), $[-0.5, 0)$ (30%), $[0, 0.5)$ (30%), $[0.5, 1]$ (20%).[9] The political party is based on mentions.[10] NTD has buckets for articles that mention 1) government (incl. supporting parties, 15%), 2) opposition parties (15%), 3) government *and* opposition parties (15%), 4) independent/foreign parties (*must* mention at least one independent/foreign party, *may* mention a majority party, 15%), and 5) no political parties (40%). We purposefully measure category performance while not including this dimension in our proposed NTD. The reason for doing so is to assess the impact of the added NTD sampling step for random walks when comparing category performance to, e.g., party performance that is part of our NTD. Finally, we would like to emphasize again that this NTD configuration is only one of many possible ways of capturing and operationalizing deliberative diversity.

We assess the model performance using the area under the ROC curve (AUC) [31] for test impressions. We follow the example of Vrijenhoek et al. [40] and calculate the top 20 predictions, as diversity metrics require a large recommendation list to calculate divergence metrics. However, the EB-NeRD small test set has a median of only 11 impression items per user session (with a median of 1 clicked item per session). We, therefore, rank and subsequently select items from among the entire test pool of $4,400+$ articles to generate the recommendation lists. This new setting makes assessing accuracy metrics other than AUC extremely challenging, as the expected value of the default scenario (e.g., predicting top 1 out of 11 impressions for EB-NeRD, cf. [24, 25]) is vastly different from our setting (top 1 out of $4,400$ items).[11]

We list the estimated energy cost to train models/create graphs (Train Cost) and to calculate the recommendations (Rec. Cost) in watt seconds.[12] The evaluation of the entire experimental procedure is done using *Informfully Recommenders* [16], an extension of the Cornac Framework [33, 36, 37].

[7]For details, please see the official implementation found on the Microsoft Recommenders GitHub repository: https://github.com/recommenders-team/recommenders
[8]We used the XLM-roBERTa model: https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment
[9]Percentages indicate the share of each bucket in the final recommendation list (e.g., 20% of all articles in the user's feed should have a positive sentiment of 0.5 or higher).
[10]EB-NeRD covers predominantly news from Denmark. We therefore model NTD according to the Danish political landscape, with governing and supporting parties as the majority and the opposition parties as the minority. Wikipedia overview: https://en.wikipedia.org/wiki/Folketing
[11]We skip AUC calculation for the re-ranking approaches, as the results for the base model have already assessed the entire item pool.
[12]We calculate the energy cost based on the power consumption (in watts for CPU and GPU components) and factor in the overall runtime [39].

Table 1: Overview of the diversity scores for the top 20 news recommendations and AUC scores for predicting user impressions. The values closest to the perfect score are highlighted in green, second closest in blue, and third closest in red.

