

## Operationalizing Framing to Support Multiperspective Recommendations of Opinion Pieces

Mulder, M.; Inel, O.; Oosterman, J.E.G.; Tintarev, N.

**DOI**

[10.1145/3442188.3445911](https://doi.org/10.1145/3442188.3445911)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

FACCT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency

**Citation (APA)**

Mulder, M., Inel, O., Oosterman, J. E. G., & Tintarev, N. (2021). Operationalizing Framing to Support Multiperspective Recommendations of Opinion Pieces. In *FACCT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 478-488). Article 3445911 (FACCT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency). <https://doi.org/10.1145/3442188.3445911>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# Operationalizing Framing to Support Multiperspective Recommendations of Opinion Pieces

Mats Mulder\*  
Delft University of  
Technology  
Delft, The Netherlands  
mats.mulder@live.nl

Oana Inel  
Delft University of  
Technology  
Delft, The Netherlands  
o.inel@tudelft.nl

Jasper Oosterman  
Blendle  
Utrecht, The Netherlands  
jasperoosterman@blendle.com

Nava Tintarev  
Maastricht University  
Maastricht  
The Netherlands  
n.tintarev@maastrichtuniversity.nl

## ABSTRACT

Diversity in personalized news recommender systems is often defined as dissimilarity, and operationalized based on topic diversity (e.g., corona versus farmers strike). Diversity in news media, however, is understood as multiperspectivity (e.g., different opinions on corona measures), and arguably a key responsibility of the press in a democratic society. While viewpoint diversity is often considered synonymous with source diversity in communication science domain, in this paper, we take a computational view. We operationalize the notion of framing, adopted from communication science. We apply this notion to a re-ranking of topic-relevant recommended lists, to form the basis of a novel viewpoint diversification method. Our offline evaluation indicates that the proposed method is capable of enhancing the viewpoint diversity of recommendation lists according to a diversity metric from literature. In an online study, on the Blendle platform, a Dutch news aggregator, with more than 2000 users, we found that *users are willing to consume viewpoint diverse news recommendations*. We also found that *presentation characteristics* significantly influence the reading behaviour of diverse recommendations. These results suggest that future research on presentation aspects of recommendations can be just as important as novel viewpoint diversification methods to truly achieve multiperspectivity in online news environments.

## CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Human-centered computing** → **User studies**; **Empirical studies in HCI**.

## KEYWORDS

recommender systems, viewpoint diversity, framing aspects

### ACM Reference Format:

Mats Mulder, Oana Inel, Jasper Oosterman, and Nava Tintarev. 2021. Operationalizing Framing to Support Multiperspective Recommendations of Opinion Pieces. In *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3442188.3445911>

\*Work done while enrolled in a master program at Delft University of Technology.



This work is licensed under a Creative Commons Attribution International 4.0 License.

FAccT '21, March 3–10, 2021, Virtual Event, Canada  
ACM ISBN 978-1-4503-8309-7/21/03.  
<https://doi.org/10.1145/3442188.3445911>

## 1 INTRODUCTION

In recent years, traditional news sources are increasingly using online news platforms to distribute their content. Digital-born news websites and *news aggregators*, which combine content from various sources in one service, are also gaining ground [29]. In 2015, 23% of survey respondents reported online media as their primary news source, and 44% considered digital and traditional sources equally relevant [29]. This change also induces a wide adoption of *news recommender systems* that automatically provide personalized news recommendations to users.

Communication studies acknowledge two important roles of media in a democratic society [17]: (1) to inform citizens about important societal and political issues and (2) to foster a diverse public sphere. Both roles are then related to multiple social-cultural objectives of democracy, such as informed decision-making, cultural pluralism and citizens welfare [24, 34]. The role of news recommender systems in promoting these democratic values is under heavy discussion in academic debate. For example, *filter bubbles* received increasing awareness, suggesting that high levels of personalisation would lock people up in bubbles of what they already know or think [30]. According to Helberger, the democratic role of news recommender systems mainly depends on the democratic theory that is being followed, such as liberal, participatory, or deliberative [17]. Thus, the development of viewpoint diversification methods can be in particular motivated in relation to the participatory and deliberative model.

Current diversification methods [21, 43], however, do not address viewpoint diversity, but define diversity as dissimilarity, or topic diversity (e.g., corona versus farmers strike). Therefore, current diversification methods are not applicable in the news domain, and novel viewpoint diversification methods are needed to maintain and assure multiperspectivity in online news environments. Moreover, to truly enable *multiperspectivity*, users should be willing to consume viewpoint-diverse recommendations. Their behaviour should be studied in real, online scenarios, to ensure that diversification criteria does not damage the adoption of the recommended items and thus, lead to the total rejection of the news recommender platform. Thus, we investigate the following research questions:

**R1:** *How is reading behaviour affected by viewpoint diverse news recommendations?*

**R2:** *How is reading behaviour affected by presentation characteristics of viewpoint diverse news recommendations?*

To answer these questions, we propose a re-ranking approach for lists of recommended articles based on aspects of news frames, a concept taken from communication studies. In particular, a news frame describes how to identify a view on an issue, in a given

article [12]. Thus, by bridging aspects from the social and the computational domains, we aim to overcome the current gap between the definition of diversity in recommender systems and news media.

During an offline evaluation, the proposed method increased the viewpoint diversity of recommended lists of news articles on several topics. Further, we measured the influence of the viewpoint diversification method on the reading behaviour of more than 2000 users, which are likely to interact with the recommended articles, in an online study on the Blendle<sup>1</sup> platform, a Dutch news aggregator. We found that reading behaviour of users that received diverse recommendations was comparable with the reading behaviour of users that received news articles optimized only for relevance. However, we did find a positive influence of two *presentation characteristics* on the click-through rate of recommendations – news articles with thumbnails and news articles with more hearts are read more often.

Therefore, we make the following contributions:<sup>2</sup>

- a novel method for *viewpoint diversification* using re-ranking of news recommendation lists, based on framing aspects;
- an online evaluation with more than 2000 users, on the Blendle platform, to understand:
  - (a) how viewpoint-diverse recommendations affect the reading behaviour of users; and
  - (b) how article's *presentation characteristics* affect the reading behaviour of users.

## 2 RELATED WORK

In this section, we first investigate how communication science understands diversity. Then, we review current approaches for diversity in recommender systems. This allows us to bridge the gap between the domains of communication and computer science, by operationalizing framing aspects in a diversification algorithm.

### 2.1 Diversity in News Media

In news media, diversity refers to multiperspectivity or diversity of viewpoints [14]. In communication science, diversity is, in general, a key measure for news quality [9, 22, 31], thus fostering democratic aspects, such as informed decision-making, cultural pluralism and citizens welfare [26, 40]. Two main approaches to assess diversity can be distinguished: source and content diversity [2, 6, 26], with most studies focusing on source diversity [1, 2, 26, 40]. When measuring source diversity, most methods follow Bennett [5]'s indexing theory, which assumes that including non-official or non-elite sources corresponds to high levels of diversity [2]. Alternatively, Napoli [26] approaches the issue as a policymaker and distinguishes three aspects of source diversity: content ownership, ownership of media outlets, and the workforce within individual media outlets.

