Encoding Social Information with Graph Convolutional Networks for Political Perspective Detection in News Media

Chang Li
Purdue University
`li1873@purdue.edu`

Dan Goldwasser
Purdue University
`dgoldwas@purdue.edu`

Abstract

Identifying the political perspective shaping the way news events are discussed in the media is an important and challenging task. In this paper, we highlight the importance of contextualizing social information, capturing how this information is disseminated in social networks. We use Graph Convolutional Networks, a recently proposed neural architecture for representing relational information, to capture the documents' social context. We show that social information can be used effectively as a source of distant supervision, and when direct supervision is available, even little social information can significantly improve performance.

1 Introduction

Over the last decade we witness a dramatic change in the way information is generated and disseminated. Instead of a few dedicated sources that employ reporters and fact checkers to ensure the validity of the information they provide, social platforms now provide the means for any user to distribute their content, resulting in a sharp increase in the number of information outlets and articles covering news events. As a direct result of this process, the information provided is often shaped by their underlying perspectives, interests and ideologies. For example, consider the following two snippets discussing the comments made by a Democratic Senator regarding the recent U.S. government shutdown.

> **thehill.com** (*Center*)
>
> Sen. Mark Warner (D-Va.) on Sunday blasted President Trump for his "inept negotiation" to bring an end to the ongoing partial government shutdown. Warner, the ranking member of the Senate Intelligence Committee, lamented the effect the shutdown has had on hundreds of thousands of federal workers who have been furloughed or forced to work without pay.

> **infowars.com** (*Right*)
>
> Senator Mark Warner (D-Va.) is being called out on social media for his statement on the partial government shutdown. Warner blamed the "suffering" of federal workers and contractors on President Trump in a Sunday tweet framing Trump as an "inept negotiator". Twitter users pointed out that Democrats are attending a Puerto Rican retreat with over 100 lobbyists and corporate executives.

Despite the fact that both articles discuss the same event, they take very different perspectives. The first reporting directly about the comments made, while the second one focuses on negative reactions to these comments. Identifying the perspective difference and making it explicit can help strengthen trust in the newly-formed information landscape and ensure that all perspectives are represented. It can also help lay the foundation for the automatic detection of false content and rumors and help identify information echo-chambers in which only a single perspective is highlighted.

Traditionally, identifying the author's perspective is studied as a text-categorization problem (Greene and Resnik, 2009; Beigman Klebanov et al., 2010; Recasens et al., 2013; Iyyer et al., 2014; Johnson and Goldwasser, 2016), focusing on linguistic indicators of bias or issue-framing phrases indicating their authors' bias. These indicators can effectively capture bias in ideologically-charged texts, such as policy documents or political debates, which do not try to hide their political leaning and use a topic-focused vocabulary. Identifying the authors' bias in news narratives can be more challenging. News articles, by their nature, cover a very large number of topics resulting in a diverse and dynamic vocabulary that is continuously updated as new events unfold. Furthermore, unlike purely political texts, news narratives attempt to maintain credibility and seem im-

partial. As a result, bias is introduced in subtle ways, usually by emphasizing different aspects of the story.

Our main insight in this paper is that the social context through which the information is propagated can be leveraged to alleviate the problem, by providing both a better representation for it, and when direct supervision is not available, a distant-supervision source based on information about users who endorse the textual content and spread it. Several recent works dealing with information dissemination analysis on social networks, focused on analyzing the interactions between news sources and users in social networks (Volkova et al., 2017; Glenski et al., 2018; Ribeiro et al., 2018). However, given the dynamic, and often adversarial setting of this domain, the true source of the news article might be hidden, unknown or masked by taking a different identity. Instead of analyzing the documents' sources, our focus is to use social information, capturing how information is shared in the network, to help guide the text representation and provide additional support when making decisions over textual content.

We construct a *socially-infused textual representation*, by embedding in a single space the news articles and the social circles in which these articles are shared so that the political biases associated with them can be predicted. Figure 1 describes these settings. The graph connects article nodes via activity-links to users nodes (*share*), and these users in turn are connected via social-links (*follow*) to politically affiliated users (e.g., the Republican or Democratic parties twitter accounts). We define an embedding objective capturing this information, by aligning the documents representation, based on content, with the representation of users who share these documents, based on their social relations. We use a recently proposed graph embedding framework, Graph Convolutional Networks (GCN) (Kipf and Welling, 2016, 2017) to capture these relationships. GCNs are neural nets operating on graphs, and similar to LSTMs on sequences, they create node embeddings based on the graph neighborhood of a given node. In the context of our problem, the embedding of a document takes into account the textual content, but also the social context of users who share it, and their relationships with other users with known political affiliations. We compare this powerful approach with traditional graph embedding methods that only capture local relationships between nodes.