Model	Re-ranking	Activ.	Cat. Calib.	Comp. Calib.	Frag.	Alt. Voices	Repr.	Cat. Gini	Sent. Gini	Party Gini	Cat. ILD	Sent. ILD	Party ILD	Train. Cost	Rec. Cost	AUC
NTD Target Values		1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.133	0.250	1.000	0.779	0.789	0.000	0.000	1.000
LSTUR [1]		0.190	0.433	0.257	0.610	0.063	0.372	0.832	0.645	0.874	0.740	0.552	0.343	351.12	69.84	0.564
NPA [45]		0.202	0.468	0.266	0.619	0.061	0.365	0.826	0.605	0.877	0.753	0.580	0.337	523.05	51.25	0.554
NRMS [46]		0.204	0.509	0.285	0.632	0.060	0.362	0.791	0.544	0.880	0.794	0.624	0.326	326.55	63.23	0.549
LSTUR	G-KL	0.282	0.444	0.240	0.612	0.104	0.546	0.828	0.133	0.250	0.754	0.779	0.789	351.12	30.15	
	PM-2	0.284	0.440	0.244	0.604	0.115	0.546	0.819	0.150	0.250	0.766	0.776	0.789	351.12	23.86	
	MMR	0.295	0.433	0.235	0.613	0.118	0.575	0.823	0.226	0.270	0.759	0.762	0.807	351.12	52.36	
NPA	G-KL	0.292	0.469	0.237	0.589	0.108	0.545	0.810	0.133	0.250	0.775	0.779	0.789	523.05	30.15	
	PM-2	0.302	0.471	0.239	0.576	0.122	0.547	0.806	0.150	0.250	0.783	0.776	0.789	523.05	23.86	
	MMR	0.314	0.463	0.236	0.598	0.110	0.575	0.808	0.219	0.270	0.772	0.763	0.807	523.05	52.36	
NRMS	G-KL	0.307	0.482	0.239	0.608	0.091	0.544	0.798	0.133	0.250	0.790	0.779	0.789	326.55	30.15	
	PM-2	0.316	0.483	0.244	0.598	0.099	0.544	0.796	0.150	0.250	0.794	0.776	0.789	326.55	23.86	
	MMR	0.315	0.477	0.238	0.616	0.099	0.571	0.798	0.204	0.270	0.790	0.767	0.807	326.55	52.36	
D-RDW		0.374	0.407	0.229	0.394	0.107	0.556	0.798	0.133	0.250	0.810	0.779	0.789	4.21	2.50	0.554
RP^3_β [9]		0.222	0.415	0.235	0.582	0.080	0.376	0.840	0.755	0.856	0.743	0.439	0.392	4.21	1.28	0.565
RWE-D [28]		0.256	0.435	0.222	0.443	0.100	0.372	0.857	0.802	0.842	0.735	0.377	0.433	1.92	0.25	0.554
Random		0.180	0.461	0.256	0.705	0.054	0.366	0.756	0.634	0.873	0.842	0.564	0.346			0.500

5 Results

Table 1 shows the results of the RADio diversity, Gini, ILD, cost estimates, and AUC. We have added a row for NTD Target Values. They show the scores that are used for ranking the top 3 results in each column. For norm-aware RADio metrics, the values consider the underlying distributions of EB-NeRD. For traditional metrics, the target values indicate the lowest achievable value for Gini and the highest one for ILD. Smaller values are better for cost estimates. And larger values are better in terms of AUC.

Activation compares article sentiment distributions, with higher values indicating a larger divergence between a user's recommendation list and the item pool. D-RDW and MMR for NPA and NRMS result in the largest divergence.

Calibration compares the category and complexity distribution in the user history and their recommendation list. A higher value signifies a greater deviation from the user's category/complexity preferences. D-RDW, RP^3_β, and LSTUR (with and without MMR) show the smallest divergence for category; RWE-D, D-RDW, and RP^3_β have the smallest divergence for complexity.

Fragmentation quantifies the difference in story distribution across users. A higher value indicates a greater variation in story chains, showing a more fragmented user base. D-RDW, RWE-D, and NPA with PM-2 provide the least fragmented user base.

Alternative Voices compares the proportion of minority and majority perspectives within the recommendations. A higher value indicates a greater disparity between the two perspectives. The random baseline, NRMS, and NPA show the smallest disparity.

Representation measures the divergence in representing political parties. A higher value indicates a larger difference between the recommendation list and the item pool. LSTUR and NPA with MMR, NRMS with MMR, and D-RDW show the largest divergence.

Gini coefficients quantify inequality in the target dimension. A higher value indicates larger inequality. D-RDW and the G-KL re-ranker outperform the base neural models across Gini of sentiment and party (with perfect scores for the NTD dimension). The random baseline achieved the best results for category Gini.

ILD calculates the average pairwise dissimilarity between items using cosine distance. A higher value means more dissimilarity (i.e., diversity) between items. D-RDW and the G-KL re-ranker have the lowest ILD for the dimensions of sentiment and party, with the random baseline performing best on category ILD.