Critics, however, state that multiple sources can still foster the same point of view and therefore, source diversity is not a direct measure for viewpoint diversity [40]. Multiple studies also indicate that power distributions in society, commercial pressure of news media and journalistic norms and practices, significantly influence which sources gain media access [2, 6]. Therefore, it is often argued that viewpoint diversity can only be achieved by fostering content diversity [3, 9, 14, 22, 26, 40]. Content diversity is defined in [37]

as “*heterogeneity of media content in terms of one or more specified characteristics*”. Baden and Springer [2] identified six common approaches to assess content diversity. The first three methods focus on the tone or political position represented in the news, *i.e.*, the inclusion of non-official positions, the diversity of political tone. These methods, however, assume that political disagreement equals viewpoint diversity [2]. Another approach uses language diversity to evaluate content diversity. However, this is again no direct measure, since different language can describe the same perspective [2].

The final two approaches assess content diversity by means of *frames*. Framing theory states that every communicative message selectively emphasizes certain aspects of the complex reality [2]. Thereby, frames enable different interpretations of the same issue [33]. Porto [31] states that news environments need to be evaluated by their ability to provide diverse frames. Baden and Springer [2] describe three frames' aspects, central to the role of viewpoint diversity in democratic media. First, frames create different interpretations of the same issue by selecting some aspects of the complex reality [13]. Second, frames are not neutral but suggest specific evaluations and courses of actions that serve some purpose better than other [12]. Third, frames are often strategically constructed to advocate particular political views and agendas. Framing, thus, can be a suitable conceptualization of viewpoint diversity.

### 2.2 Diversity in Recommender Systems

Traditionally, research on recommender systems focused on evaluating their performance in terms of accuracy [43]. Such focus, however, induced a problem which is known as *over-fitting*, *e.g.*, a model is fitted so strongly to a user that it is unable to detect any other interests [21]. Additionally, there is a need for more user-centric evaluations of recommender systems. Thus, diversity has become one of the most prominent *beyond-accuracy metrics* for recommender systems [43]. In this context, diversity is generally defined as the opposite of similarity [21], and it is often based on topic diversity (*e.g.*, corona versus farmers strike). For example, Ziegler et al. [43] proposed a topic diversification method based in the *intra-list diversity* metric.

Current diversification methods for recommender systems, thus, do not focus on viewpoint diversity and are not applicable in the news domain. To the best of our knowledge, only one study for viewpoint diversification has been proposed so far [36]. They propose a new distance measure for viewpoint diversity based on linguistic representations of news articles. This diversity measure was then applied in a post-processing re-ranking algorithm [8] to a list of news articles. These allowed optimizing for the balance between topic relevance and viewpoint diversity. In a small scale user study [36], readers indicated a lower intent to consume diversified content, motivating the need to study behavioural measures for newsreaders on a larger scale. Thus, we argue that more research is required to understand the relationship between the metric and the influence on readers behaviour.

In this work, we bridge the current gap between the notion of framing in communication science and computational approaches. Additionally, we study how viewpoint diversification affects newsreaders' behaviour in an applied setting. The next section justifies the operationalization of *framing* in the computational domain.

<sup>1</sup><https://blendle.com/>

<sup>2</sup>Repository: <https://github.com/mats-mulder/FAccT2021-OperationalizingFraming>

### 3 FRAMING FOR VIEWPOINT DIVERSITY

Framing is an extensively researched concept in different domains, including psychology, communication and sociology, having its roots in the latter domain. Bateson [4] states that communication only gets meaning in its context and by the way the message is constructed. Later, frame theory gained increasing momentum and was generally understood as follows: every communicative message selectively emphasizes certain aspects of a complex reality [2]. Thus, every news article (unintentionally) comprises some form of framing [2]. Frames are often deliberately used to construct strategic, often political, views on a topic. Consequently, frames enable different interpretations of the same issue [2]. However, every frame inevitably deselected other, equally plausible and relevant frame [2].

When considering frames in news articles, multiple definitions exist [10, 13, 15]. However, the definition of Entman [12] is the most commonly adopted in the literature. It states that framing includes the selection of “*some aspects of perceived reality and make the more salient in a communicating text, in such a way as to promote a particular definition of a problem, causal interpretation, moral evaluation and treatment recommendation for the item described*”. Within this definition, the problem describes *four framing functions*:

- (1) **Problem Definition** : “what a causal agent is doing with what costs and benefits”; e.g., *a second Coronavirus wave is approaching*;
- (2) **Causal Attribution** : “identifying the forces creating the problem”; e.g., *(it is due to the) government policy response*;
- (3) **Moral Evaluation** : “evaluate causal agents and their effects”; e.g., *response to approaching second wave came too late (negative evaluation)*;
- (4) **Treatment Recommendation** : “offer and justify treatments for the problems and predict their likely effects”; e.g., *there must be predefined measures to be deployed at a critical threshold of virus spread*.

Frames can be found at different levels of analysis, including single sentences, paragraphs or articles as a whole [12]. A frame, however, may not necessarily include all the four functions.

Most framing analysis approaches focus on manual analysis of articles [20, 23, 39]. Only recently, some computer-assisted methods gained interest [7, 16, 41]. As a result, the identification of frames often falls into a methodological black box [23]. Thereby, the main issue includes the ambiguity of “*which elements should be present in an article or news story to signify the existence of a frame*” [23]. To overcome this problem, some recent studies [2, 23, 39] propose a novel identification method based on the extraction of the four aforementioned framing aspects in the definition of Entman [12].

#### 3.1 Focus Group Setup

To guide the operationalization of framing, we began with a qualitative analysis. In a small focus group, we aimed to gain insights into how the four framing functions of the main frame of an article manifest in its content and how to identify them computationally.

*Participants.* We invited three experts in the field of news framing analysis. All experts had a background in journalism, communication, or news media, and multiple years of relevant experience.

*Materials.* As a basis for discussion, we used opinion pieces on the topic of *Dutch farmers protests*. Opinion pieces refer to news articles that reflect the authors opinion and thus, do not claim to be objective. Our domain experts indicated that this type of news article is the most suitable to identify framing functions.

*Procedure.* The focus group procedure consisted of two steps.

1. *Annotation session:* First, the participants were asked to perform framing analysis on an opinion piece, using the four framing functions as described by Entman [12]. In particular, the participants had to individually highlight parts of the article, such as word clauses or sentences, that can be related to one of the four framing functions of the main frame of the news article.

2. *Review session:* Second, the results were discussed, together with some general questions on news article analysis and framing. For every highlighted part, we asked the participants to motivate why the highlighted part is related to one of the four framing functions. Besides, we used the results as input to a broader discussion on news article analysis and framing, such as:

- What main heuristic did you use to analyze the article?
- What procedure did you follow to analyze the framing functions of the article?
- Can you derive any patterns in the way framing functions manifest in opinion pieces?

#### 3.2 Results of Framing Analysis

In the review session, experts indicated the structure of the article as the main heuristic to find the framing functions of the main frame. They also mentioned that opinion pieces are strongly shaped by journalistic values on how an article should be structured. We then analyzed this heuristic according to the four framing functions:

- (1) **Problem Definition** : In opinion pieces, the first part of the article often presents the main problem that the author addresses and includes the title, the lede, and the first  $x$  paragraphs. Work on manual frame analysis [20] supports this finding. The number of introductory paragraphs,  $x$ , can be different per source, author, or article.
- (2) **Causal Attribution + Moral Evaluation** : The body of an article is used to analyze the main problem and usually contains different factors that contribute to the problem under investigation and their evaluation. We can match this with: a) the causal attribution of a frame (forces creating the problem), and b) the moral judgements (evaluate the causal attribution and their effect) [12].
- (3) **Treatment Recommendation** : Treatment recommendations can be seen as suggestions to improve or solve the issue described by the problem definition of the main frame. They normally appear in the concluding paragraphs, according to the focus group members.

This structure, however, is only a heuristic, applying just to opinion pieces. Other types, *i.e.* interviews, are structured differently.