Given the difficulty of providing direct supervision in this highly dynamic domain, we study this problem both when direct supervision over the documents is available, and when using *distant-supervision*, in which the document level classification depends on propagating political tendencies through social network, which is often incomplete and provides conflicting information.

To study these settings we focus on U.S. news coverage. Our corpus consists of over 10k articles, covering more than 2k different news events, about 94 different topics, taking place over a period of 8 years. We remove any information about the source of the article (both meta-data and in the text) and rely only on the text and the reactions to it on social media. To capture this information, we collected a set of 1.6k users who share the news articles on Twitter and a handful of politically-affiliated users followed by the sharing users, which provide the distant supervision. We cast the problem as a 3-class prediction problem, capturing left-leaning bias, right-leaning bias or no bias (center).

Our experimental results demonstrate the strength of our approach. We compare direct text classification or node classification methods to our embedding-based approach in both the fully supervised and distant supervised settings, showing the importance of socially infused representations.
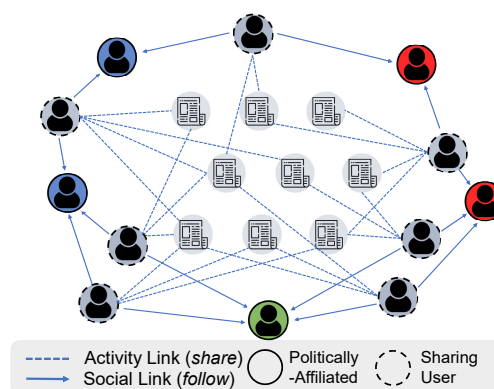

Figure 1: Information Flow Graph

2595

2 Related Work

The problem of perspective identification is typically studied as a supervised learning task (Lin et al., 2006; Greene and Resnik, 2009), in which a classifier is trained to differentiate between two specific perspectives. For example, the *bitterlemons* dataset consisting of 594 documents describing the Israeli and Palestinian perspectives. More recently, in SemEval-2019, a hyperpartisan news article detection task was suggested[1]. The current reported results on their dataset are comparable to ours, when using text information alone, demonstrating that it is indeed a challenging task. Other works use linguistic indicators of bias and expressions of implicit sentiment (Greene and Resnik, 2009; Recasens et al., 2013; Choi and Wiebe, 2014; Elfardy et al., 2015). In recent years several works looked at indications of framing bias in news articles (Baumer et al., 2015; Budak et al., 2016; Card et al., 2016; Field et al., 2018; Morstatter et al., 2018). We build on these work to help shape our text representation approach.

Recent works looked at false content identification (Volkova et al., 2017; Patwari et al., 2017), including a recent challenge[2] identifying the relationship between an article's title and its body. Unlike these, we do not assume the content is false, instead we ask if it reflects a different perspective.

Using social information when learning text representations was studied in the context of graph embedding (Pan et al., 2016), extending traditional approaches that rely on graph relations alone (Perozzi et al., 2014; Tang et al., 2015; Grover and Leskovec, 2016) and information extraction and sentiment tasks (Yang et al., 2016a; West et al., 2014). In this work we focus on GCNs (Kipf and Welling, 2017; Schlichtkrull et al., 2018), a recent framework for representing relational data, that adapts the idea of convolutional networks to graphs. Distant supervision for NLP tasks typically relies on using knowledge-bases (Mintz et al., 2009), unlike our setting that uses social information. Using user activity and *known* user biases was explored in (Zhou et al., 2011), our settings are far more challenging as we do not have access to this information.

Articles	10,385	Twitter Users	1,604
-Left	3,931	Pol. Users	135
-Right	2,290	Left Pol. Users	49
-Center	4,164	Right Pol. Users	51
Sources	86	Center Pol. Users	35
Types	94	Avg # shared per Article	23.29
Events	2,020	Avg # pol. users followed	20.36

Table 1: Dataset Statistics

3 Dataset Description

We collected 10,385 news articles from two news aggregation websites[3] on 2,020 different events discussing 94 event types, such as elections, terrorism, etc. The websites provide news coverage from multiple perspectives, indicating the bias of each article using crowdsourced and editorial reviewed approaches[4]. We preprocessed all the documents to remove any information about the source of the article.

