Credibility Assessment in the News: Do we need to read?

James Fairbanks
Georgia Tech Research Institute
Atlanta, GA
james.fairbanks@gtri.gatech.edu

Natalie Fitch
Georgia Tech Research Institute
Atlanta, GA
natalie.fitch@gtri.gatech.edu

Nathan Knauf
Georgia Institute of Technology
Atlanta, GA
nate.knauf@gatech.edu

Erica Briscoe
Georgia Tech Research Institute
Atlanta, GA
erica.briscoe@gtri.gatech.edu

ABSTRACT

While news media biases and propaganda are a persistent problem for interpreting the true state of world affairs, increasing reliance on the internet as a primary news source has enabled the formation of hyper-partisan echo chambers and an industry where outlets benefit from purveying “fake news”. The presence of intentionally adversarial news sources challenges linguistic modeling of news article text. While modeling text content of articles is sufficient to identify bias, it is not capable of determining credibility. A structural model based on web links outperforms text models for fake news detection. Analysis of text based methods for bias detection reveals the existence of liberal words and conservative words, but there is no analogue for fake news words versus real news words.

CCS CONCEPTS

• **Machine Learning**; • **Natural Language Processing**;

KEYWORDS

Machine Learning, Natural Language Processing, Graphical Models, Fake News Detection, Online Media, Social Media, Media Bias, Political Bias Detection, WWW, News articles

ACM Reference Format:
James Fairbanks, Natalie Fitch, Nathan Knauf, and Erica Briscoe. 2018. Credibility Assessment in the News: Do we need to read?. In *Proceedings of WSDM workshop on Misinformation and Misbehavior Mining on the Web (MIS2).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

The adage that “a lie gets halfway around the world before the truth has a chance to get its pants on” has never been more true than in the age of online media, where information (and misinformation) spreads widely and quickly. Although *fake news* is not a new problem, its recent reach is unprecedented and must be mitigated with novel strategies to preserve valuable public and private institutions. Our recent work focuses on the problem of detecting fake news, which we contrast to the problem of detecting mere *bias* in online media.

While the term “Fake News” may be contested and includes many varieties, such as fabricated stories, clickbait, and negative coverage, our focus is concerned with two distinct problems in the study of problematic journalism. The first problem we denote as bias detection, which identifies the political bias of an article or publisher (conservative vs liberal). The second problem we denote as credibility assessment, which determines the truthfulness (fact-based reporting) of the article or publisher (credible vs not credible). We frame these problems as binary classification tasks. We use news articles and their metadata from the Global Database of Events Language and Tone (GDELT) to train, validate, and test our models [20]. For each task, we evaluate the performance of two different classification methods, one based on article content and the other based on structural properties. The content model serves as a baseline approach, using Natural Language Processing (NLP) methods to construct textual features and supervised learning to predict an article’s credibility and bias. The structural method is a probabilistic graphical model approach, leveraging the link citation network among articles and domains in order to make predictions using a belief propagation algorithm.

This paper demonstrates the following: *a*) fake news articles can be detected sans text using Belief Propagation on the link structure, *b*) while biased articles can be detected using text or links, only links can reveal the fake news articles and *c*) this biased article detection model for online media focuses on specific keywords. The following sections detail current research in automated fake news detection, the GDELT dataset, our classification methodology, and conclusions.

2 RELATED WORK

Fake news can be described as articles written in the style of a newspaper that is false and written with the intent to deceive or mislead [1], but the form it takes may exhibit a large degree of variability. A variety of forms compose the fake news genre, such as clickbait, (low quality journalism intended to attract advertising revenue), news stories using digitally altered images or fabricated facts, stories which erroneously describing a photo or video, mispairing a photo with written content, reporting factually on only one side of a story, counterfeit news sources or twitter accounts, articles that cite questionable sources, satire/irony, and conspiracy

MIS2, 2018, Marina Del Rey, CA, USA

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

theories, among other examples. Due to this variety and the high-profile and ubiquitous nature of fake news, especially in politics, researchers are studying methods to mitigate this problem.

2.1 Political Bias Detection Methods

Although problematic journalism adversely affects many different areas, one area that is particularly vulnerable is US politics. The 2016 US presidential campaign provides salient examples of fake news. Both social media and traditional news outlets perpetuated fake news stories during this time. Many survey respondents in one study admitting to believing false news stories [1]. Furthermore, Allicott and Gentzkow [1] show that fake news headlines from the 2016 campaign were believed at similar rates as well-known conspiracy theories, such as "President Obama was born in another country" or the "US government knew the 9/11 attacks were coming but consciously let them proceed."

