A Dynamic Embedding Model of the Media Landscape

Jérémie Rappaz*
EPFL
Switzerland
jeremie.rappaz@epfl.ch

Dylan Bourgeois*
EPFL
Switzerland
contact@dtsbourg.me

Karl Aberer
EPFL
Switzerland
karl.aberer@epfl.ch

ABSTRACT

Information about world events is disseminated through a wide variety of news channels, each with specific considerations in the choice of their reporting. Although the multiplicity of these outlets should ensure a variety of viewpoints, recent reports suggest that the rising concentration of media ownership may void this assumption. This observation motivates the study of the impact of ownership on the global media landscape and its influence on the coverage the actual viewer receives. To this end, the selection of reported events has been shown to be informative about the high-level structure of the news ecosystem. However, existing methods only provide a static view into an inherently dynamic system, providing underperforming statistical models and hindering our understanding of the media landscape as a whole.

In this work, we present a dynamic embedding method that learns to capture the decision process of individual news sources in their selection of reported events while also enabling the systematic detection of large-scale transformations in the media landscape over prolonged periods of time. In an experiment covering over 580M real-world event mentions, we show our approach to outperform static embedding methods in predictive terms. We demonstrate the potential of the method for news monitoring applications and investigative journalism by shedding light on important changes in programming induced by mergers and acquisitions, policy changes, or network-wide content diffusion. These findings offer evidence of strong content convergence trends inside large broadcasting groups, influencing the news ecosystem in a time of increasing media ownership concentration.

CCS CONCEPTS

• **Information systems** → **Data mining**; • **Computing methodologies** → **Learning latent representations**.

KEYWORDS

media pluralism; factorization methods; ranking methods

*Both authors contributed equally to the paper
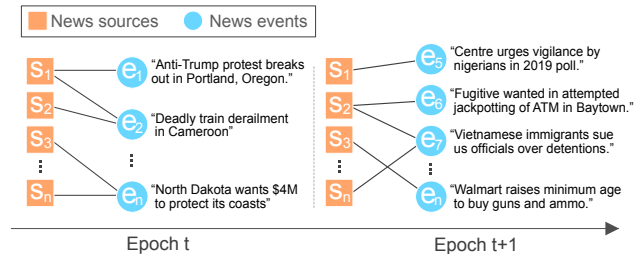
Figure 1: Illustration of the setting: we observe the coverage of news events from a fixed set of sources over several time epochs. Example events are extracted from the GDELT database.

1 INTRODUCTION

Information about world events is reported as a continuous stream, processed then broadcast to an audience through a large number of channels and across various mediums. As no coverage could claim to be exhaustive, sources must first filter the stories that will be disseminated. The selection made by each channel is partial by nature. However, the variety of outlets, each with a wide array of considerations in the choice of its reporting, is assumed to ensure the diversity of news to which the reader is exposed to. This principle is often referred to as the *external pluralism* assumption. It should ensure heterogeneity in the media space, encapsulating anything from the diversity of ownership to the independence of the editorial board. [1] *External pluralism* is also known as the "supplier" pluralism, since it should exclude the possibility of large broadcast groups exerting influence on downstream reporting. Yet in practice, this assumption does not always hold. News channels are often owned or operated by commercial, private entities, implying that the ecosystem as a whole is influenced by economically motivated forces, such as mergers, acquisitions, or regulatory actions. The increase in concentration of ownership has been observed to be the dominating force in the media landscape, as reported by the Pew Research Institute in a 2017 study on the acquisition of local television stations. [2]

While the literature, discussed in more detail in Section 2, still debates the causal effects of market ownership structures on the diversity of offerings, the Federal Communications Commission (FCC) [3] has defended the idea that "there is a positive correlation between viewpoints expressed and ownership of an outlet". [4] This

---

[1] As opposed to *internal pluralism*, where sources are assumed to present a wide variety of ideological viewpoints, communicated through different mediums. [8]

[2] http://www.pewresearch.org/fact-tank/2017/05/11/buying-spree-brings-more-local-tv-stations-to-fewer-big-companies/

[3] The FCC is the regulating body for multimedia communications in the United States.

[4] FCC, Biennial Media Ownership Order (2003)

uncertainty strongly motivates the development of novel and interpretable methods, which would allow the observation of large-scale movements in the media landscape and allow correlating the observations with real-world factors.

Recent studies have tackled the problem of evaluating similarities across news sources [5, 38] and have found that the analysis of coverage patterns can highlight non-obvious relationships between sources, such as the aforementioned corporate ownership structures. However, they have only captured a static snapshot of the media landscape. This hinders the ability to observe changes in content diffusion, and limits attempts to correlate the observations with potential factors in the news space. These models' predictive power is also limited by their inability to incorporate temporal dependencies. In other words, it restricts them to an informative but partial and static view of the news ecosystem. Indeed, the model of any dynamic system requires a time-dependent formulation, necessary to detect variations and measure trends. Existing approaches are blind to such transient effects.

Studying the effect of real-world factors on content diffusion is a challenging task. It requires the detection of subtle variations in the coverage patterns of individual channels, which can occur over long time spans. Any attempt to characterize these effects would suffer from the absence of a ground truth, prompting the need for a comparison across news channels. While similarities between channels are straightforward to compute for short, consistent periods of time, the continuous occurrence of new and unpredictable events breaks the ability to reason across successive snapshots of the news. A successful estimate of similarity over time should hence maintain temporal consistency of pairwise distances across news channels. At any given point in time, it should ideally do so without compromising its predictive accuracy compared to static methods. These challenges and constraints motivate the use of the models presented here.

In this work, we propose a dynamic embedding model of the media landscape. By predicting news sources' coverage, the model captures their similarities throughout the observed period. The embedding space is maintained consistent over time by augmenting the model with a knowledge of previous time steps, and by adding a temporal regularization on the model's parameters. This improves the model's predictive capabilities and provides a temporally coherent source-wise similarity metric, allowing the visualization and analysis of long-ranging fluctuations in the news ecosystem. We also propose a systematic method to detect abrupt transition patterns in this similarity space. This enables the analysis of the news ecosystem beyond domain-specific knowledge and hand-crafted analysis.

In the context of news reporting, temporally-aware models provide a powerful tool for correlating content programming changes and real-world factors, from individual news outlet's behaviors to large-scale dynamics. Shining a light on these movements would naturally guide the search for source-level context, helping to characterize and identify the observation's driving forces. This information could then provide context to journalists investigating the coverage of a particular story or feed watchdog-like processes that monitor the health and evolution of the media landscape. To illustrate, we provide some prototypical questions which could arise in a journalistic probe, and that could be elucidated by information derived from the proposed model:

- What effect does the ownership of a news channel have on its content diffusion?
- Which sources are most similar (resp. dissimilar) to a sample source, and how has this similarity evolved over time?
- Which are the most consistent (or varying) news channels in terms of broadcast content?
- Which broadcast groups exert a large influence on the content of their respective channels?

The rest of this paper is organized as follows. In Section 2, we present relevant work from the literature. In Section 3, we describe the dataset used in our study. In Section 4, we introduce our model. In Section 5, we discuss the predictive performance of the model. In Section 6, we illustrate the use of our model in the identification of abrupt changes the media landscape, attempting to interpret these changes in coverage in Section 7. In Section 8, we conclude and propose future research directions.

