Higher Ground? How Groundtruth Labeling Impacts Our Understanding of the Spread of Fake News During the 2016 Election

ABSTRACT

The spread of fake news in online social media platforms has garnered much public attention and apprehension. Consequently, both the industry and academia alike are demonstrating increased effort in building fact-checking websites, classifiers, and platforms to detect and curb fake news.

Yet, sources differ in what they consider to be fake news sites. In this paper, we first aggregate 5 distinct lists of fake news sites, and 3 lists of real news sites published by experts and reputable organizations. Then, using each pair of fake and real lists as an independent "groundtruth", we examine i) attributes that contribute to a site being marked as fake by a source or sources, ii) the pervasiveness of fake news on social media, and iii) the agenda-setting differences between fake and real news sites. We observe that depending on the groundtruth, the prevalence of fake news varies significantly but temporal patterns do not. In addition, we show that the age, popularity, and ideological leaning of fake news sites are uncorrelated with whether it is more likely to be listed as fake news by a given source comparing to another. Finally, the agenda-setting difference between fake and real news sites remain relatively constant across different groundtruth lists.

CCS CONCEPTS

• **Social and professional topics** → **Computing / technology policy**; • **General and reference** → *Cross-computing tools and techniques*; • **Applied computing** → *Law, social and behavioral sciences*;

KEYWORDS

fake news, fake news labeling, 2016 election

ACM Reference Format:

. 1997. Higher Ground? How Groundtruth Labeling Impacts Our Understanding of the Spread of Fake News During the 2016 Election . In *Proceedings of ACM Woodstock conference (WOODSTOCK'97).* ACM, New York, NY, USA, 12 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

In the surprising aftermath of the 2016 election, fake news swiftly become a topic of interest, scrutiny, fear and apprehension for political pundits, media scholars, and the general public alike. Many worry that fake news will threaten the legitimacy of the U.S. government, weaken public trust in both established political institutions and mainstream news media, and even jeopardize the very fabric of American democracy itself [1, 8, 26, 65]. Outside of the States, the fake news endemic is also culpable of contributing to Brexit in Europe [38], the rising hate, violence, and nationalism in Indonesia [39], and endangering the election integrity of nations in Europe and Latin America [5, 19].

Indeed, recent studies centered around the influence of fake news have painted a pretty bleak picture. Fake news, backed by armies of social bots, disseminates significantly faster and deeper than real news [8, 65]. Or, as Mark Twain puts it, "a lie travels around the globe while the truth is putting on its shoes". In addition, subsequent research also indicate that it's difficult for the general public to tell fake news apart from real news, and that repeated exposure to the same untrue information cause readers to perceive it as more accurate [9, 56]. Moreover, the current administration has failed to prioritize combating false information [3].

Fortunately, both the academia and industry are putting in substantial resources to study and understand the characteristics of fake news in the hope of neutralizing its influence [16, 26, 28, 33, 72]. In the industry, leading social media and tech firms including Facebook and Twitter [28, 33] rolled out features to use third party independent fact-checkers to assess the factualness of news articles and to alert users to "disputed information". In the academia, computer and data science researchers, and political science and communications scholars have taken steps to i) understand fake news' political attributes and influence [26, 72] (e.g. partisanship, targetted audience, agenda-setting priority and power, et cetera), ii) build predictors and platforms to preemptively detect fake news before it gains traction [30, 31], iii) study the news consumption behaviors of various population groups and draft conducive curriculums to raise the general public's overall information literacy [16].

While these are all applaudable endeavors, a review of domain-level labels of fake and real news sites used by these studies exposes groundtruth labeling inconsistency. In other words, one study may label or use a dataset that labels breitbart.com as a fake news site whereas another does not. A further examination of these groundtruth labels reveal inconsistencies in both the labeling process and label quality evaluation stage. Thus, for our work, we assess whether and to what extent does difference in groundtruth labels affect downstream analysis.

Our paper makes the following contributions:

- We aggregate 5 lists of fake news sites and 3 lists of real news sites contributed by both the academia and other reputable sources [34, 41, 45, 73, 76, 78, 81]. We provide meta-analysis of these news lists from 6 dimensions: i) number of news outlets in each list, ii) method of annotation, iii) labeling process, iv) labeling granularity, v) quality evaluation, vi) and

WOODSTOCK'97, July 1997, El Paso, Texas USA

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

use by other work in the academia. We observe a significant difference in domains listed. Moreover, a subset of these lists lack clarification on their labeling process and quality evaluation procedure. In addition, lists provided by Zimdars and mediabiasfactcheck.com (see detail in Section 4) are by far most popular.

- Focusing on the list of fake news sites, we examine the extent and potential causes of their dissimilarities. We observe that the popularity and age of fake new sites is not significant in determining whether it is more likely to be listed by a given source. Additionally, all sources share a similar preference in how much they choose to list fake news with a liberal or conservative-leaning. Moreover, we observe that domains labeled by Zimdars as *fake* are more likely to be listed in other sources comparing to sites labeled as *rumor* or *junksci*.
- Based on the groundtruth labels used, the prevalence of fake news, calculated as the fraction of news articles in our dataset that are from fake news sites, vary considerably: from as high as over 30% to as low as less than 5%. This observation is likely due to the inconsistency in labeling of high-traffic domains; it suggests that the perceived prevalence of fake news may well be a by-product of groundtruth labels.
- Additionally, unlike prevalence, we also observe that the time-series of fake news generated using different groundtruth have a high correlation with each other, suggesting that temporal patterns of fake news are robust against groundtruth labels.
- Finally, using topic modeling to study the agenda-setting priority difference between fake and real news sites, we observe that regardless of the groundtruth used, the trend we would observe stays comparatively constant. That is, the assessment as to whether a topic was more central to the coverage of fake news outlets (compared to the traditional news sites) is robust to the choice of groundtruth.

To summarize, the use of different lists of fake news have varied impact on the subsequent analysis. While selecting one fake news source over another may significantly alter the perceived prevalence of fake news, temporal patterns and agenda-setting priorities remain constant. For future work, we urge researchers to consider the difference in groundtruths while conducting studies in the area of fake news.

2 RELATED WORK

We review three areas of research most relevant to our work. First, we discuss existing theories and frameworks that define and operationalize the term "fake news". Second, we summarize prior work that focuses on examining the difference in characteristic of online fake news in comparison to real news, or studying the impact of fake news on the political process and society at large. Finally, we inspect existing platforms and automated tools that classify, predict, track, or curb the spread of fake news.

A significant amount of work has been done to define and categorize fake news [67, 71, 81]. Tondoc et al., for instance, identifies 2 primary dimensions of fake news: levels of facticity and deception. The former refers to the degree to which fake news relies on actual facts. Political satires and parodies, for instance, are often a humorous take on real-life events. Therefore, they rely more on real news than, say, junk science about space aliens. The latter dimension, deception, focuses on the intention of the fake news provider. A deliberate effort to spread information known to be false is different from an honest mistake. Ward [77], on the other hand, conceptualized fake news using 3 distinct dimensions: type of content, motivation, and dissemination method. In terms of categorization, Zimdars [81] divide fake news into multiple subtypes whereas other researchers including Allcott and Gentzkow [6] provide a more straight-forward, a single-category definition of fake news: "news articles that are intentionally and verifiably false, and could mislead readers".

Focusing on examining the difference between fake and real news, prior research identify various traits unique to or are more prominent in fake news articles in comparison to real news articles: i) certain linguistic patterns[58] (such as punctuations, part-of-speech, word choices), ii) auxiliary data associated with the articles [67, 68] (such as the source of the article, date, and place of publishing), iii) network cascading patterns [7, 8, 65, 75] (e.g. depth, breadth, speed et cetera).

Research have also extensively documented the negative impact fake news has on individuals and society at large including: health-care, stock market, disaster management [44, 54, 60], both within the United State [32, 69] and in the international stage [5, 32]. Furthermore, studies on cognitive information process suggest that people in general are bad at separating false information apart from the truth [2], and are susceptible to the influence of fake news [9].