It comes as no surprise then that there has been an explosion of academic research efforts to combat fake news by specifically tackling the task of political bias and propaganda identification. Most often political bias prediction methods are built around dissecting news article text alone, akin to how a human might detect bias when reading an article. Attempts to simulate this complex process of human reasoning usually utilize deep learning methods. Convolutional Neural Nets (CNN) and Recurrent Neural Networks (RNN) are popular frameworks to train models for bias detection in words or sentences as previously reported [10, 13, 22].

News articles can be classified by their providers using CNNs in combination with a bidirectional RNN to detect which sentences in particular are "bias-heavy" and are the most informative features for the classification task [10]. Iyyer et al. points out how simplistic linguistic models such as bag-of-words ignores sentence structure and phrasing with respect to political ideology detection. Instead, they use RNNs to combine semantic and syntactic information at the sentence level and demonstrate how the composition of a sentence is a better predictor of its political ideology [13]. Rao et al. use Long Short-Term Memory (LSTM) to classify U.S. political bias in twitter messages as supporting either Democratic or Republican viewpoints using word embeddings and neural networks [22].

Word embedding methods are also an important component to any NLP type task where text must be transformed into features. Both studies in [10] and [22] use various methods such as bag-of-words, and pre-trained models such as "bag-of-vectors" using GloVe vectors [19]. We create a baseline content-based model using term frequency inverse document frequency (TF-IDF) weighted matrix of singular words and bigrams and also use paragraph vectors for comparison [16].

2.2 Credibility Assessment Methods

Prior to the rise of the internet, the only people with access to large audiences were authors working through editorial systems such as academic journals, book publishers, and newspaper editors. These editorial systems control access to audiences and enforce ethical norms such as honesty and objectivity within the academic, scholarly, and journalistic communities. The democratization of online media enables anyone to set up a website and publish content, with much of this content being published on social media. However, the desire to control advertising revenue and an influx of political campaign funding has led to the proliferation of websites disguised as newspaper sites that are directed to push political agendas. This phenomenon is present on both the political left and right with differences in the issues covered.

Formal fact checking processes are a modern invention. "A Nexis search suggests that as recently as 2000 and 2001, no news outlet ran a 'fact check' after the State of the Union address. In 2003 and 2004, only the Associated Press fact-checked the annual speech to Congress. In 2010 and 2011, by contrast, the address drew dozens of fact-checking articles and segments" [8]. Manual fact checking efforts appear to suffer from slow reaction times which allow falsehoods to spread further than truths, and accusations of bias that prevent readers from changing their minds in light of fact checker's evidence.

Most research into online media credibility has been focused on social media, where content is generated by any user without evaluation by knowledgeable gatekeepers or complying with editorial standards. There are several logistical benefits of studying social media including the availability of the data from standard "streams" offered by the social media service, the homogeneity of the data, and the structured metadata available on social media posts. General online media is a more heterogeneous environment with larger engineering burdens on researchers.

Prior to the advent of modern social media, scholars studied the formation of trust networks in the construction of semantic web resources, this research focused on trust networks in the authorship network [7]. Like the traditional media gatekeepers, attention to trust on the web focuses on identifying sources such as publishers, editors, and authors that are trustworthy and reputable sources, rather than verifying the accuracy of individual articles, stories, or facts. NLP models for detecting rumors on Twitter are quite good using the text of the tweet [21].

Therefore, we like other researchers in this field view link analysis as an important tool for fact checking news reports and political statements. Related research shows that it is possible to verify the accuracy of statements by politicians about facts in various domains such as history and geography [6] by using the Wikipedia knowledge graph and shortest path based "semantic proximity" metric. The idea of reputation in linked media has also been explored with connections to linear algebra [9], and probabilistic models [3]. The Web of Trust is an internet browser extension that uses a reputation scheme to protect internet users from cyber security threats. However, using link information requires link resolution as a preprocessing step. For this task we turn to Hoaxy for guidance. Hoaxy is a platform for tracking online misinformation, which uses the Twitter API to identify differences between posts that spread fake news and those that combat it [23]. In this paper we adopt Hoaxy's method for defining canonical urls: "To convert the URLs to canonical form, we perform the following steps: first, we transform all text into lower case; then we remove the protocol schema (e.g. 'http://'); then we remove, if present, any prefix instance of the strings 'www.' or 'm.'; finally, we remove all URL query parameters." Like all URL canonicalization procedures, this is a heuristic.[1]

[1]One can find online services that route content using urls embedded in query parameters of links for example the oembed standard, which means this heuristic can remove potentially useful information.

3 DATASET DESCRIPTION

The dataset that we use for this study is a collection of articles from The Global Database of Events Language and Tone (GDELT) Project [20], whose mission is to monitor the world's broadcast, print, and web news information in over 100 languages in order to create a free and open platform for computing. The GDELT Event Database contains over 431 million events from 1979 to present day and accounts for supranational, state, sub-state, and non-state actors. We augment the GDELT database, which only stores article metadata, with the text and links from the article sources. A PostgreSQL database stores metadata from new articles obtained from the GDELT event stream every 15 minutes. The source url is then accessed and the downloaded content parsed using the Newspaper library.[2] This semistructured information is stored in a Mongo database for later retrieval and analysis [5]. In particular, the content based approach queries for article text, while the structural approach queries for article links.