The results of the annotation session also indicate that each framing function related to the main frame of an article can normally be found within one paragraph. Additionally, a paragraph can include multiple framing functions, but words, clauses, and sentences generally represent a single framing function.

## 4 DATASET

The experimental dataset used in our studies consists of opinion pieces, in Dutch. The choice of article type is motivated by the focus group session presented in Section 3, in which the structure of this article type is put forward as the primary heuristic to find framing aspects. We picked topics that we expected **a)** to be present on the Blendle platform at the time when we performed the online user study; **b)** to contain different viewpoints addressed in the news; and **c)** to balance issues that more current versus long-standing. The dataset consists of four ongoing topics: *Black Lives Matter*, *Coronavirus*, *U.S. Elections* - as more current topics, and the dominance and privacy issues around *Big Tech* - as a long-standing topic.

We collected our dataset from an archive containing more than 5 million Dutch news articles. The archive is known to undergo checks for articles quality, to remove undesirable content, such as the weather or short actualities. For each topic, we used the search terms (queries) and restrictions shown in our GitHub repository<sup>3</sup>. We provide the list of search terms in Dutch, because we do not want to add additional bias through translation. Additionally, since the proposed method heavily relies on the structure of the article, we set up a filter for the minimum number of words to 450 and a filter for the minimum number of paragraphs to 5.

Table 1 provides an overview of the dataset, per topic. While the length of the articles varies across topics, they are usually far longer than the 450-word limit we chose. Four publishers are present for all topics: De Volkskrant, De Standaard, Trouw and Het Algemeen Dagblad. Furthermore, De Volkskrant is the most prominent publisher for all topics, except for the *U.S. Elections* topic. The inclusion of other, less frequent, publishers varies per topic. Overall, our dataset covers a set of 15 unique publishers.

We also present some properties concerning the presentation characteristics of the articles on the news aggregator website. We observe that the ratio of articles that contains a thumbnail image depends on the topic. For the *Black Lives Matter* and *Coronavirus* topics, more than half of the articles have a thumbnail image, while the opposite holds for the other two topics. The number of custom titles from the editorial team and the average title length also differ considerable per topic. Only a few articles have an editorial title, and they usually appear for the *Big Tech* and *U.S. Elections* topics.

**Table 1: Overview of the experimental dataset, per topic.**

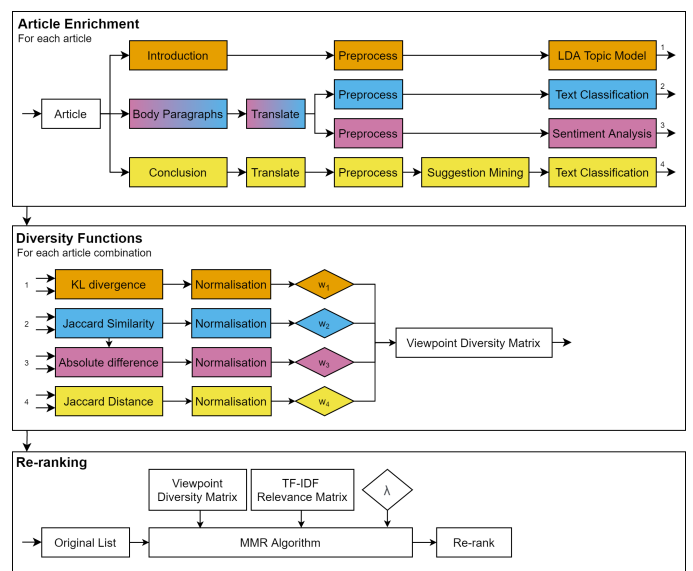
Topic	Articles	Publishers	Avg #Words	With thumb.	With ed. title	Avg title length
Black Lives Matter	69	10	697	39	1	6.3
Coronavirus	52	7	608	27	4	5.2
U.S. Elections	42	6	744	20	8	9.6
Big Tech	51	10	761	17	10	8.1

## 5 VIEWPOINT DIVERSITY METHODOLOGY

We proposed a novel diversification method based on framing aspects, using the insights from the focus group. First, we describe the extraction pipeline, which supports the structure heuristic described in the results of the focus group session (Section 3). The

<sup>3</sup>Search queries: <https://git.io/JttZu>

pipeline forms the basis for the generation of recommendation lists that we use in the offline evaluation (Section 6) and the online study (Section 7). We implemented the pipeline<sup>4</sup> using methods employed by the news aggregator platform and off-the-shelf natural language processing toolkits, such as the IBM Watson Natural Language Understanding<sup>5</sup> (IBM Watson NLU). We chose to use state-of-the-art and off-the-shelf methods used by Blendle to ensure output quality. Then we describe the distance function, which combines the metadata related to each framing aspect in a measure for viewpoint diversity for news articles. Finally, we present the re-ranking algorithm based on this viewpoint diversity measure. Our contribution stands in the novelty of the overall diversification framework, rather than the implementation of specific components. Figure 1 shows an overview of the end-to-end pipeline.



**Figure 1: Viewpoint diversification pipeline describing a) article enrichment, b) diversity functions, and c) re-ranking. Color coding indicates the four different framing functions, and color blending is applied where the pipeline contributes to both causal attribution and moral evaluation.**

### 5.1 Metadata Extraction

For each framing aspect, as described in the definition of Entman [12], we implemented an extraction pipeline:

**Problem Definition**. As described in Section 2, the problem definition can be understood as the central issue or topic under investigation [23]. Therefore, we decided to use a topic model as the main extraction method for this framing aspect. The model, provided by the research partner, at Blendle, included a 1000-topic latent Dirichlet allocation (LDA) model trained on 900k Dutch news articles. Based on the conclusions from the focus group described in Section 3, the title and the first  $x$  paragraphs are used to retrieve

<sup>4</sup>Diversification pipeline: <https://git.io/JttZm>

<sup>5</sup><https://www.ibm.com/nl-en/cloud/watson-natural-language-understanding>



metadata related to this framing aspect. We also applied multiple pre-processing steps on the content, including cleaning, chunking, tokenization, lemmatization and stop-word removal.

**Causal Attribution** + **Moral Evaluation**. The causal attribution of a frame relates to the forces creating the problem, while the moral judgements evaluate the causal attribution and their effect [12]. In the focus group session, Section 3, we concluded that the body of an article usually elaborates on these aspects. Additionally, paragraph-level seems to be the most suitable level of analysis. Thus, we first translated the news articles in English using Google Translate and then we applied a text-classification algorithm using IBM Watson NLU. The service returns a category for each paragraph according to a predefined five-level taxonomy, from the most general category (e.g., level 1 - technology and computing), to the most specific one (e.g., level 5 - portable computer). To extract information related to the evaluation of these attributions, we also analyze the sentiment of these paragraphs (in English), using IBM Watson NLU. Thereby, it would be able to identify if two articles evaluate the same aspects of a problem differently. The content of interest for this task includes all paragraphs except the  $x$  introductory and  $y$  concluding paragraphs. We optimize these variables during the offline evaluation.

**Treatment Recommendation**. A treatment recommendation suggests remedies for problems and predicts their likely effect [12]. The research domain of suggestion mining, the task of retrieving sentences that contain advice, tips, warnings and recommendations from opinionated texts [27], was found to be highly relevant for this framing aspect [28]. However, the state-of-the-art models are topic-specific [28], and can not be easily applicable to our domain. Thus, a more naive rule-based approach was applied for this study, being more generally applicable. In a crowdsourcing task with domain experts, we evaluated and optimized the generally applicable rules from the literature on the news article content. Afterwards, we implemented the method to extract sentences that contain suggestions from the article content (translated in English using Google Translate beforehand). To obtain comparable information between the suggestions of two articles, the suggestion sentences were classified using the same text-classification algorithm as for causal attribution. According to the focus group discussion, the content of interest for this framing aspect includes the  $y$  concluding paragraphs of an article. We optimize this variable in the offline evaluation.