Computational Cost lists the average watt seconds for calculating the recommendation list of 20 items per user.[13] We see that the random walk models require two orders of magnitude less energy for training compared to the neural models and one to almost three orders of magnitude less for the recommendations process.

AUC assesses model classification quality in predicting item clicks, with 0.5 being the expected value of random guessing. We see RP^3_βhas the highest AUC, followed by LSTUR in the second place, and a tie between D-RDW, RWE-D, and NPA in third place.

6 Discussion

There are four dimensions to our performance comparison: 1) normative RADio diversity, 2) traditional Gini/ILD diversity, 3) energy cost, and 4) AUC. The results in Table 1 show that D-RDW has consistent scores in the top 3 across all four of these dimensions. Furthermore, the NTD re-ranking approaches outperform the original neural models on non-normative diversity.

Interpreting RADio is challenging, as the metrics express a divergence between two underlying distributions. For example, if the dataset is diverse, a small divergence is desirable, and the other way around. Therefore, we need to first know the effective NTD-relevant distributions in EB-NeRD. For political party mentions, we have 2.66% for the governing parties and their supporters, 0.88% for the opposition parties, 0.41% for both governing and opposition parties, 16.96% for independent and foreign parties, and 79.09% of articles have no political entities. In terms of sentiment, 9.51% are negative, 58.01% somewhat negative, 30.01% somewhat positive, and 2.48% positive.

[13]The training phase for the random walk models is the graph construction (incl. graph augmentation). The recommendation phase for graph-based algorithms includes random walks, applying the distribution, and item re-ranking.

Looking now at the results for RADio, we see that random walk models show the best calibration for user preferences and minimize overall fragmentation. For traditional diversity, the NTD in-process resampling of D-RDW and the NTD post-processing for re-rankers allow targeting and boosting of specific metrics to achieve perfect scores across Gini as well as ILD for sentiment and party. As such, applying NTD to the recommender pipeline is an effective way for editors and journalists to boost specific Gini and ILD scores.

While neural models with and without NTD re-ranking allow for combining AUC gains with excellent Gini and ILD performance, it comes at around 400-1000 times the cost of random walk algorithms. Even if one disregards ecological concerns and diversity optimization, there is an economic argument to be made to use random walks. They allow for cheaper/more frequent recommendations, with cheaper updates [29], which is of particular importance in the news domain where items have a relatively short shelf life.

Lastly, we see that the random walk model RP_{β}^{3} can outperform AUC-optimized neural models on EB-NeRD, with D-RDW and RWE-D achieving scores comparable with LSTUR and NPA. While AUC values below 0.6 indicate poor discriminatory power, it is important to emphasize that the news presents a difficult domain. Reference AUC values for state-of-the-art algorithms are around this value for MIND [48], EB-NeRD [24], and NeMig [21].

7 Limitations and Future Work

Since the dataset heavily influences the performance of recommender models and re-rankers, future work should better focus on examining the relationship between dataset characteristics and diversity outcomes. Second, further experiments could focus on better understanding the normative target distributions for multiple diversity dimensions on user engagement and satisfaction. Third, integrating personalized approaches to diversity is a promising direction for future work (i.e., what level of diversity is suitable for what type of user). To facilitate this, a user modeling module could be developed to enhance the system's ability to understand and adapt to user attributes, such as preferences for news topics, political orientation, or desired levels of diversity. Lastly, user studies are required to understand what users perceive/accept as diverse and satisfying [15, 41], and what would be desirable approaches for visualizing diversity-driven recommended item lists (i.e., assessing the effectiveness of the graph coloring step of D-RDW).