Since we use supervised learning, labels were crawled from the website Media Bias Fact Check (MBFC) [4]. MBFC is a volunteer-run fact checking site rating websites based on political/idealogical bias and credibility of factual reporting. Ratings are subjective but are based on a structured rubric and numerical scoring system to assign labels. For each source (domain), a minimum of 10 headlines and 5 stories are reviewed against four categories:

> *a*) Biased Wording/Headlines (Does the source use loaded words to convey emotion to sway the reader? Do headlines match the story?), *b*) Factual/Sourcing (Does the source report facts and back up claims with well sourced evidence?), *c*) Story Choices (Does the source report news from both sides or do they only publish one side?), and *d*) Political Affiliation (How strongly does the source endorse a particular political ideology? In other words how extreme are their views?).

MBFC computes the average score across the four categories and converts it to five categorical labels for political bias ("Right," "Right Center," "Center," "Left Center," and "Left"). Conversely, only the numerical score for the second category "Factual/Sourcing" is used to produce the five categorical labels for credibility ("Very Low," "Low," "Mixed," "High," and "Very High"). Defining the credible labels in this way acknowledges that credibility in this paper is defined according to this single, objective metric while related metrics such as "fairness" or "impartiality" are used to measure bias. Note, each of these labels are assigned at the publisher or domain level and therefore every article originating from the same source will have the same set of labels. This method is similar to how people subscribe to (or ignore) entire publications rather than individual articles or authors.

Since our classifiers predict a binary label, we combine labels "Right"/"Right Center" and "Left"/"Left Center" to form our labels "conservative" and "liberal," respectively, for the bias problem. Similarly, we combine "Low"/"Very Low" and "High"/"Very High" to form our classifier labels denoted as "credible" and "not credible," respectively, for the credibility assessment problem . Although the GDELT database is very large, only a fraction of articles could qualify as useful in the content based model. The number of articles that had both a set of labels scraped from MBFC and textual information was 124,300 from 242 domains. Since the structural method does not rely on article text, articles with link data were collected to create a graph with 19,786 nodes (domains) and 32,632 edges (links).[3] One benefit of structural methods is the ability to learn from articles that do not have annotated labels.

[2] https://github.com/codelucas/newspaper/

4 METHODS

Two separate approaches are developed for bias detection and credibility assessment. The content model establishes baseline performance and uses more traditional text-based features and classification methods. The structural method leverages more sophisticated graphical and statistical analysis in which we show improved performance over the content model.

4.1 Modeling Article Text

The content model relies on traditional natural language processing methods to extract textual features, then using classification methods to establish baseline performance for our problem. All text processing and model training, validation, and testing occur using Python and various helpful packages such as *scikit-learn* [18], *pymongo* [12], and *spacy* [11]. Textual information is aggregated to form a corpus and represented as a TF-IDF matrix, which is input for the classifier.

Each article is treated as an individual data point (i.e. a document). For each article, text is cleaned via the following pipeline:

(1) foreign character removal/replacement
(2) contraction expansion (i.e. "can't" becomes "can not")
(3) punctuation removal
(4) stop word removal
(5) word lemmatization

The remaining words and bi-grams are the vocabulary in the TF-IDF matrix representation of the corpus. For this task there are a total of 124,300 articles and about 7.5 million items in the vocabulary after text cleaning. Then, these document TF-IDF vectors act as features for two different classification models: Logistic Regression and Random Forest.

We also apply a deep learning doc2vec model to this problem [16] to contrast the TF-IDF approach. The specific model is the Wikipedia Distributed Bag of Words from https://github.com/jhlau/doc2vec. A Distributed Bag of Words is used in accordance with the literature on empirical document vectors [15]. These pre-trained paragraph vectors are used as features in a Logistic Regression classifier.

The total dataset is split into 80% for the training set and 20% for the holdout test set. During the training phase, 3-fold cross-validation was performed using the training set to construct the best classifier. Then the final classifier was tested against the holdout test set. Finally, receiver operating characteristic (ROC) curves and area under the curve (AUC) scores were computed to quantify classifier prediction performance for all models.

[3] Prior to filtering out social media buttons and sites that only link to themselves there are 29,692 unique domains.