Finally, focusing on available datasets and large-scale automated approaches, we note numerous lists of real and fake news sites created by various sources [34, 41, 45, 73, 76, 78, 81], all of which will be discussed in detail in Section 4. Other datasets worth noting here are i) the set of popular fake news articles aggregated and published by buzzfeed.com on Facebook during 2016 US elections [69], ii) LIAR[76] which includes 12.8K annotated short statements sampled from politifact.com, and iii) CREDBANK [49], a large scale crowd-sourced dataset of approximately 60 million tweets with annotated credibility scores. Additionally, various fake news classifiers and platforms exist [67, 68] including: i) Hoaxy,[64] a platform for the visualization of the spread of both factual and doubtable claims, ii) NELA-toolkit [31], a tool used to check whether an online news article is fake or real, iii) BS Detector[82] which searches all the links on a given webpage for reference to unreliable sources.

Unlike these related valuable work, our paper focuses on aggregating various lists of fake and real news sites (i.e. diverging groundtruth labels) and conduct cross-groundtruth analysis from multiple different dimensions. We assess whether and to what extent difference in groundtruth labels affect downstream studies.

3 DATA

Our paper contains 3 types of data: (i) lists of fake and real news sites, (ii) the set tweets shared about the two presidential candidates during the 2016 election campaign season, (iii) the web pages corresponding to the URLs shared in those tweets.

Fake and Real News Site Lists: We collect 5 distinct fake news lists and 3 real news list from both the academia and the Press [6, 26, 45, 57, 64, 81], resulting in 1593 total number of aggregated fake news sites, and 8562 real news sites. We describe the details of these list in Section 4.

Twitter Data: Between May 23, 2014, and January 1, 2017, on a daily basis, we collect i) 5,000 tweets that were uniformly randomly sampled from all tweets on that day that included the keyword Trump, ii) 5,000 random tweets that are similarly sampled from tweets that mention Clinton. The data collection was performed using Sysomos MAP - a social media search engine that includes access to all tweets (Twitter firehose) going back one year as well as other social media sites and online news. The resulting dataset includes approximately 4.8 million tweets about Donald Trump and similarly, approximately 4.8 million tweets about Hillary Clinton.

Web pages: Next, we identify unique shortened URLs shared in these tweets and built automated scripts to scrape their content. We remove URLs internal to Twitter since such URLs correspond to quoted retweets as opposed to external news content. The resulting dataset includes approximately 1.7 million unique URLs. For each share (i.e. tweet including a valid URL), we record the following data: i) the shortened URL, ii) the original URL, iii) domain name (e.g. nytimes.com), iv) title of the document, v) text body of the document, (vi) the date of the tweet, vii) Twitter id of the user sharing the URL, and viii) a binary categorization that indicates whether this tweet is in our Clinton or Trump sample.

Data Preprocessing: We remove all web pages that belong to other popular social media domains [1] and filter out the subset that are from tweets posted earlier than 12/01/2015 or later than 01/01/2017. For the remaining web pages, we use `BeautifulSoup` to extract all texts embedded in $< p >$ tags [2]. Next, we drop duplicated articles matching for original URL, title and body content. We then apply `gensim`'s [3] preprocessing pipeline to remove stop words and hyperlinks in addition to lemmatizing and stemming non-stop words. We remove all articles that have fewer than 100 or more than 800 tokens [4]. In sum, we derive approximately 174K unique articles after preprocessing shared approximately 711K times (measured using the number of tweets) on Twitter.

4 META-REVIEW

Prior to assessing how difference in groundtruth labels affect downstream analysis of fake news, we first begin by analyzing the publicly available lists of fake and real news domains (focusing on how these lists were curated, its applications and how often these are referenced by other companies and researchers) follow by examining consistencies and inconsistencies amongst the different lists.

4.1 Available Sources and Applications

4.1.1 *Lists of Fake News Sites:*

We aggregated 5 distinct fake news lists: i) Zimdars (`ZDR`), ii) mediabiasfactcheck.com (`MBFC`), iii) politifact.com (`POLTI`), iv) dailydot.com (`DDOT`), and v) Allcott and Gentzkow (`GZKOW`).

[1]Here, we remove web pages from facebook.com,youtube.com,linkis.com,reddit.com, vine.com, etc.

[2]If the last paragraph contains the keywords "contact us", "email us", or "follow us", it's excluded.

[3]`gensim` is a popular Python library used for natural language processing [61]

[4]Web pages with less than 100 words after preprocessing account for 26.9% of all web pages; web pages with over 800 words account for 1.8% of total.

Zimdars' list (ZDR): To elaborate, Zimdars' list `ZDR` is a set of fake news websites compiled by Melissa Zimdars, professor of communication at Merrimack College [81]. The domains listed in Zimdars list have following tags fake, satire, bias, conspiracy, rumor, state, junksci, hate, clickbait, unreliable, political, reliable and unidentified. The list was manually annotated by Prof.Zimdars and her collaborators using their standard [82] [5]. This list is widely popular both in the academia and the Press.

Mediabiasfactcheck (MBFC): mediabiasfactcheck.com is an independent online media outlet maintained by a small team of researchers and journalists [45]. They adhere to International Fact-Checking Network Fact-checkers Code of Principles for tagging different websites [35]. In addition to archiving all fake news sites, `MBFC` also maintains real news sites and provides an independent assessment of the ideological leaning of each website documented. The website also runs a voting poll to provide media consumers with two perspectives and to analyze how much their analysis of the website matches the opinion of the public.

Politiact (POLTI): politifact.com is a nonpartisan fact-checking website owned by Poynter Institute, a nonprofit journalism school [57]. In collaboration with Facebook, its staff generated the top fake news sites most-shared on people's Facebook news feeds during the 2016 election.

DailyDot (DDOT): dailydot.com is one of the mainstream online news sites that focus on entertainment and social issues [78]. The Dailydot list comprises of fake domains whose posts are often shared across social media. It was largely created by referencing other pre-existing fake news lists like Zimdars list.

Allcott and Gentzkow (GZKOW): Finally, we have Allcott and Gentzkow [21] who accumulated a list of fake news sites from both in and outside of the academia including Grinberg [24], Silverman [69] from buzzfeed.com, Schaedel [63] from factcheck.org, a subset of websites from Zimdar's list and snopes list [34].

4.1.2 *Lists of Real News Sites:*

Three real news lists were aggregated from the following: i) Alexa (`ALEXA`), ii) mediabiasfactcheck.com `MBFC`, and iii) Guo and Vargo (`VARGO`). Details on `MBFC` is provided in the previous section. Here, we describe `ALEXA` and `VARGO`.

Alexa (ALEXA): (alexa.com) is an online domain directory owned by Amazon, we crawl for all the websites listed under Alexa's *News* category when the data was still available to the public for free.

Guo and Vargo (VARGO): `VARGO` is from Guo and Vargo[26] who assembled a list of real news sources and used existing fake news sources mentioned in the previous section to study the agenda-setting power of fake news and fact-checkers. The list contains fact-based news websites which were compiled through a manual content analysis of the top 2,760 U.S. news media websites found in GDELT 's global knowledge graph. The coders who did the manual annotations reached an intercoder reliability of 0.988 Krippendorff for general media categorization and 1 for partisanship.

[5]Zimdars' list also contain domains that are categorized as reliable, political, and unknown. These 3 categories are all removed for subsequent analyses.