## 5.2 Distance Functions

Having defined the extraction pipeline for each framing aspect, i.e., problem definition, causal attribution, moral evaluation and treatment recommendation [12], we now define our distance function. We compare the extracted metadata for every pair of articles. Thus, we implement a distance function for each framing aspect.

**Problem Definition**. The metadata regarding the problem definition framing aspect involves a probability distribution over 1000 topics. Thus, we need a statistical distance measure. We chose the Kullback-Leibler divergence because it is one of the most commonly used statistical distance measures for LDA-models, and it is used in comparable work on viewpoint diversification [36].

**Causal Attribution** and **Moral Evaluation**. We compare the five-level taxonomy categories extracted from the pipeline described in the previous section, to obtain a distance measure for the causal attribution framing function of the primary frame. Thus, we use the weighted Jaccard index, which measures the similarity (or diversity) of two sets [18]. The index is calculated for each level of detail in the five-level taxonomy, such that we apply weight factors per taxonomy level. Thereby, overlap in higher levels of detail can contribute more to the overall similarity score. In the offline evaluation, we compare different weight factors per taxonomy-levels.

For the moral evaluation framing aspect, we implement the distance function by multiplying the Jaccard distance and the absolute sentiment difference between each paragraph combination of two articles. Thus, paragraphs with no overlapping categories yield a value of zero, while highly similar paragraphs, with different sentiment scores, lead to high levels of diversity related to the moral evaluation framing aspect.

**Treatment Recommendation**. For the treatment recommendation, the distance function is identical to the distance function related to the causal attribution. Thus, we used the five-level taxonomy classification, i.e., from the most general to the most specific category, as returned by IBM Watson NLU, and the Jaccard index.

## 5.3 Re-ranking

We implement the re-ranking of the input list of articles using the Maximal Marginal Relevance (MMR) algorithm [8]. In our case, the re-ranking consists of ranking news articles that are more diverse higher. First, we normalize the output of the distance functions related to each framing aspect using a min-max normalization, and then we combine them in a diversity score through a weighted sum. We optimize the weight factors during the offline evaluation. Note here that we re-rank news articles that are known to also be relevant for the given topic. Where most re-ranking algorithms for recommender systems order lists only on relevance, the MMR algorithm provides a linear combination between diversity, in our case viewpoint diversity, and relevance, set by the parameter  $\lambda$ . Thus, the re-ranking algorithm is defined as follows:

$$MMR \equiv \max_{i \in R \setminus S} [\lambda(Rel(i) - (1 - \lambda)\max_{j \in S} (1 - Div(i||j)))] \quad (1)$$

In Eq. (1),  $R$  is the ranked list of relevant articles,  $S$  is the list of selected articles in  $R$ , and  $R \setminus S$  is the list of articles in  $R$  that are not yet selected. Since this work proposes a measure for viewpoint diversity rather than a relevance measure, we implemented the relevance score ( $Rel$ ) using a simple frequency-inverse document frequency (TF-IDF) score. In Eq. (1), the  $\lambda$  parameter takes values in  $[0,1]$ , where  $\lambda = 1$  returns a list of articles ranked only on relevance, while  $\lambda = 0$  returns a list of articles ranked on maximum diversity.

## 6 OFFLINE EVALUATION

In this section, we describe the offline evaluation of our viewpoint diversity-driven approach for re-ranking lists of news articles.

### 6.1 Materials

For our offline experiment, we used the news dataset introduced in Section 4, which covers 214 news articles on four topics.

## 6.2 Procedure

The experimental procedure consists of four main steps that we detail as follows. First, we process and enrich all the news articles in our dataset according to the four framing aspects [12]: problem definition, causal attribution, moral evaluation, and treatment recommendations (for details see Section 5.1).

Second, we generate the diversity matrix by comparing all combinations of two articles, based on the enrichment described in Section 5.1. Thus, using the distance function defined in Section 5.2 we measure the dissimilarity of two articles based on the framing aspects. Finally, since the MMR algorithm re-ranks a list of news articles based on a linear combination between diversity and relevance, we calculate the TF-IDF relevance matrix, including a relevance score for each two article combination.

Third, we optimize the model variables and evaluate the performance on the diversity metric in Section 6.3, using cross-validation. For each article  $i$  in the dataset, we calculate a set of  $s$  recommendations by re-ranking the remainder articles in the dataset. To prevent over-fitting, we use cross-validation. We split the dataset into  $k$  distinct sets. We experimented with different values of  $k = 5, 10, 20$  and  $s = 3, 6, 9$ . For every set, we take the following steps:

- (1) Grid search of model variables on training set: The training set contains the  $k - 1$  subsets of articles. We obtain the optimal combination of the model variables for the training set using a grid search. An overview of the model variables can be found in Table 2 and in Section 6.2.1.
- (2) Evaluation on test set: After the variables are trained on the  $k - 1$  subsets, the model is evaluated on the test set for different values of  $\lambda$ , between 0 and 1 with a step of 0.1. As described before, for each article in the test set, a set of  $s$  recommendations is calculated by re-ranking the remaining articles in the dataset.

And finally, we combined the results of all  $k$  cross-validations.

**Table 2: Overview of possible values of model variables**

Variable	Values
Weight Framing function - Problem Definition	[0.1, 0.2, 0.3, 0.4]*
Weight Framing function - Causal Attribution	[0.1, 0.2, 0.3, 0.4]*
Weight Framing function - Moral Evaluation	[0.1, 0.2, 0.3, 0.4]*
Weight Framing function - Treatment Recommendation	[0.1, 0.2, 0.3, 0.4]*
Taxonomy level weight	[equal, ascending]
Number of introducing paragraphs	[1, 2]
Number of concluding paragraphs	[1, 2]
$\lambda$	[0.0, 0.1, ..., 0.9]

\*Note that all framing function weight factors should sum up to 1

**6.2.1 Model variables.** Table 2 shows the model variables that we optimize during the offline evaluation. We choose the variation of the weights for each framing aspect such that no single framing aspect can have the majority. Additionally, a step-size of 0.1 is assumed to bring enough variation. We consider two variations for the taxonomy level weights: equal weights for each taxonomy level or ascending weights. Finally, the number of introductory and concluding paragraphs can be either 1 or 2.

## 6.3 Evaluation Metrics

We assess the performance of the viewpoint diversification method using a metric from literature [36], based on the Intra-List Diversity metric [36, 38, 42, 43]. It is defined as the average distance between all pairs of articles  $i$  and  $j$ , such that  $i \neq j$ . Thereby, the distance between a pair is defined by the articles' channels (predefined taxonomy of 20 high-level topics) and the articles' LDA topic-distribution, as derived from the enrichment methods in Section 5:

$$Distance(i, j) = 0.5 \times Distance_{Channels} + 0.5 \times Distance_{LDA} \quad (2)$$

The channel distance is calculated using the cosine distance, whereas the LDA distance using the Kullback-Leibler divergence.

*Additional metrics.* We also measure the effectiveness of the diversification model on other properties, as follows:

*Relevance:* We measure the TF-IDF relevance for the recommendation lists, such that we can measure the effectiveness of the viewpoint diversification method.