4.2 Reputation in the Graph Structure

Unlike print newspaper articles which have unstructured citations to other articles, online media features structured links in the form of HTML tags. The great advances in search engine technology, Pagerank and HITS, was using link structure instead of textual content to determine web site importance [14, 17]. We apply a semi-supervised graphical learning algorithm called loopy belief propagation (BP) [3] to the information contained in the web structure of online media. Our implementation is written in the Julia Programming Language [2].

In general, the BP approach treats each node as a random variable $x_i \in \{0, 1\}$ where the output is a marginal probability $p(x_i)$ quantifying the *belief* that a node i belongs to class x_i. A node's *belief* denoted $b(x_i)$ (or class label probability) is inferred from both the prior knowledge of a node $i's$ class ("conservative/liberal" or "credible"/"not credible") and also neighbors of node i, $N(i)$. The BP algorithm is iterative and intuitively works by passing *messages* denoted by $m_{ij}(x_j)$, which is the message from node i to node j about node $j's$ likelihood of being in class x_j. More formally, the message update Equation 1 is given below:

$$m_{ij}(x_j) \leftarrow \sum_{x_i \epsilon X} \phi(x_i)\psi_{ij}(x_i, x_j) \prod_{k \epsilon N(i)/j} m_{ki}(x_i) \tag{1}$$

The function $\phi(x_i)$ represents the *a priori* belief that node i belongs to class x_i, and is used to encode the known labels for the training set. The posterior beliefs are calculated after the message propagation is complete. The function $\psi_{ij}(x_i, x_j)$ is a hyper-parameter that determines the conditional probability that if a neighboring node i is of class x_i, then its neighbor j will be of class x_j. Table 1 shows the choice of the affinity matrix ψ, for $\epsilon > 0$ this choice of ψ assumes homophily of the labels.

Table 1: Edge Potentials between Neighboring Nodes

$\psi_{ij}(x_i, x_j)$	x_i	x_j
x_i	1-ϵ	ϵ
x_j	ϵ	1-ϵ

Finally, the posterior node beliefs are computed from the final messages according to the following Equation 2:

$$b_i(x_i) = k\phi(x_i) \prod_{x_j \epsilon N(i)} m_{ji}(x_i) \tag{2}$$

The total number of articles used for this task is larger than the content based approach since there are many more articles that link to ones that do not exist in the GDELT Event Database. Therefore, link information is captured for these articles but no text information. For this task, there are articles from a total 19,786 domains with 32,632 links of the types described in Table 2 to create the graph.

The graph $G = (V, E)$ is undirected and unweighted where the set of domain names are the nodes V and each link shared between a source and destination domain corresponds to the set of edges E. After the graph is constructed, 3-fold cross-validation is used to evaluate prediction performance of the BP algorithm. Specifically, a third of the nodes' labels are withheld and assigned an *a priori* probability of 50% likely to be in either class, while the other two-thirds of the nodes are initialized to have 99% probability as belonging to their true class label. AUC scores are computed for each test fold and the final AUC score is the average across all three folds. An example of the structural bias model is seen in Fig. 1. Nodes are color-coded according to the computed posterior beliefs (more blue for liberal and more red for conservative) after the BP algorithm has terminated.

Table 2: Link Types used in Graph Construction

HTML Tag	Description
`<a>`	Mutually linked sites (text content)
`<link>`	Shared CSS (visual style)
`<script>`	Shared JavaScript files (user interaction)
`<img>`	Common images, logos, or icons (visual content)

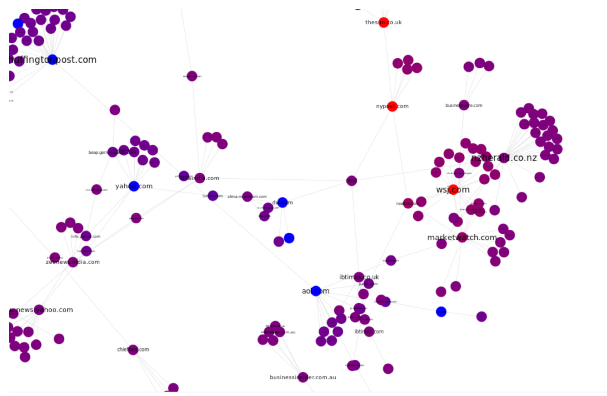

Figure 1: A subset of the bias graph model color-coded to show bias truth labels (blue=liberal, red=conservative, purple=centered) illustrating mainstream both left of center (huffingtonpost, yahoo news, ibtimes.co.uk) and right of center news sources (wall street journal, marketwatch, businesswire).