2 RELATED WORK

The study of the evolution of the media landscape has led to prolific research lines spanning various fields. Early work had a tendency to examine the media ecosystem in isolation. For example Steiner's seminal work [43] studies the interplay of consumer preferences and in-market competition on the diversity of radio broadcasting. This study omitted the role of external driving forces, which were only later modeled by Anderson & Coate. [3] They predicted that media consolidation, while economically beneficial for the market, would reduce competition and hence diversity for the viewer. This insight was later formulated in terms of ideological bias by Gentzkow & Shapiro [10] in the case of newspapers.

The tendency to integrate external driving factors has picked up steam in recent years, most notably with a theory of convergence in the media ecosystem. This convergence expresses itself through two seemingly contradictory features. On one hand, information is being delivered through an ever-increasing number of channels and means of diffusion. On the other hand, media ownership concentration has seen an upwards trend, with a large proportion of channels being owned by only a handful of media conglomerates. [31] This dichotomy has been studied by, among others, Jenkins. [20] The author proposed a sketch of the phenomenon that looks further than the sole technological influence, reaching for larger cultural factors. Vizcarrondo et al. [45] have more specifically investigated the concentration of media ownership. They reported on changes in the diversity of ownership within the media industry covering the 1976 through 2009 time period.

A large body of work has also been introduced to study the effect of an ideologically slanted press. For instance in a large-scale observational study, DellaVigna et al. [6] measured the effect of the introduction of a conservative-oriented channel (Fox News) led to gains of 0.4 to 0.6 percentage points in Republican voting in the towns where the channel was being broadcast. While specific to a particular orientation, this work is in line with studies showing the profound influence of the media in voters' political awareness [27] and their participation in the electoral process. [11, 12, 28]

Notation	Description
R	Interaction matrix $\in \mathbb{R}^{|S| \times |E|}$
S / E	News source / Event set
$E_{s_i}^+$	Set of events covered by source s_i
s_i	Source $s_i \in S$
e_k	Event $e_k \in E$
K	Number of latent factors
$\hat{x}_{s_i, e_k}$	Predicted preference of source s_i for event e_k
D / D_s	Training / Testing set
γ	STD of the Gaussian random walk
α	Learning rate
λ_Θ	ℓ_2-regularization weight
λ_T	Temporal regularization weight
Θ	Set of model parameters

Table 1: Notation

The Federal Communications Commission (FCC) regularly issues studies regarding the state of the news ecosystem. Specifically, some of these studies focus on the effect of ownership on local news stations' content programming behaviors. However, by the authors' own admission [5], these works often lack the breadth required by a large-scale empirical study. For example, Pritchard [32] conducted a study of the diversity of coverage for cross-owned media outlets during the 2000 presidential campaign but on a sample of only 10 newspapers. Groseclose & Milyo [16] proposed a measure of media bias which was evaluated on a set of 8 newspapers. Djankov et al. [7] did survey the news ecosystem on large scale, building a map of media ownership in 97 countries around the world, but this work dates back to 2003.

Dimensionality reduction methods have been extensively used to model different perspectives or opinions. For example, Lahoti et al. [24] rely on Non-Negative Matrix Factorization methods to learn a liberal-conservative ideology space on Twitter. In particular, the authors propose to approximate the users' ideology based on their online news consumption.

The study performed by Saez-Trumper et al. [39] relies on a Principal Component Analysis (PCA) to detect similarities across news channels. In particular, they performed their analysis by covering three different types of biases in the press: *gatekeeping bias*, *coverage bias* and *statement bias*. The authors' bias-capturing method does present similarities to our approach, but is formulated as an unsupervised approach and does not tackle temporal variability. Recently, Bourgeois et al. [5] proposed a supervised, embedding-based method to capture similarities across news sources based on their respective coverage. This model was also designed statically and is hence unable to accurately model temporal variations in news coverage.

Learning to produce a personalized ranking from positive interactions only is a well-studied problem that has been investigated in the context of recommendations from implicit feedback [30] and as One-Class feedback [19] in the context of recommender systems.

Temporally-aware methods have received increasing attention and many previous models have now been adapted to the temporal setting. The *dynamic embedding*, proposed by Rudolph et al. [36] as a variation of traditional embedding methods, is generally aimed toward temporal consistency. The method is introduced in the context of word embeddings, which are used to characterize the evolution of English language. The model is built upon the initial *exponential family embeddings* model. [37]

The field of personalization has many examples of temporally-aware models since human preferences tend to evolve over time. For example, influential work from Koren et al. [22] models the changing nature of preference through a linear drifting term. Another approach relies on the use of Tensor Factorization (TF), [1, 9, 46] in which the extra dimension models temporal patterns in the data. We do not consider TF-based methods as valid candidate approaches since we focus on the problem of grounding representations over time by penalizing unnecessary differences between successive solutions of the model. The temporal modelling capabilities of TF-based methods would predict the evolution of sources and introduce additional temporal variations, consequently degrading interpretability.

He et al. [18] introduced a temporally-aware model of a recommender system in order to capture the evolution of fashion trends. Similar to Koren et al. [22], the authors also proposed the addition a drifting term to the model. The authors later proposed the use of a higher-order Markov chain that captures both short- and long-term dynamics [17]. Note that both models make use of Bayesian Personalized Ranking (BPR [35]) for their respective optimization procedure. In the context of networks, Yu et al. [48] proposed a temporal factorization for analyzing the evolution of network structures.

The media response to global events has been studied, with many using the same data-source as our work. However, these approaches have focused mainly on content-level analysis, leveraging the multimodal capabilities of the GDELT dataset (please refer to Section 3 for more detailed information about the dataset). These works have monitored the media response to events such as protests, [33] natural disasters [23] or conflicts. [14, 21, 47] This data-source was also used to get a world-wide view on coverage of global issues like climate change. [29]

Research Questions: Given the work above, several research questions are of our interest and have remained unanswered:

RQ1: How can we model the global media landscape over time?
RQ2: Is media consolidation highlighted by the resulting latent representation?
RQ3: How to systematically detect abrupt deviations in content diffusion at a source-level?

3 DATA

In this section, we present the selected data source and explicit our data collection process, providing general statistics about the resulting dataset.

---

[5]"[a] larger number of independent owners will tend to generate a wider array of viewpoints in the media than would a comparatively smaller number of owners. We believe this proposition, even without the benefit of conclusive empirical evidence." FCC, Biennial Media Ownership Order (2003)

	GDelt Dataset
# sources	7 278
# unique events	174M
# interactions	588M
time span	Feb 2015 - May 2018

Table 2: Meta-data about the dataset (after preprocessing)

Figure 2: Number of sources covering an event.

Figure 3: Number of events covered per source.

3.1 Data Source

Recently, several event collection databases have emerged on the Web, making accessible to the general public a global view of the daily world events. In this work, we rely on the *Global Database of Events, Language, and Tone* (GDELT [26]), a large database of annotated news events. GDELT [6] was selected since it is publicly available and covers a reasonably large timespan and geography, offering a larger set of sources [2] to study compared to alternatives such as EventRegistry. [7]

For decades, event coding was performed manually. In the 1990s, the first automated systems started to gain traction in the academic community, with initiatives such as the KEDS system. [40] Its successor was proposed in the form of *Text Analysis By Augmented Replacement Instructions* (TABARI), which is the engine that runs event coding for GDELT. This framework is designed to process large amounts of text to extract the presence of pairs of *actors* and *verbs*. To do so it matches elements from user-provided dictionaries, which contain a massive collection of event protagonists (i.e. actors) ranging from recognizable named entities (e.g. *Barack Obama*) to functional placeholders (e.g. *a local woman*). These actors are able to interact with the world through *verbs* (i.e. actions), which can be self-contained (e.g. *announces their intent to*) or involve a second actor (e.g. *criticizes their opponent*). Several standards exist for these dictionaries. GDELT uses the *Conflict and Mediation Event Observations* (CAMEO [13]). [8] Note that, as a remnant of previous hand-curated event annotation frameworks [25], TABARI also provides an interface for manual hand-off to domain experts if the sentences become too complex. This reinforces GDELT's ability to uniquely annotate even the most fine-grained events.

