NeMig - A Bilingual News Collection and Knowledge Graph about Migration

Andreea Iana[1,*], Mehwish Alam[2], Alexander Grote[3], Nevena Nikolajevic[3], Katharina Ludwig[4], Philipp Müller[4], Christof Weinhardt[3] and Heiko Paulheim[1]

[1]*Data and Web Science, University of Mannheim, Germany*

[2]*Télécom Paris, Institut Polytechique de Paris, France*

[3]*Karlsruhe Institute of Technology, Germany*

[4]*Institute for Media and Communication Studies, University of Mannheim, Germany*

Abstract

News recommendation plays a critical role in shaping the public's worldviews through the way in which it filters and disseminates information about different topics. Given the crucial impact that media plays in opinion formation, especially for sensitive topics, understanding the effects of personalized recommendation beyond accuracy has become essential in today's digital society. In this work, we present *NeMig*, a bilingual news collection on the topic of migration, and corresponding rich user data. In comparison to existing news recommendation datasets, which comprise a large variety of monolingual news, NeMig covers articles on a single controversial topic, published in both Germany and the US. We annotate the sentiment polarization of the articles and the political leanings of the media outlets, in addition to extracting subtopics and named entities disambiguated through Wikidata. These features can be used to analyze the effects of algorithmic news curation beyond accuracy-based performance, such as recommender biases and the creation of filter bubbles. We construct domain-specific knowledge graphs from the news text and metadata, thus encoding knowledge-level connections between articles. Importantly, while existing datasets include only click behavior, we collect user socio-demographic and political information in addition to explicit click feedback. We demonstrate the utility of NeMig through experiments on the tasks of news recommenders benchmarking, analysis of biases in recommenders, and news trends analysis. NeMig aims to provide a useful resource for the news recommendation community and to foster interdisciplinary research into the multidimensional effects of algorithmic news curation.

Keywords

news corpora, user data, recommender system, bias analysis, aspect diversity, knowledge graph

1. Introduction

The large volume of news published online each day creates an information overload that exceeds the consumptive capacities of readers. At the same time, the digitalization of news consumption has sparked an unprecedented development of personalized recommendation

algorithms, used by online news platforms to process the continuously growing quantities of news and offer readers personalized suggestions. However, models which are optimized to maximize congruity to users' preferences and past behavior, tend to produce recommendations which are highly similar in content to previously read/clicked ones [1, 2]. By filtering out news deemed irrelevant to the user, news recommenders control how and which information is disseminated to the public [3]. Consequently, algorithmic news curation has a large impact on opinion formation and voting behavior [4, 5], as well as the potential to fuel conflicts and polarization [6, 7, 8, 2, 9], especially when dealing with controversial subjects such as war, climate change, or migration. In addition to the power of recommender systems to shape people's perception of the world, the fundamental role that media plays in today's society as a public forum [10, 11], have deemed analyzing and understanding the implications of news curation algorithms necessary for modern democracies.

While plethora of recommendation algorithms have been proposed in recent years, resources available to analyze their effects beyond accuracy performance and in multilingual scenarios are scarce. The existing body of work exhibits two main shortcomings: (1) news datasets for recommendation (i) mostly focus on general or less sensitive topics (e.g., entertainment, fashion, sports) [12, 13] and (ii) leave largely unexplored features which are critical for analyzing the algorithmic creation of "filter bubbles" [3] or underlying recommender biases (e.g., sentiment or political orientation); (2) news collections for media analysis (e.g., fake news detection [14, 15], news narratives analysis [16, 17], news bias detection [18, 19]) lack user information, and thus, cannot easily be used in a recommendation scenario.

In this paper, we introduce *NeMig* a bilingual news collection and knowledge graphs (KGs) about refugees and migration, in German and English, as well as corresponding real-world user data. We collect over 7K, and respectively, 10K articles from German and US news outlets. The articles span a large political spectrum in order to cover various viewpoints on the topic. **1)** Compared to existing resources, NeMig (i) targets the same polarizing topic in two languages and (ii) is annotated with the sentiment polarization of articles and political leanings of the media outlets, in addition to subtopics and disambiguated named entities. **2)** We construct corresponding knowledge graphs (KGs) from the news' text, metadata, and extracted named entities, which we further expand with up to two-hop neighbors from Wikidata of the named entities. We provide the KGs in different variants. **3)** We collect and publish in an anonymized fashion both explicit feedback and socio-demographic data for 3K users for each of the two datasets. Although our user dataset is small compared to large benchmark news datasets, it contains information about the users' media consumption, political attitudes and interests, personality traits and demographics, which makes the dataset valuable for studying the effects of recommender systems beyond accuracy-based performance, e.g., on political polarization. Furthermore, the user data constitutes a starting point for the generation of synthetic user datasets, which comprise not only click behavior information, but also explicit data about the users' background and preferences.

We demonstrate the utility of the resource through experiments on different downstream tasks. Firstly, we benchmark several news recommenders on NeMig. Secondly, we investigate biases and polarization effects of the benchmarked recommendation models, and evaluate the contribution of various features to the quality of the KG. Lastly, we analyze news trends by examining the evolution over time and political orientation of the most frequent entities in our

corpora, identifying correlations between the patterns of evolution and worldwide events.

NeMig is available in two versions, for research purposes: i) we release the full news datasets in N-Triples format, under restricted access on Zenodo [20], as the news bodies can only be provided upon request due to copyright policies, ii) NeMig, with all features but without the news bodies, is freely available, under a CC BY-NC-SA 4.0 license, in tabular format on GitHub.[1] The user data is available in anonymized tabular format on both Zenodo [20] and GitHub.[2]

2. Related Work

News Recommendation Benchmarks. Various datasets have been constructed for developing and benchmarking news recommender systems (NRS). Plista [21] comprises a collection of over 70K German articles gathered from 13 news portals, as well as click data for over 14 million users. Globo [22, 23] is a Portuguese dataset for news recommendations collected from `Globo.com`, and contains data for over 300K users and 46K news distributed in 1.2 million sessions. The Norwegian dataset Adressa [12] contains news collected from the Adresseavisen's news portal and is provided in two variants: a light version with over 11K articles and click data for 561K users, and a larger one with more than 48K articles and clicks for 3 million users. In addition to metadata (e.g., categories, authors), the articles are annotated with named entities. MIND [13], the most recent large-scale dataset, is constructed from user click logs of Microsoft News. It covers 1 million users and more than 161K news articles in English. Moreover, the dataset is enriched with named entities linked to Wikidata.

News Knowledge Graphs. Knowledge graphs have been shown to model knowledge-level connections between news that cannot be captured by purely text-based models [24, 25]. In the context of the Newsreader[3] project, focusing on the multilingual processing of news articles, the authors generate an event-centric KG from a multilingual news dataset that describes which events took place, where, when, and who was involved in them [26, 27]. News Graph [28] is a graph constructed specifically for news recommendation. It is generated from a news corpus collected from MSN News containing over 621k articles, 594k news entities, as well as user-item interaction logs. The graph contains news content, user behaviors, and news topic entities enriched with neighboring triples from Microsoft Satori, along with three types of collaborative relations for entities, i.e., co-occurring in the same news, clicked by the same user, and clicked by the same user in the same browsing session. The use of semantic KGs for the production, distribution, and consumption of news is summarized in [29].

News Datasets for Media Discourse Analysis. The advent of news platforms as an ubiquitous means of information for Internet users has established online news as a fundamental source for various media discourse analysis tasks. For instance, Horne et al. [15] has created a large dataset of political news collected from mainstream and alternative sources, enriched with content-based and social media engagement features, that can be used for news and engagement characterization, news attribution, and content copying, or exploration of news narratives.

---

[1]https://github.com/andreeaiana/nemig_nrs/tree/main/data
[2]https://github.com/andreeaiana/nemig/tree/main/data/user_data
[3]http://www.newsreader-project.eu

Similarly, Lim et al. [19] and Färber et al. [18] collected and annotated articles with different types of bias at the sentence level for analyzing news biases. In a related line of work, researchers have designed datasets for monolingual [14, 16] and multilingual cross-domain [30] fake news detection, or for distinguishing between fake and satire news published online [31, 32].

Limitations of Current News Resources. Existing news datasets have several shortcomings. On the one hand, the large monolingual recommendation benchmarks benefit from sizable user interaction data, but fail to explore news (e.g., sentiment orientation, political leaning) or user features (e.g., demographics, political interests) which can be exploited in the analysis of recommender system effects. On the other hand, the news collections for media discourse analysis contain rich annotations (e.g., news biases, social media features, social context, spatio-temporal information), but are ill-equipped to be seamlessly utilized in training recommendation algorithms due to the lack of user feedback data. In contrast to these resources, NeMig comprises a (i) bilingual news collection, annotated not only with text and metadata information, but also with subtopic, sentiment and political orientation information, and (ii) implicit user feedback, demographic, and political data. Moreover, we use Wikidata as an open source for entity linking, and not a commercial knowledge base, as done in [28].