Table 1: List of Real and Fake News Sources and Their Use in Academia

List	No. of Outlets	Annotation	Labeling Process	Source	Domain Subtypes
DDOT(F)	175	Zimdars, research team	as followed by referenceed lists	subset of zimdars and other lists (not mentioned)	fake only
GZKOW(F)	673	Politifact, snopes, buzzfeed	as followed by referenceed lists	3 sources (politifact list, snopes.com and buzzfeed list)	fake only
MBFC(F)	804	Manual	International Fact-Checking Network Fact-checkers Code of Principles.	collection of media outlets (not mentioned)	conspiracy, pseudoscience, questionable sources
POLTI(F)	325	Manual	Manual	politifact.com	fake, imposter, some fake, parody
ZDR(F)	779	Manual	Open Sources steps for analyzing websites	Prof. Zimdars and OpenSource.co	fake, satire, bias, conspiracy, rumor, state, junksci, hate, clickbait, unreliable,political,reliable, unidentified
MBFC(R)	1685	see MBFC(F)	see MBFC(F)	see MBFC(F)	real(left bias, left-center-bias, least biased, right-center bias, right bias, pro-science)
ALEXA(R)	5497	Amazon	not mentioned	Amazon Alexa Website Directory	real only
VARGO(R)	2649	Manual	not mentioned	GDELT Global Knowledge graph	real only

List	Quality.Evaluation	Use in Academia
DDOT(F)	depends on the reference list	build automated fake news tracking platforms [29, 64], assess agenda setting power of fake and real news sites [26, 50, 72]
GZKOW(F)	depends on the reference list	assess impact on election, examine network cascading behaviour of fake news [6, 8]
MBFC(F)	not mentioned	analyze the rise of the Alt-right in the States [43], Globalism [70], and the virality of fake and real news [13], information literacy [16], rising polarization [12], information quality [52]
POLTI(F)	not mentioned	study the diffusion of fake news on social media [21], information literacy [51], automate fake news detection [23]
ZDR(F)	Title Analysis, Source Analysis, Writing Style Analysis, Aesthetic Analysis, Social Media Analysis	examine network cascading behaviour difference between fake and real news articles during the 2016 election [6, 8], build automated fake news classification and tracking platforms [30, 31, 64], assess agenda setting power of fake and real news sites [26, 72], impact assessment [14, 18, 62], ethics and policy [17, 37]
MBFC(R)	not mentioned	see MBFC(F)
ALEXA(R)	no information	examine network cascading behaviour difference between fake and real news articles[6, 8], news sharing behaviour in right-leaning echo chambers [42]
VARGO(R)	intercoder reliability of 0.988 Krippendorffs alpha (α) for general media categorization and 1 for partisanship.	assess agenda-setting power of fake and real news sites [26, 72]

Labeling Process and Quality Assessment: As shown in Table 1, the vast majority of sources are manually selected and annotated. In addition, there seems to be a lack of description of how datasets were annotated or how the quality of the labels were evaluated. In fact, out of all sources listed, only 2 have meaningful quality evaluations: `ZDR` and `MBFC`. To elaborate `ZDR`, performs their analysis on the following six fronts - 'Domain Analysis' (if the domain name is a variant of a major news agency, its highly likely to be fake), 'About Us Analysis' (Has anyone ever listed the domain as fake before and why?), 'Source Analysis' (are there sufficient references to support their claim), 'Writing style analysis'(Lack of proper style guide is usually indicative of lack of editorial procedure), 'Aesthetic analysis' (fake websites usually tend to have bad designs), and 'Social media Analysis'(are the social media pages of the website linked to other unreliable sources?). And `MBFC` scores the website in four different areas - Biased Wording(are the words in the articles written to convey any emotion to the readers?), Factual/Sourcing (are there sufficient references to back up the claim mentioned in the article), Story Choices(does the source report news from different perspectives?) and Political Affiliation(does the source endorse a particular political ideology?). Even though the methodology of annotations is mentioned, the score on individual categories is still abstract.

Use of Fake and Real News Lists in Research: Use of each fake and real news sources are summarized in Table 1 [6]. As shown, for

[6]Allcott and Gentzkow's list can be further broken down into sublists they gathered GZKOW-BF(F) 223, GZKOW-FC(F) 61, GZKOW-GB(F) 92, GZKOW-GNR(F) 382, GZKOW-GO(F) 47, and GZKOW-GR(F) 61.

instance, more than 10 distinct articles/studies cited Zimdars' list for various purposes including: building automated fake news classification and prediction platforms [30, 31, 64], assessing fake news' impact on the 2016 election [6, 14], determining fake news sites' agenda-setting priorities and power [26], examining network cascading difference between real and fake news articles [8], discussions on information literacy and policy [37], general calls for action, et cetera. It's ostensibly the most widely used list in the academia for the study of fake news. We also note that, in addition to labeling news sites as real or fake, mediabiasfactcheck includes the ideological leaning of websites (extremely liberal, liberal, center, right, extremely right). Many research [43, 70] leverage this ideological rating to study the relationship between news quality and partisanship.

Here, we highlight the fact that these lists are used by a vast array of studies, therefore, it's important to understand how groundtruth selection (i.e. choose which fake and/or real list to use) may impact downstream analysis.

4.2 Dissimilarities of Fake News Sources and What's Causing Them

Given the wide range of applications of these sources (particularly the fake news sources), in this section, we first examine how similar or dissimilar they are with each other, we then explore and highlight domain-level attributes that are predictive of whether a specific fake news site is likely to be listed by a given source.

List Similarity: We use 2 metrics to determine the similarity between different sources of fake news sites. We first calculate the fraction of hyperlinks being present in at least 2 of the 5 distinct sources, then 3, then 4, and so on. We observe that close to 50% of total domains are only listed in a single list. In fact, only 5.7% of all domains listed by at least 1 fake source are listed by all fake sources. Furthermore, we use Jaccard index [22] to determine the similarity between different lists. We write F as the set of fake news lists, we then calculate Jaccard score using $|u \cap v|/|u \cup v|$ where $u \in F, v \in F$ and $u \neq v$. We observe a similar trend. More than half of the pairs have a Jaccard similarity of $<= 0.1$ and approximately 75% of all pairs have a similarity score of $<= 0.2$. Both metrics suggest that domains listed by different sources are quite different from each other. We note that `MBFC` and `DDOT` have the lowest Jaccard similarity score of 0.08; `GTKOW` and `POLTI` have the highest score of 0.48.

Domain Characteristics: Next, we explore characteristics that may be significant in predicting how likely a fake news domain is to be present in a given source. Referring back to Table 1, we note that lists such as `DDOT` and `GZKOW` referenced to other lists such as `ZDR` for fake news sites. It's somewhat surprising then, that we do not see a higher similarity between the lists. One possibility is that `DDOT` and `GZKOW` used an earlier version of `ZDR` that does not include the younger and newly accumulated fake news sites. Another possibility is that mainstream news sites such as Dailydot and Politifact focus only on the most prominent fake news sites and ignore the lesser well-known ones. Moreover, we also see that `ZDR` further categorizes fake news sites into subtypes whereas `DDOT` and `GZKOW` have a single `fake` category. Thus, it's likely that certain subtypes of fake news in `ZDR` are more likely to be included in other sources (e.g. Horte et al. [31] purposefully excluded satirical news sites when training a certain version of their fake news classifier). Finally, a number of conservative elites has long criticized both the Press and academia as having a liberal bias [25]; thus, here we also examine whether liberal fake news sites are less likely to be present in sources.

In sum, for a given fake news site i that's listed by at least one of the 5 fake news lists, we identify 4 attributes of i that are of interest: i) ideological leanings (liberal, conservative, center, or unknown) [7] obtained from `MBFC`, ii) domain subtype (biased, fake, unreliable, hate, clickbait, conspiracy, hate, junksci, rumour, satire, and unknown) obtained from `ZDR`, iii) website age (approximated using the timestamp of the earliest tweet that contained a link to i), and iv) popularity (approximately using the total number of tweet shares within our dataset). We hypothesize that older domains with higher popularity are more likely to be present in sources. In addition, domains listed by `ZDR` as *fake* and *unreliable* should also be more likely to be present than domains listed as *hate* or *junksci*. This is likely considering that while fake information websites in the *junksci* may spread the untruth about certain scientific topics (e.g. vaccine), research centered on political misinformation and disinformation (e.g. Allcott and Gentzkow) is unlikely to include these in their list.

Let the binary variable $y_{i,s} = \{0, 1\}$ denote whether domain i exists in the list of fake news sites provided by source s, and where $S = \{ZDR, MBFC, GZKOW, DDOT, POLTI\}$. We fit the following logistic regression model:

$$y_s = \beta_0 + \beta_1 ideology + \beta_2 subtype + \beta_3 popularity + \beta_4 age + \varepsilon_i$$

Logistic regression results [8] are summarized in Table 2. Surprisingly, the popularity and age of a fake news site is not indicative of whether it's more or less likely to be listed by various sources except for Zimdars' list (`ZDR`), which indicates that older and more popular fake news sites are more likely to be listed (consistent with our prior belief). Ideological leaning is not statistically significant across all models, suggesting that amongst the sources, no any one source is preferentially choosing fake news sources that are labeled as conservative or liberal leaning. Note, however, it's still possible that all sources have a preference for liberal or conservative fake news sites. In addition, we observe that domains categorized as *fake* by Zimdars are uniformly more likely to be present in all the others sources. Finally, we note that the R^2 values for all models except for `DDOT` (which has the smallest number of domains) are low, suggesting that our models did not capture a significant portion of the variance.