*Kendall's  $\tau$ :* We compute the *Kendall's  $\tau$  rank correlation coefficient* [19] to measure the similarity between two ranks of recommended items.

*Average number of words:* We compute the average number of words for the recommended article lists as a measure of quality (*i.e.*, longer news articles can be considered to be higher quality).

*Publisher Ratio:* We measure the publisher ratio for the recommendation lists because this could potentially provide insights on the effect of the content diversity on the source diversity.

## 6.4 Baseline

To assess if the proposed diversification method can increase the viewpoint diversity based on the presented metric, we compare it with a baseline, consisting of a full relevance MMR, where  $\lambda = 1$ , such that we rank the recommendations purely on the TF-IDF relevance. We chose this baseline because it has minimal effects on the recommendations in terms of viewpoint diversity.

## 6.5 Results

In Figure 2, we show the performance of the model in terms of viewpoint diversity and relevance for different values of  $\lambda$ , and the optimal setting of the model variables (see Table 3 for the optimal model variables values, per topic). Variations of the cross-validation variable  $k$  did not yield significant differences between the results, and thus, we fixed  $k = 10$ . The list size  $s$  did show to influence the number of publishers included in the recommended list, but the results were not significant. Thus, we fixed the list size to  $s = 3$ , to align with the online evaluation set up, where only 3 recommended articles can be shown at a time.

Across all topics, the proposed diversification method is capable of increasing the viewpoint diversity of recommendation lists. According to the metric, the viewpoint diversity increases on average from 0.55 to 0.79 between  $\lambda = 1$  and  $\lambda = 0$ . Additionally, the average relevance score decreases from 0.58 to 0.27. In our project repository,<sup>6</sup> we provide examples of ranked lists of articles based on relevance and diversity, for the topic of *Coronavirus*.

<sup>6</sup>Ranked list of relevant and diverse news articles: <https://git.io/JttGA>

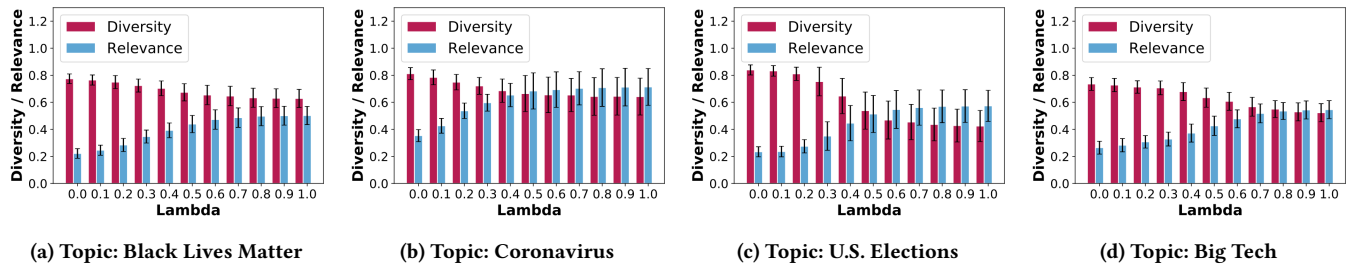
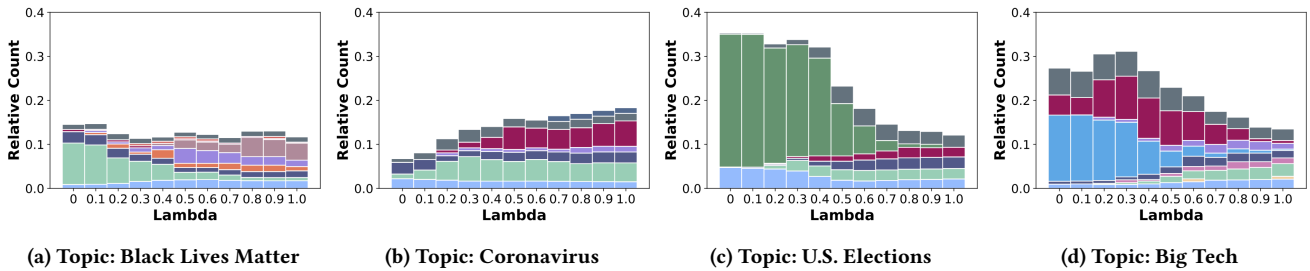
Figure 2: Diversity and relevance scores for different values of  $\lambda$  per topic.

Figure 3: Average number of publishers in recommendation lists, normalised by the input ratio, for all topics.

**Table 3: Overview of model variables used during the offline and online evaluation for each topic: cross validation folds ( $k$ ), recommended list size ( $s$ ), number of introductory paragraphs, number of concluding paragraphs, general weights for the four framing aspects, category weights and  $\lambda$ .**

Topic	$k$	$s$	intro. par.	concl. par.	general weight	cat. weight	$\lambda$
Black Lives Matter	10	3	2	1	[0.2, 0.4, 0.1, 0.3]	eq	0
Coronavirus	10	3	2	1	[0.1, 0.4, 0.1, 0.4]	eq	0
U.S. Elections	10	3	1	2	[0.1, 0.4, 0.1, 0.4]	eq	0
Big Tech	10	3	1	2	[0.2, 0.4, 0.1, 0.3]	asc	0

*Kendall's  $\tau$ .* We computed the Kendall's  $\tau$  rank correlation to assess whether the proposed diversification method is capable of providing different recommendation lists compared to the baseline. We computed the coefficient between the baseline ( $\lambda = 1$ ) and each other value of  $\lambda = [0.0, 0.1, \dots, 0.9]$ . Overall, we observed that the re-ranking of the set of recommendations based on viewpoint diversity results in different recommendation lists compared to the baseline. The coefficient decreases for smaller values of  $\lambda$ , but it is bounded around  $\tau = 0$  for decreasing values of  $\lambda$ .

*Average number of words.* We observe no consistent pattern in the average number of words for different values of  $\lambda$  across topics. For the *Black Lives Matter* and *Big Tech* topics, the average number of words increases for larger values of  $\lambda$ , for the *U.S. Elections* topic the average decreases and for *Coronavirus* the average is stable.

*Publisher ratio.* Figure 3 shows the average number of articles in the recommended lists, normalized by the input ratio, for each value of  $\lambda$ . For every topic, the number of publishers increases for larger values of  $\lambda$  and the number of different publishers for the baseline

recommendation list is larger than the one in the diverse recommendation list. Thus, we observe that the diversification method influences the publisher ratio. For small  $\lambda$  values, some publishers get amplified, while others are excluded. We see this effect primarily for the topics of *U.S. Elections* and *Big Tech*. *Coronavirus* seems to be the only exception. We further discuss this in Section 8.1.

## 7 ONLINE STUDY

We conducted a between-subjects online study on the Blendle platform to compare the reading behaviour of users who receive news articles optimized only for relevance, versus news articles that are also diverse on viewpoint.<sup>7</sup>

### 7.1 Materials

In the online study, we used the articles collected in Section 4.

### 7.2 Participants

We selected 2076 active users of the news aggregator platform. These users were assumed to most likely see and use the recommendation functionality. We included only users who clicked at least four times on a recommended article below any article read, in the last 14 days before the study. Groups for baseline and diversified recommendations were created by randomly splitting the users.

### 7.3 Independent Variables

In the between-subjects user study we manipulated the following conditions, referring to the recommended list of news articles:

- **baseline recommendation:** was implemented using a MMR that was based only on relevance ( $\lambda = 1.0$ ).

<sup>7</sup>Online study interface and setup: <https://git.io/JttGb>



- **diversified recommendation:** was implemented using a MMR that maximized viewpoint diversity ( $\lambda = 0.0$ ).