3. The NeMig News Corpora

3.1. Data Collection

We construct the news corpora by crawling news articles from 40 German[4], and respectively 45 US[5], legacy and alternative media outlets, selected by a team of researchers from the media and communication domain. The news are sampled from the entire political spectrum, such as to cover different viewpoints on the topic. We include German articles published between Jan. 1, 2019 and Dec. 31, 2021[6], and English ones between Jan. 1, 2021 and July 1, 2022[7]. For each language, we identify relevant articles using 13 keyword stems representative of the migration topic, such as *flüchtl**, *asyl**, or *migration** for the German outlets, and *refugee**, *asylum seeker**, or *migrant** for the US sources. Moreover, only articles written in either German or English (for the German and English corpora, respectively), with a minimum length of 150 words, and containing at least two keywords stems have been included in the two corpora in order to avoid foreign language articles, disclaimers, advertisements, or reader comments. From the resulting raw datasets, we further excluded duplicates, videos, live tickers (i.e., articles continuously updated with news headlines), and outliers[8] (i.e., articles which are abnormally long or short). The datasets contain 7,346 (German), and respectively 10,814 (English) filtered articles. Figs. 1a and 1b show the distribution of the articles over time and over media outlets. We observe that the news is unevenly distributed over various publishers, with smaller or niche media outlets being underrepresented. In comparison, with the exception of a few outlets, the articles are relatively evenly distributed over the different time periods.

---

[4]The implementation of the German news crawlers is available at https://github.com/andreeaiana/german-news
[5]The implementation of the English news crawlers is available at https://github.com/andreeaiana/us-news
[6]The data was collected in Dec. 2020 and Jan. 2021, and updated in Jan. 2022.
[7]The data was collected in Aug. 2022.
[8]Outliers were identified by using two standard deviations from the mean.

(a) German news corpus.
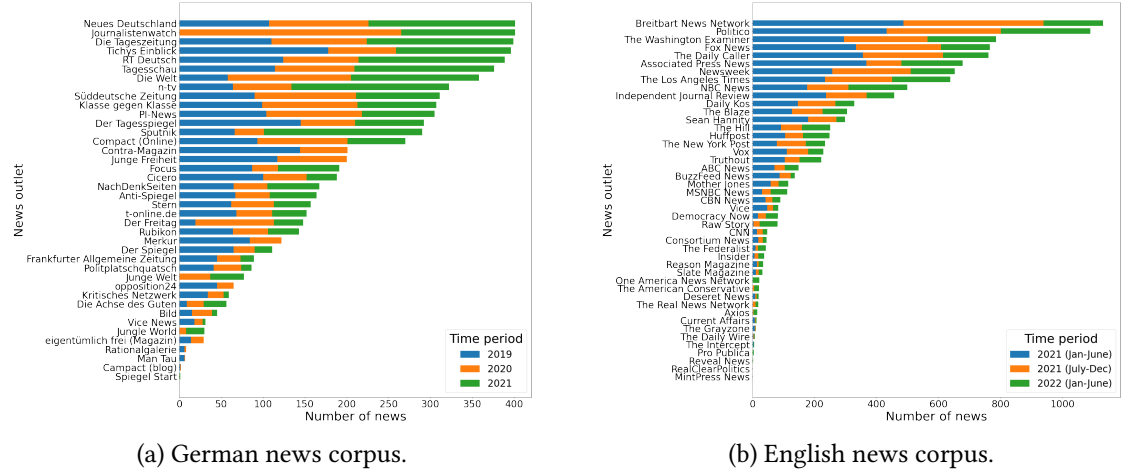
(b) English news corpus.

Figure 1: Distribution of news over time and over outlets.

Each news article contains a title, a body, and if available, an abstract. Additionally, its metadata specifies its provenance (news outlet, URL), publishing and modification dates, authors, and keywords provided by the media outlet. Furthermore, we classify the political leaning of the news outlets. Concretely, we categorize the US sources into three political orientation classes, namely *left*, *center*, and *right*, based on the AllSides Media Bias Chart[9]. Since no corresponding classification exists for the German media, the same researcher team classifies the outlets into four political orientation groups: *left*, *center*, *right*, *conspiracy*.[10] We find (see Fig. 2) that the distribution of news outlets over the political orientation classes differs between the German and the English datasets. While the majority of outlets for German are center media, the majority of outlets for English are left. On the article level, however, the distribution for the English dataset is drastically different and similar between the two corpora.

3.2. Data Annotation

Fig. 3 illustrates the corpora annotation and KG construction pipeline.[11] Given an input dataset, the majority of steps can be run end-to-end.[12]

Sentiment Analysis. Sentiment polarization information is leveraged by fairness-aware news recommenders to produce sentiment-diverse or agnostic recommendations [1, 33]. Furthermore, such information can be used to investigate whether algorithmic news curation propagates or encourages the sentiment polarization of readers [34, 3, 35] or, more generally, the sentiment bias of different news outlets [36, 37]. We use a multilingual *XLM-R* [38] language model,

---

[9]https://www.allsides.com/media-bias/media-bias-chart

[10]Disclaimer: We excluded the outlet *Man Tau* from further analysis regarding the political orientation as it cannot be categorized in any of the four groups, but we include the corresponding news in the dataset.

[11]The implementation of the annotation and KG construction pipeline, and the intermediary files produced by the annotation process are available at https://github.com/andreeaiana/nemig

[12]Sub-topic extraction constitutes the only exception, as it requires setting parameters (e.g., threshold for pruning topics) based on intermediary results. We run this process separately and integrate its output in the pipeline.
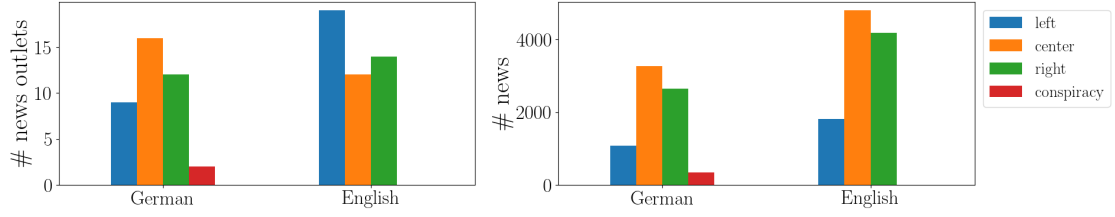
Figure 2: Distribution of outlets and news over political orientation classes.
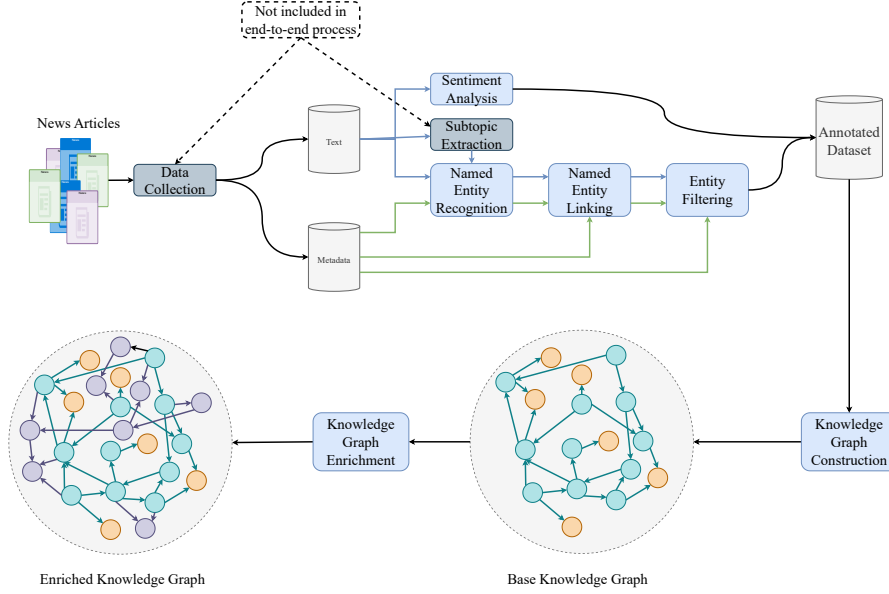

Figure 3: Corpora annotation and knowledge graph construction pipeline. Arrow colors denote the applicability of steps to: all components (black), text features (blue), and metadata (green). Node colors indicate: custom or linked resources (blue), literals (orange), and Wikidata neighboring nodes (purple). Edge colors denote properties in the original KG (blue), or extracted from Wikidata (purple).

trained on 198M tweets and fine-tuned for sentiment analysis [39], to classify news into one of three sentiment classes: positive, neutral, or negative. We perform sentiment classification on the concatenation of the news' title and abstract.[13] Table 1 shows the distribution of news over sentiment classes in the two datasets. For both corpora, only a small number of articles have a positive sentiment. Moreover, while the distribution of articles with neutral and negative sentiments is relatively balanced in the German corpus, the English corpus has significantly more negative articles.