GDELT also augments every news event it tags by extracting meta-information about the article including, but not limited to, its location, its tone, its Goldstein Scale [15] and refences the URL the event was scanned at. It scours a wide array of sources, from television stations to blogs, news wires and papers. Thanks to the information provided by this augmented event coding framework, GDELT assigns, for each news event, a global identifier, which makes it possible to link the same event's coverage across different news sources. Beyond the rich annotations provided by GDELT, this tracking is central to our study given that we only work at the coverage level, without considering the content itself.
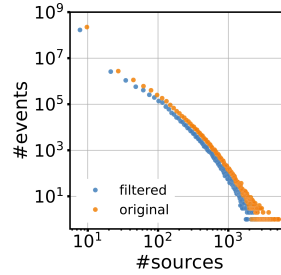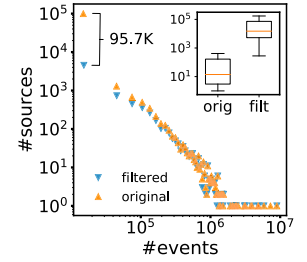
3.2 Data Preprocessing

From the massive resource maintained by GDELT we can gather a dataset of interactions between sources and events, recording which sources covered which uniquely identifiable events. We focus our data collection campaign on the publicly available dumps of GDELT 2.0, released every 15 minutes since February 2015. General statistics about the dataset are presented in Table 2 as well as Fig. 2 and Fig. 3.

When considering the full time-span, GDELT references more than 105K different news sources. This represents a considerable increase compared to the 63K sources reported by Kwak et al. [2] in 2016. However, as is shown in Fig. 3, most of the sources have only published a few articles over the relevant stretch. To maintain a consistent number of channels over time, we discard all channels inactive in any one of the time slices from our dataset. This retains around 7 278 news sources in our dataset. The filtering step does remove a large fraction of available channels, but it mostly affects sources with a very low publishing rate: despite preserving only 7% of the total channels, the selection still accounts for more than 76% of interactions in the dataset (see Fig. 3).

The dataset is split into slices with a duration of one month. This allows for a decent trade-off between having a significant amount of events covered in the training set, while also providing enough samples to observe time-dependent changes over the considered period. In principle, this scale could be modified to study the media landscape at different granularity levels. For example, a more fine-grained split might allow the observation of changes correlated with specific events. However, we choose to leave this analysis as future work given the significant amount of computation that the model requires.

4 MODEL

In this section, we first motivate our approach. Then, we briefly introduce the use of Matrix Factorization methods in a ranking setting by formally describing the task and its objectives. Finally, we propose two strategies to extend the model to a temporal setting.

4.1 Motivation

The main focus of our work is to establish a dynamic model of coverage similarities across news channels in order (i) to uncover congruent coverage patterns in the pool of available channels, (ii) to establish a dynamic, predictive model of coverage and (iii) to

---

[6]https://www.gdeltproject.org/

[7]https://eventregistry.org/

[8]An exhaustive list of the considered categories can be found at http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf

provide a systematic methodology to identify and interpret structural changes in the media landscape. Recent political and societal focus on media accountability and transparency provide strong motivators in the development of such tools (see Section 1).

We model the interrelationships between sources and events by relying on a Matrix Factorization (MF) method. MF methods represent a natural way of projecting two disjoint sets of items in a common latent space of K dimensions in order to model their interactions. Such personalized methods are commonly use by recommender systems, which routinely aim to model retail purchasing decisions. Used in our context, they model coverage decisions instead. We cast the problem to a One-Class learning setting [30] since we observe positive interactions only. The One-class formulation avoids making assumptions about negative examples: we do not distinguish between real negatives (i.e. the source purposely didn't cover an event) and unobserved interactions (i.e. the source wasn't aware of the event).

4.2 Problem Statement

Let us consider a set of news sources S and a set of events E. Interactions between the two are represented by an interaction matrix $R \in \mathbb{R}^{|S| \times |E|}$. Observations take the form of dyadic interactions $(s_i, e_j), s_i \in S, e_j \in E$ which express source s_i's coverage of event e_j. Equivalently, we define in matrix form that $R_{i,j} = 1$ if source s_i has covered event e_j and $R_{i,j} = 0$ otherwise.

Predicting the unobserved entries of matrix R is achieved by taking the dot-product of two low rank matrices, such that $R \approx P^T \cdot Q$, where $P \in \mathbb{R}^{K \times |S|}$ and $Q \in \mathbb{R}^{K \times |E|}$ with $K << |S|, |E|$. Every source s_i (resp. every event e_j) is represented by a column in P (resp. Q). We will refer to these columns as an embedding vector throughout the remainder of this work. We will refer to Θ as the set of parameters for our MF predictor, such that $\Theta = \{P, Q\}$.

Objectives: The model is trained with the objective of predicting the likelihood of a source covering a particular event. The predicted likelihood $\hat{x}_{s_i, e_j}$ of source s_i covering event e_j is computed as the dot-product between the two respective embedding vectors,

$$\hat{x}_{s_i, e_j} = p_{s_i}^T \cdot q_{e_j}. \tag{1}$$

Instead of best approximating the reconstruction of matrix R, this objective is stated as a ranking problem in which positive examples should obtain a higher rank than negative ones, i.e. to predict a higher score for an event that has been covered than for random negative samples. Optimizing a MF model with a ranking criterion is equivalent to maximizing the following probability,

$$
\begin{aligned}
\Pr(e_j >_{s_i} e_k | \Theta) &:= H(\hat{x}_{s_i, e_j} - \hat{x}_{s_i, e_k}) \\
&\approx \sigma(\hat{x}_{s_i, e_j} - \hat{x}_{s_i, e_k})
\end{aligned}
\tag{2}
$$

where e_j is an event covered by source s_i and e_k is a randomly sampled negative event; formally, $e_j \in E_{s_i}^+$ and $e_k \in E \setminus E_{s_i}^+$. We adopt the notation $>_{s_i}$ to denote a source s_i preferring to cover e_j over e_k and model the observation of this preference using $H(\cdot)$, the Heaviside step function: $H(\cdot)$ is equal to $\mathbf{1.0}$ for positive inputs and to $\mathbf{0.0}$ otherwise. Therefore, $H(\hat{x}_{s_i, e_j} - \hat{x}_{s_i, e_k})$ would always be equal to $\mathbf{1.0}$ for an ideal predictor. In practice, $H(\cdot)$ is approximated by the differentiable logistic sigmoid $\sigma(\cdot)$.

Finally, we maximize BPR, our log-likelihood criterion

$$\text{BPR} := \sum_{(s_i, e_j, e_k) \in D} \ln \sigma(\hat{x}_{s_i, e_j} - \hat{x}_{s_i, e_k}) - \lambda_\Theta \, ||\Theta||^2. \tag{3}$$

Note the inclusion of an ℓ_2-regularization term over the set of parameters Θ. We please refer the reader to the work of Rendle et al. [35] for more details about this optimization scheme.