5 GROUNDTRUTH IMPACT

We explore the consequences of using different groundtruth labels from 3 distinct dimensions: i) impact on the perceived prevalence of fake news, ii) the similarities and dissimilarities of fake news time-series generated using different groundtruth pairs, and finally

[7] Here we collapse `MBFC`'s *extreme left* and *left* categories into single *liberal*class. Same for *conservative*.

[8] Here, we domain subtype for `ZDR` when fitting logistic regression because given labels come from this particular list, it's guaranteed that a domain with labeling is definitely present.

Table 2: Determing Factors For Inclusion in Sources.

	Dependent variable:				
	(ZDR)	(MBFC)	(GZKOW)	(DDOT)	(POLTI)
conservative	14.094	15.696	−13.878	15.216	−15.626
	(535.411)	(535.411)	(535.411)	(10,754.010)	(535.411)
liberal	14.661	14.341	−14.465	14.446	−16.216
	(535.411)	(535.411)	(535.411)	(10,754.010)	(535.411)
ideology_unknown	14.704	13.884	−14.541	15.030	−16.278
	(535.411)	(535.411)	(535.411)	(10,754.010)	(535.411)
clickbait		−0.989**	0.522	−16.038	−0.201
		(0.495)	(0.441)	(1,772.006)	(1.125)
conspiracy		1.225***	0.179	−0.394	1.094*
		(0.325)	(0.370)	(0.867)	(0.641)
fake		0.758***	1.599***	3.871***	2.975***
		(0.272)	(0.296)	(0.516)	(0.506)
hate		0.492	−0.175	−16.526	0.628
		(0.635)	(0.692)	(2,851.494)	(1.139)
junksci		1.840***	0.592	−16.597	0.706
		(0.565)	(0.593)	(2,684.427)	(1.145)
rumor		0.592	0.662	−16.672	1.225
		(0.823)	(0.773)	(3,796.827)	(1.175)
satire		1.672***	1.235***	−16.542	2.543***
		(0.396)	(0.383)	(1,582.943)	(0.589)
subtype_unknown		1.909***	0.533**	−1.364**	0.916*
		(0.267)	(0.251)	(0.688)	(0.501)
unreliable		−0.032	0.036	0.908	0.684
		(0.380)	(0.430)	(0.706)	(0.765)
popularity	0.0001*	−0.00000	0.00000	−0.0002	−0.00000
	(0.00003)	(0.00001)	(0.00001)	(0.0001)	(0.00002)
age	−62.164***	−4.340	23.539	27.125	38.076
	(16.112)	(17.016)	(16.941)	(35.390)	(23.343)
Constant	555.606	39.239	−202.497	−266.505	−335.989
	(555.365)	(155.867)	(557.420)	(10,758.890)	(576.478)
Observations	763	763	763	763	763
R^2	0.04	0.082	0.067	0.55	0.18
Note:					*p<0.1; **p<0.05; ***p<0.01

iii) observational difference in agenda-setting priorities of fake and real news sites.

5.1 Prevalence

There has been significant interest and diverging opinions in the prevalence of fake news in both academia and the industry. For instance, prior work by Silverman [1] shows that selective fake news articles "garnered hundreds of thousands of shares – sometimes more than twice as many as legitimate news scoops about the candidates in major outlets like the New York Times", and work by Howard et al indicates that share of fake news actually outperformed real news the day right before the 2016 election in Michigan [32]. Whereas study by Allcott and Gentkow [6], using their aggregated list, suggests that average adult saw and remembered only 1.14 fake news article during the course of the election.

Understanding the "real" pervasiveness of fake news is of vital importance given that recent work by Pennycook [56] suggests that repeated exposure to the same untrue information cause readers to perceive it as more accurate. Thus, the prevalence of fake news likely has a direct impact on the general public's interpretation of the true state of the world. Here, we assess to what extent groundtruth difference impacts perceived pervasiveness of fake news from 2 distinct view points: i) given a pair of lists (one for real news sites, and the other fake), what fraction of total news articles are fake news, and ii) aggregate all the fake news lists together into a single master list, what fraction of total fake news can an individual fake news source capture (i.e. recall).

Figure 1: Prevalence of Fake News Based on Source Pairs

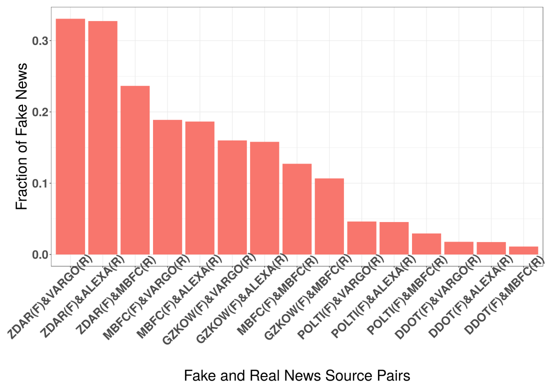
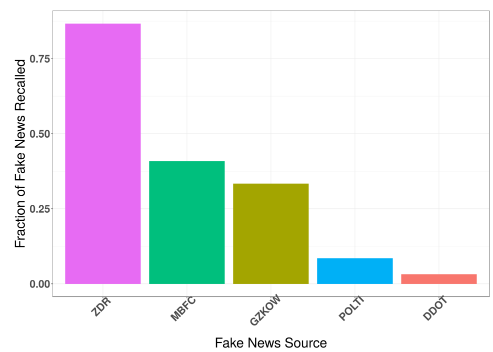


Figure 2: Recall of Fake News Based on Fake News Source



Fake News Prevalence as Fraction of Shared Articles: In this section, we calculate the prevalence of fake news as the fraction of shared news articles that are from fake news sites. As shown in Figure 1 (the x-axis denotes different groundtruth pairs and the y-axis indicates the fraction of shared fake news articles), a simple check of fake news prevalence using different groundtruth pairs reveals that, based on the groundtruth, fake news could amount to be more than 30% of total news shares (using `ZDAR` and `VARGO` as the fake and real news source pair) on Twitter or as low as less than 5% (using `DDOT`, and `MBFC`). In addition, referring back to Allcott and Gentkow's work, we see that using their fake news list, the total fraction of fake news ranges from approximately 10% to more than 15%.

We also measure fake news prevalence using the fraction of unique Twitter accounts that have shared at least 1 fake news article. We observe a similar pattern. This suggests that the "pervasiveness" of fake news may well be significantly dependent on which groundtruth labels are used.

Measuring Fake Prevalence Using Recall: First, we aggregate all domains from the 5 lists of fake news sites into a single master list (i.e. a domain is included and marked as fake if it is mentioned by at least 1 of the fake news sources). We then determine all the tweets in our dataset that contain an url to one of the domains in the master list. For the subset of domains listed in an individual fake news source, we calculate the fraction of the tweets this particular subset of domains is able to recall.

As shown on Figure 2, we see a considerable difference between the sources: `ZDR` is able to recall 87% of total fake news from the master list whereas `MBFC` recalls 41%, and `POLTI` is only able to recall 8.9%. This drastic difference is somewhat surprising given that the number of unique domains in `ZDR` is 779, whereas the number is 804 for `MBFC`, and 325 for `POLTI`. In other words, despite of having a smaller list, `ZDR` recalls a much higher fraction of total fake news shares. A closer examination of the most popular domains reveals a possible explanation: `ZDR` listed some well known domains as *unreliable*, or as *propaganda* while `MBFC` labels them as *highly reliable*, or *mixed* [9].

5.2 Temporal Series Correlation Analysis

Many prior works in fake news are invested in understanding its temporal trends (e.g. assessing whether the pervasiveness of fake news as a whole is on the rise [79]), some even use the patterns observed to do other things (e.g. training models to detect fake news [4, 20, 66]). In this section, we examine whether the time-series extracted of fake news vary based on different groundtruth pairs. We use 2 methods to determine the similarities or dissimilarities of these time-series: Pearson's Correlation [40] and principal component analysis (PCA) [80].