5 DISCUSSION

5.1 Content vs Structure

The simple construction of the content model provides a performance baseline for both bias detection and credibility assessment of articles. For both the bias and credibility problems, Logistic Regression using TF-IDF matrix features out-performed both Random Forest and Logistic Regression using pre-trained doc2vec features. For the bias problem, the class label distribution is approximately 60/40 for liberal/conservative articles, respectively. The best AUC score from the content model is 0.926, which is achieved using Logistic Regression (TF-IDF), as can be seen in Fig. 2. However, results for the credibility problem in the content-based approach proved to be overly optimistic due to extreme class imbalance: 1,107 "not credible" to 99,969 "credible" articles. This class imbalance led to a classifier trained to almost always predict the majority class, which

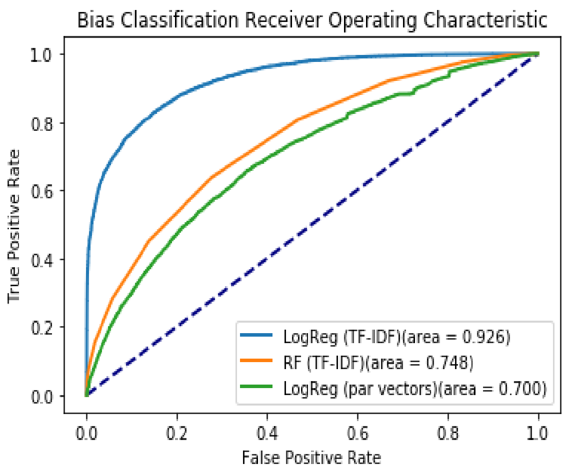

Figure 2: ROC Curves for Content Model Bias Detection

leads to an inflated AUC score of 0.973. When sample weights were adjusted relative to the distribution of labels in each training set fold and when randomly under-sampling the holdout test set to include balanced counts of each label, the holdout test set AUC score drops to 0.358

On the other hand, the structural method achieves improved performance for the credibility problem over the content based approach despite the class imbalance. The ROC curves for each of the 3 folds and average AUC score is shown in Figure 3 for bias and Figure 4 for credibility. The fact that credibility assessment is improved in the structural approach validates our intuition that detecting a source's credibility of factual reporting is more difficult to do based on text alone, since "real" and "fake" news articles use similar words and phrases to report their respective narratives. The application of more meaningful linguistic features such as sentence structure and sentiment would improve automatic credibility assessment of news articles.

Table 3: Summary of AUC Results

Task/ModelAUC	Content (LogReg)	Structural
bias	0.926	0.931
credibility	0.358	0.889

5.2 Word Play: Informative Words in the Bias and Credibility Problems

5.2.1 Liberal and Conservative Words. Since the content model performed well in classifying bias, we assume that certain words in this dataset may indicate a "liberal" vs "conservative" bias, which allows us to evaluate the most informative words according to the magnitude of coefficients in the logistic regression model. A selection of the top 1% of informative conservative and liberal words are featured in Table 4. After some scrutiny, these terms may divulge more about the peculiarities and properties of the dataset

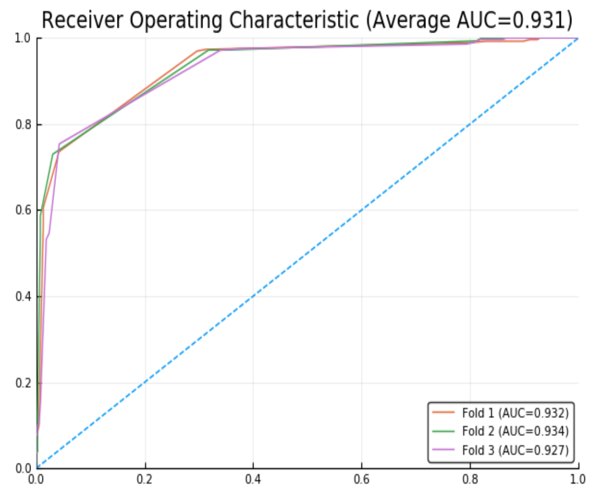

Figure 3: Structure-Based Method: Bias Detection Performance

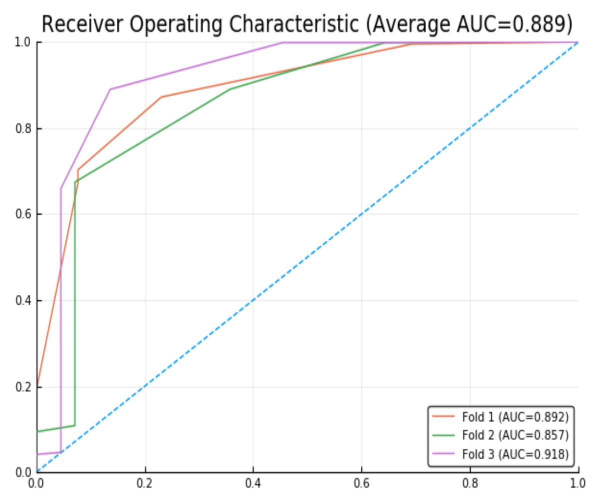

Figure 4: Structure-Based Method: Credibility Assessment Performance

used rather than any objective truth about which words or phrases are good indicators of political bias.