4.3 Temporal Setting

In this section, we describe the adoption of the dynamic embedding scheme proposed by Rudolph et al. [36] in the context of news coverage modeling. In particular, we adopt two strategies to maintain temporal consistency across time slices, respectively based (i) on a Gaussian random walk (RW) and (ii) on the addition of a temporal regularization term (RG). We adopt the notation $p_{s_i}^{(t)}$ to denote the embedding vector of news source s_i at the t-th time step.

Prior on the embedding vectors: Without information about former time slices, existing methods typically initialize embedding vectors to small, randomly distributed values. However, such approaches do not take advantage of any prior knowledge acquired during anterior training steps. The addition of a prior on embedding vectors represents a simple, yet powerful strategy to leverage previously acquired knowledge about sources. In particular, embedding vectors at the t-th time step are initialized using a Gaussian random walk around their final values at time step $(t-1)$. The Gaussian random walk is expressed as follows:

$$p_{s_i}^{(t)} \sim \begin{cases} \mathcal{N}(0, \gamma^{-1} I), & \text{if t=0}. \\ \mathcal{N}(p_{s_i}^{(t-1)}, \gamma^{-1} I), & \text{otherwise}. \end{cases} \tag{4}$$

This initialization scheme ensures a smooth transition of the parameter set learned in two consecutive time slices. This yields a more stable embedding space, offering a coherent expression of divergence across time-steps. Since events are inherently much more volatile than sources, we initialized their embedding vectors at random at each new time slice.

Optimization using temporal regularization: The second part of the dynamic scheme takes the form of a temporal regularization term. The newly introduced term penalizes large variations across time steps by minimizing the distance of an embedding vector at the t-th step to its final value at step $(t-1)$.

The final log-likelihood criterion, BPR-T, can then be formulated as follows for the t-th time split

$$
\begin{aligned}
\text{BPR-T}^{(t)} := &\sum_{(s_i, e_j, e_k) \in D} \ln \sigma(\hat{x}_{s_i, e_j} - \hat{x}_{s_i, e_k}) \\
&- \lambda_\Theta \, ||\Theta||^2 \\
&- \underbrace{\lambda_T ||p_{s_i}^{(t)} - p_{s_i}^{(t-1)}||^2}_{\text{temporal regularization}}
\end{aligned}
\tag{5}
$$

The model is optimized using stochastic gradient ascent and is fitted once for every time split. Update steps are defined as follows:

$$q_{e_j}^{(t)} \leftarrow q_{e_j}^{(t)} + \alpha(\sigma(-\hat{x}_{s_i,e_j,e_k}) \cdot \quad p_{s_i}^{(t)} \quad - \lambda_\Theta \, q_{e_j}^{(t)})$$

$$q_{e_k}^{(t)} \leftarrow q_{e_k}^{(t)} + \alpha(\sigma(-\hat{x}_{s_i,e_j,e_k}) \cdot (-p_{s_i}^{(t)}) - \lambda_\Theta \, q_{e_k}^{(t)})$$

$$p_{s_i}^{(t)} \leftarrow p_{s_i}^{(t)} + \alpha(\sigma(-\hat{x}_{s_i,e_j,e_k}) \cdot (q_{e_j}^{(t)} - q_{e_k}^{(t)})) \qquad (6)$$

$$- \lambda_\Theta \, p_{s_i}^{(t)}$$

$$- \lambda_T (p_{s_i}^{(t)} - p_{s_i}^{(t-1)}))$$

where α is our learning rate. We use the notation $\hat{x}_{s_i,e_j,e_k}$ to denote the quantity $(\hat{x}_{s_i,e_j} - \hat{x}_{s_i,e_k})$. Note that triplets (s_i, e_j, e_k), $e_j \in E_{s_i}^+$ and $e_k \in E \setminus E_{s_i}^+$ forming the training dataset D are randomly sampled during the optimization. [9]

4.4 Evaluation

To assess the performance of the different methods, we adopt a *leave-one-out* methodology, in which a single event per source is withheld at random from the training set D to constitute the test set D_s. This approach ensures that all sources have similar weights in the evaluation. We adopt the widely used *Area Under the Curve* (AUC) as a measure of performance. In the context of this work, the evaluation procedure is formally defined as follows

$$\text{AUC} = \frac{1}{|D_s|} \sum_{(s_i, e_j, e_k) \in D_s} H(\hat{x}_{s_i,e_j} - \hat{x}_{s_i,e_k}). \qquad (7)$$

where e_j is an event covered by s_i and e_k is an event that s_i hasn't covered, randomly sampled at testing time. Negative samples are drawn uniformly at random across all unique event of the current time slice (we omitted the time indices for the sake of brevity).

4.5 Experimental Setting

The code for experiment and analysis will be made available at publication time under an open-source license. All experiment-related code was run on a 6-core machine, equipped with an `Intel(R) Xeon(R) CPU E5-2630 @ 2.30GHz`, for a total training time of approximately 3 days. [10] We restricted the tuning of hyper-parameter λ_T to a subset of values $\in \{0.001, 0.01, 0.1, 1.0\}$. All scores and figures are reported using $\lambda_T = 0.1$, which we found to provide the highest cross-validated accuracy. For computational reasons, the other parameters were coarse-tuned on a static snapshot and were set to $K = 20$ and $\lambda_\Theta = 0.1$, $\gamma = 0.01$ and $\alpha = 0.1$.

5 RESULTS

In this section, we compare the existing static embedding model with the dynamic method presented in Section 4. The effect of both the prior on the embedding vectors and the temporal regularization are then measured in isolation.

Table 3 summarizes the performances of the various approaches by taking the mean AUC scores obtained, for each month, over the considered period. *BPR* denotes the core of the algorithm without

---

[9]We sampled both positive and negative examples uniformly. More complex sampling approaches exist [34] but are outside of the scope of this work.

[10]The training time is reported for the full 3-year period; a production-ready application would typically be optimized and use incremental updates instead.
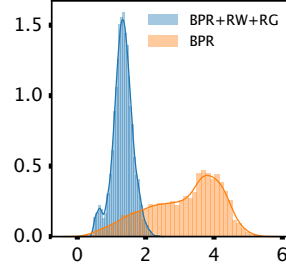

Figure 4: Average displacement of embeddings

	AUC
POP	0.6509
BPR	0.8959
BPR + RG	0.9089
BPR + RW	0.9318
BPR + RW + RG	**0.9337**

Table 3: Method contribution to performance

any temporal component. *RG* denotes the use of temporal regularization and *RW* denotes the use of a Gaussian random walk for embedding initialization. To compare with a non-parametrized approach, we also include *POP*, a popularity-based baseline: the score for a given event is a function of its frequency in the training set.

Several observations can be made in light of these results. Firstly, the combination of the two strategies *RG* and *RW* is shown to provide the best observed predictive performances. Secondly, the individual strategies do not provide the same performance improvements. Results suggest that a proper initialization contributes much more to an accurate prediction than strong temporal regularization. In parallel experiments, we even observed that an increase of λ_T decreases performances. We hypothesize that this is due to the model's inability to handle abrupt changes in source behavior, since a strong regularizer would penalize a large difference with respect to the previous time step.