Sub-topic Modeling. We observe that news cover discourses about refugee migration from different geographical areas or related to various political events. In order to extract sub-topics from each dataset, we use the neural topic modeling approach proposed in [40], with a pre-trained English Sentence Transformer for the English news, and multilingual Sentence

---

[13]Note that we use the concatenation of title and the article's first sentences as input when abstracts are not provided.

(a) German news corpus.
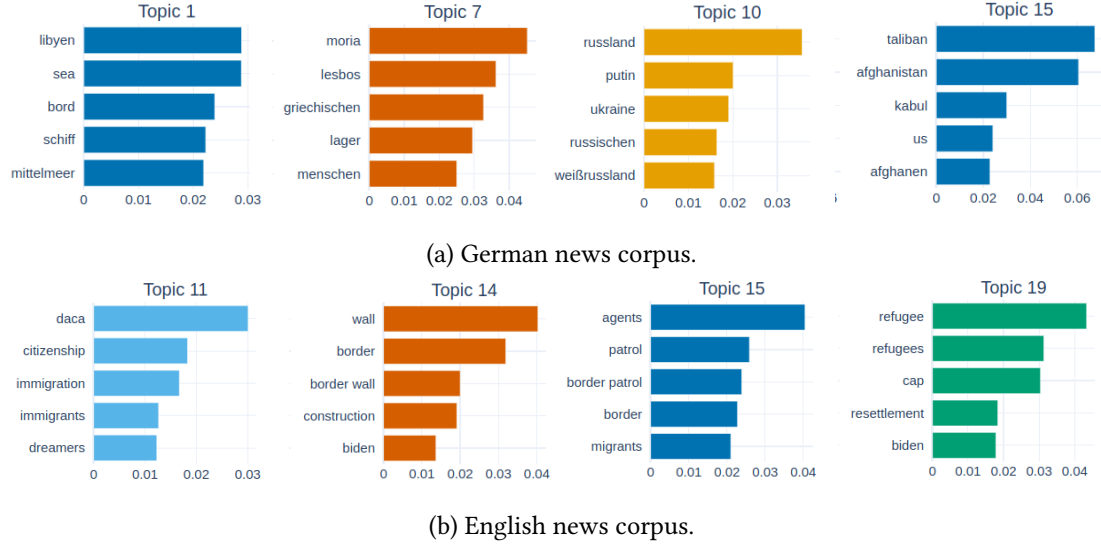

(b) English news corpus.

Figure 4: Examples of words assigned to different sub-topics. The X-axis indicates the c-TF-IDF scores [40] of the words representative of each sampled topic.

Transformer [41] for the German ones. We assign labels to the resulting topics based on the topmost representative terms of the respective cluster of documents. We group articles that cannot be assigned to any topic appearing in at least 15 news into a separate cluster. We extract 25 sub-topics from the German corpora, and 40 from the English one, respectively. We find, as shown in Fig. 4, that the sub-topics indicate migration from (e.g., Libya, Ukraine) or to (e.g., Greece, US) different areas, migration-related political events (e.g., Russian invasion of Ukraine, US army retreating from Afghanistan) or policies (e.g., US immigration and border control laws).

Named Entity Recognition. Events are generally described in news by means of named entities (NEs) that indicate what, when, and where it happened, or who was involved in it [42]. We extract NEs from both the textual content, and the metadata of articles, and classify them into four classes: persons (PER), organizations (ORG), locations (LOC), and miscellaneous (MISC). We use two *XLM-R* [38] language models to perform named entity recognition, one fine-tuned on the German [14] and the other on the English subset[15] of the CoNLL03 dataset [43]. Concretely, we input the title, abstract, or sentence-segmented news body [44] for text components, whereas for metadata, we use author names, lists of keywords from the news outlets, and those used to describe sub-topics. In the former case, we output only the NEs recognized in these original input. In the latter scenario, the output consists not only of NEs extracted by the model but also of the original inputs, where no entity was extracted. In the KG construction step, we include NEs extracted from metadata, as well as other information (e.g., frequent keywords, unrecognized authors) to model connections between articles.

Entity Linking. Next, we disambiguate the NEs through named entity linking to Wikidata [45] using a multilingual entity linking model [46], which is based on a sequence-to-sequence

---

[14]https://huggingface.co/xlm-roberta-large-finetuned-conll03-german
[15]https://huggingface.co/xlm-roberta-large-finetuned-conll03-english

Table 1

News corpora statistics.

		German	English
	raw	8,647	12,220
	filtered	7,346	10,814
# News	positive	144	590
	negative	3,476	2,933
	neutral	3,726	7,291
	title	7.9	11.8
Avg #words	abstract	29.4	24.9
	body	898.4	1,013.9

Table 2

Statistics of extracted named entities

	German			English		
	Total	Linked	Not linked	Total	Linked	Not linked
Title	1,075	1,075	0	1,542	1,542	0
Abstract	2,383	2,383	0	2,365	2,365	0
Body	19,683	19,683	0	33,857	33,857	0
Publishers	40	34	6	45	44	1
Authors	490	193	297	1,242	386	856
Keywords	4,481	2,395	2,086	5,175	2,636	2,539

architecture [47] that generates entity names in over 100 languages. As input, we use a text sequence containing one named entity annotated with special start and end tokens. The input sequence is either the title, abstract, or sentence-segmented news body, or for metadata, either the extracted entity from the previous step or the original keyword or author name. For both input types, the model outputs the linked entity, the corresponding Wikidata QID, the language of the Wikidata entity label, as well as a score. We select Wikidata as the external knowledge base for linkage and graph expansion due to (i) being open source, (ii) its wide coverage, and (iii) its up-to-dateness [48], which is particularly important in the context of news.

Entity Filtering. The previous step can yield both entities not identified in Wikidata although they exist (e.g., keyword *academia* corresponds to Wikidata QID *Q1211427*), as well as incorrect links. Therefore, we filter incorrectly extracted or linked entities. For NEs originating from textual components, we (1) threshold on the confidence score of the named entity recognition model; (2) remove entities without a Wikidata page or linked to a Wikimedia disambiguation page; (3) remove entities whose type does not correspond to the entity type in Wikidata based on type-specific properties[16]; (4) threshold on the score of the named entity linking model; (5) remove entities whose Wikidata label is in a foreign language and are only observed once in the dataset. For authors and keywords, we use similar filtering pipelines. Table 2 shows the statistics for the named entity recognition and linking steps.

4. Knowledge Graph Construction

Wikidata [45] is a large knowledge graph comprising a vast amount of information that is often used to model and integrate common-sense knowledge into recommender systems [49, 25]. Nevertheless, it also contains data which can be irrelevant for domain-specific recommendation

---

[16]We used the following properties to detect the type of an entity in Wikidata: PER is *instance of human*; ORG has *headquarters location, inception*, or is *founded by*; LOC has *coordinate location*.
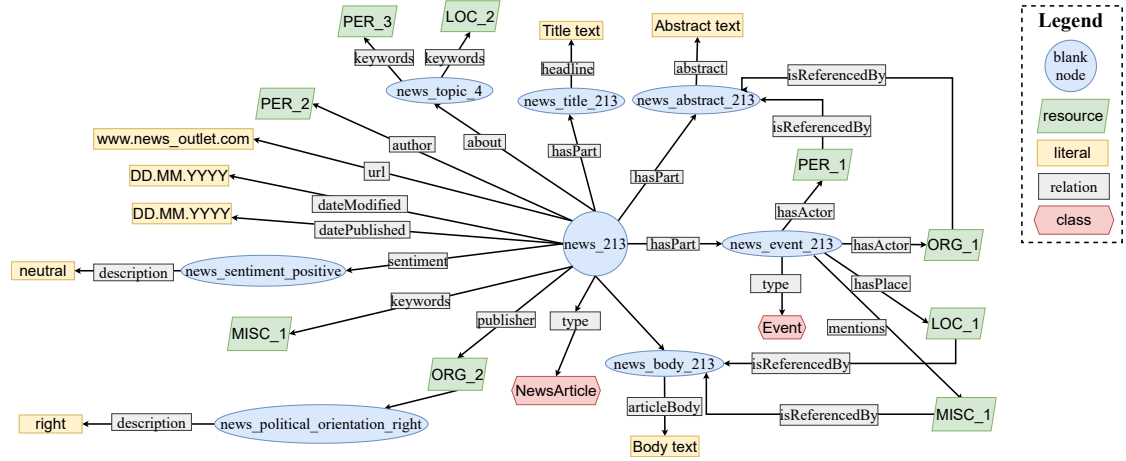
Figure 5: Example of an article's representation in NeMigKG. The node colors indicate: (i) blank nodes (blue); (ii) disambiguated or custom resources (green); (iii) literals (yellow); (iv) classes (red). The gray boxes denote relation types.

tasks. Thus, we build a domain-specific KG from the annotated corpora, combining textual content and metadata from news with Wikidata triples of the NEs and their *k*-hop neighbors.