8 Conclusion

In this paper, we present D-RDW, a first operationalization of deliberative recommenders from the normative framework of Helberger [19]. It is the best-performing solution when looking at the combined dimensions of RADio, Gini/ILD, energy cost, and AUC when compared to state-of-the-art neural models and random walks. Furthermore, we provide a stand-alone version of the NTD sampler of D-RDW as a re-ranker for existing neural models. The re-rankers allow for achieving perfect results when added to the neural models across the traditional ILD and Gini diversity metrics.

The NTD sampling at the in- and post-processing steps is an effective and easily explainable approach to achieving a normative goal by virtue of adjusting the desired normative distribution based on enforcing distributions of article dimensions in news feeds.

Taken together, this enables the input of editorial and journalistic values into recommendations. We hope D-RDW and random walks can become a cornerstone of exploring norm-aware NRSs that go beyond accuracy objectives by facilitating experiments with norms and values together with the assessment of their societal impact.

Acknowledgments

This work was partially funded by the Digital Society Initiative (DSI) of the University of Zurich (UZH) under a grant from the DSI Excellence Program, the Hasler Foundation, and the Swiss Federal Office of Communications (OFCOM).

References

[1] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural news recommendation with long-and short-term user representations. In *Proceedings of the 57th annual meeting of the association for computational linguistics*. 336–345.

[2] Philipp Bachmann, Mark Eisenegger, and Diana Ingenhoff. 2022. Defining and measuring news media quality: Comparing the content perspective and the audience perspective. *The International Journal of Press/Politics* 27, 1 (2022), 9–37.

[3] Christian Baden and Nina Springer. 2017. Conceptualizing viewpoint diversity in news discourse. *Journalism* 18, 2 (2017), 176–194.

[4] Shumeet Baluja, Rohan Seth, D. Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, and Mohamed Aly. 2008. Video suggestion and discovery for youtube: taking random walks through the view graph. In *Proceedings of the 17th International Conference on World Wide Web* (Beijing, China) *(WWW '08)*. Association for Computing Machinery, New York, NY, USA, 895–904. doi:10.1145/1367497.1367618

[5] Christine Bauer, Chandni Bagchi, Olusanmi A. Hundogan, and Karin van Es. 2024. Where Are the Values? A Systematic Literature Review on News Recommender Systems. *ACM Trans. Recomm. Syst.* 2, 3, Article 23 (jun 2024), 40 pages. doi:10.1145/3654805

[6] Abraham Bernstein, Claes De Vreese, Natali Helberger, Wolfgang Schulz, Katharina Zweig, Lucien Heitz, Suzanne Tolmeijer, et al. 2021. Diversity in News Recommendation (Dagstuhl Perspectives Workshop 19482). *Dagstuhl Manifestos* 9, 1 (2021), 43–61. doi:10.4230/DagMan.9.1.43

[7] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. 335–336.

[8] Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. 2017. Optimizing the recency-relevancy trade-off in online news recommendations. In *Proceedings of the 26th International Conference on World Wide Web*. 837–846.

[9] Fabian Christoffel, Bibek Paudel, Chris Newell, and Abraham Bernstein. 2015. Blockbusters and Wallflowers: Accurate, Diverse, and Scalable Recommendations with Random Walks. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria) *(RecSys '15)*. Association for Computing Machinery, New York, NY, USA, 163–170. doi:10.1145/2792838.2800180

[10] Van Dang and W Bruce Croft. 2012. Diversity by proportionality: an election-based approach to search result diversification. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 65–74.

[11] Marco De Gemmis, Pasquale Lops, Giovanni Semeraro, and Cataldo Musto. 2015. An investigation on the serendipity problem in recommender systems. *Information Processing & Management* 51, 5 (2015), 695–717.

[12] Chantat Eksombatchai, Pranav Jindal, Jerry Zitao Liu, Yuchen Liu, Rahul Sharma, Charles Sugnet, Mark Ulrich, and Jure Leskovec. 2018. Pixie: A system for recommending 3+ billion items to 200+ million users in real-time. In *Proceedings of the 2018 world wide web conference*. 1775–1784.