The use of the proposed dynamic strategies (*RG* and *RW*) provides better overall consistency of the latent space across time slices. This is visible by measuring the average displacement of sources in embedding space. As shown in Fig. 4, sources are much more stable in the dynamic setting compared to the static embedding procedure. The added stability of the embedding space provides a usable expression of divergence across time-steps. This means source similarity can be coherently compared across the entire observed period, while also providing an overall improvement of the AUC scores (see Fig. 5).
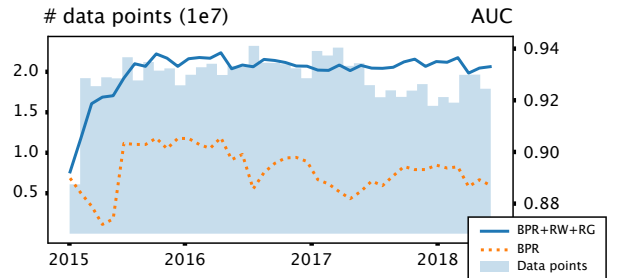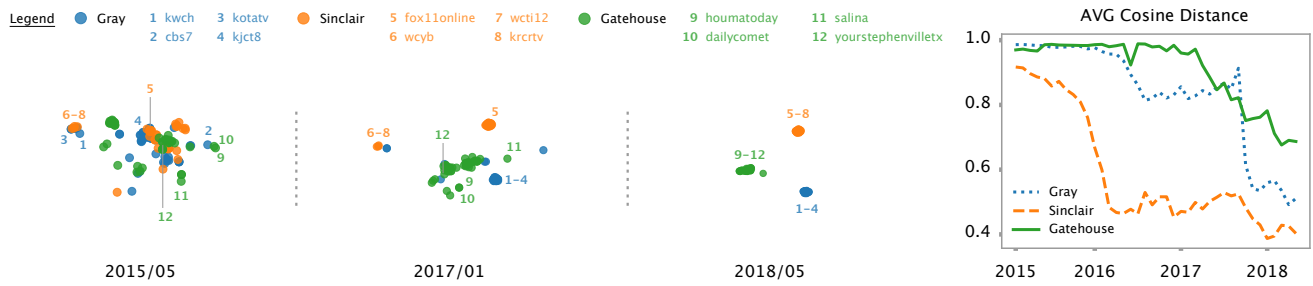

Figure 5: Performances (AUC) and dataset size per month

Figure 6: Case-study of the influence of ownership on the selection of covered events (best seen in color)

Left: Sample of the temporal evolution of the media landscape, learned with dynamic embeddings, illustrated with t-SNE. [44] To test the effect of ownership on diversity of coverage, we monitor the evolution of a set of sources all owned by the same broadcast group in 2018, backtracking their evolution in embedding space. Starting with an initial seed source known to be operated by - or affiliated to - one of three large American media conglomerates (Gray Television Inc., Sinclair Broadcasting Group and GateHouse Media), we build a set of 50 sources per group so as to have a comparably sized sample for each. These sources are also all verified to be owned - or operated by - one of these groups at the last observed time-frame of our dataset (May 2018) by cross-checking publicly available information. We project their positions from the embedding space onto a display-friendly plane, showcasing snapshots of their movement over time, as they collapse into highly similar and cohesive clusters in the similarity space.

Right: Average inter-source cosine distance between sources of each group in the dynamic embedding space, over time.

6 ANALYSIS

In the following section, we discuss the interpretation of the model introduced in Section 4. We first describe an approach to visualize the evolution of the news ecosystem. Then, we propose a systematic, unsupervised way to detect abrupt deviations in this space.

6.1 Visualizing the Media Landscape

We start by introducing an example case-study, which should illustrate the usefulness of the presented model in facilitating the understanding of the news ecosystem. The study is centered on three representative media conglomerates that we track throughout the 40-month period covered by our dataset: *Gray Television Inc.* (over 100 television stations), *Sinclair Broadcasting Group* (over 190 television stations) and *GateHouse Media* (over 140 newspapers). [11]

As a starting point, the model described in Section 4 is optimized in its most successful setting (*RW+RG*). Dimensionality reduction is performed in order to have a more interpretable view into the embedding vectors. In Fig. 6 - (*left*) we use a t-distributed Stochastic Neighbor Embedding (t-SNE [44]), a popular method for the visualization of high-dimensional data. In order to maintain consistency across time steps in the projection, we initialize the parameters of the t-SNE optimization procedure at time slice t with the final parameters of time step $t-1$. This seeds the projected points' positions instead of assigning them to a random initial position, allowing for easier tracking between time steps.

Additionally, in order to avoid interpreting from model parameters only, which might be misleading due to optimization artifacts, we correlate sources' trajectories with the average pairwise cosine distance between sources of each group. These distances are computed using sources' respective sets of covered events for each

month. Overall, this procedure allows us to coherently visualize the evolution of the media landscape over time, uncovering non-obvious dynamics at several scales, from the ecosystem as whole (e.g. convergence phenomenons) down to individual sources (e.g. shift toward a group).

The designed visualization of the media landscape is presented in Fig. 6. Please refer to Section 7 for a more detailed discussion.

6.2 Detecting Fluctuations in the Media Landscape

Even with extensive domain knowledge, tracking the evolution of a group of sources belonging to specific entities remains a tedious task, since it requires the manual identification of sources of interest and the validation of their common factors. Therefore, in the following section, we propose an unsupervised method which leverages the models' a priori knowledge to identify abrupt changes in sources' content diffusion patterns. In particular, the proposed framework aims to identify attractors, e.g. sources that tend of attract others in latent space, suggesting an alignment of coverage.

Attractors: News channels involved in a consolidation of resources typically tend to have increasingly similar coverage patterns. As seen in Fig. 6 - (*right*), the phenomenon manifests itself as the convergence of a subset of sources toward a common position in embedding space. Systematically detecting such gatherings around a common location, that we will loosely refer to as attractors, would allow to interpret each of these patterns in isolation. We propose a method in two steps. Firstly, we identify sources whose distances to other channels are abruptly reduced at any point in time. Secondly, we identify the absolute position towards which those sources tend to converge.

---

[11]as of August 2018.
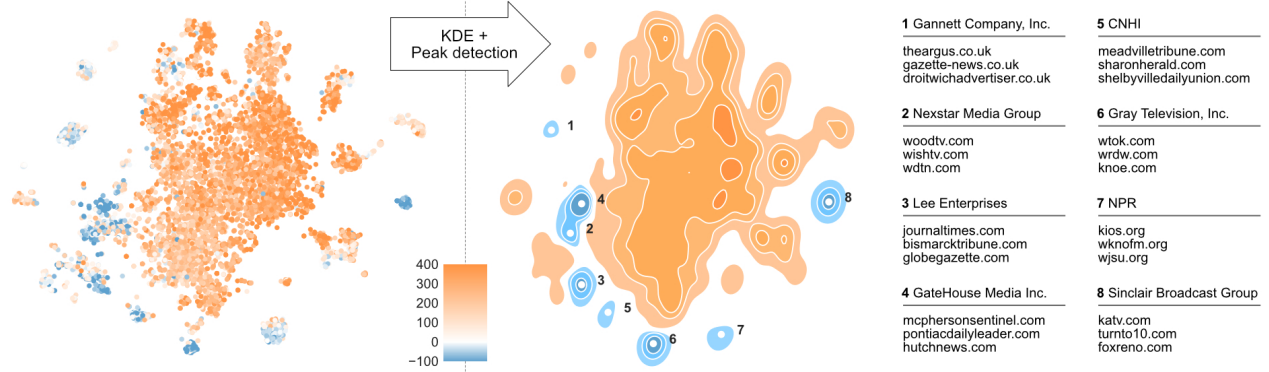
Figure 7: Left: t-SNE [44] projection of the embedding space for May 2018, colors represent the weight w_{s_i} of each source s_i. **Center:** Identified attractor map built from the attraction potential map ($\sigma = 1.9$, $k = 500$). **Right:** Detail and affiliation of attractors identified in (*center*), with the set of 3 sources closest to the uncovered poles. Best seen in color.