First, a large majority of the top terms (most not repeated here) are simply words that point to a specific domain or publisher. For example, the conservative terms "sputnik," "caller," ""wa," and "tlr" refer to "sputniknews.com," "dailycaller.com," "thewest.com.au," and "thelibertarianrepublic.com," respectively, and all of which are labeled as "Right" or "Right Center" by Media Bias Fact Check. Similarly, on the liberal side, the terms "thomson"/"reuters" and "dailys" refer to "news.trust.org" (the news arm of the Thomson Reuters Foundation) and "elitedaily.com," respectively, and all of which are labeled as "Left" or "Left Center".

Second, many of the top terms can usually be explained by their proximity to their sources in that they are either buzzwords or represent broader topics heavily reviewed by them. For example, the liberal words "rs" and "crore" are terms for Indian currency (where "Rs" is the symbol for the rupee and "crore" is a short-hand term indicating a large amount of Indian currency). This makes sense when we consider that almost 25% of the articles queried are sourced from either "timesofindia.indiatimes.com" or "economictimes.indiatimes.com", which are liberal leaning sites. Similarly, the conservative term "perth" is the capitol city of Western Australia that is mentioned frequently in the conservative-leaning site "thewest.com.au". In other words, the resulting distribution of sources in which a top word appears is dominated by either liberal or conservative sources.

Third, sometimes terms suggest how a publisher operates or the style of the publisher. For example, the term "paywall", which is when a site restricts access to certain content with a subscription, shows up frequently on one liberal-leaning site "qz.com". The liberal terms "getty" and "image" come together to suggest that a number of liberal (and probably conservative) leaning sites use the American photo agency Getty Images, Inc to support their reports. On the conservative side, the term "afp" refers to the Agence France-Presse, which is the third largest global news agency and of which Getty Images, Inc is also a partner.

Fourth, fortunately there are also a couple of interesting terms that do appear to make a statement connecting political bias to the content of an article. For example, one of the top conservative words is "wire" or "wires". There are a total of 337 articles that mention the conservative term "wire" or "wires", which notably, is only a fraction of the 124,300 articles in the dataset. Therefore we wanted to know, why is this term a good predictor of bias? Nearly half of the articles containing the term "wire(s)" (142/337) originate from the conservatively labeled domain "nypost.com" because many of their articles contain the phrase "post wires". However, of the remaining 195 articles containing the term "wires" or "wire", in the liberal-leaning sources there is a 2:1 mention of the term within the context of "Obama wire tapping Trump" vs referring to a theme of terrorism to include buzzwords such as "explosives" or "terrorists" or "bomb". In other words, liberal sources including the term "wire" more likely refers to the talking point concerning "Obama wire tapping Trump" than the topic of "terrorism". On the other hand, the remaining conservative sources containing the term "wire" or "wires" usually refer to it within the context of "terrorism" but had zero instances referring to it within the context of "Obama wire tapping Trump", at least in this dataset. This trend also pairs well with the conservative term "daesh" (appearing frequently on sputniknews.com), which is a derogatory alternative to the term "ISIS" meant to delegitimize the terrorist group. One more example is the liberal term "aug". This term is simply the abbreviation of the month of August. Besides August apparently being a newsworthy month, a couple of dates jumped out over and over again. One is August 12th and refers to the recent politically-charged Charlottesville protest/riot and the other date is August 25th, which refers to reports of over 400,000 Rohingya fleeing from insurgent attacks in Myanmar. It turns out that both conservative and liberal sites mention these events, however in this dataset there are 3,382 liberal leaning articles to only 839 conservative leaning articles that talk about them.

Table 4: Most Informative Bias Words in Content-Based Model

Conservative	Liberal
sputnik	advertisement
afp	aug
daesh	rs, crore
tlr	getty, image
caller	thomson, reuters
wa, perth	dailys
wire	paywall

5.2.2 Credible vs Not Credible Words. The logistic regression model for credible vs non-credible articles produces interpretable lists of words that indicate whether an article is more or less likely to be credible. For full results and coefficients see Table 6 and Table 7. Table 5 contains selected words from these lists. Credible news articles mostly contain words that are typical of newspaper style, such as “photo,” ”image,” ”support,” and ”campaign”. The non-credible word list contains more highly specialized nouns indicating that they refer to specific conspiracy theories, rather than a general style of writing. For example “wikileaks,” ”dnc,” and ”fbi” refers to the specific conspiracy theories surrounding the 2016 US presidential race where the FBI investigated wikileaks publications of Democratic National Committee emails. Also, the South American criminal gang MS13 is a subject of right wing conspiracy theories. Another category of words associated with non-credible articles are the last names of specific public figures and places such as “Beck,” ”Girod,” ”Arroyo,” ”Bohlender,” and ”Greece”. This analysis shows that one challenge to any content (text) based model of fake news articles will be the constantly changing landscape of conspiracy theories and current events that reference specific people, places, or organizations.