First, let F and R represent the sets of fake and real news sources as shown on Table 1. For each unique pair combinations of (f,r) where $f \in F$ and $r \in R$, we calculate the fraction of news articles shared each day that are from fake news sites in the time period between 2015-12-01 to 2017-01-01 and denote the time series as $P(f,r) = \{p^1_{f,r}, p^2_{f,r}, p^3_{f,r} ...\}$ where $p^t_{f,r}$ is the fraction of fake news URL shared t days after 2015-12-01. In addition, we apply 7-day rolling mean to smooth $P_{F,R}$.

We calculate Pearson's Correlation for all groundtruth pairs $P(f_i, r_i)$ and $P(f_j, r_j)$. As shown in Figure 3, unlike the significant difference we see earlier for prevalence, Pearson's Correlation between time-series generated using different groundtruth is relatively high for the majority of pairs (the median coefficient as depicted is 0.8).

Next, we apply PCA to $P_{F,R}$ and extract the first 2 principal components. We observe that the first component explains 98.8% of the total variance, then the second component 0.8%. Resulting clusters from PCA are plotted in Figure 4. Data points are colored according to their fake news source (e.g. red dots are from Dailydot `DDOT`) and shape represents the size of the list–if a list contains

Figure 3: Time-series Correlation between Source Pairs

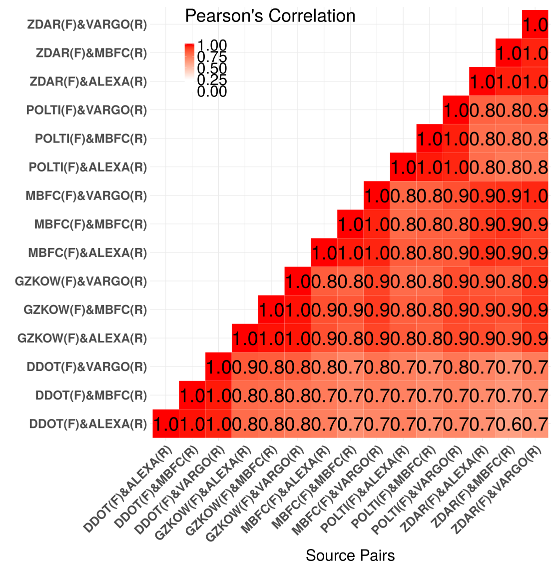
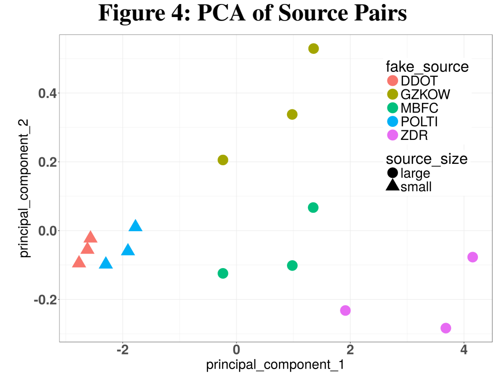


Figure 4: PCA of Source Pairs



less than 300 domains, it's labeled as *small*, otherwise *large*. For instance, Dailydot contains less than 300 domains in its fake news list, it is therefore categorized as small. Focusing on the x-axis (i.e. the first principal component), we see that data points are spread out based on both size and fake news source. We also examine the relationship between PCA clusters and the real news source, and do not see a clear association (plot omitted). These observations imply that difference in fake news source is the primary driver behind variance in the time-series. Or put it in another way, it doesn't matter much which list of real news sites we choose, it's which list of fake news sites we use that determines the time-series.

5.3 Agenda-setting Priorities

Agenda-setting theory postulates that the most frequently covered social and political issues by news media are what the general public consider the most important. Or, in the words of Cohen, the Press "may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to

[9] The domains we observe include washingtonexaminer.com, politicususa.com, dailycaller.com, et cetera

think about" [47]. Within the context of fake news and its relation to the 2016 election, fake news may lead voters to re-evaluate issue priority and candidate viability. Fortunately, researchers including Guo and Vargo are already undertaking studies to understand the agenda-priority difference between fake and real news sites [26, 72].

In this section, we focus on examining whether difference in groundtruth labels lead to diverging observations in agenda-setting. We first use an iterative topic modeling process to extract issues, or topics, being reported on by both fake and real news. We then assign all news articles within our dataset to its corresponding issue category, or topic. Finally, we examine whether difference in groundtruth labels leads to inconsistent conclusions (e.g. one pair of groundtruth leads to the interpretation that fake news sites report more on terrorism than real news sites or vice versa, whereas another pair of groundtruth does not).

5.3.1 Topic Modeling of News Articles Using Vanilla and Guided LDA.

The latent Dirichlet allocation (LDA) model is a generative model used extensively by researchers within the social computing field to "cluster" documents into human readable categories for the study of a wide range of social and political topics [15, 55, 59]. Guided LDA is an extension of the original LDA that allow sets of keywords to guide document topic assignment by increasing their "confidence" or weights. Each set of keywords is used to guide documents into a unique topic. For details of these models, we refer readers to the original papers [36, 46].

We first leverage the vanilla LDA to extract important seed words. To do so, we use `gensim` to generate 5 different models with the number of topics equal to $\{50, 75, 100, 125, 150\}$ respectively. In addition, we set all models to ignore words that have a frequency of less than 100 or occur in more than 50% of total documents, and to generate bigrams for words that co-occurred at least 100 times or more. We then select the model with the optimal coherence score [10]. For the optimal model, we obtain the top 30 most representative words for each topic. We then inspect words and categorize them into coherent sets (i.e. topics). Using this approach, we obtain 409 unique seed words divided into 33 different sets. Finally, we run the guided LDA implemented by Singh [74] using the derived seed word sets, adjusting model's seed confidence to 0.25 and setting the number of total topics to 125 [11].

We assign each document into a single topic according to the maximum probability of its topic distribution (this topic is also later referred to as the document's *actual* topic label).

5.3.2 Topic Modeling Quality Assessment and Selection.

Topic Pre-selection: Iterating over every single topic and using the most weighted keywords from each as cues, we first filter out the subset of topics that lacked coherent themes. We then collapse topics that share the same human-interpretable theme into a single topic. This process results in 19 distinct topics. Then, for each topic, we randomly sample 0.2% of its total documents (or 10 if the size of a topic is small) and obtain 434 documents. Additionally, we also sample 0.2% documents from the pool not included in the 19 topics. This gives us a total of 525 documents. We shuffle and publish these documents on Amazon Mechanical Turk for crowdsourced labeling.

Table 3: List of Topics, Fraction of Total Documents Accounted for, Most Weighted Keywords, and F1 Scores

topic	doc frac	most weighted tokens	f1
abortion	0.96%	woman abort life plan_parenthood issu punish femal	0.87
benghazi	0.60%	attack benghazi libya committe report secretari secur	0.75
c-health	0.86%	medic doctor releas report mental suffer letter pneumonia	0.75
climate	1.40%	climat coal environment industri land administr regul power	0.89
clinton-wst	0.30%	speech wall_street talk ask question issu transcript releas	0.82
d&i	0.75%	commun lgbt issu group equal woman discrimin anti marriag	0.78
economy	4.40%	trade job china deal compani manufactur econom	0.79
election	20.27%	sander berni primari voter percent poll voter cruz	0.77
email	5.76%	email depart investig inform server classifi comey secretari	0.84
immigration	2.28%	immigr border mexico wall illeg deport mexican build latino	0.85
mid-east	3.86%	muslim islam israel isi terror terrorist attack unit syria obama	0.76
religion	1.14%	christian evangel church faith religi leader pastor religion pope	0.78
russia	1.81%	russia russian putin intellig hack offici govern vladimir_putin	0.76
security	1.70%	iran china nuclear polici foreign deal unit world nato secur	0.78
sexual	1.93%	woman accus alleg rape husband sexual claim sexual_assault	0.82

Independent Quality Assessment using MTurk: First and foremost, the success of a crowdsourcing task relies heavily on the right mechanisms to ensure worker qualifications [10]. We require that workers: 1) reside in the U.S. 2) have successfully completed at least 1,000 HITs; and 3) have an approval rate of at least 98%.