We first define the matrix $Z \in \mathbb{R}^{|S| \times |S|}$ that represents the cumulative difference of distances between any two news sources over time. We compute such distances in relative terms in order to avoid any drift component from the evolving latent space. More formally, we define Z as follows.

$$Z_{i,j} = \sum_t \quad \|p_{s_i}^{(t)} - p_{s_j}^{(t)}\| - \|p_{s_i}^{(t-1)} - p_{s_j}^{(t-1)}\|, \qquad (8)$$

where $\|\cdot\|$ is the Euclidean norm. In this setting, two sources whose distances are consistently reduced would produce a negative value of large magnitude in their corresponding entries of Z. Therefore, we rely on the matrix Z to identify channels having undergone large reductions of distance with other sources. In more details, we retrieve from Z the k-sources having the largest negative cumulative difference with any other source, i.e. the minimal value of each row in $Z_{i,j}$. By taking into consideration only the top-k, we can capture large shifts only and avoid considering small movements due to random factors. Once identified, the relative displacement of these k sources can be visualized in latent space. In particular, each source s_i in our dataset will be qualified by a single weight w_{s_i}, computed as the sum of the cumulative difference in Z with respect to the k considered sources. As shown in Fig. 7 - (*left*), negative values reveal sources that tend to exhibit agglomerative behaviors.

Until now, we observed strong fluctuations in terms of inter-source distances. The next step is to define a systematic way of identifying the centers around which these shifts occur, in absolute terms and at any given point in time. On a 2D projection of sources, we apply a weighted Kernel Density Estimator (KDE), a non-parametric method for estimating the Probability Density Function from a set of samples, under weak smoothness assumptions. The objective is to detect areas containing a high density of sources with negative weights and locate their density peaks. We use the weights w_{s_i} as input of the estimator. The bandwidth selection is a function of the data's covariance, multiplied by a constant factor

σ. [42] An example of the resulting density is presented in Fig. 7 - (*center*). Finally, local extrema are collected using a local minimum filter, a simple method routinely used in computer vision. The set of identified poles and the top-3 closest sources surrounding them is shown in Fig. 7 - (*right*).

Attractees: Having identified a set of attractors in latent space, the dual observation can be made: the identification of sources that experienced large movements in latent space toward any of the previously identified attraction poles. These are sources that have been strongly influenced by external forces, for example in the content consolidation phase after an acquisition. The detection of these phenomenon is done relative to a specific pole of attraction. In order to track the distance to a pole over time, we start with a set of seed sources. An obvious choice is to study the top-3 closest sources from the poles, detailed in Fig. 7 - (*right*). The ranking of sources having undergone a large shift can once again be made systematic. In particular, we rank sources according to the largest difference in distance to the pole between two consecutive time steps. The distance to the centroid of these sources yields the distance maps shown in Fig. 8 for the top-4 sources with the largest shifts.

7 DISCUSSION

The structure of the news landscape is in a constant state of flux. It is often difficult to follow the evolution of its organizational structure and even more so to determine what influenced these changes. In this section, we discuss how the fluctuations of broadcast patterns can be informative about channels' organizational structure. We report that important changes in this structure are identifiable through abrupt shifts in content diffusion, and showcase the models' ability to systematically highlight this variability in the coverage space.
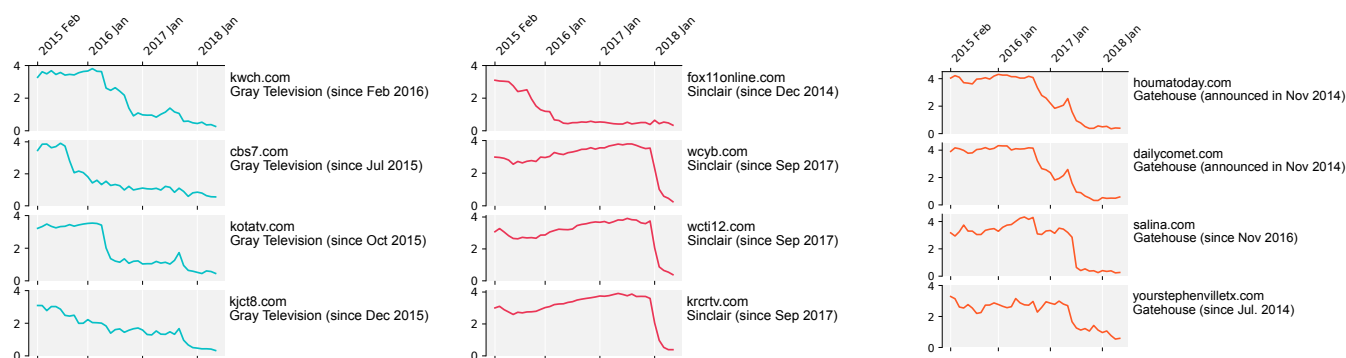
Figure 8: Euclidean distance from channels to their respective affiliations, in latent space, over time. Affiliation positions are computed as the centroid of 3 seed sources (taken from Figure 7 - (*right*)). Those patterns are extracted without any supervision: retrieved channels are the ones having undertaken the largest shift of content over the considered period (2015-2018). Channels acquisition dates are given for reference.

7.1 Observing the effects of ownership on the media landscape

The selection and diffusion of events by individual news channels could be influenced by a large number of factors, from obfuscated economic drivers to convoluted distribution schemes. Theoretically, the external pluralism assumption would prevent large-scale organizational changes in news outlets from inducing significant shifts in coverage. Our findings question the validity of this assumption.

We provide evidence that ownership can indeed exert its influence on the content being distributed downstream. The most distinct and recurring pattern pointing to this conclusion is the subsequent alignment of coverage patterns after an outlet's acquisition. Some examples are clearly observable in Fig. 6, such as the acquisition of 14 stations from the Bonten Media Group by Sinclair Broadcasting Group (SBG) in a deal completed on September 1st, 2017. [12] Visible through the lens of a decrease in the average inter-source distance of the Sinclair stations, this consolidation of coverage can also be tracked in the embedding spaces' visualization in Fig. 6 - (*left*). This can be observed in Fig. 8 as well, albeit with a slight delay with a sharp decrease in the distance from channels like `wcyb.com`, `wcti12.com` or `krcrtv.com` to the center of the Sinclair attraction pole (#8 in Fig. 7).

We observe similar behaviors during others large-scale acquisitions, for example in the purchase of a group of assets from the Morris Publishing Group by GateHouse in August 2017. [13] Other observations of this phenomenon include the sudden increase in coverage similarity of Gatehouse-owned stations around April 2017 (see Fig. 8 - *right*). While not directly correlated to a specific merger or acquisition, these movements could hint at a company-wide content alignment campaign.

Such observations also support the convergence hypothesis. The visualization in Fig. 6 exemplifies this effect: many of the sources that are present in one of these group's media portfolios start out from vastly different regions in embedding space. This is visible in their high initial average cosine distance in Fig. 6 - (*right*) and their dispersed placement in Fig. 6 - (*left*). In the last frame however, these same sources form highly coherent, tight groups in embedding space - and in the visualization. Despite the fact that this case-study back-tracks the evolution of sources across time, explaining the density of the last frame, their convergence points to a unification of coverage patterns over time.