6 CONCLUSIONS

Structural analysis of online media articles can identify fake news articles given a fraction of labeled samples. Textual analysis successfully identifies bias in online news articles, but is insufficient to determine credibility. For bias determination, the content model’s most informative terms reveal patterns and peculiarities in the underlying dataset, but do not always reveal associations with conservative and liberal bias. Relevant words for credibility assessment, especially with respect to non-credible articles, appear to be highly tailored to the specific conspiracy theories found in the training set. We believe this is a result of the adversarial writing process where fake news authors are trying to convince readers that the article if real. Future work should focus on developing more robust and novel features that can generalize to works from unseen publishers and topics.

Further research should study the credibility problem from a generative process perspective, understanding how fake news authors write the articles with an intent to deceive the reader. Taking the economic perspective of click streams and advertising revenue is critical to countering the propagation of fake news. We posit that the best techniques for solving the fake news problem will

Table 5: Selected words associated with credible vs non-credible articles. Notice that the credible words are mostly generic journalistic words, while the noncredible words are highly specific referring to a particular person, organization, or location.

Credible	Noncredible
said	follow
photo	investwatchblog
image	views
july	antimedia
women	[daily] caller
support	revolutionizing
podcast	wikileaks
india	greece
campaign	christian
owner	arroyo
cent	beck
picture	graviola
percent	antifa
care	dnc
did	ms13
images	wolves
politics	fbi
app	girod
indian	bohlender

likely combine structural information from the web network with the content information in the article text. One example of this combined approach is to insert the predicted probability from the content model as the *a priori* $\phi(x_i)$ probability of the article node's class label before running Belief Propagation.

In order to facilitate research into fake news, it is important to capture the dynamic aspects of the rapidly changing propaganda networks. Fake articles are edited, challenged, posted at multiple sites, and taken down. These dynamics cannot be captured without accessing links multiple times and analyzing changes to content. A collaborative network of researchers building a shared dataset will be required to progress research in this field.

A APPENDIX

For completeness we include tables of coefficients for credible vs non credible words.

ACKNOWLEDGMENTS

The authors would like to acknowledge Joel Schlosser for his advice as well as the feedback from the anonymous reviewers.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 2 (May 2017), 211–236. https://doi.org/10.1257/jep.31.2.211

[2] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah. 2017. Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.* 59, 1 (Jan. 2017), 65–98. https://doi.org/10.1137/141000671

[3] D. Chau, C. Nachenberg, J. Wilhelm, A. Wright, and C. Faloutsos. 2011. Polonium: Tera-Scale Graph Mining and Inference for Malware Detection. In *Proceedings of the 2011 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 131–142. http://epubs.siam.org/doi/abs/10.1137/1.9781611972818.12 DOI: 10.1137/1.9781611972818.12.

[4] Media Bias Fact Check. 2017. The Most Comprehensive Media Bias Resource. (2017). Retrieved October 31, 2017 from https://mediabiasfactcheck.com/

[5] Kristina Chodorow and Michael Dirolf. 2010. *MongoDB: The Definitive Guide* (1st ed.). O'Reilly Media, Inc., Sebastopol, CA.

[6] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational Fact Checking from Knowledge Networks. *PLOS ONE* 10, 6 (June 2015), e0128193. https://doi.org/10.1371/journal.pone.0128193

[7] Jennifer Golbeck, Bijan Parsia, and James Hendler. 2003. Trust Networks on the Semantic Web. In *Cooperative Information Agents VII (Lecture Notes in Computer Science)*. Springer, Berlin, Heidelberg, 238–249. https://doi.org/10.1007/978-3-540-45217-1_18

[8] Lucas Graves and Tom Glaisyer. 2012. *The Fact-Checking Universe in Spring 2012: An Overview.* The New America Foundation, Washington, DC, USA. https://www.issuelab.org/resource/the-fact-checking-universe-in-spring-2012-an-overview.html

[9] R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2004. Propagation of Trust and Distrust. In *Proceedings of the 13th International Conference on World Wide Web (WWW '04)*. ACM, New York, NY, USA, 403–412. https://doi.org/10.1145/988672.988727

[10] Nicholas P Hirning, Andy Chen, and Shreya Shankar. 2017. Detecting and Identifying Bias-Heavy Sentences in News Articles. (2017).