We collected social information consisting of Twitter users who share links to the collected articles. We focused on Twitter users who follow political users and share news articles frequently (100 articles minimum). We found 1,604 such Twitter users. The list of political users was created by collecting information about active politically affiliated users. It consists of 135 Twitter users who are mainly politicians, political journalists and political organizations. The set of political users and Twitter users are disjoint. The summary of the dataset is shown in Table 1.

Data Folds We created several data splits to evaluate our model in the supervised settings, based on three criteria: *randomly* separated, *event* separated and *time* separated splits. In the event-separated case, we divide the news articles such that all articles covering the same news event will appear in a single fold. For the time-separated case, we sort the publication dates (from oldest to latest) and divide them in three folds. Each time one fold is used as training data (33%) and the other two combined as test data (66%). We use the same folds throughout the experiment of supervised classification for evaluation purpose.

Constructing the Social Information Graph We represent the relevant relationships as an information graph, similar to the one depicted in Figure 1. The social information graph $G = \{V, E\}$,

---

consisting of several different types of vertices and edges, is defined as follows:

- Let $P \subset V$ denote the set of the *political users*. These are Twitter users with a clear, self-reported, political bias. They may be the accounts of politicians (e.g., Sarah Palin, Nancy Pelosi), political writers in leading newspapers (e.g., Anderson Cooper) or political organizations (e.g., GOP, House Democrats). Note that even political users that share a general political ideology can differ significantly in the type of issues and agenda they would pursue, which would be reflected in their followers.

- Let $U \subset V$ denote the set of *Twitter users* that actively spread content by sharing news articles. The political bias of these users is not directly known, only indicated indirectly through the *political users* they follow on Twitter.

- Let $A \subset V$ denote the set of news articles shared by the *Twitter users* (U).

The graph vertices are connected via a set of edges described hierarchically, as follows:

- $E_{UP} \subset E$: All the Twitter users are connected to the political users whom they follow. Note that a Twitter user may be connected to many different political users.

- $E_{AU} \subset E$: All the articles are connected to the Twitter users who share them. Note that an article may be shared by many different Twitter users.

4 Text and Graph Model

Our goal is to classify news articles into 3-classes corresponding to their bias. Since we have both the textual and social information for the news articles, we can obtain representations for them using either the text or graph models. In this section, we briefly go through the text representation methods, and then move to describe the graph based models we considered in this paper.

4.1 Text Representations and Linguistic Bias Indicators

To predict the bias of the news articles, we can consider it as a document classification task. We use the textual content of a news article to generate a feature representation. Deciding on the appropriate representation for this content is one of the key design choices. Previous works either use traditional, manually engineered representations for capturing bias (Recasens et al., 2013) or use latent representations learned using deep learning methods (Iyyer et al., 2014). We experimented with several different choices of the two alternatives, and compared them by training a classifier for bias prediction over the document directly. The results of these experiments are summarized in Table 2. Due to space constraints, we provide a brief overview of these alternatives, and point to the full description in the relevant papers.

Linear BoW Unigram features were used. The articles consist of 77,772 unique tokens. We used TFIDF vectors as unigram features obtained by using scikit-learn (Pedregosa et al., 2011).

Bias Features These are content based features drawn from a wide range of approaches described in the literature on political bias, persuasion, and misinformation, capturing structure, sentiment, topic, complexity, bias and morality in the text. We used the resources in (Horne et al., 2018b) to generate 141 features based on the news article text, which were shown to work well for the binary hyper-partisan task (Horne et al., 2018a).

Averaged Word Embedding (WE) The simplest approach for using pre-trained word embeddings. An averaged vector of all the document's words using the pre-trained GloVe word embeddings (Pennington et al., 2014) were used to represent the entire article.

Skip-Thought Embedding Unlike the Averaged word vector that does not capture context, we also used a sentence level encoder, Skip-Thought (Kiros et al., 2015), to generate text representations. We regard each document as a long sentence and map it directly to a 4800-dimension vector.