7.2 Detecting highly influential broadcast groups

News outlets present complex content distribution schemes, as is particularly visible in television broadcasting: the on-air content is produced by a wide range of affiliates, from well-known household names to in-house teams, [14] being distributed through channels with another, often different set of owners. While the consolidation of broadcast material for economies of scale or investigative resources for economies of scope can be economically beneficial for the broadcaster, it is also potentially deceitful for a news consumer, as the exact origin of the broadcast content is not always known. By extension, the unique slant it carries in its selection of news is not clearly obvious. Not only does it carry is unique biases in terms of the way in which it covers the content, a topic not discussed in this work, but it also has the ability to over-emphasize or under-report certain events with little accountability.

This influence on coverage can be observed when interpreting the agglomeration dynamics highlighted by Fig. 7. Information sinks can be highlighted through the discovery of attraction poles in the embedding space.

---

[12]https://tvnewscheck.com/article/103465/sinclair-buying-bonten-stations-for-240m/

[13]https://www.poynter.org/news/gatehouse-acquires-morris-publishings-11-daily-newspapers

---

[14]A study conducted by Pew Research in 2014 already demonstrated a steady decline in locally produced content, with 1 in 4 local news stations not producing any of their own content [31].

Such sets of highly accretive sources, i.e. sources that draw other sources to align with their coverage, cluster neatly into large broadcast entities, some of which have been mentioned before. Fig. 7-(right) presents these groupings more exhaustively. The three large media groups chosen for analysis in Fig. 6 are present (*Gray Television Inc.*, *Sinclair Broadcasting Group* and *GateHouse Media Inc.*), along with several other large players in the American media landscape.

Previous studies [41] and [4] have studied differences in terms of the types of content covered in television and newspapers outlets, finding that TV stations cover proportionally more "global" news than newspapers. Television is traditionally thought to be more impacted by media consolidation for this reason: content is costly to produce, hence it makes sense for large entities to share their footage at scale. In our model, this should intuitively lead to the flagging of television conglomerates as the strongest attractors, with high content similarities. However, we observe in Fig. 7 that all mediums are represented and impacted by the convergence phenomenon. This could hint to the effect a convergence of mediums can have on the media landscape, with the efficacy co-ownership regulations being jeopardized by the all-encompassing nature of online content delivery.

7.3 Interpretation of the temporal consistency

None of this qualitative analysis would be possible without a temporal consistency constraint on the embedding space. Without such stability, the model could take advantage of an unnecessarily large number of degrees of freedom to align sources. In consequence, it would converge to very different solutions from one epoch to another. Due to the stochastic nature of the procedure, coverage changes would be rendered indistinguishable from optimization artifacts (see Fig. 3). By penalizing sources that deviate from their previous positions, only significant coverage changes can force a source to migrate to a different region in space. In other words, in order to provoke a displacement, the channels' coverage should differ enough from the previous time step to outweigh the temporal constraint. If this condition is met, the source will converge towards a different neighborhood that better fits its coverage patterns, typically getting closer to similar channels.

This variability in time can be tuned through the regularization parameter λ_T, as detailed in Section 4, providing a way to highlight more global dynamics - in the case of strong regularization - or more individual variations - with weak regularization. We also observe that the constraint provides predictive gains. This can be explained by the accumulation of knowledge about sources over time. This last hypothesis is corroborated by the pattern observed in Fig. 5, in which the accuracy reaches its maximum after the first few epochs before stabilizing until the end of the considered period.

8 CONCLUSION

This work tackles the problem of dynamically modeling the filtering decisions of individual actors of the news ecosystem. Beyond the predictive capabilities of the approach, the knowledge gathered by the model is leveraged to characterize the evolution of the media landscape over time. Firstly, we report performance benefits of adding a temporal component to a coverage prediction model. In particular, we show that a dynamic embedding model is able to outperform existing approaches thanks to its ability to propagate knowledge obtained from former time slices to the current prediction step. Secondly, we demonstrate the application of this model as a framework in which to reason about the latent structure of media landscape, modeling the temporal evolution of news outlets' decision processes. Maintaining a consistent latent representation of sources' preferences enables powerful interpretation and visualization methods, highly effective in investigating shifts in the media ecosystem, at a large scale but also at the individual source level.

We demonstrate the potential of the method on several channel acquisition campaigns. We show drastic post-acquisition content alignment in channels belonging to large, well-known broadcast conglomerates. This corroborates the hypothesis of deep consolidation of broadcast material inside news networks. Our work highlights the fragility of the external plurality assumption, which is supposed to guarantee a diversity of ownership and hence viewpoints. Finally, we automate this investigative process and explore several strategies to systemically identify abrupt variations in the news ecosystem, fingerprints of sharp changes in media programming.

Future work: The main focus of this work was to provide a clear and insightful representation of the news landscape, and we foresee several directions to pursue this research effort. Firstly, the method is ripe for the addition of content-based refinements. Our approach focuses on the information's propagation patterns but it is still blind to the content itself and the way it is handled by news outlets. We could for example mention in this vein a semantic analysis of the covered articles. The fact that the model is content-agnostic can however be of great use for its application to other domains. Such methods could be applied to other dynamic systems in which information propagates and evolves, such as online social-networks, knowledge bases or citation networks. Lastly, as has been highlighted throughout this work, particularly in the context of the case-study, we believe that our work would strongly benefit from an interdisciplinary approach, providing tools to journalists, policy-makers or economists, whose expertise would also add great insight to our analysis.

REFERENCES

[1] Evrim Acar, Daniel M Dunlavy, and Tamara G Kolda. 2009. Link prediction on evolving data using matrix and tensor factorizations. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on*. IEEE, 262–269.

[2] Jisun An and Haewon Kwak. 2016. Two Tales of the World: Comparison of Widely Used World News Datasets GDELT and EventRegistry. In *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM16)*. 619–622.

[3] Simon P. Anderson and Stephen Coate. 2005. Market Provision of Broadcasting: A Welfare Analysis. *The Review of Economic Studies* 72, 4 (2005), 947–972. http://www.jstor.org/stable/3700696

[4] Bergan D. Fico F. Lacy S. Wildman S.S. Baldwin, T. 2010. News Media Coverage of City Governments in 2009. (2010).

[5] Dylan Bourgeois, Jérémie Rappaz, and Karl Aberer. 2018. Selection Bias in News Coverage: Learning it, Fighting it. In *The International World Wide Web Conference 2018*.

[6] Stefano DellaVigna, Ethan Kaplan, Alan B. Krueger, Marco Manacorda, Enrico Moretti, Torsten Persson, Sam Popkin, Riccardo Puglisi, Matthew Rabin, Jesse M. Shapiro, Uri Simonsohn, Laura Stoker, David Stromberg, Tatyana Deryugina, Monica Deza, Dylan Fox, Melissa Galicia, Calvin Wai-Loon Ho, Sudhamas Khanchanawong, Richard M. Kim, Martin Kohan, Vipul Surender Kumar, Jonathan J. Leung, Clarice Li, Tze Yang Lim, Ming Mai, Sameer Parekh, Sharmini Radakrishnan, Rohan Relan, Dan Acland, Saurabh Bhargava, Avi Ebenstein, and Devin G. Pope. 2005. The Fox News Effect: Media Bias and Voting.