4.1. Base graph construction

NeMigKG is a KG in which the nodes denote news content and real-world entities (e.g., authors, locations), while the edges indicate different relation types between them. Fig. 5 illustrates an example of an article in NeMigKG.

Relations. We model relations using several schemas and vocabularies. Specifically, we represent edges between articles and their textual content or metadata with *schema.org*[17] relationships, and those involving extracted NEs with the *Simple Event Model* [50]. Furthermore, we encode information about miscellaneous NEs with the *schema:mentions* relation, and map the provenance of NEs to the news text with the *dcterms:isReferencedBy*.[18] Additionally, we create the relations `nemig:sentiment` and `nemig:political_orientation` to represent edges between news and their sentiment labels, and between media outlets and their political leaning, respectively. Lastly, we identify pairs of related news[19], i.e., articles with identical titles, overlapping bodies, but different provenance, which we map with the *schema:isBasedOn* relation. We find that such related news are based on one another, with the most recent published one representing an update, extension, or longer discussion of the original article.

Nodes. NeMigKG contains two kinds of nodes: (i) *literals* (e.g., title, dates), and (ii) *resources* which encode NEs extracted from different parts of an article's content (e.g. persons, locations) or metadata (i.e., publisher, author). We further distinguish the latter into (i) disambiguated entities from Wikidata (denoted as *Wikidata resources*) and (ii) *custom resources* created from entities that

---

Table 3

Statistics of the German and English NeMigKG.

| | | German NeMigKG | | | | | English NeMigKG | | | | |
		Base	Entities	Enriched ($k=1$)	Enriched ($k=2$)	Complete	Base	Entities	Enriched ($k=1$)	Enriched ($k=2$)	Complete
Triples	Nodes	89,601	59,480	97,021	367,514	397,635	134,775	91,833	138,734	433,965	476,907
	Relations	19	11	768	1,172	1,180	19	11	828	1,195	1,203
	Triples	516,948	458,482	821,529	2,846,849	2,905,315	933,801	8 47,695	1,354,454	3,599,561	3,685,667
Nodes	% resources	66.38	100	100	100	92.42	68.14	100	100	100	91.00
	% literals	33.62	0	0	0	7.58	31.86	0	0	0	9.00
Resources	% blank	61.28	61.29	37.57	9.92	9.92	58.48	58.48	38.71	12.38	12.38
	% custom (not linked)	4.00	4.00	2.45	0.65	0.65	3.68	3.68	2.43	0.78	0.78
	% Wikidata (linked)	34.71	34.71	59.97	89.43	89.43	37.84	37.84	58.86	86.85	86.85

could not be identified or linked to Wikidata, but which still encode meaningful information and provide knowledge-level connections between news (e.g., authors not in Wikidata, or frequent keywords). We represent each article in NeMigKG with a unique identifier. For each title, abstract, body, topic, sentiment, political leaning, or event mapped in the KG, we create additional nodes, as shown in Fig. 5.

Enrichment with External Information. We extend NeMigKG with up to *2*-hop neighbors from Wikidata of all *Wikidata resources* already included in our graph to inject additional information from external knowledge bases. Firstly, we extract, for all *Wikidata resources*, all triples in Wikidata containing another entity as their tail. We ignore literal neighbors. Secondly, we incorporate the matching triples into the graph and retrieve a new neighbors set for the added tail entities. We repeat this process iteratively until we enrich NeMigKG with all relevant triples two hops away from the original *Wikidata resources*. We post-process the resulting graph by removing all sink entities, i.e., nodes with a degree smaller than two, which cannot be used to derive further connections between the news as they are leaf nodes or not linked to Wikidata.

4.2. Graph types

The majority of KG embedding models used in downstream tasks encode only entities and relations, while ignoring literals. Therefore, we create four variants of NeMigKG for each of the two datasets, as follows: (1) *base* - contains literals and entities from the corresponding news corpus; (2) *entities* - contains only resource nodes from the *base* graph, and no literals; (3) *enriched entities* - contains only resource nodes and *k*-hop triples from Wikidata; (4) *complete* - represents the *enriched entities* graph with the literals re-added. Table 3 shows the statistics of the different KGs for both languages.

5. User Data

We collect user data in terms of (i) explicit click feedback, (ii) demographics, and (iii) political information. For each corpus, we conduct one online user study aimed at measuring the political polarization effects of NRS. The participants are recruited through online-access panels and selected using a quote procedure to create a representative sample of German, and respectively, US Internet users aged 18 to 74. Among the German users, 50.2% are male and 49.6% are female, while 44.6% have an Abitur (passed the secondary school final examinations). 46.2% of the US users are male and 52.5% are female, and 33.2% have at least a high school degree.

Table 4

A sample of questions used during the online user study conducted in the US for collecting user data.

	Variable	Question	Answer Scale
Demographics	Gender	Please indicate your gender.	0 = male, 1 = female , 2 = other, 3 = no answer
	Nationality	What is your citizenship?	0 = U.S. citizenship, 1 = U.S. and another non-U.S. citizenship, 2 = Only non-U.S. citizenship, 3 = No Answer
Empathy	EMP1	How strongly do you agree with the following statements?	When someone else is feeling excited, I tend to get excited too.
	EMP2		It upsets me to see someone being treated disrespectfully.
Ideological Polarization	IPO1	In the following, you will see a series of opposing statements. Please indicate how strongly you agree or disagree with the statements. There are no right or wrong answers.	The U.S. should take in more refugees.
	IPO2		The U.S. has already taken in too many refugees.
	IPO5		Immigrants strive for peaceful cohabitation with U.S.-Americans.
	IPO6		Immigrants are hostile toward U.S.-Americans.
			7-point Likert scale: 1=Strongly disagree to 7=Strongly agree
Emotions	EMO1	Which emotions do you feel towards refugees and immigrants in the USA?	Anger
	EMO2		Fear
	EMO9		Compassion
	EMO13		Gratitude
Media Usage	MED1	I receive information about US-American politics via:	newspapers and magazines or their websites (e.g. New York Times, The Wallstreet Journal...)
	MED2		TV networks or their websites (e.g. Fox News, CNN...)
	MED4		Facebook (for political information)
			7-point Likert scale: 1=Never to 7=Very Often
Perceived Polarization	PRO1	How strongly do you agree or disagree with the following statements?	Democratic and Republican partisans in the U.S. are increasingly hostile to one another.
	PRO3		Democratic and Republican partisans in the U.S. are very polarized.
			7-point Likert scale: 1=Strongly disagree to 7=Strongly agree
Affective Polarization	REP	In the following we would like to know about your party identification. Please mark on the scale how warm or cold you feel towards the respective parties.	Republican
	DEM		Democrat
			0 (negative) to 100 (positive)
Political Scale	POL1	Where on the scale would you place your political point of view?	11-point Likert Scale: 1 = Left, 6 = Center, 11 = Right
Political Topics	POL2	Please indicate how strongly you agree or disagree with the following statements.	I am generally very interested in politics.
	POL3		I regularly inform myself about current political affairs in the U.S.
			7-point Likert scale: 1=Strongly disagree to 7=Strongly agree
Participation	PPA1	Please indicate how likely it is that you will engage in the following activities in the near future.	Participating in an online political discussion on the topic of immigration in the U.S.
	PPA6		Participating in a political demonstration on the topic of immigration in the U.S.
			7-point Likert scale: 1= Not likely at all to 7=Very likely
Prosocial Behavior	PRO3	Please indicate how strongly you agree or disagree with the following statements.	I am willing to make a one-time donation of _ $ for refugees in the U.S.
	PRO4		How often do you have professional contact (e.g. at work or at school) with immigrants?
			float number >= 0
			7-point Likert Scale: 1 = Never to 7 = Very often

In the study, participants are firstly asked to rate 10 articles as *worthy* or *not to be read further* (i.e., explicit feedback). We use the IDs of the news deemed worthy of reading to build the user *Click History*, which is later used to construct each user's profile. Afterwards, each participant is shown a list of six news (impressions), for which a binary rating is recorded. This process was repeated five times. We use the information gathered in the second step to construct the *Impressions* log. Similar to [13], we structure the anonymized user behaviors in the format *[impression ID, user ID, Click History, Impression Log]*, where the *Impression Log* contains the IDs of the news shown to the user and the label indicating whether the user clicked on them.