Hierarchical LSTM over tokens and sentences We used a simplified version of the Hierarchical LSTM model (Yang et al., 2016b). In this case documents are first tokenized into sentences, then each sentence was tokenized into words. We used a word-level LSTM to construct a vector representation for each sentence, by taking the average of all the hidden states. Then, we ran another single layer unidirectional LSTM over the sentence representations to get the document representation by taking average of all the hidden states.

4.2 Graph-Based Representations

In addition to the textual information, the news articles are also part of the information network defined in Section 3. Intuitively, news articles shared by the same Twitter users are likely to have the same bias, and users who share a lot of news in common are close in their political preferences. A similar intuition connects users who follow similar politically affiliated users. Capturing this information allows us to predict the bias of a news article, given its social context. We design our embedding function to map all graph nodes into a low dimensional vector space, such that the graph relationships are preserved in the embedding space. In the shared embedding space, nodes that are connected (or close) in the graph should have higher similarity scores between their vector representations.

4.2.1 Directly Observed Relationships in Graph (DOR)

Our first embedding approach aims to preserve the local pairwise proximity between two vertices directly. This is similar to first-order graph embedding methods (Tang et al., 2015). There are two different relations observed in the graph: Twitter user to political user (follow) and news article to Twitter user (share). We construct our embedding over multiple views of the data, each view w corresponds to a specific type of graph relation. We can then define an loss function L_w for each view w as follows:

- Twitter User to Political User (UP): This objective maximizes the similarity of a Twitter user, u and all the political users in the set $P_u \subset P$, where P_u is the set of political users that u follows.

$$L_{UP} = - \sum_{u \in U} \sum_{p \in P_u} log P(p|u) \quad (1)$$

- News Article to Twitter User (AU): This objective maximizes the similarity of a news articles, a and all the Twitter users in the set $U_a \subset U$, where U_a is the set of Twitter users who shared news article a on Twitter.

$$L_{AU} = - \sum_{a \in A} \sum_{u \in U_a} log P(u|a) \quad (2)$$

All the conditional probabilities can be computed using a softmax function. Taking $P(p|u)$

as an example:

$$P(p|u) = \frac{exp(e_u^T e_p)}{\sum_{q \in P} exp(e_u^T e_q)} \quad (3)$$

where e_u and e_p are embeddings of twitter user u and political user p respectively.

Computing Eq. 1 and Eq. 2 can be expensive due to the size of the network. To address this problem, we refer to the popular negative sampling approach (Mikolov et al., 2013), which reduce the time complexity to be proportional to the number of positive example pairs (i.e. number of edges in our case).

The loss defined for the two views are summed with the classification loss defined in Eq. 9 as the final loss function to be optimized in DOR embedding model.

$$L_{DOR} = L_{clf} + L_{UP} + L_{AU} \quad (4)$$

4.2.2 Graph Convolutional Networks (GCN)

Graph Convolutional Networks is an efficient variant of convolutional neural networks which operate directly on graphs. It can be regarded as special cases of a simple differentiable message-passing framework (Gilmer et al., 2017):

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in N(i)} M^{(l)}(h_i^{(l)}, h_j^{(l)}) \right) \quad (5)$$

where $h_i^{(l)} \in \mathbb{R}^{d^{(l)}}$ is the hidden state of node v_i in the l-th layer of the neural network, with $d^{(l)}$ as the dimensionality of representation at layer l. $N(i)$ is the set of direct neighbors of node v_i (usually also include itself). Incoming messages from the local neighborhood are aggregated together and passed through the activation function $\sigma(\cdot)$, such as $\tanh(\cdot)$. $M^{(l)}$ is typically chosen to be a (layer-specific) neural network function. Kipf and Welling (2017) used a simple linear transformation $M^{(l)}(h_i^t, h_j^t) = W^{(l)} h_j$ where $W^{(l)}$ is a layer-specific weight matrix.

This linear transformation has been shown to propagate information effectively on graphs. It leads to significant improvements in node classification (Kipf and Welling, 2017), link prediction (Kipf and Welling, 2016), and graph classification (Duvenaud et al., 2015).

One GCN layer can be expressed as follows:

$$H^{(l+1)} = \sigma(\hat{A} H^{(l)} W^{(l)}) \quad (6)$$

where $\hat{A}$ is the normalized adjacency matrix, and $W^{(l)}$ is the layer-specific trainable weight matrix. $H^{(l)} \in \mathbb{R}^{N \times D^{(l)}}$ is the matrix of hidden states in the l-th layer. $H^{(0)} = X$ is the input vectors. It can either be one-hot representations of nodes or features of the nodes if available. $\sigma(\cdot)$ is the activation function.