## 7.4 Procedure

During the two-week experiment, six days per week, we provided recommendations for two articles featured on the selected users' homepage. We provided sets of three recommendations below the content on the reading page of the original article. Every morning, we chose these two articles manually, to match the selected topics (*Black Lives Matter*, *Big Tech*, *Coronavirus*, and *U.S. Elections*). Afterwards, both the baseline and diversified recommendation sets were calculated for both articles and shown on the Blendle platform.

## 7.5 Dependent Variables

To analyze the reading behaviour of the two different user groups and answer **RQ1**, we measure specific events on the news aggregator platform (*i.e.*, check whether the user opened the article and if the user finished reading the article). Based on these available events, we observe multiple implicit (click-through-rate per news article, click-through-rate per recommendation set and completion rate of recommendation); and explicit (heart ratio) measures of the reading behaviour. To answer **RQ2**, we look into presentation characteristics of the recommended articles (*i.e.*, presence of editorial title, presence of thumbnail and counting number of hearts).

1. *Click-through rate per article:* The number of clicks on a news article is divided by the total number of users who finished one of the original news articles for which that article was recommended. The completion of an original news article is registered using a scroll-position.

2. *Click-through rate per recommendation set:* The total number of clicks on either of the three news articles in the recommendation set is divided by the number of users who finished the original news article (using scroll-position) for which the recommendation set was presented.

3. *Completion rate of recommendation:* Is implemented as the number of users that read the full recommended article (using scroll-position) divided by the number of users who opened the news article. The completion rate is assumed to be a measure for the user satisfaction with the recommendations. We can argue that short news articles are more likely to be completed than long news articles. Thus, we also analyze the completion rate of a news article in relation to the number of words in the news article.

4. *Favourite ratio:* The news aggregator platform allows users to mark an article as a favourite, illustrated by an icon of a heart. The users can click this icon at the end of the article content. We implemented the measure as the number of users of the user group (baseline or diverse) that clicked on the icon, divided by the number of users in the same group that completed the article. The metric is assumed to be a marker of user satisfaction with the article.

5. *Presentation characteristics:* We measured three additional properties of a recommended article during the experiment, which referred to the presentation characteristics of recommended news articles. First, the editorial team can replace the original title of a news article with a custom, editorial title. In general, these custom titles are longer and more explanatory than the original ones. Second, articles can be presented with or without a thumbnail image.

Third, the number of users who selected the article as a favourite is visualised by a counting number of hearts in the left-upper corner of an article banner. All three properties are assumed to potentially influence the click-through rate and are, therefore, measured during the experiment.

6. *Source diversity:* Finally, we also measured the influence of the source diversity of the recommendation set on the click-through rate. As seen in Section 6, higher levels of viewpoint diversity showed to influence the number of times a publisher is included in the recommendation.

## 7.6 Results

The online study ran six days a week for two weeks. Thus, we provided recommendations below 24 articles. During the experiment, the topic of *Coronavirus* became extremely prominent, so we provided recommendations below 18 out of 24 news articles on this topic. In contrast, the *Black Lives Matter* topic lost all actuality, resulting in no recommendations for this topic. For the *U.S. Elections* topic, we provided recommendations below four articles, and for the *Big Tech* topic, below two news articles.

*Click-through rate per recommended article.* The mean click-through rate per recommended article for the baseline was 0.11 (stderr. = 0.011) while for the diversified recommendations was 0.087 (stderr. = 0.0083), when looking at all topics. Furthermore, c.f. Mann-Whitney U test ( $U=570$ ,  $p\text{-val}>0.05$ ), we did not find a significant difference between the two user groups in terms of click-through rate per recommended article (same result per topic).

*Click-through rate per recommended set.* The mean click-through rate per recommended set for the baseline recommendations was 0.31 (stderr. = 0.016) while for the diversified recommendations was 0.25 (stderr. = 0.016) when looking at all topics (Figure 4a). According to the Mann-Whitney U test ( $U=2.9$ ,  $p\text{-val}<0.05$ ), we find a significant difference between the mean click-through rate per recommended sets for the two user groups. Per topic, we only find a significant difference for *Coronavirus*, Figure 4b, with a click-through rate per recommended set of 0.32 (stderr. = 0.018) for the baseline and 0.25 (stderr. = 0.018) for the diversified recommendations ( $U=80.0$ ,  $p\text{-val}<0.05$ ). For the other topics, we found no significant difference between the two user groups.

*Completion rate.* We found no significant difference in terms of completion rate for the two users groups, c.f. Mann-Whitney U test ( $U=600.0$ ,  $p\text{-val}>0.05$ ). We also applied the Spearman's rank correlation to see whether the completion rate is correlated with the length of the articles. However, we found no correlation in either of the two conditions (baseline group:  $\rho=-0.26$ ,  $p\text{-val}>0.05$ ; diverse group:  $\rho=-0.19$ ,  $p\text{-val}>0.05$ ).

*Heart ratio.* We found no significant difference, c.f. Mann-Whitney U test, for all topics and across topics ( $U=580.0$ ,  $p\text{-val}>0.05$ ), in terms of heart ratio for the two user groups. Thus, the recommendations quality was comparable between the two conditions.

7.6.1 *Influence of presentation characteristics.* We measured the influence of three factors, namely the presence of an editorial title, the presence of a thumbnail and the number of users that chose the article as a favourite on the click-through rate of an article.

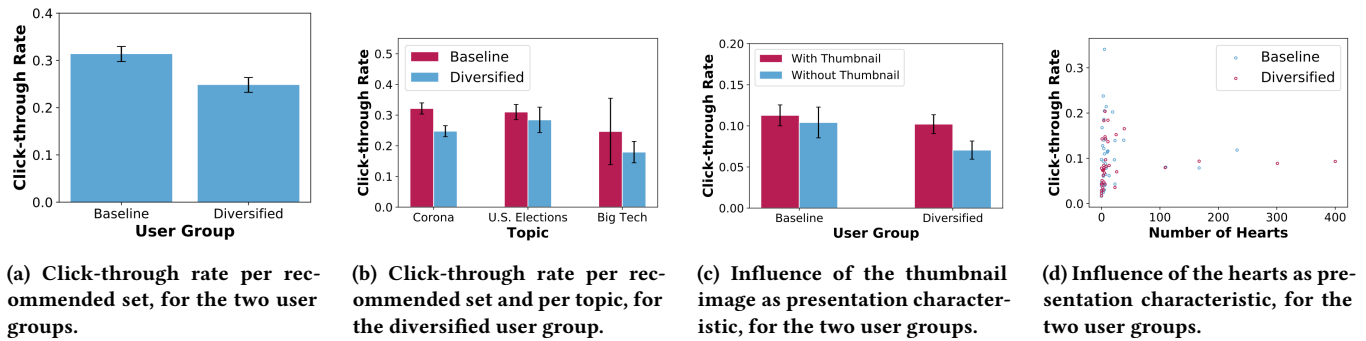


Figure 4: Overview of significant results in the online study

*Editorial title.* Regarding the influence of the inclusion of an editorial title on the click-through rate, no statistical significance was found for neither user groups, c.f. Mann-Whitney U test (baseline:  $U=57.0$ ,  $p\text{-val}>0.05$ ; diverse:  $U=43.0$ ,  $p\text{-val}>0.05$ ).

*Thumbnail image.* We found no statistically significant influence of the inclusion of a thumbnail image on the click-through rate for baseline users, c.f. Mann-Whitney U test ( $U=150.0$ ,  $p\text{-val}>0.05$ ). In contrast, we found a statistically significant difference for diverse users ( $U=88.0$ ,  $p\text{-val}<0.05$ ). Recommendations with a thumbnail are 3.1% more times opened than recommendations without a thumbnail for diverse users, as seen in Figure 4c.