For each document, we assign 3 independent workers to categorize it. Workers are given a list of prior categories (19 aforementioned topics + 1 *none of the above* option) to choose from and are instructed to identify and select a single *primary* category of a given article. They are also instructed to choose *none of the above* option if the primary topic of an article is unclear. We use Krippendorff's alpha [27] to measure intercoder reliability given that it's robust against missing data (many of our crowdsourced workers only label a subset of all documents). We observe an inter-coder reliability of 0.62, which is considered to have a moderate agreement.

For each article, we mark its *perceived* topic according to the majority vote (e.g. 2 out of 3 workers selected the same primary category). Articles that do not have a majority is labeled as *unknown*. We observe 46, or 8.6% *unknown* documents. Next, for each topic, we calculate its precision, recall, and f1 scores using the *perceived* and *actual* topic labels. We filter out the topics that have an f1 score of < 0.75. This process produces 15 distinct topics which are all listed in Table 3. As a whole, the 15 selected topics accounts for 49% of total news articles and spans across both substantive policy issues (e.g. economy, immigration, national security et cetera) and political scandals with less immediate policy implications (e.g. sexual scandals from both Democrats and Republicans, Clinton's paid Wall St. speech controversy, et cetera).

5.3.3 Agenda-setting Priorities as Measured Using Different Groundtruth Pairs.

As pointed out previously, understanding the agenda-setting differences between traditional and fake news producers is an important open research question [26, 47, 72]. Here, we provide an in-depth analysis of whether and to what extent groundtruth list choices affect the perceived difference in agenda-setting.

We first describe the overall distribution of topics in all news articles. Then, we present the subset of topics that fakes news sites

[10]Coherence score for a topic is defined as the average of the pairwise word-similarity scores of the words in the topic (e.g. PMI) [53]. A model's coherence score is calculated as the aggregation of its topic coherence scores

[11]Here, we use perplexity score [48], defined as log-likelihood per word, to determine the optimal number of topics given that `gensim` does not support coherence calculation for guided LDA.

have reported significantly more or less on than real news sites and discuss whether difference in groundtruth leads to inconsistent observations. Finally, we focus on examining topics that contribute most to the agenda-setting difference between fake and real news sites.

Topic Overview: Table 3 provides a summary on the list of topics (ordered alphabetically). For each topic, we provide its name, its total contribution, most weighted keywords, and f1 score. As depicted, the most prevalent topic, `election`, includes words such as *polling, voting, rallies*, and *debates*. This topic accounts for 20.3% of total news articles. It is followed by Clinton's email scandal, `email`, which amounts to 5.76% of total documents, the `economy` at 4.4%, and the Middle East (`mid-east`) at 3.86%. In addition, we also identify multiple small but interesting topics such as candidates' health, `c-health`, that correspond to Clinton's pneumonia case and Trump's physician's health assessment.

Topic Distribution as a Function of Groundtruth Pairs: For each groundtruth pair (f, r) and topic i from the set of 15 topics $\mathscr{S}$, we derive distributions $\{k^1_{f,r}, k^2_{f,r}, k^3_{f,r} ... k^{16}_{f,r}\}$ and $\{l^1_{f,r}, l^2_{f,r}, l^3_{f,r} ... l^{16}_{f,r}\}$ where $k^i_{f,r}$ is the fraction of fake news articles that are on topic i, $l^t_{f,r}$ is the fraction of real news articles that are on topic i, and $\sum_{i=1}^{16} k^i_{f,r} = 1, \sum_{i=1}^{16} l^i_{f,r} = 1$. Note that topic 16, `OTHER`, contains all documents not in any of the 15 selected topics.

Topic	Principal Comp1	Principal Comp2
abortion	0.032	**0.117**
benghazi	-0.08	-0.08
c-health	-0.01	-0.08
climate	-0.04	**0.15**
clinton-wst	0.02	0.04
economy	0.04	**0.53**
election	**0.88**	**-0.26**
email	**-0.25**	**-0.74**
immigration	-0.02	**0.12**
mid-east	**-0.38**	-0.04
religion	0.01	**0.17**
sexual	-0.04	0.04

Table 4: Factor Loading Analysis of Topics Most Correlated with Principal Components

For each topic i and all groundtruth pairs (F, R), we apply Student's T-test on $K^i_{F,R}$ and $L^i_{F,R}$ to determine whether the difference in the means of $K^i_{F,R}$ and $L^i_{F,R}$ are statistically significant. In other words, we assess whether fake news sites publish relatively (measured using normalized fractions) more or fewer articles on certain topics than real news sites and vice versa. We observe a significance for 12 topics. For the 12 topics, we plot the corresponding distributions for $K^{\mathscr{S}}_{F,R}$ and $L^{\mathscr{S}}_{F,R}$ in Figure 5. As shown, topic distributions for fake news are colored in red and real news in blue. We observe that real news sites are clearly more invested in reporting on campaigns and elections comparing to fake news sites. The average fraction of real news articles across all source pairs focusing on `election` is 22.5%, while the average is less than 15% for fake news articles. Real news sites are also more concentrated on topics including `economy`, `abortion`, and `climate`. Fake news sites, on the other hand, spend a considerable fraction, approximately 10%, of all articles on Clinton's email scandal alone, twice that of real news sites. Fake news sites also place stronger emphasis on topics such as the Middle East (terrorism, Muslims, Syria, et cetera), sexual scandals (a manual inspection reveals these are mostly related to Bill Clinton), Benghazi in its relation to Clinton when she served as the Secretary of State, candidate's health issues (e.g. reporting Clinton's pneumonia and claims of early onset dementia).

In addition, we also calculate the difference between $K^{\mathscr{S}}_{F,R}$ and $L^{\mathscr{S}}_{F,R}$ and denote it as $Z^{\mathscr{S}}_{F,R}$. As an example, $z^1_{f,r} = k^1_{f,r} - l^1_{f,r}$ is the fractional difference for topic 1 between fake and real news sites when using (f, r) as the groundtruth pair. We also plot $Z^{\mathscr{S}}_{F,R}$ in green in the same Figure 5. One notable finding is that, for all distributions of Z, most—if not all—of the data points consistently stay above or below the horizontal $y = 0$ line. In other words, regardless of the groundtruth pair we use, the trend we would observe stays comparatively constant. That is, the assessment as to whether a topic was more central to the coverage of fake news outlets (compared to the traditional news sites) is robust to the choice of groundtruth sets.

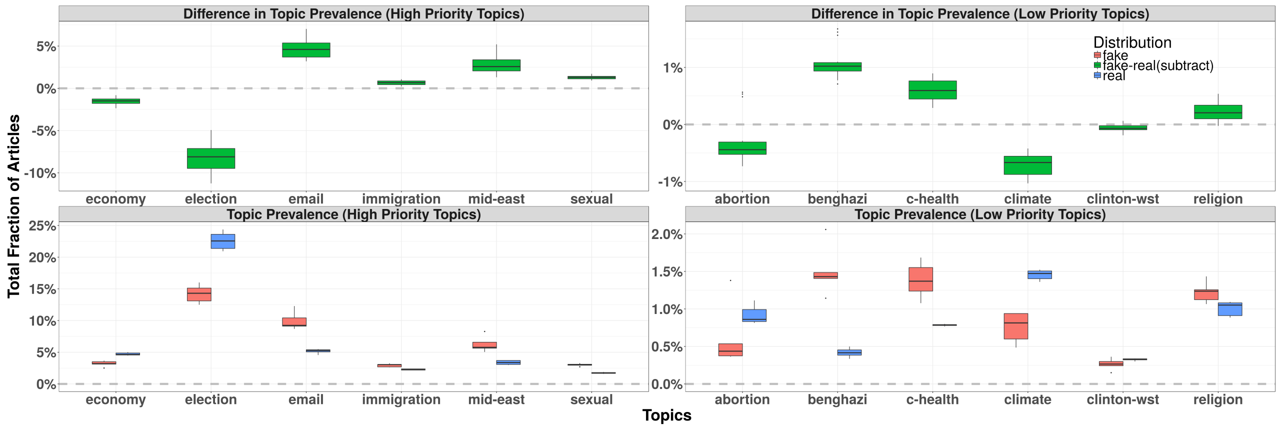
Describe Topic Distribution Difference Using Factor Analysis: Here, we provide a more in-depth analysis of variables (i.e. issue or topics) that contribute most to the variance in reporting by fake and real news sites of different ground pairs. We apply PCA to $K^{\mathscr{S}}_{F,R}$ and extract the first 2 principal components. We observe that the first component explains 48.1% of the total variance whereas the second component 38.2%. The resulting clusters are plotted in Figure 6. Similar to the PCA results from the time-series analysis in Section 5.2, we see a clear association between the clusters and the source of fake news sites. Though here, we don't see any apparent association between the clusters and how large the size of a fake news list is.