Multiple GCN layers can be stacked in order to capture high-order relations in the graph. We consider a two-layer GCN in this paper for semi-supervised node classification. Our forward model takes the form:

$$V = \tanh\left(\hat{A}\tanh\left(\hat{A}XW^{(0)}\right)W^{(1)}\right) \quad (7)$$

where X is the input matrix with one-hot representations and V is the representation matrix for all nodes in the graph.

Figure 2 shows an example of how our GCN model aggregates information from a node's local neighborhood. The orange document is the node of interest. Blue edges link to first order neighbors and green edges link to second order neighbors.
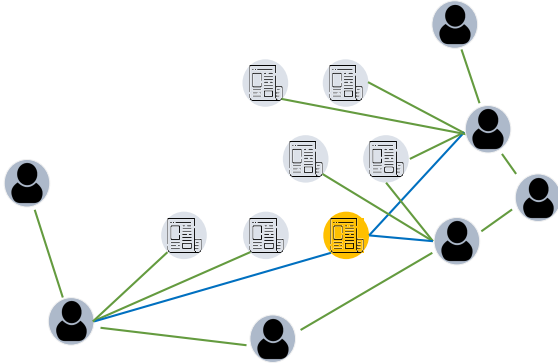

Figure 2: Example of Unfolding of GCN Computational Graph

4.3 Document Classification

The representation v of a news article (obtained with text models or graph models) captures the high level information of the document. It can be used as features for predicting the bias label with a feed-forward network.

$$p = \text{softmax}(W_c v + b_c) \quad (8)$$

We use the negative log likelihood of the correct labels as classification training loss:

$$L_{clf} = -\sum_a \log p_{aj} \quad (9)$$

where j is the bias label of news article a.

5 Joint Model

Given that we have two representations available for news articles, namely the textual one and social one, it is natural to make the prediction combining both of them. We propose to align the representations of the same document from graph and text models in a joint training fashion as shown in Figure 3. The objective function for the alignment is:

$$L_{align} = -\sum_{a \in A} \log P(e_a^G | e_a^T) \quad (10)$$

where e_a^T is the embedding for document a based on its content, and e_a^G is the embedding for document a based on graph structures. $P(e_a^G | e_a^T)$ is defined the same way as in Eq. 3. Negative sampling is again utilized to reduce time complexity.

Connecting the text and graph embedding of the same news articles, allows the bias signal to flow between the two sides. Therefore the text model may learn from the social signal and the graph model may use textual content to adjust its representation as well. We describe the loss function for the joint model in two settings - full supervision (i.e., labels associated with *documents* directly) and distant supervision, when bias information is only provided for a handful of *users*, which do not actively share documents.

Full Supervision In the full supervision case, the loss consists of three parts, namely the classification loss of text model (L_{clf}^T), the classification loss of graph model (L_{clf}^G), and the loss for aligning the embeddings of the text and the graph models (L_{align}).

$$L_{joint} = \alpha L_{clf}^T + \beta L_{clf}^G + \gamma L_{align} \quad (11)$$

Here α, β and γ are hyper-parameters to adjust the contribution of the three parts. We set all of them to default value 1 in experiments in this paper.

Distant Supervision Unlike the full supervision case where we have training labels for documents, we only have access to the labels of political users. However, since the text and social representation use the same space, user bias information can be propagated to the document representation, acting as a distant supervision source.