Before the study's stimulus phase, we collect information regarding the users' media usage (*MED[1-9]*), political attitudes (*POL[1-9]*), and empathy levels (*EMP[1-8]*). After the stimulus, we ask participants questions regarding their socio-demographic status (gender, age, qualification, nationality, born in Germany or the US, parents born in Germany or the US, income), emotions (*EMO[1-14]*), levels of ideological (*IPO[1-12]*), affective (i.e, identification with the main political parties), and perceived (*PPO*[1-6]) polarization, of political participation (*PPA[1-8]*), and of prosocial behavior (*PRO[1-5]*). The resulting user datasets contain information about 3,432 German and 3,000 US users, respectively. The German dataset has a sparsity of only 0.02%, meaning that out of the news included in the online user study, less than 1% were never seen by the participants, in either of the two study phases. In contrast, the English dataset is much sparser, with 9.27% of the news never interacted with by any of the users. Table 4 shows an example of questions used during the user data collection process in the US. Please refer to the project page[20] or Zenodo [20] for a full set of questions used in both studies, as well as an extensive explanation of each question, and the corresponding answer scale.

6. Experiments

We analyze the performance of different news recommenders on NeMig on a range of recommendation tasks. Afterwards, we perform news trends analysis using the named entities identified in the corpora.

---

[20]https://github.com/andreeaiana/nemig/tree/main/data/user_data

6.1. Recommendation Models

We benchmark various recommendation models, which differ in their news and user encoders.[21] The former converts input features (e.g., title, topics, entities) into a news embedding by contextualizing pretrained word embeddings [51]. The later aggregates the embeddings of the clicked news into a user-level representation.

- *NRMS* [52] embeds news titles with a two-layer encoder consisting of multi-head self-attention [53], followed by additive attention [54]; a similar architecture is used to learn user representations.
- *NAML* [55] uses a sequence of convolutional neural network (CNN) [56] and additive attention to learn representations from news titles, and abstracts, and additionally leverages categories through a linear category encoder; its user encoder consists of an additive attention layer.
- *MINS* [57] embeds titles and abstracts of news as NRMS [52], and categories through a linear embedding layer; it learns user representations with a combination of multi-head self-attention, multi-channel GRU-based recurrent network [58], and additive attention.
- *CAUM* [59] combines the text encoder used by NRMS [52] with an entity embedder composed of attention layers in order to learn news representations; it produces candidate-aware user representations with a candidate-aware self-attention network which models long-range dependencies between clicked news, conditioned on the candidate, combined with a candidate-aware CNN that captures short-term user interests from adjacent clicks, also conditioned on the candidate's content.
- *DKN* [24] is a knowledge-aware recommender which learns representations from news titles with a knowledge-aware CNN [56] over aligned embeddings of words and entities, and of users with a candidate-aware attention network.
- *TANR* [60] has the same architecture as NAML [55], but does not use a category embedding layer; additionally, it injects information on topical categories, by jointly optimizing the recommender for content personalization and topic classification.
- *SentiDebias* [33], built on the architecture of NRMS [52], addresses the problem of sentiment debiasing using adversarial learning to reduce the model's sentiment bias (originating from the user data) and generate sentiment-diverse recommendations.

We evaluate the aforementioned NRS not only in terms of the standard content personalization performance, but also w.r.t. aspect-based diversity and personalization of results, meaning that we investigate how diverse or faithful the recommendations are to the user's consumed news in terms of aspect A_p. Concretely, following [61], we define *aspect-based diversification* as the level of uniformity of an aspect's distribution among the recommended news. In contrast, we define *aspect-based personalization* as the level of homogeneity between a user's recommendations and clicked news w.r.t. the distribution of an aspect (e.g., sentiment). We experiment with three aspects: *political leaning*, *topical categories*, and the *sentiment* of news. Note that we use the political leaning of a news outlet as proxy for determining the political leaning of an article originating from that outlet.

---

[21]The implementation of the news recommenders is available at https://github.com/andreeaiana/nemig_nrs

6.2. Evaluation Metrics

We use the common metrics AUC, MRR, nDCG@3, and nDCG@6 to report content personalization performance. Following [61], we measure aspect-based diversity of recommendations at position k using the normalized entropy of aspects A_p's distribution in the recommendation list:

$$D_{A_p}@k = -\sum_{j \in A_p} \frac{p(j) \log p(j)}{\log(|A_p|)}, \tag{1}$$

where $|A_p|$ denotes the number of classes of aspect A_p. We evaluate aspect-based personalization[22] with the generalized Jaccard similarity [62]:

$$PS_{A_p}@k = \frac{\sum_{j=1}^{|A_p|} \min(\mathscr{R}_j, \mathscr{H}_j)}{\sum_{j=1}^{|A_p|} \max(\mathscr{R}_j, \mathscr{H}_j)}, \tag{2}$$

where R_j and H_j represent the probability of a news with class j of aspect A_p to be contained in the recommendations list $\mathscr{R}$, and, respectively, in the user history $\mathscr{H}$. All metrics are bounded to the $[0, 1]$ range.

6.3. Experimental Setup

We use pre-trained 300-dimensional German[23] and English[24] GloVe embeddings [51] and 100-dimensional TransD embeddings [63] pretrained on NeMigKG to initialize the word and entity embeddings of the NRS. Note that we use the entity embeddings trained on the variant of NeMigKG enriched with 1-hop Wikidata neighbors. Moreover, we use the extracted sub-topics as topical categories in NAML, MINS, and TANR, and embed them with 100-dimensional vectors. Following [64], we sample four negatives per positive example. We train all models for 10 epochs, using a batch size of 4.[25] We optimize with the Adam algorithm [65], with the learning rate set to 1e-5. We set all other model-specific hyperparameters to optimal values reported in the respective papers. We use 70% randomly sampled user interactions for training the NRS, 10% as validation data, and the remaining 20% as a test set. We repeat each experiment five times with different random seeds and report averages and standard deviations. We normalize the scores to the $[0, 100]$ range.

6.4. Results and Discussion

We first discuss the recommendation performance of the aforementioned NRS on NeMig. We then analyze whether they are prone to implicit biases in terms of three aspects: *political leaning*, *topical categories*, and *sentiment orientation*. Lastly, we investigate the influence of different input features on the quality of NeMigKG.

---

[22]Note that perfect aspect-based personalization would imply identical distributions of aspect A_p's in the recommendations list and user history.

[23]https://www.deepset.ai/german-word-embeddings

[24]https://nlp.stanford.edu/projects/glove/

[25]We trained each model on a single NVIDIA RTX 2080Ti GPU.

Table 5

Recommendation performance of different NRS. The best results per column are highlighted in bold, the second best are underlined.

Model	German				English			
	AUC	MRR	nDCG@3	nDCG@6	AUC	MRR	nDCG@3	nDCG@6
NRMS	57.96±0.66	47.83±0.46	44.90±0.65	60.49±0.35	52.40±0.89	41.46±0.59	37.49±1.07	54.21±0.53
NAML	50.49±0.13	47.37±0.84	44.27±1.20	60.13±0.65	50.02±0.20	41.82±0.70	38.14±0.84	54.49±0.62
MINS	57.70±0.53	47.85±0.78	44.96±0.33	60.50±0.71	52.91±0.68	_41.96±1.13_	38.19±0.41	_54.59±0.41_
CAUM	_58.18±0.85_	_48.17±1.08_	_45.27±0.60_	_60.74±0.55_	_53.10±0.78_	41.83±0.93	38.16±1.42	54.50±0.32
DKN	**58.73±0.39**	**48.20±0.39**	**45.48±0.43**	**60.78±0.29**	**53.69±0.66**	**42.14±0.66**	**38.57±0.83**	**54.74±0.55**
TANR	50.62±0.14	47.28±0.44	44.22±0.51	60.06±0.33	50.17±0.10	41.82±0.12	_38.26±0.33_	54.49±0.17
SentiDebias	56.80±0.30	47.27±0.44	44.09±0.18	60.04±0.44	52.63±0.90	41.63±1.34	37.58±0.13	54.33±0.58

Content Personalization. Table 5 summarizes the results on content personalization for both datasets. We find that, despite the small dataset sizes, the state-of-the-art neural models achieve high predictive performance. We notice a high similarity w.r.t. content personalization performance of NAML and TANR, as well as of NRMS and SentiDebias. Both pairs of models share nearly the same news and user encoder architectures, and differ only in their optimization objectives. These results indicate that secondary optimization goals (i.e., topic classification in the case of TANR, or sentiment debiasing in SentiDebias) have little effect on pure recommendation performance.