[13] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. 2019. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM conference on recommender systems*. 101–109.

[14] Lucien Heitz. 2023. Classification of Normative Recommender Systems. In *Proceedings of the First Workshop on the Normative Design and Evaluation of Recommender Systems*.

[15] Lucien Heitz, Oana Inel, and Sanne Vrijenhoek. 2024. Recommendations for the Recommenders: Reflections on Prioritizing Diversity in the RecSys Challenge. In *Proceedings of the Recommender Systems Challenge 2024*. 22–26.

[16] Lucien Heitz, Runze Li, Oana Inel, and Abraham Bernstein. 2025. Informfully Recommenders – Reproducibility Framework for Diversity-aware Intra-session

Recommendations. In *Proceedings of the 19th ACM Conference on Recommender Systems*.

[17] Lucien Heitz, Juliane A Lischka, Rana Abdullah, Laura Laugwitz, Hendrik Meyer, and Abraham Bernstein. 2023. Deliberative Diversity for News Recommendations: Operationalization and Experimental User Study. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 813–819.

[18] 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* 10, 10 (2022), 1710–1730.

[19] Natali Helberger. 2019. On the Democratic Role of News Recommenders. *Digital Journalism* 7, 8 (2019), 993–1012.

[20] Natali Helberger, Kari Karppinen, and Lucia D'acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society* 21, 2 (2018), 191–207.

[21] Andreea Iana, Mehwish Alam, Alexander Grote, Nevena Nikolajevic, Katharina Ludwig, Philipp Müller, Christof Weinhardt, and Heiko Paulheim. 2023. NeMig-A Bilingual News Collection and Knowledge Graph about Migration. (2023).

[22] Glen Joris, Camiel Colruyt, Judith Vermeulen, Stefaan Vercoutere, Frederik De Grove, Kristin Van Damme, Orphée De Clercq, Cynthia Van Hee, Lieven De Marez, Veronique Hoste, et al. 2020. News diversity and recommendation systems: Setting the interdisciplinary scene. *Privacy and Identity Management. Data for Better Living: AI and Privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2. 2 International Summer School, Windisch, Switzerland, August 19–23, 2019, Revised Selected Papers 14* (2020), 90–105.

[23] Hal A Kierstead. 2000. A simple competitive graph coloring algorithm. *Journal of Combinatorial Theory, Series B* 78, 1 (2000), 57–68.

[24] Johannes Kruse, Kasper Lindskow, Saikishore Kalloori, Marco Polignano, Claudio Pomo, Abhishek Srivastava, Anshuk Uppal, Michael Riis Andersen, and Jes Frellsen. 2024. EB-NeRD a large-scale dataset for news recommendation. In *Proceedings of the Recommender Systems Challenge 2024*. 1–11.

[25] Johannes Kruse, Kasper Lindskow, Saikishore Kalloori, Marco Polignano, Claudio Pomo, Abhishek Srivastava, Anshuk Uppal, Michael Riis Andersen, and Jes Frellsen. 2024. RecSys Challenge 2024: Balancing Accuracy and Editorial Values in News Recommendations. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 1195–1199.

[26] Jian-Guo Liu, Kerui Shi, and Qiang Guo. 2012. Solving the accuracy-diversity dilemma via directed random walks. *Physical Review E* 85, 1 (2012), 016118.

[27] Nathan N Liu and Qiang Yang. 2008. Eigenrank: a ranking-oriented approach to collaborative filtering. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. 83–90.

[28] Bibek Paudel and Abraham Bernstein. 2021. Random walks with erasure: Diversifying personalized recommendations on social and information networks. In *Proceedings of the Web Conference 2021*. 2046–2057.

[29] Bibek Paudel, Fabian Christoffel, Chris Newell, and Abraham Bernstein. 2016. Updatable, accurate, diverse, and scalable recommendations for interactive applications. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (2016), 1–34.