Inference Given the graph representation, decisions can be made in multiple ways. Each document has a dual representation, as a text node and a social node. Also, given the social context of a
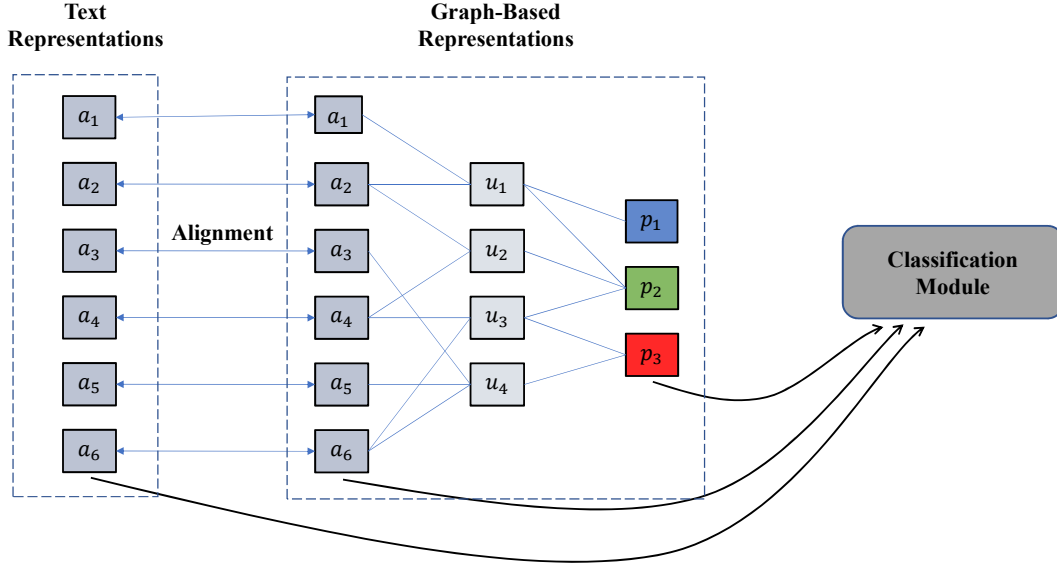
Figure 3: Overall Architecture: Representations are learned for news articles based on textual information and graph structure; these two representations are aligned in our joint model; only labels of political users are available during training in distant supervision case

document, decision can be defined over the users that share it (assuming that users tend to share information which agrees with their biases). To take advantage of that fact, we define a simplified inference process. At test time, we can predict the bias of a news article with the embeddings from text model (Text), the embeddings from graph model (Graph), and the embeddings of sharing users who shared this article (User). The last method (User) works by averaging bias prediction scores s_u^b for all Twitter users that shared an article a. The bias prediction score is computed in Eq. 8 before the $\text{softmax}(\cdot)$ applied.

$$\arg\max_b \frac{\sum_{u \in U_a} s_u^b}{|U_a|} \qquad (12)$$

Finally, two or three of the scores listed above can be combined to make the decision.

6 Experiments

We designed our experiments to evaluate the contribution of social information in both the fully supervised setting, and when only distant supervision is available through the social graph. We begin by evaluating several text classification models that help contextualize the social information. Finally, we evaluate our model's ability to make predictions when very little social information is available at test time.

6.1 Implementation Details

We used the spaCy toolkit for preprocessing the documents. All models are implemented with PyTorch (Paszke et al., 2017). Hyperbolic tangent (tanh) is used as non-linear activation function. We use feed-forward neural network with one hidden layer for the bias prediction task given textual or social representation. The sizes of LSTM hidden states for both word level and sentence level are 64. The sizes of hidden states for both GCN layers are 16. For the training of the neural network, we used the Adam optimizer (Kingma and Ba, 2014) to update the parameters. We use 5% of the training data as the validation set. We run the training for 200 epochs (50 epochs for HLSTM models), and select the best model based on performance on validation set. Other parameters in our model includes negative sample size k=5, mini-batch size b=30 (mini-batch update only used for HLSTM models). The learning rate is 0.001 for HLSTM models and 0.01 otherwise.

6.2 Experimental Results

Text Classification Results The result of supervised text classification is summarized in Table 2. We report the accuracy of bias prediction. Results clearly show that HLSTM outperforms the other methods in supervised text classification setting. Also, adding the hand engineered bias features with HLSTM representation does not help to

2600

improve performance.

Model	Split	Text
Majority	Rand	40.10
	Event	40.10
	Time	40.50
Linear BoW	Rand	58.47
	Event	59.88
	Time	55.41
Bias Feat.	Rand	54.06
	Event	53.51
	Time	52.96
Avg WE	Rand	59.37
	Event	59.37
	Time	53.46
SkipThought	Rand	68.67
	Event	66.35
	Time	60.89
HLSTM	Rand	**74.59**
	Event	**73.55**
	Time	**66.98**
HLSTM + Bias Feat.	Rand	69.32
	Event	69.87
	Time	66.79

Table 2: Supervised Classification Using Textual Features

Network Classification Results   We show the results of predicting bias using graph information alone, without text, in Table 3. The GCN model outperforms DOR significantly in each of the four settings. Similar to the text classification results, performance on random and event splits are comparable. However, there is a sharp drop in performance for time split. This can be explained by the fact that temporally separated news events will discuss different entities and world events and as a result will have very different word distributions. Event-separated splits are less susceptible to this problem, as similar figures and topics are likely to be discussed in different events.