[11] Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, 1373–1378. https://aclweb.org/anthology/D/D15/D15-1162

[12] David Hows, Peter Membrey, Eelco Plugge, and Tim Hawkins. 2013. *Python and MongoDB.* Apress, Berkeley, CA, 139–169. https://doi.org/10.1007/978-1-4302-5822-3_7

[13] 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*. Association for Computational Linguistics, Baltimore, Maryland, USA, 1113–1122.

[14] Jon M. Kleinberg. 1999. Authoritative Sources in a Hyperlinked Environment. *J. ACM* 46, 5 (Sept. 1999), 604–632. https://doi.org/10.1145/324133.324140

[15] Jey Han Lau and Timothy Baldwin. 2016. An empirical evaluation of doc2vec with practical insights into document embedding generation. In *Proceedings of the Workshop on Representation Learning for NLP*, Vol. 1. Association for Computational Linguistics, Berlin, Germany, 78–86.

[16] Quoc Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32 (ICML'14)*. JMLR.org, II–1188–II–1196. http://dl.acm.org/citation.cfm?id=3044805.3045025

[17] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking: Bringing Order to the Web. In *Stanford InfoLab*. Citeseer, Stanford, Palo Alto, CA, 1–17.

[18] 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 (2011), 2825–2830.

[19] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1532–1543. http://www.aclweb.org/anthology/D14-1162

[20] The GDELT Project. [n. d.]. Watching Our World Unfold. ([n. d.]). Retrieved October 31, 2017 from https://www.gdeltproject.org/

[21] Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. 2011. Rumor Has It: Identifying Misinformation in Microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 1589–1599. http://dl.acm.org/citation.cfm?id=2145432.2145602

[22] Adithya Rao and Nemanja Spasojevic. 2016. Actionable and Political Text Classification using Word Embeddings and LSTM. In *Proceedings of the Fifth International Workshop on Issues of Sentiment Discovery and Opinion Mining, WISDOM 2016, San Francisco, CA, USA, August 14, 2016.* ACM.

[23] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A Platform for Tracking Online Misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 745–750. http://dl.acm.org/citation.cfm?id=2890098

Table 6: Words that indicate an article is credible according to a logistic regression model.

word	coefficient value
tlrs	5.32914771988
said	4.39662937235
photo	4.22859888024
image	3.65477754753
july	3.12585411501
women	2.91325593852
support	2.82099051005
podcast	2.77402098585
india	2.73290112857
campaign	2.67759733424
owner	2.64117853155
cent	2.63603955584
picture	2.55722733958
like	2.54639855492
climate	2.54514709161
percent	2.54333486858
care	2.53517993637
did	2.43262750028
images	2.41333556765
politics	2.40167721869
app	2.39699744523
indian	2.37924520364
business	2.3766925319
network	2.35517915282
latest	2.33009229501
getty	2.22811232574
canada	2.21911409169
express	2.11779836847
al	2.11420217184
qatar	2.09988911117
coverage	2.09562171716
comments	2.08320274736
pardon	2.0746413354
political	2.04672331226
work	2.0322289001
editing	1.96870039433
rs	1.96399859445
trade	1.95351708094
putin	1.94553348346
build	1.93696184507
guam	1.90871135341
advertisement	1.89738511546
reuters	1.87704689555
delhi	1.86179895342
leadership	1.85136497359
given	1.84560067318
june	1.83083067076
tweets	1.82907466155
party	1.82851962563
stories	1.81999064765

Table 7: Words that indicate an article is not credible according to a logistic regression model.

word	coefficient value
follow	-26.6313808837
iwb	-18.7858211472
investwatchblog	-18.6102589482
views	-12.4077059541
read	-9.61882848711
antimedia	-9.17618739645
caller	-6.36197119866
caring	-6.02746402192
revolutionizing	-5.44279586416
wikileaks	-5.3874427841
greece	-5.37036137789
christian	-5.26378928804
bible	-5.18255221419
fact	-4.83635067973
arroyo	-4.72986532666
facebook	-4.69708655641
alquds	-4.64891841962
content	-4.58275487795
abortion	-4.4480731547
licensing	-4.40700423334
article	-4.32826287624
time	-4.32189511592
miles	-4.2941344646
beck	-4.22598133714
graviola	-4.22161030684
foundation	-4.20427374104
publisher	-4.13008518306
antifa	-4.1085136531
barton	-4.05063906078
typo	-4.04743287728
created	-4.04199860908
twitter	-4.03992320386
protected	-3.98228806112
sharing	-3.93042439758
dnc	-3.91364202056
ms13	-3.87528857916
debt	-3.8589909956
report	-3.820507377
state	-3.81578985852
moments	-3.79004924637
gospel	-3.78399434845
wolves	-3.77616588737
fbi	-3.76882487816
population	-3.73133511986
girod	-3.72891863502
bohlender	-3.66746513815
herero	-3.65703245664
smirnov	-3.65245101356
jongun	-3.63304263077
freedom	-3.61125797726