In addition, we also use factor loading to assess the correlation between the observed variables (i.e. topic distributions) and the latent variables (i.e. principal components) [11]. We summarize the results in Table 4. Coefficients for the topics that are most significantly correlated with either principal components are in **bold**. As shown, `election` is the topic most correlated with the first principal component followed by `email` and `mid-east`. For the second principal component, we see a correlation with a larger set of topics including `abortion`, `climate`, `economy`, `immigration`, `religion`, et cetera. This observation suggests that the difference in how much more or less fake and real news sites report about the election, Clinton's email scandal, and terrorism loads most into the first principal component. In other words, these topics are most significant in making the distinction between fake and real news sites. The remaining variance explained by principal component 2 is loaded by some of the other topics including the economy, and immigration, et cetera.

6 DISCUSSION

In this paper we first provided a comprehensive overview of the publicly available lists of fake and real news sites including their size, annotation process, label evaluation procedure, use in academia, et cetera. We observed that while a few of the lists are well documented, a considerable fraction did not clarify the pipeline they used to generate the lists. Further, we also compared and contrasted various fake news sources. We observed that a fake news sites' age and

Figure 5: Topic Prevalence Distribution Overview. The x-axis is the list of topics. A topic is a high priority if it accounts for at least 3% of total fake or real articles. The red and blue box plots respectively indicate the fractions of articles from fake and real news sites in each topic, calculated using different groundtruth source pairs. Finally, green boxplots represent the fractional difference between fake and real news sites.



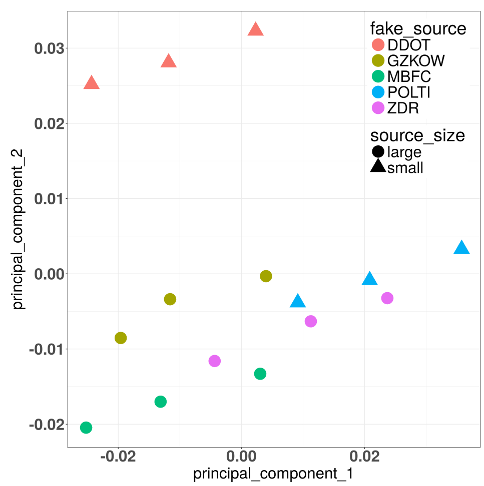


Figure 6: PCA Plot for Topic Distribution Differece Between Fake and Real News Sites Based on Distinct Groundtruth.

popularity are not significant in determining whether it would more likely to be incorporated by sources. In addition, we also showed that all sources are comparable in how much they choose to list conservative-leaning fake news sites over liberal-leaning ones or vice versa. In addition, we also explored whether and to what extent groundtruth for fake and real news sites matter to downstream analysis. We demonstrated that the perceived prevalence of fake news vary substantially based on which groundtruth is selected: fake news can constitute to more than 30% of total news or less than 5%. Further, we also showed that diverging groundtruth does not lead to a significant difference in time-series patterns of fake news. Finally, after an iterative topic modeling process with considerable manual efforts from the researchers, we showed that agenda-setting priority difference between fake and real news sites remain relatively robust to the groundtruth difference.

There are several caveats to our study. First, our analysis of groundtruth difference and its impact on subsequent analysis is limited to the domain-level labels. There are more granular labels that annotate at article-level or even at sentence-level. Future work should take into account the difference in the granularity of labels. Second, there is currently a considerable effort from both the academia and industry in building large-scale automated pipelines to classify, predict, and track fake news, as well as neutralize its influence. Concurrently, there is an ongoing conversation on how black-boxed, unaccountable machine learning models lead to increased inequality. Thus, an effort to make predictive models more transparent and human interpretable is also in progress. Here, we argue that understanding whether and how groundtruth difference affect the performance of automated fake news classifiers as well as their feature weights is of vital importance and may also provide valuable insights.

REFERENCES

[1] [n. d.]. https://www.politico.com/interactives/2017/politico50/craig-silverman/

[2] [n. d.]. Stanford researchers find students have trouble judging the credibility of information online. https://ed.stanford.edu/news/stanford-researchers-find-students-have-trouble-judging-credibility-information-online. Accessed: 2019-11-05.

[3] [n. d.]. Trumps State Department lacks money, clear mandate to fight Russian disinformation. https://www.usatoday.com/story/news/world/2018/09/21/trump-administration-lacks-resources-fight-russian-fake-news/1292089002/

[4] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2018. Fake news identification on twitter with hybrid cnn and rnn models. In *Proceedings of the 9th International Conference on Social Media and Society*. ACM, 226–230.

[5] Katitza Rodriguez Alimonti and Veridiana. 2018. Fake News Offers Latin American Consolidated Powers An Opportunity to Censor Opponents. https://www.eff.org/deeplinks/2018/03/fake-news-offers-latin-american-consolidated-powers-opportunity-censor-opponents