Model	Split	Graph	User	G+U
DOR	Rand	74.74	72.02	74.57
	Event	74.87	72.74	75.18
	Time	65.65	65.07	65.36
	Dist	56.45	56.95	56.54
GCN	Rand	88.65	78.83	**88.89**
	Event	**88.78**	76.11	88.70
	Time	81.14	71.31	**82.00**
	Dist	63.72	40.08	**67.03**

Table 3: Classification Results Using Social Relations in Full Supervised and Distant Supervised Setting

Joint Model Results   Table 4 shows the results of our joint model. When aligning the text and graph embeddings using joint training, both show improvement, and prediction with text or graph representations alone is better than those listed in Table 2 and 3, especially for text. Note that the increase in accuracy is much greater for the more expressive HLSTM model. Making prediction with the aggregation of multiple scores usually leads to better accuracy.

Interestingly, the model's distant supervision performance is almost comparable with fully supervised text classification results. This demonstrates the strength of our joint model, and its ability to effectively propagate label information from users down to documents.

We also evaluated our model when smaller amounts of social information was available at test time. We tested our joint model with only 50% and 10% of the links for test articles kept. The results are summarized in Table 5. Clearly the performance improves as more social links are available. However, even with little social links provided in the latter case, our joint model propagates information effectively and results in an increase in performance compared to text classification.

Qualitative Analysis   In Table 6, we compared the bias prediction by our text and joint model on several news articles (only titles shown in the table). These examples demonstrate the subtlety of bias expression in text, which helps motivate social representations to support the decision.

7   Conclusion

In this paper we follow the intuition that the political perspectives expressed in news articles will also be reflected in the way the documents spread and the identity of the users who endorse them. We suggest a GCN-based model capturing this social information, and show that it provides a distant supervision signal, resulting in a model performing comparably to supervised text classification models. We also study this approach in the supervised setting and show that it can significantly enhance a text-only classification model.

Modeling the broader context in which text is consumed is a vital step towards getting a better understanding of its perspective. We intend to study fine-grained political perspectives, capturing how different events are framed.

Acknowledgements

Model	Split	Graph	User	G+U	Text	G+T	G+U+T
GCN + SkipThought	Rand	89.95	81.49	89.75	70.61	90.34	91.02
	Event	89.40	79.06	89.64	69.16	90.15	90.78
	Time	84.95	76.59	85.30	64.12	84.09	86.25
	Dist	67.78	45.30	70.03	58.68	69.82	70.66
GCN + HLSTM	Rand	89.03	83.66	88.57	86.84	91.48	**91.74**
	Event	89.34	80.22	88.62	88.39	91.69	**91.72**
	Time	84.83	74.50	85.09	81.36	85.57	**86.21**
	Dist	71.74	69.39	71.16	61.13	**72.16**	71.85

Table 4: Results of Joint Model Combining Text and Graph Relations

Model	Split	Graph	User	G+U	Text	G+T	G+U+T
GCN + HLSTM (50%)	Rand	86.73	78.62	86.24	85.62	89.31	89.35
	Event	86.55	78.34	85.89	84.52	89.21	89.51
	Time	82.25	70.93	81.45	80.05	85.57	85.48
GCN + HLSTM (10%)	Rand	76.13	57.76	75.55	78.61	81.35	81.49
	Event	76.58	57.10	75.75	77.60	80.55	80.93
	Time	73.24	54.09	72.48	72.92	76.52	76.75

Table 5: Results of Joint Model with Reduced Links for Test Documents

Text	Joint	Gold	Title
Right	Right	Right	Hacked Powell email reveals Hillary hates Obama for 2008
Right	Right	Right	Donald Trump will let James Comey testify
Center	Center	Center	Clinton: I am done with being a candidate
Center	Center	Center	Senate confirms Sessions as attorney general
Left	Left	Left	Clinton: Trump Doesn't See President Obama as an American Video
Left	Left	Left	Trump uses Twitter to promote leaked intelligence on North Korea
Center	Left	Left	Hillary Clinton's Campaign Says It Will Participate In Wisconsin Recount
Left	Center	Center	Supreme Court justices hint at striking Voting Rights Act provision
Left	Center	Right	Boston Marathon bombs: how investigators use technology to identify suspects
Right	Right	Left	Israel risks becoming apartheid state if peace talks fail, says John Kerry

Table 6: Examples of Bias Prediction by Text and Joint Model

References

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482, Denver, Colorado. Association for Computational Linguistics.