*Favorite articles.* We applied the Spearman’s rank correlation to see whether we find a correlation between the click-through rate and the number of hearts. Figure 4d shows the distribution of click-through rates and the number of hearts. We only found a moderate positive correlation of 0.57, also statistically significant ( $p\text{-val}<<0.05$ ) for the diversified user group.

**7.6.2 Source diversity.** As seen in the offline evaluation, higher levels of viewpoint diversity turned out to have remarkable effects on the publisher ratio. Therefore, we also evaluated the effect of the source diversity of a recommendation set on the click-through rate. For each recommendation set, we computed the number of different publishers and we found recommendation sets in which all articles are from a different publisher and sets in which two articles are from the same publisher. Afterwards, the click-through was calculated for each category. The results for both the baseline users and diverse users show that no statistically significant difference can be found in the click-through rate between two or three different publishers in the recommendation set for neither baseline nor diversified users.

## 8 DISCUSSION

We first discuss the results of the offline and online evaluation and then provide an overview of the limitations of our approach. We conclude with directions for future work.

### 8.1 Offline Evaluation

The offline evaluation indicated that the proposed method is capable of increasing the viewpoint diversity of recommendation sets according to the metric defined in previous literature [36]. The average viewpoint diversity scores across all topics increased from 0.55

to 0.79 for an increasing level of diversity in the MMR algorithm. Simultaneously, the average relevance score decreased from 0.58 to 0.27. Remarkably, the diversity score of 0.41 in [36] is considerably smaller than the maximum average value of 0.79 found in this work. A possible factor could be that in [36], the LDA topic model was excluded from the diversification method to prevent any interference with the evaluation metric, whereas the diversification method in this work still depends on an LDA topic-model. Therefore, the difference in viewpoint diversity scores between the methods can possibly appear due to the interference of metadata between the viewpoint diversity metric and diversification method in this work.

In the offline evaluation, we saw that the *publisher ratio* decreases for larger diversity in the MMR (lower values of  $\lambda$ ). We found that the maximum frequency of an article in the recommendation lists is around 2 to 4 times higher at  $\lambda = 0$ , compared to  $\lambda = 1$ . Thus, for larger contributions of diversity, the algorithm increasingly selects the same article recommendation. This suggests that some news articles in the dataset get amplified, thereby suppressing the inclusion of different sources. To study this effect thoroughly, the offline evaluation should allow for the assessment of the contribution of individual framing aspects to the global viewpoint diversity score, per article. These results also suggest caution for very low values of  $\lambda$ , since this may consistently amplify single articles representing particularly diverse (and potentially more extreme) views.

Although approaches that use source diversity are more popular, scholars generally agree that viewpoint diversity can only be achieved by fostering content diversity, because, multiple sources can still refer to the same point of view [40]. Based on these findings, this study used a content-based approach. In the offline evaluation, it became clear that increasing levels of content diversity exclude multiple publishers and thus, decreases source diversity. Moreover, some specific publishers got amplified remarkably for high levels of content diversity. Thus, viewpoint diversification methods could benefit from considering both content and source diversity.

### 8.2 Online Evaluation

No major influence of viewpoint diversification on the reading behaviour was found, except for the click-through rate calculated per recommendation set, which indicated a statistically significant difference between baseline and diverse users of 6.5% (in favour

for baseline recommendations). However, the results of the click-through rate calculated per recommendation indicated no significant difference between the two user groups. Likewise, the other two measurements of the reading behaviour, including the completion rate of recommendations and the ratio of users who selected a recommendation as a favourite, showed no significant difference between baseline and diverse users.

In reflection on the motivation of this study, the proposed diversification for news media is capable of enhancing the viewpoint diversity of news recommendation, while maintaining comparable measures of the reading behaviour of users. The results thus suggest that recommender systems are capable of preserving the quality standards of multiperspectivity in online news environments. Thereby, situations of extreme low diversity, known as filter bubbles, could also be mitigated.

These results are in contrast with the most comparable study, which found a negative effect on intent to read diversified news articles [36]. The authors proposed a viewpoint diversification method based on the MMR-algorithm with linguistic features, such as gravity, complexity, and emotional tone. During a user study, 15 participants were asked to make a forced choice between a recommendation from the diverse set and a recommendation from the baseline set, after reading an article on the same topic. It was found that 66% of the participants chose the baseline article, compared with 33% who chose the diverse article. However, in the current study, we observed the actual reading behaviour of both user groups and argue that the present setup holds a higher level of ecological validity.

Additionally, the results shed light on the importance of how a recommendation is presented. Multiple presentation properties, such as the inclusion of a thumbnail image and the number of times an article is marked as favourite, were shown to have a significant influence on the click-through rate of recommendations. Future research, thus, should not only address the capability of a model to enhance viewpoint diversity according to an offline metric but also evaluate what presentation characteristics could impact the users' willingness to read multiperspectival news. Related research on *viewpoint-aware interfaces*, which aim to explain the recommendation choices to users [25, 35], and research on interfaces for cognitive bias mitigation in recommender systems [32] are valuable.

### 8.3 Limitations

We further discuss the limitations of our approach.

*Choice of participants in the online study.* Only users who frequently followed recommendations below articles were selected for the experiment. Thus, the click-through rates presented in this study are higher than for average news readers.

*Limited number of topics and articles.* For both the online and offline evaluations, we used only opinion pieces. Furthermore, each evaluation had a limited number of topics, namely four, as well as a limited number of news articles. New topics could reveal additional results that hold across topics.

*Missing user perceptions.* While we could study user behavior at a reasonable scale, a notable omission is users' qualitative judgement of viewpoint diversity in the resulting recommendations. We plan to continue collaborating with the news aggregator platform to refine the proposed framework, *i.e.*, to improve viewpoint extraction.

*Presentation characteristics.* Some presentation characteristics, and in particular the heart ratio, could also be markers of quality. Further qualitative analysis is needed to *e.g.*, understand how much of user behavior is directed by quality. We also saw that for some topics the presence of thumbnail was more common than for other topics, and it would be relevant to study whether this also interacted with user perceptions of relevance or quality.

*Relevance metric.* The offline study could use a more sophisticated relevance measure between the recommendation and the original article. The relevance score was based on a simple TF-IDF score, limited to the terms in a handcrafted search query.

*Influence of  $\lambda$ .* Given limited time for online testing, we only compared against a maximum viewpoint diversity score.

*Influence of publishers.* In Figure 3 we see that, although 15 publishers are represented in the datasets, three publishers are predominant. Due to the limited number of articles and the unbalance in terms of publishers, the inclusion of a wide variety of perspectives on a topic can be challenged.

### 8.4 Future Work

As future work, we plan to investigate further the presentation characteristics and how they influence user experience, in addition to behaviour. In more controlled settings, we will study the relative effects of actual (*e.g.*, as judged by experts) versus perceived quality (*e.g.*, number of hearts in the interface) of recommended news items. Future work will also focus on defining a better metric to measure viewpoint diversity, as opposed to topic diversity, *c.f.*, [11]. Additionally, we learnt that contextual information, general knowledge about a topic (*e.g.*, current measures in place to stop the virus spread) can also be essential to reveal a specific frame. We hope this work will encourage further research on defining, conceptualizing, and evaluating framing in the computational domain.