[6] 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

[7] 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

[8] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2018. Trends in the Diffusion of Misinformation on Social Media. *arXiv preprint arXiv:1809.05901* (2018).
[9] Meital Balmas. 2014. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. *Communication Research* 41, 3 (2014), 430–454.
[10] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science* 6, 1 (2011), 3–5.
[11] Anna B Costello and Jason W Osborne. 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation* 10, 7 (2005), 1–9.
[12] Michelle Croft and Raeal Moore. 2017. Checking What Students Know about Checking the News. (2017).
[13] Kareem Darwish, Walid Magdy, and Tahar Zanouda. 2017. Trump vs. Hillary: What went viral during the 2016 US presidential election. In *International Conference on Social Informatics*. Springer, 143–161.
[14] Anil R Doshi, Sharat Raghavan, Rebecca Weiss, and Eric Petitt. 2018. The Impact of the Supply of Fake News on Consumer Behavior During the 2016 US Election. (2018).
[15] Dag Elgesem, Ingo Feinerer, and Lubos Steskal. 2016. Bloggers responses to the snowden affair: Combining automated and manual methods in the analysis of news blogging. *Computer Supported Cooperative Work (CSCW)* 25, 2-3 (2016), 167–191.
[16] Lesley SJ Farmer. [n. d.]. Dont Get Faked Out by the News: Becoming an Informed Citizen. In *The Fifth European Conference on Information Literacy (ECIL)*. 174.
[17] Gheorghe-Ilie Farte and Daniel-Rares Obada. 2018. Reactive Public Relations Strategies for Managing Fake News in the Online Environment. (2018).
[18] Álvaro Figueira and Luciana Oliveira. 2017. The current state of fake news: challenges and opportunities. *Procedia Computer Science* 121 (2017), 817–825.
[19] Richard Fletcher, Alessio Cornia, Lucas Graves, and Rasmus Kleis Nielsen. 2018. Measuring the reach of fake news and online disinformation in Europe. *Reuters Institute Factsheet* (2018).
[20] Adam Fourney, Miklos Z Racz, Gireeja Ranade, Markus Mobius, and Eric Horvitz. 2017. Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2071–2074.
[21] Matthew Gentzkow and NBER Chuan Yu. 2018. Trends in the Diffusion of Misinformation on Social Media. (2018).
[22] David W Goodall. 1966. A new similarity index based on probability. *Biometrics* (1966), 882–907.
[23] Torstein Granskogen. 2018. *Automatic Detection of Fake News in Social Media using Contextual Information*. Master's thesis. NTNU.
[24] N Grinberg, K Joseph, L Friedland, B Swire-Thompson, and D Lazer. 2018. *Fake news on Twitter during the 2016 US presidential election*. Technical Report. Working Paper. Available from the authors.
[25] Neil Gross and Solon Simmons. 2006. Americans views of political bias in the academy and academic freedom. In *annual meeting of the American Association of University Professors*.
[26] Lei Guo and Chris Vargo. 2018. "Fake News" and Emerging Online Media Ecosystem: An Integrated Intermedia Agenda-Setting Analysis of the 2016 U.S. Presidential Election. *Communication Research* (June 2018), 009365021877717. https://doi.org/10.1177/0093650218777177
[27] Andrew F Hayes and Klaus Krippendorff. 2007. Answering the call for a standard reliability measure for coding data. *Communication methods and measures* 1, 1 (2007), 77–89.
[28] Alex Heath. 2016. Facebook is going to use snopes and other fact-checkers to combat and buryfake news. *Business Insider* (2016).
[29] Stefan Helmstetter and Heiko Paulheim. 2018. Weakly Supervised Learning for Fake News Detection on Twitter. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 274–277.
[30] Benjamin D. Horne and Sibel Adali. 2017. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. *arXiv:1703.09398 [cs]* (March 2017). http://arxiv.org/abs/1703.09398 arXiv: 1703.09398.
[31] Benjamin D. Horne, William Dron, Sara Khedr, and Sibel Adali. 2018. Assessing the News Landscape: A Multi-Module Toolkit for Evaluating the Credibility of News. In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*. ACM Press, Lyon, France, 235–238. https://doi.org/10.1145/3184558.3186987
[32] Philip N Howard, Samantha Bradshaw, Bence Kollanyi, and Gillian Bolsolver. 2017. Junk News and Bots during the French Presidential Election: What Are French Voters Sharing Over Twitter In Round Two?
[33] Elle Hunt. 2017. Disputed by multiple fact-checkers: Facebook rolls out new alert to combat fake news. *The Guardian* 21 (2017).
[34] Poynter Institute. 2019. International Fact Checking Network. https://ifcncodeofprinciples.poynter.org/. [Online; accessed -04-November-2019].
[35] Poynter Institute. 2019. International Fact Checking Network. https://ifcncodeofprinciples.poynter.org/. [Online; accessed -04-November-2019].
[36] Jagadeesh Jagarlamudi, Hal Daumé III, and Raghavendra Udupa. 2012. Incorporating lexical priors into topic models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 204–213.
[37] Vasilis Koulolias, Gideon Mekonnen Jonathan, Miriam Fernandez, and Dimitris Sotirchos. 2018. Combating Misinformation: An ecosystem in co-creation.
[38] Adam Kucharski. 2016. Post-truth: Study epidemiology of fake news. *Nature* 540, 7634 (2016), 525.
[39] Yenni Kwok. 2017. Where Memes Could Kill: Indonesias worsening problem of fake news. *Time, January* 6 (2017).
[40] I Lawrence and Kuei Lin. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* (1989), 255–268.
[41] Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, Vol. 2. Citeseer, 1–49.
[42] Lucas Lima, Julio Reis, Philipe Melo, Fabricio Murai, Leandro Araújo, Pantelis Vikatos, and Fabrício Benevenuto. 2018. Inside the Right-Leaning Echo Chambers: Characterizing Gab, an Unmoderated Social System. *arXiv preprint arXiv:1807.03688* (2018).
[43] Thomas J Main. 2018. *The Rise of the Alt-Right*. Brookings Institution Press.
[44] Alessandro R Marcon, Blake Murdoch, and Timothy Caulfield. 2017. Fake news portrayals of stem cells and stem cell research. *Regenerative medicine* 12, 7 (2017), 765–775.
[45] MBFC contributors. 2018. Media Bias/Fact Check (MBFC News) About. https://mediabiasfactcheck.com/about/
[46] Jon D Mcauliffe and David M Blei. 2008. Supervised topic models. In *Advances in neural information processing systems*. 121–128.
[47] Maxwell E McCombs and Donald L Shaw. 1972. The agenda-setting function of mass media. *Public opinion quarterly* 36, 2 (1972), 176–187.
[48] Hemant Misra, Olivier Cappé, and François Yvon. 2008. Using LDA to detect semantically incoherent documents. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 41–48.
[49] Tanushree Mitra and Eric Gilbert. [n. d.]. CREDBANK: A Large-Scale Social Media Corpus with Associated Credibility Annotations. ([n. d.]), 10.
[50] Nikil S Mukerji. [n. d.]. A conceptual analysis of fake news. ([n. d.]).
[51] Nikil S Mukerji. [n. d.]. What Is Fake News? ([n. d.]).
[52] Matti Nelimarkka, Salla-Maaria Laaksonen, and Bryan Semaan. 2018. Social Media Is Polarized, Social Media Is Polarized: Towards a New Design Agenda for Mitigating Polarization. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*. ACM, 957–970.
[53] David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 100–108.
[54] Leysia Palen and Amanda L Hughes. 2018. Social Media in Disaster Communication. In *Handbook of Disaster Research*. Springer, 497–518.
[55] Marco Pennacchiotti and Ana-Maria Popescu. 2011. A Machine Learning Approach to Twitter User Classification. *Icwsm* 11, 1 (2011), 281–288.
[56] Gordon Pennycook, Tyrone Cannon, and David G Rand. 2018. Prior exposure increases perceived accuracy of fake news. (2018).
[57] Politifact contributors. 2018. Who pays for PolitiFact? https://www.politifact.com/truth-o-meter/blog/2011/oct/06/who-pays-for-politifact/
[58] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638* (2017).
[59] Daniel Ramage, Susan T Dumais, and Daniel J Liebling. 2010. Characterizing microblogs with topic models. *ICWSM* 10, 1 (2010), 16.
[60] K Rapoza. 2017. Can fake news impact the stock market?
[61] R Rehurek and P Sojka. 2011. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic* 3, 2 (2011).
[62] Regina Rini. 2017. Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal* 27, 2 (2017), E–43.
[63] S Schaedel. 2018. Websites that post fake and satirical stories. FactCheck.
[64] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A Platform for Tracking Online Misinformation. *Proceedings of the 25th International Conference Companion on World Wide Web - WWW '16 Companion* (2016), 745–750. https://doi.org/10.1145/2872518.2890098 arXiv: 1603.01511.
[65] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592* (2017), 96–104.
[66] Kai Shu, H Russell Bernard, and Huan Liu. 2019. Studying fake news via network analysis: detection and mitigation. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*. Springer, 43–65.

[67] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

[68] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding user profiles on social media for fake news detection. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 430–435.

[69] Craig Silverman. 2016. Here are 50 of the biggest fake news hits on Facebook from 2016. *BuzzFeed, https://www. buzzfeed. com/craigsilverman/top-fake-news-of-2016* (2016).

[70] Kate Starbird. 2017. Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter.. In *ICWSM*. 230–239.

[71] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. 2018. Defining fake news A typology of scholarly definitions. *Digital Journalism* 6, 2 (2018), 137–153.

[72] Chris J Vargo, Lei Guo, and Michelle A Amazeen. 2018. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *new media & society* 20, 5 (2018), 2028–2049.

[73] Chris J Vargo, Lei Guo, and Michelle A Amazeen. 2018. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society* 20, 5 (May 2018), 2028–2049. https://doi.org/10.1177/1461444817712086

[74] Singh Vikash. 2018. GuidedLDA. https://github.com/vi3k6i5/GuidedLDA

[75] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[76] William Yang Wang. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. *arXiv:1705.00648 [cs]* (May 2017). http://arxiv.org/abs/1705.00648 arXiv: 1705.00648.

[77] Claire Wardle. 2017. Fake news. Its complicated. *First Draft News* (2017).

[78] Wikipedia contributors. 2018. The Daily Dot — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=The_Daily_Dot&oldid=865156488 [Online; accessed 27-October-2018].

[79] Alison Willmore. [n. d.]. The Oscars Were Not Actually A Battle Against Trumps America. ([n. d.]), 7.

[80] Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems* 2, 1-3 (1987), 37–52.

[81] Melissa Zimdars. 2016. My fake news list went viral. But made-up stories are only part of the problem. *The Washington Post* (2016).

[82] Melissa Zimdars. 2019. Open Sources. http://www.opensources.co/. [Online; accessed -04-November-2019].