Beata Beigman Klebanov, Eyal Beigman, and Daniel Diermeier. 2010. Vocabulary choice as an indicator of perspective. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 253–257, Uppsala, Sweden. Association for Computational Linguistics.

Ceren Budak, Sharad Goel, and Justin M Rao. 2016. Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271.

Dallas Card, Justin Gross, Amber Boydstun, and Noah A. Smith. 2016. Analyzing framing through the casts of characters in the news. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1410–1420, Austin, Texas. Association for Computational Linguistics.

Yoonjung Choi and Janyce Wiebe. 2014. +/- EffectWordNet: Sense-level lexicon acquisition for opinion inference. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1181–1191, Doha, Qatar. Association for Computational Linguistics.

David Duvenaud, Dougal Maclaurin, Jorge Aguilera-Iparraguirre, Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P. Adams. 2015. Convolutional networks on graphs for learning molecular fingerprints. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, pages 2224–2232, Cambridge, MA, USA. MIT Press.

Heba Elfardy, Mona Diab, and Chris Callison-Burch. 2015. Ideological perspective detection using semantic features. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 137–146, Denver, Colorado. Association for Computational Linguistics.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in Russian news: a computational analysis of intricate political strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*,

pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272, International Convention Centre, Sydney, Australia. PMLR.

Maria Glenski, Tim Weninger, and Svitlana Volkova. 2018. Identifying and understanding user reactions to deceptive and trusted social news sources. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 176–181, Melbourne, Australia. Association for Computational Linguistics.

Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 503–511, Boulder, Colorado. Association for Computational Linguistics.

Aditya Grover and Jure Leskovec. 2016. Node2vec: Scalable feature learning for networks. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 855–864, New York, NY, USA. ACM.

Benjamin D. Horne, William Dron, Sara Khedr, and Sibel Adali. 2018a. Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, pages 235–238, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Benjamin D. Horne, Sara Khedr, and Sibel Adali. 2018b. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018.*, pages 518–527.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122, Baltimore, Maryland. Association for Computational Linguistics.

Kristen Johnson and Dan Goldwasser. 2016. "all I know about politics is what I read in twitter": Weakly supervised models for extracting politicians' stances from twitter. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2966–2977, Osaka, Japan. The COLING 2016 Organizing Committee.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Thomas N. Kipf and Max Welling. 2016. Variational graph auto-encoders. *CoRR*, abs/1611.07308.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Skip-thought vectors. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 3294–3302.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on?: Identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, CoNLL-X '06, pages 109–116, Stroudsburg, PA, USA. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 3111–3119, USA. Curran Associates Inc.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011, Suntec, Singapore. Association for Computational Linguistics.

Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R. Corman, and Huan Liu. 2018. Identifying framing bias in online news. *Trans. Soc. Comput.*, 1(2):5:1–5:18.

Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, and Yang Wang. 2016. Tri-party deep network representation. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 1895–1901. AAAI Press.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch.

Ayush Patwari, Dan Goldwasser, and Saurabh Bagchi. 2017. TATHYA: A multi-classifier system for detecting check-worthy statements in political debates. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 2259–2262.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 701–710, New York, NY, USA. ACM.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659, Sofia, Bulgaria. Association for Computational Linguistics.

Filipe Nunes Ribeiro, Lucas Henrique, Fabrício Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P. Gummadi. 2018. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018.*, pages 290–299.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, pages 593–607.

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 1067–1077, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 647–653, Vancouver, Canada. Association for Computational Linguistics.

Robert West, Hristo S. Paskov, Jure Leskovec, and Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2.

Yi Yang, Ming-Wei Chang, and Jacob Eisenstein. 2016a. Toward socially-infused information extraction: Embedding authors, mentions, and entities. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1452–1461, Austin, Texas. Association for Computational Linguistics.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016b. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, San Diego, California. Association for Computational Linguistics.

Daniel Xiaodan Zhou, Paul Resnick, and Qiaozhu Mei. 2011. Classifying the political leaning of news articles and users from user votes. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*.