Generally, we find that the two knowledge-aware models, DKN and CAUM, outperform all the other recommenders. This suggests that injecting additional external knowledge into the model can help detect additional, knowledge-level connections between news, thus resulting in better news representations. The results are partly at odds with the findings of [66], which observed that while CAUM indeed outperforms other models with candidate-agnostic user encoders, DKN underperforms. However, [66] conducted experiments on the MIND dataset [13], which is magnitudes larger than NeMig. This points to the fact that both knowledge-awareness and candidate-awareness user modeling might be particularly beneficial in scenarios where little training data is available. Lastly, we observe that the quality of content personalization is higher for the German than for the English dataset. We hypothesize that this is due to the higher sparsity of the English dataset, in which significantly more news have never been seen by the users.

Aspect-based Diversity and Personalization. Next, we analyze the diversity and personalization of results, using topical categories (ctg), sentiment (snt), and political leaning (pol) as aspects. Concretely, we report the performance for content personalization (nDCG@k), aspect-based diversity (D_{ctg} for topical categories, D_{snt} for sentiment polarization, and D_{pol} for political leaning) and aspect-based personalization (PS_{ctg} for topical categories, PS_{snt} for sentiment polarization, and PS_{pol} for political leaning). We report results for $k = 3$.

Table 6 summarizes the results on content personalization and aspect diversity. As expected, SentiDebias [33] achieves the highest sentiment diversity on both datasets. NAML [55] outperforms, by a relatively large margin, the other models in terms of category diversity of results. These results are somewhat surprising and at odds with those obtained on other datasets [61], as NAML leverages information on topical categories to customize results to the users' preferences, and not to address diversity. Furthermore, compared to the first two aspects, we

Table 6

Content personalization and aspect diversity (in terms of topical categories, sentiments, and political leaning) performance of different NRS. The best results per column are highlighted in bold, the second best are underlined.

Model	German				English			
	nDCG@3	D_{ctg}@3	D_{snt}@3	D_{pol}@3	nDCG@3	D_{ctg}@3	D_{snt}@3	D_{pol}@3
NRMS	44.90±0.65	18.55±0.25	33.75±0.54	30.68±0.32	37.49±1.07	_22.19±0.34_	32.50±0.27	26.31±0.94
NAML	44.27±1.20	**19.31±0.21**	33.63±0.18	30.41±0.37	38.14±0.84	**23.89±0.65**	32.08±0.64	25.66±1.24
MINS	44.96±0.33	18.18±0.44	33.79±0.34	30.10±0.44	38.19±0.41	21.94±1.11	32.34±0.16	25.22±1.11
CAUM	_45.27±0.60_	18.09±0.33	33.13±0.43	30.42±0.33	38.16±1.42	21.90±0.63	32.09±0.26	25.90±0.63
DKN	**45.48±0.43**	18.46±0.16	_33.79±0.44_	_30.92±0.59_	**38.57±0.83**	22.04±0.40	32.46±0.66	**26.58±0.90**
TANR	44.22±0.51	18.52±0.18	33.70±0.36	30.39±0.38	_38.26±0.33_	21.59±0.19	32.41±0.68	_26.47±0.82_
SentiDebias	44.09±0.18	_18.66±0.30_	**33.84±0.34**	**30.99±0.30**	37.58±0.13	22.10±1.11	**33.26±0.34**	25.81±1.11

Table 7

Content and aspect personalization (in terms of topical categories, sentiments, and political leaning) performance of different NRS. The best results per column are highlighted in bold, the second best are underlined.

Model	German				English			
	nDCG@3	PS_{ctg}@3	PS_{snt}@3	PS_{pol}@3	nDCG@3	PS_{ctg}@3	PS_{snt}@3	PS_{pol}@3
NRMS	44.90±0.65	20.72±0.45	**42.66±0.30**	36.55±0.09	37.49±1.07	18.48±0.18	41.38±0.40	40.66±0.22
NAML	44.27±1.20	20.42±0.31	42.59±0.25	36.23±0.32	38.14±0.84	18.24±0.81	**41.43±0.19**	**41.17±0.30**
MINS	44.96±0.33	**21.10±0.60**	42.56±0.34	**36.74±0.27**	38.19±0.41	_18.83±0.41_	41.20±0.16	_41.15±0.21_
CAUM	_45.27±0.60_	20.85±0.79	42.48±0.43	_36.59±0.36_	38.16±1.42	18.67±1.45	_41.41±0.26_	40.95±0.41
DKN	**45.48±0.43**	_20.95±0.35_	42.51±0.36	36.08±0.24	**38.57±0.83**	18.56±0.35	41.25±0.14	40.70±0.15
TANR	44.22±0.51	20.88±0.24	_42.64±0.19_	36.42±0.34	_38.26±0.33_	**19.10±0.29**	41.27±0.24	40.98±0.20
SentiDebias	44.09±0.18	20.50±0.39	42.64±0.34	36.40±0.24	37.58±0.13	18.57±0.33	41.29±0.34	40.84±0.20

find that all recommenders perform nearly identical w.r.t. degree of political diversification of recommendations. This could be due to the fact that none of the models explicitly target political diversification.

Aspect-based diversity is tightly correlated with aspect-based personalization [61]. More specifically, higher levels of diversity come at the cost of personalization, as can be seen in Table 7, which illustrates the results on content and aspect-based personalization. We observe that categorical and political personalization are much more aligned with content personalization performance, thqn with the sentiment of news, as discussed also in [61]. We find this to be intuitive, as users tend to choose which news to read based on their topics/categories of interest, as well as political preferences, and not on their sentiment polarization.

An earlier version of the German dataset [67] has been used in [68] and [69] to examine sentiment and stance recommender bias, and to investigate polarization and filter bubble creation through online studies, respectively. While in this study we have analyzed political biases of recommenders only from the perspective of users' click histories, in the future their political information can be further explored in conjunction with the models' predictions to understand whether any existing biases are reinforced or amplified by the algorithms. Similarly, NeMig can be used in future works to analyze these trends also in multilingual or cross-lingual recommendation scenarios, as it comprises information on the same topic in both German and English.

Table 8

Content personalization and aspect-based diversity (in terms of topical categories, sentiment, and political leaning) performance of DKN with different variants of NeMigKG.

Model	German					English				
	AUC	nDCG@3	D_{ctg}@3	D_{snt}@3	D_{pol}@3	AUC	nDCG@3	D_{ctg}@3	D_{snt}@3	D_{pol}@3
content entities	58.72±0.45	45.57±0.44	18.44±0.16	33.67±0.38	31.17±0.56	53.67±0.63	38.60±1.01	21.99±0.45	32.33±0.34	26.71±0.75
+ topical categories	58.82±0.23	45.74±0.31	18.43±0.09	33.83±0.39	30.93±0.64	53.67±0.50	38.68±0.72	22.02±0.38	32.20±0.45	26.47±1.12
+ sentiment	58.81±0.26	45.69±0.38	18.48±0.08	33.77±0.44	30.98±0.65	53.70±0.53	38.53±0.73	21.95±0.37	32.17±0.18	26.89±0.72
+ political leaning	58.73±0.42	45.55±0.43	18.45±0.16	33.68±0.47	31.00±0.45	53.73±0.72	38.66±0.79	22.00±0.46	32.28±0.36	26.74±0.52
+ 1-hop neighbors	58.82±0.33	45.74±0.46	18.50±0.09	33.77±0.40	31.20±0.50	53.78±0.44	38.76±0.65	21.95±0.37	32.16±0.25	26.60±0.59
+ 2-hop neighbors	58.73±0.39	45.48±0.43	18.46±0.16	33.79±0.44	30.92±0.59	53.69±0.66	38.57±0.83	22.04±0.40	32.46±0.66	26.58±0.90

Knowledge Graph Ablations. The type of data contained in the knowledge graph is paramount for training high quality knowledge graph embeddings, which are further exploited by knowledge-aware recommenders to improve the accuracy of predictions. Thus, we analyze the impact of input features on NeMigKG. Specifically, we pre-train entity embeddings on different versions of NeMigKG using TransD [63], and study the recommendation and aspect-based diversity performance of DKN using these embeddings. We summarize the results in Table 8 in terms of content personalization (nDCG@k), and aspect-based diversity for topical categories (D_{ctg}), sentiment (D_{snt}), and political leaning (D_{pol}). Adding topical categories, sentiment and political information in NeMigKG increases the diversity of recommendations w.r.t. these aspects, while having minor effects on accuracy-based performance. We find that extending NeMigKG with k-hop neighbors from Wikidata is beneficial only up to $k = 1$ hops. This indicates that a certain degree of contextualization of named entities with general knowledge is helpful when the data size is small. However, injecting too much external knowledge (i.e., $k = 2$) can dilute the original information and have detrimental effects on the recommender's downstream performance.