## 9 CONCLUSIONS

In this paper, we proposed a novel method for enhancing the diversity of *viewpoints* in lists of news recommendations. Inspired by research in communication science, we identified frames as the most suitable conceptualization for news content diversity. We operationalized this concept as a computational measure, and we applied it in a re-ranking of topic relevant recommended lists, to form the basis of a novel viewpoint diversification method.

In an *offline evaluation*, we found that the proposed method improved the diversity of the recommended items considerably, according to a viewpoint diversity metric from literature. We also conducted an *online study* with more than 2000 users, on the Blendle platform, a Dutch news aggregator. The reading behaviour of users receiving diversified recommendations was largely comparable to those in the baseline. Besides, the results suggest that presentation characteristics (thumbnail image, and the number of hearts) lead to significant differences in reading behaviour. These results suggest that research on presentation aspects for recommendations may be just as relevant as novel viewpoint diversification methods, to achieve multiperspectivity in automated online news environments.

*Acknowledgments.* We would like to thank the team at Blendle for their support in this research, as well as the anonymous users that participated in our user study.

## REFERENCES

- [1] Christian Baden and Nina Springer. 2014. Com (ple) menting the news on the financial crisis: The contribution of news users' commentary to the diversity of viewpoints in the public debate. *European journal of communication* 29, 5 (2014), 529–548.
- [2] Christian Baden and Nina Springer. 2017. Conceptualizing viewpoint diversity in news discourse. *Journalism* 18, 2 (2017), 176–194.
- [3] C Edwin Baker. 2001. *Media, markets, and democracy*. Cambridge University Press.
- [4] Gregory Bateson. 1955. A theory of play and fantasy; a report on theoretical aspects of the project of study of the role of the paradoxes of abstraction in communication. *Psychiatric research reports* 2 (1955), 39–51.
- [5] W Lance Bennett. 1996. An introduction to journalism norms and representations of politics. (1996).
- [6] Rodney Benson. 2009. What makes news more multiperspectival? A field analysis. *Poetics* 37, 5-6 (2009), 402–418.
- [7] Björn Burscher, Daan Odijk, Rens Vliegthart, Maarten De Rijke, and Claes H De Vreese. 2014. Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures* 8, 3 (2014), 190–206.
- [8] 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.
- [9] Jihyang Choi. 2009. Diversity in foreign news in US newspapers before and after the invasion of Iraq. *International Communication Gazette* 71, 6 (2009), 525–542.
- [10] Claes H De Vreese. 2005. News framing: Theory and typology. *Information design journal & document design* 13, 1 (2005).
- [11] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. 2020. Assessing Viewpoint Diversity in Search Results Using Ranking Fairness Metrics. In *BLAS Workshop in association with ECMLPKDD'2020*.
- [12] Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication* 43, 4 (1993), 51–58.
- [13] William A Gamson and Andre Modigliani. 1989. Media discourse and public opinion on nuclear power: A constructionist approach. *American journal of sociology* 95, 1 (1989), 1–37.
- [14] Herbert J Gans. 2003. *Democracy and the News*. Oxford University Press.
- [15] Todd Giltin. 1980. *The whole world is watching: Mass media in the making and unmaking of the new left*. McGraw-Hill.
- [16] Esther Greussing and Hajo G Boomgaarden. 2017. Shifting the refugee narrative? An automated frame analysis of Europe's 2015 refugee crisis. *Journal of Ethnic and Migration Studies* 43, 11 (2017), 1749–1774.
- [17] Natali Helberger. 2019. On the democratic role of news recommenders. *Digital Journalism* 7, 8 (2019), 993–1012.
- [18] Paul Jaccard. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat* 37 (1901), 547–579.
- [19] Maurice George Kendall. 1948. Rank correlation methods. (1948).
- [20] Anne C Kroon, Alena Kluknavska, Rens Vliegthart, and Hajo G Boomgaarden. 2016. Victims or perpetrators? Explaining media framing of Roma across Europe. *European Journal of Communication* 31, 4 (2016), 375–392.
- [21] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems—A survey. *Knowledge-Based Systems* 123 (2017), 154–162.
- [22] Andrea Masini, Peter Van Aelst, Thomas Zerback, Carsten Reinemann, Paolo Mancini, Marco Mazzoni, Marco Damiani, and Sharon Coen. 2018. Measuring and explaining the diversity of voices and viewpoints in the news: A comparative study on the determinants of content diversity of immigration news. *Journalism Studies* 19, 15 (2018), 2324–2343.
- [23] Jörg Matthes and Matthias Kohring. 2008. The content analysis of media frames: Toward improving reliability and validity. *Journal of communication* 58, 2 (2008), 258–279.
- [24] Denis McQuail. 1992. *Media performance: Mass communication and the public interest*. Vol. 144. Sage London.
- [25] Sayooran Nagulendra and Julita Vassileva. 2014. Understanding and controlling the filter bubble through interactive visualization: a user study. In *Proceedings of the 25th ACM conference on Hypertext and social media*. 107–115.
- [26] Philip M Napoli. 1999. Deconstructing the diversity principle. *Journal of communication* 49, 4 (1999), 7–34.
- [27] Sapna Negi. 2019. *Suggestion mining from text*. Ph.D. Dissertation. NUI Galway.
- [28] Sapna Negi, Kartik Asooja, Shubham Mehrotra, and Paul Buitelaar. 2016. A study of suggestions in opinionated texts and their automatic detection. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*. 170–178.
- [29] Nic Newman, David AL Levy, and Rasmus Kleis Nielsen. 2015. *Reuters Institute digital news report 2015: Tracking the future of news*. Reuters Institute for the Study of Journalism.
- [30] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- [31] Mauro P Porto. 2007. Frame diversity and citizen competence: Towards a critical approach to news quality. *Critical Studies in Media Communication* 24, 4 (2007), 303–321.
- [32] Alisa Rieger, Mariët Theune, and Nava Tintarev. 2020. Toward Natural Language Mitigation Strategies for Cognitive Biases in Recommender Systems. In *2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence 2020*. Association for Computational Linguistics (ACL).
- [33] Dietram A Scheufele. 1999. Framing as a theory of media effects. *Journal of communication* 49, 1 (1999), 103–122.
- [34] Jesper Strömbäck. 2005. In search of a standard: Four models of democracy and their normative implications for journalism. *Journalism studies* 6, 3 (2005), 331–345.
- [35] Nava Tintarev. 2017. Presenting diversity aware recommendations: Making challenging news acceptable. (2017).
- [36] Nava Tintarev, Emily Sullivan, Dror Guldin, Sihang Qiu, and Daan Odijk. 2018. Same, same, but different: algorithmic diversification of viewpoints in news. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. 7–13.
- [37] Jan Van Cuilenburg. 1999. On competition, access and diversity in media, old and new: Some remarks for communications policy in the information age. *New media & society* 1, 2 (1999), 183–207.
- [38] Saül Vargas and Pablo Castells. 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 109–116.
- [39] Rens Vliegthart. 2012. Framing in mass communication research—An overview and assessment. *Sociology Compass* 6, 12 (2012), 937–948.
- [40] Paul S Voakes, Jack Kapfer, David Kurpius, and David Shano-yeon Chern. 1996. Diversity in the news: A conceptual and methodological framework. *Journalism & Mass Communication Quarterly* 73, 3 (1996), 582–593.
- [41] Hong Tien Vu and Nyan Lynn. 2020. When the news takes sides: Automated framing analysis of news coverage of the Rohingya crisis by the elite press from three countries. *Journalism Studies* (2020), 1–21.
- [42] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM conference on Recommender systems*. ACM, 123–130.
- [43] 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*. ACM, 22–32.