[30] Felix Petersen, Debarghya Mukherjee, Yuekai Sun, and Mikhail Yurochkin. 2021. Post-processing for individual fairness. *Advances in Neural Information Processing Systems* 34 (2021), 25944–25955.

[31] Foster Provost and Tom Fawcett. 2001. Robust classification for imprecise environments. *Machine learning* 42 (2001), 203–231.

[32] Junyang Rao, Aixia Jia, Yansong Feng, and Dongyan Zhao. 2013. Taxonomy based personalized news recommendation: Novelty and diversity. In *International Conference on Web Information Systems Engineering*. Springer, 209–218.

[33] Aghiles Salah, Quoc-Tuan Truong, and Hady W Lauw. 2020. Cornac: A Comparative Framework for Multimodal Recommender Systems. *Journal of Machine Learning Research* 21, 95 (2020), 1–5.

[34] Holli Sargeant, Eliska Pirkova, Matthias C Kettemann, Marlena Wisniak, Martin Scheinin, Emmi Bevensee, Katie Pentney, Lorna Woods, Lucien Heitz, Bojana Kostic, et al. 2022. *Spotlight on Artificial Intelligence and Freedom of Expression: A Policy Manual.* Organization for Security and Co-operation in Europe.

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

[36] Quoc-Tuan Truong, Aghiles Salah, and Hady Lauw. 2021. Multi-modal recommender systems: Hands-on exploration. In *Fifteenth ACM Conference on Recommender Systems*. 834–837.

[37] Quoc-Tuan Truong, Aghiles Salah, Thanh-Binh Tran, Jingyao Guo, and Hady W Lauw. 2021. Exploring Cross-Modality Utilization in Recommender Systems. *IEEE Internet Computing* (2021).

[38] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. 2014. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems*. 209–216.

[39] Tobias Vente, Lukas Wegmeth, Alan Said, and Joeran Beel. 2024. From clicks to carbon: The environmental toll of recommender systems. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 580–590.

[40] Sanne Vrijenhoek, Gabriel Bénédict, Mateo Gutierrez Granada, Daan Odijk, and Maarten De Rijke. 2022. Radio–rank-aware divergence metrics to measure normative diversity in news recommendations. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 208–219.

[41] Sanne Vrijenhoek, Savvina Daniil, Jorden Sandel, and Laura Hollink. 2024. Diversity of What? On the Different Conceptualizations of Diversity in Recommender Systems. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 573–584. doi:10.1145/3630106.3658926

[42] Sanne Vrijenhoek, Lien Michiels, Johannes Kruse, Alain Starke, Nava Tintarev, and Jordi Viader Guerrero. 2023. Normalize: The first workshop on normative design and evaluation of recommender systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1252–1254.

[43] Mingyang Wan, Daochen Zha, Ninghao Liu, and Na Zou. 2023. In-processing modeling techniques for machine learning fairness: A survey. *ACM Transactions on Knowledge Discovery from Data* 17, 3 (2023), 1–27.

[44] Mengsha Wang, Yingyuan Xiao, Wenguang Zheng, and Xu Jiao. 2018. RNDM: A random walk method for music recommendation by considering novelty, diversity, and mainstream. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 177–183.

[45] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019. NPA: neural news recommendation with personalized attention. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2576–2584.

[46] Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019. Neural news recommendation with multi-head self-attention. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*. 6389–6394.

[47] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. Sentirec: Sentiment diversity-aware neural news recommendation. In *Proceedings of the 1st conference of the Asia-Pacific chapter of the association for computational linguistics and the 10th international joint conference on natural language processing*. 44–53.

[48] 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*. 3597–3606.

[49] Ting Yu, Junpeng Guo, Wenhua Li, Harry Jiannan Wang, and Ling Fan. 2019. Recommendation with diversity: An adaptive trust-aware model. *Decision Support Systems* 123 (2019), 113073.