6.5. News Trends Analysis

Lastly, we demonstrate how NeMig can be used to analyze trends and correlations between entities used in the news discourse and events. Concretely, we analyze the evolution of named entities overtime and political groups. Figs. 6-7 show the evolution over time and political orientation of the source outlets of some of the top-20 most frequent entities in our corpora. We find that generally the most frequently sampled entities are covered predominantly in center media, which is also the source of the majority of news in both corpora. Notable exceptions in the German corpus (Fig. 6) are *Angela Merkel*, appearing equally or more often in right-wing media, and *Die Linke*, with a more equal distribution among all political classes, although left-wing coverage dominates the discourse. The evolution of entities over time reveals correlations with major events. In terms of organizations, German news report most frequently about the main political parties, whose evolution coincides with elections on the federal and European level, or about central political figures in international politics (e.g., the coverage of *Recep Tayyip Erdoğan* appears to be potentially correlated with the Turkish offensive in north-eastern Syria). In contrast to German media, the representation of the top entities from the English corpus in the left US media is generally lower.
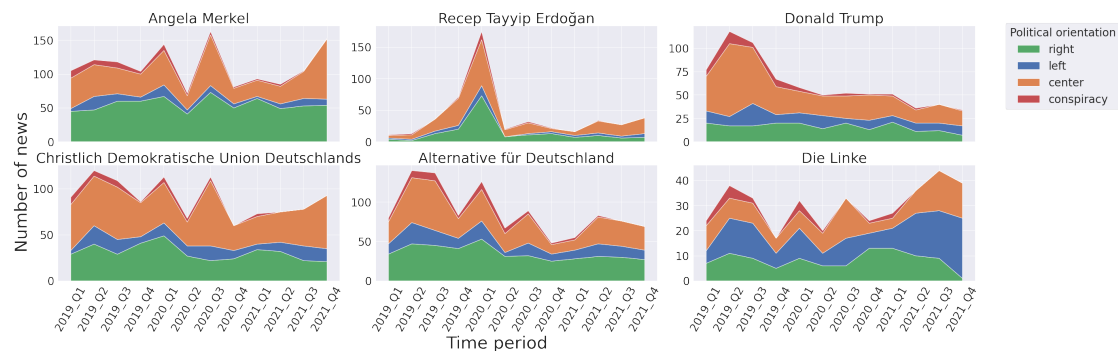
Figure 6: Evolution over time and political orientation of a sample of the top-20 most popular entities from the German dataset. The first row contains entities of type PER, and the second of type ORG.
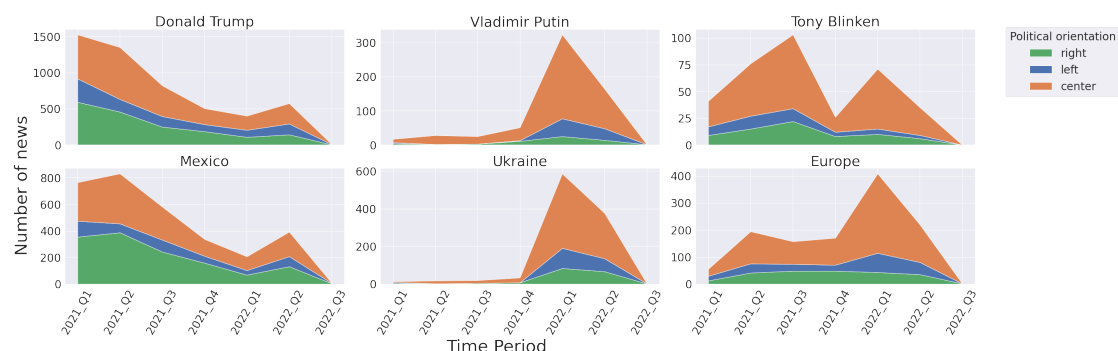

Figure 7: Evolution over time and political orientation of a sample of the top-20 most popular entities from the English dataset. The first row contains entities of type PER, and the second of type LOC.

7. Conclusion

We introduced NeMig, a bilingual news collection and knowledge graphs on the topic of migration, along with user data, including demographic and political information. The news is collected from mainstream and alternative German and US media outlets spanning a wide political scale. We extract sub-topics and named entities linked to Wikidata from the news text and metadata. In contrast to existing datasets, we provide sentiment and political annotations, and use Wikidata for incorporating background knowledge. We hope that this resource will inspire future research on analyzing the multidimensional implications of news curation algorithms, in both monolingual and cross-lingual settings.

Acknowledgments

References

[1] C. Wu, F. Wu, T. Qi, Y. Huang, 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, 2020, pp. 44–53. URL: https://aclanthology.org/2020.aacl-main.6.

[2] P. Liu, K. Shivaram, A. Culotta, M. A. Shapiro, M. Bilgic, The interaction between political typology and filter bubbles in news recommendation algorithms, in: Proceedings of the Web Conference 2021, 2021, pp. 3791–3801.

[3] E. Pariser, The filter bubble: What the Internet is hiding from you, Penguin UK, 2011.

[4] L. M. Bartels, Messages received: The political impact of media exposure, American political science review 87 (1993) 267–285.

[5] R. Dewenter, M. Linder, T. Thomas, Can media drive the electorate? the impact of media coverage on voting intentions, European Journal of Political Economy 58 (2019) 245–261.

[6] J. Jensen, S. Naidu, E. Kaplan, L. Wilse-Samson, D. Gergen, M. Zuckerman, A. Spirling, Political polarization and the dynamics of political language: Evidence from 130 years of partisan speech [with comments and discussion], Brookings Papers on Economic Activity (2012) 1–81.

[7] E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on facebook, Science 348 (2015) 1130–1132.

[8] A. Boutyline, R. Willer, The social structure of political echo chambers: Variation in ideological homophily in online networks, Political psychology 38 (2017) 551–569.

[9] K. Ludwig, P. Müller, Does social media use promote political mass polarization? a structured literature review, Questions of Communicative Change and Continuity. In Memory of Wolfram Peiser (2022) 118–166.

[10] J. M. Balkin, Free speech in the algorithmic society: Big data, private governance, and new school speech regulation, UCDL rev. 51 (2017) 1149.

[11] N. Helberger, On the democratic role of news recommenders, Digital Journalism 7 (2019) 993–1012.

[12] J. A. Gulla, L. Zhang, P. Liu, O. Özgöbek, X. Su, The adressa dataset for news recommendation, WI '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1042–1048.

[13] F. Wu, Y. Qiao, J. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, M. Zhou, MIND: A large-scale dataset for news recommendation, in: D. Jurafsky, J. Chai, N. Schluter, J. R. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 3597–3606.

[14] W. Y. Wang, "liar, liar pants on fire": A new benchmark dataset for fake news detection, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2017, pp. 422–426.

[15] B. Horne, S. Khedr, S. Adali, Sampling the news producers: A large news and feature data set for the study of the complex media landscape, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 12, 2018.

[16] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with

news content, social context, and spatiotemporal information for studying fake news on social media, Big data 8 (2020) 171–188.

[17] E. Levi, G. Mor, S. Shenhav, T. Sheafer, Compres: A dataset for narrative structure in news, arXiv preprint arXiv:2007.04874 (2020).

[18] M. Färber, V. Burkard, A. Jatowt, S. Lim, A multidimensional dataset based on crowdsourcing for analyzing and detecting news bias, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 3007–3014.

[19] S. Lim, A. Jatowt, M. Färber, M. Yoshikawa, Annotating and analyzing biased sentences in news articles using crowdsourcing, in: Proceedings of the Twelfth Language Resources and Evaluation Conference, 2020, pp. 1478–1484.