[7] Simeon Djankov, Caralee McLiesh, Tatiana Nenova, and Andrei Shleifer. 2003. Who Owns the Media? *Journal of Law and Economics* 46, 2 (2003), 341–381.

[8] Gillian Doyle. 2002. *Media Ownership: The Economics and Politics of Convergence and Concentration in the UK and European Media*. https://doi.org/10.4135/9781446219942

[9] Daniel M Dunlavy, Tamara G Kolda, and Evrim Acar. 2011. Temporal link prediction using matrix and tensor factorizations. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5, 2 (2011), 10.

[10] Matthew Gentzkow and Jesse Shapiro. 2010. What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica* 78, 1 (2010), 35–71. https://EconPapers.repec.org/RePEc:ecm:emetrp:v:78:y:2010:i:1:p:35-71

[11] Matthew Gentzkow, Jesse M Shapiro, and Michael Sinkinson. 2009. *The Effect of Newspaper Entry and Exit on Electoral Politics*. Working Paper 15544. National Bureau of Economic Research. https://doi.org/10.3386/w15544

[12] L.M. George and Joel Waldfogel. 2008. National Media and Local Political Participation: The Case of the New York Times. (01 2008), 33–48.

[13] Deborah J. Gerner, Rajaa Abu-Jabr, Philip A. Schrodt, and ÃÚmÃijr Yilmaz. 2002. Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions. In *of Foreign Policy Interactions. Paper presented at the International Studies Association*.

[14] Kristian Skrede Gleditsch, Nils W Metternich, and Andrea Ruggeri. 2014. Data and progress in peace and conflict research. *Journal of Peace Research* 51, 2 (2014), 301–314. https://doi.org/10.1177/0022343313496803 arXiv:https://doi.org/10.1177/0022343313496803

[15] Joshua S. Goldstein. 1992. A Conflict-Cooperation Scale for WEIS Events Data. *The Journal of Conflict Resolution* 36, 2 (1992), 369–385. http://www.jstor.org/stable/174480

[16] Tim Groseclose and Jeffrey Milyo. 2005. A Measure of Media Bias. *The Quarterly Journal of Economics* 120, 4 (2005), 1191–1237. https://EconPapers.repec.org/RePEc:oup:qjecon:v:120:y:2005:i:4:p:1191-1237.

[17] Ruining He, Chen Fang, Zhaowen Wang, and Julian McAuley. 2016. Vista: a visually, socially, and temporally-aware model for artistic recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 309–316.

[18] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*. International World Wide Web Conferences Steering Committee, 507–517.

[19] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. *2008 Eighth IEEE International Conference on Data Mining* (2008), 263–272.

[20] Henry Jenkins. 2004. The cultural logic of media convergence. *International journal of cultural studies* 7, 1 (2004), 33–43.

[21] Swetha Keertipati, Bastin Tony Roy Savarimuthu, Maryam Purvis, and Martin K. Purvis. 2014. Multi-level Analysis of Peace and Conflict Data in GDELT. In *MLSDA@PRICAI*.

[22] Yehuda Koren. 2009. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 447–456.

[23] Haewoon Kwak and Jisun An. 2014. A first look at global news coverage of disasters by using the gdelt dataset. In *International Conference on Social Informatics*. Springer, 300–308.

[24] Preethi Lahoti, Kiran Garimella, and Aristides Gionis. 2018. Joint non-negative matrix factorization for learning ideological leaning on Twitter. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 351–359.

[25] Edward J. Laurance. 1990. Events data and policy analysis:. *Policy Sciences* 23, 2 (01 May 1990), 111–132. https://doi.org/10.1007/BF00175597

[26] Kalev Leetaru and Philip A. Schrodt. 2013. GDELT: Global data on events, location, and tone. *ISA Annual Convention* (2013).

[27] Jeffery J. Mondak. 1995. Media exposure and political discussion in u.s. elections. *Journal of Politics* 57, 1 (1 5 1995), 62–85. https://doi.org/10.2307/2960271

[28] Felix Oberholzer-Gee and Joel Waldfogel. 2006. *Media Markets and Localism: Does Local News en EspaÃśol Boost Hispanic Voter Turnout?* Working Paper 12317. National Bureau of Economic Research. https://doi.org/10.3386/w12317

[29] Alexandra Olteanu, Carlos Castillo, Nicholas Diakopoulos, and Karl Aberer. 2015. Comparing Events Coverage in Online News and Social Media: The Case of Climate Change. In *ICWSM*. AAAI Press, 288–297.

[30] Rong Pan, Yunhong Zhou, Bin Cao, Nathan N Liu, Rajan Lukose, Martin Scholz, and Qiang Yang. 2008. One-class collaborative filtering. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 502–511.

[31] Deborah Potter and Katerina Eva Matsa. 2014. *State of the News Media 2014: A Boom in Acquisitions and Content Sharing Shapes Local TV News in 2013*. Technical Report. Pew Research Center.

[32] David Pritchard. 2002. Viewpoint Diversity in Cross-Owned Newspapers and Television Stations: A Study of News Coverage of the 2000 Presidential Campaign. (September 2002). https://docs.fcc.gov/public/attachments/DOC-226838A7.pdf

[33] Fengcai Qiao, Pei Li, Jingsheng Deng, Zhaoyun Ding, and Hui Wang. 2015. Graph-Based Method for Detecting Occupy Protest Events Using GDELT Dataset. In *Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CYBERC '15)*. IEEE Computer Society, Washington, DC, USA, 164–168. https://doi.org/10.1109/CyberC.2015.77

[34] Steffen Rendle and Christoph Freudenthaler. 2014. Improving pairwise learning for item recommendation from implicit feedback. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 273–282.

[35] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI*.

[36] Maja Rudolph and David Blei. 2017. Dynamic Bernoulli embeddings for language evolution. *arXiv preprint arXiv:1703.08052* (2017).

[37] Maja Rudolph, Francisco Ruiz, Stephan Mandt, and David Blei. 2016. Exponential family embeddings. In *Advances in Neural Information Processing Systems*. 478–486.

[38] Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 1679–1684.

[39] Diego Sáez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *CIKM*.

[40] P. A. Schrodt, S. G. Davis, and J. L. Weddle. 1994. Political Science: KEDS–A Program for the Machine Coding of Event Data. *Social Science Computer Review* 12, 4 (1994), 561.

[41] Holli Semetko and Patti Valkenburg. 2000. Framing European Politics: A Content Analysis of Press and Television News. 50 (06 2000), 93 – 109.

[42] B. W. Silverman. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

[43] Peter O. Steiner. 1952. Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting. *The Quarterly Journal of Economics* 66, 2 (1952), 194–223. https://EconPapers.repec.org/RePEc:oup:qjecon:v:66:y:1952:i:2:p:194-223.

[44] L.J.P van der Maaten and G.E. Hinton. Nov 2008. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 9: 2579âĂŞ2605 (Nov 2008).

[45] Tom Vizcarrondo. 2013. Measuring concentration of media ownership: 1976–2009. *International Journal on Media Management* 15, 3 (2013), 177–195.

[46] Liang Xiong, Xi Chen, Tzu-Kuo Huang, Jeff Schneider, and Jaime G Carbonell. 2010. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM, 211–222.

[47] James Edward Yonamine. 2013. A nuanced study of political conflict using the global datasets of events location and tone (GDELT) dataset. (2013).

[48] Wenchao Yu, Charu C Aggarwal, and Wei Wang. 2017. Temporally factorized network modeling for evolutionary network analysis. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 455–464.