[20] A. Iana, M. Alam, A. Grote, N. Nikolajevic, K. Ludwig, P. Müller, C. Weinhardt, H. Paulheim, NeMig - A Bilingual News Collection and Knowledge Graph about Migration, 2022. doi:10.5281/zenodo.7442424.

[21] B. Kille, F. Hopfgartner, T. Brodt, T. Heintz, The plista dataset, in: Proceedings of the 2013 International News Recommender Systems Workshop and Challenge, NRS '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 16–23.

[22] G. de Souza Pereira Moreira, F. Ferreira, A. M. da Cunha, News session-based recommendations using deep neural networks, in: B. Hidasi, A. Karatzoglou, O. S. Shalom, B. Shapira, D. Tikk, F. Vasile, S. Dieleman (Eds.), Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2018, Vancouver, BC, Canada, October 6, 2018, ACM, 2018, pp. 15–23.

[23] G. de Souza Pereira Moreira, D. Jannach, A. M. da Cunha, Contextual hybrid session-based news recommendation with recurrent neural networks, IEEE Access 7 (2019) 169185–169203.

[24] H. Wang, F. Zhang, X. Xie, M. Guo, Dkn: Deep knowledge-aware network for news recommendation, in: Proceedings of the 2018 world wide web conference, 2018, pp. 1835–1844.

[25] A. Iana, M. Alam, H. Paulheim, A survey on knowledge-aware news recommender systems, Semantic Web (2022) 1–62.

[26] P. Vossen, R. Agerri, I. Aldabe, A. Cybulska, M. van Erp, A. Fokkens, E. Laparra, A.-L. Minard, A. P. Aprosio, G. Rigau, M. Rospocher, R. Segers, Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news, Special Issue Knowledge-Based Systems, Elsevier (2016).

[27] M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger, T. Bogaard, Building event-centric knowledge graphs from news, Journal of Web Semantics (2016).

[28] D. Liu, T. Bai, J. Lian, X. Zhao, G. Sun, J.-R. Wen, X. Xie, News graph: An enhanced knowledge graph for news recommendation., in: KaRS@ CIKM, 2019, pp. 1–7.

[29] A. L. Opdahl, T. Al-Moslmi, D.-T. Dang-Nguyen, M. Gallofré Ocaña, B. Tessem, C. Veres, Semantic knowledge graphs for the news: A review, ACM Comput. Surv. (2022). doi:10.1145/3543508.

[30] G. K. Shahi, D. Nandini, Fakecovid–a multilingual cross-domain fact check news dataset for covid-19, arXiv preprint arXiv:2006.11343 (2020).

[31] J. Golbeck, M. Mauriello, B. Auxier, K. H. Bhanushali, C. Bonk, M. A. Bouzaghrane, C. Bun-

tain, R. Chanduka, P. Cheakalos, J. B. Everett, et al., Fake news vs satire: A dataset and analysis, in: Proceedings of the 10th ACM Conference on Web Science, 2018, pp. 17–21.

[32] Z. Liu, S. Shabani, N. G. Balet, M. Sokhn, Detection of satiric news on social media: Analysis of the phenomenon with a french dataset, in: 2019 28th International Conference on Computer Communication and Networks (ICCCN), IEEE, 2019, pp. 1–6.

[33] C. Wu, F. Wu, T. Qi, W.-Q. Zhang, X. Xie, Y. Huang, Removing ai's sentiment manipulation of personalized news delivery, Humanities and Social Sciences Communications 9 (2022) 1–9.

[34] J. McCoy, T. Rahman, M. Somer, Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities, American Behavioral Scientist 62 (2018) 16–42.

[35] J. Cho, S. Ahmed, M. Hilbert, B. Liu, J. Luu, Do search algorithms endanger democracy? an experimental investigation of algorithm effects on political polarization, Journal of Broadcasting & Electronic Media 64 (2020) 150–172.

[36] R. R. R. Gangula, S. R. Duggenpudi, R. Mamidi, Detecting political bias in news articles using headline attention, in: Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, 2019, pp. 77–84.

[37] K. Lazaridou, R. Krestel, Identifying political bias in news articles, Bulletin of the IEEE TCDL 12 (2016).

[38] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8440–8451.

[39] F. Barbieri, L. E. Anke, J. Camacho-Collados, Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 258–266.

[40] M. Grootendorst, Bertopic: Neural topic modeling with a class-based tf-idf procedure, arXiv preprint arXiv:2203.05794 (2022).

[41] N. Reimers, I. Gurevych, Making monolingual sentence embeddings multilingual using knowledge distillation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2020.

[42] L. Li, D.-D. Wang, S.-Z. Zhu, T. Li, Personalized news recommendation: a review and an experimental investigation, Journal of computer science and technology 26 (2011) 754–766.

[43] E. F. Sang, F. De Meulder, Introduction to the conll-2003 shared task: Language-independent named entity recognition, arXiv preprint cs/0306050 (2003).

[44] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A python natural language processing toolkit for many human languages, arXiv preprint arXiv:2003.07082 (2020).

[45] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (2014) 78–85.

[46] N. De Cao, L. Wu, K. Popat, M. Artetxe, N. Goyal, M. Plekhanov, L. Zettlemoyer, N. Cancedda, S. Riedel, F. Petroni, Multilingual autoregressive entity linking, Transactions of the Association for Computational Linguistics 10 (2022) 274–290.

[47] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, L. Zettlemoyer, Multilingual denoising pre-training for neural machine translation, Transactions of the

Association for Computational Linguistics 8 (2020) 726–742.

[48] N. Heist, S. Hertling, D. Ringler, H. Paulheim, Knowledge graphs on the web-an overview., 2020.

[49] Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, Q. He, A survey on knowledge graph-based recommender systems, IEEE Transactions on Knowledge and Data Engineering 34 (2020) 3549–3568.

[50] W. R. Van Hage, V. Malaisé, R. Segers, L. Hollink, G. Schreiber, Design and use of the simple event model (sem), Journal of Web Semantics 9 (2011) 128–136.

[51] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[52] C. Wu, F. Wu, S. Ge, T. Qi, Y. Huang, X. Xie, 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), 2019, pp. 6389–6394.

[53] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[54] D. Bahdanau, K. H. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: 3rd International Conference on Learning Representations, ICLR 2015, 2015.

[55] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, X. Xie, Neural news recommendation with attentive multi-view learning, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 3863–3869.

[56] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1746–1751.

[57] R. Wang, S. Wang, W. Lu, X. Peng, News recommendation via multi-interest news sequence modelling, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 7942–7946.

[58] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder–decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, 2014, p. 1724.

[59] T. Qi, F. Wu, C. Wu, Y. Huang, News recommendation with candidate-aware user modeling, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 1917–1921.

[60] C. Wu, F. Wu, M. An, Y. Huang, X. Xie, Neural news recommendation with topic-aware news representation, in: Proceedings of the 57th Annual meeting of the association for computational linguistics, 2019, pp. 1154–1159.

[61] A. Iana, G. Glavaš, H. Paulheim, Train once, use flexibly: A modular framework for multi-aspect neural news recommendation, 2023. arXiv:2307.16089.

[62] V. Bonnici, Kullback-leibler divergence between quantum distributions, and its upper-bound, arXiv preprint arXiv:2008.05932 (2020).

[63] G. Ji, S. He, L. Xu, K. Liu, J. Zhao, Knowledge graph embedding via dynamic mapping matrix, in: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers), 2015, pp. 687–696.

[64] C. Wu, F. Wu, Y. Huang, Rethinking infonce: How many negative samples do you need?, in: L. D. Raedt (Ed.), Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, International Joint Conferences on Artificial Intelligence Organization, 2022, pp. 2509–2515.

[65] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, ICLR (2014).

[66] A. Iana, G. Glavas, H. Paulheim, Simplifying content-based neural news recommendation: On user modeling and training objectives, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023, pp. 2384–2388.

[67] A. Iana, A. Grote, K. Ludwig, M. Alam, P. Müller, C. Weinhardt, H. Sack, H. Paulheim, Geneg: German news knowledge graph, 2022. doi:10.5281/zenodo.5913171.

[68] M. Alam, A. Iana, A. Grote, K. Ludwig, P. Müller, H. Paulheim, Towards analyzing the bias of news recommender systems using sentiment and stance detection, in: Companion Proceedings of the Web Conference 2022, 2022, pp. 448–457.

[69] K. Ludwig, A. Grote, A. Iana, M. Alam, H. Paulheim, H. Sack, C. Weinhardt, P. Müller, Divided by the algorithm? the (limited) effects of content-and sentiment-based news recommendation on affective, ideological, and perceived polarization, Social Science Computer Review (2023) 08944393221149290.