COVID-19 in Bulgarian Social Media: Factuality, Harmfulness, Propaganda, and Framing

Preslav Nakov, 1 Firoj Alam, 1 Shaden Shaar, 1
Giovanni Da San Martino 2 and Yifan Zhang 1
1 Qatar Computing Research Institute, HBKU, Qatar
2 University of Padova, Italy
{pnakov, fialam, sshaar, yzhang}@hbku.edu.qa, dasan@math.unipd.it

Abstract
With the emergence of the COVID-19 pandemic, the political and the medical aspects of disinformation merged as the problem got elevated to a whole new level to become the first global infodemic. Fighting this infodemic is currently ranked very high on the list of priorities of the World Health Organization, with dangers ranging from promoting fake cures, rumors, and conspiracy theories to spreading xenophobia and panic. With this in mind, we studied how COVID-19 is discussed in Bulgarian social media in terms of factuality, harmfulness, propaganda, and framing. We found that most Bulgarian tweets contain verifiable factual claims, are factually true, are of potential public interest, are not harmful, and are too trivial to fact-check; moreover, zooming into harmful tweets, we found that they spread not only rumors but also panic. We further analyzed articles shared in Bulgarian partisan pro/con-COVID-19 Facebook groups and found that propaganda is more prevalent in skeptical articles, which use doubt, flag waving, and slogans to convey their message; in contrast, concerned ones appeal to emotions, fear, and authority; moreover, skeptical articles frame the issue as one of quality of life, policy, legality, economy, and politics, while concerned articles focus on health & safety.

1 Introduction
The ongoing global COVID-19 pandemic has brought an unprecedented situation with a lot of uncertainty: as this was a new disease, very little was known about it. This created an information void, where there was a lot of demand but little supply of reliable new information: a perfect breeding ground for all kinds of rumors and conspiracy theories, whose spread was facilitated by social media, which in turn optimized for user engagement (yet, later, they did put serious efforts in trying to limit the spread of false claims about COVID-19).

Unlike previous events that attracted a lot of disinformation, the emergence of the COVID-19 pandemic gave rise to a new powerful blending of medical and political disinformation, which resulted in the first global infodemic. Indeed, shortly after having declared the COVID-19 outbreak a pandemic, the World Health Organization had to engage in counter-measures against the growing infodemic, which it ranked among its top priorities in the fight against the COVID-19 pandemic. 1

Figure 1 shows some tweets that demonstrate how COVID-19 is discussed in Bulgarian social media. We can see that the problem goes beyond factuality: while some tweets spread rumors (Figure 1a), other discuss cure (Figure 1b). Indeed, the infodemic quickly extended to promoting bad cure, instilling panic, xenophobia, racism, and distrust in authorities, among others. (Alam et al., 2021b)

1 https://www.who.int/health-topics/infodemic

(a) rumor

I just saw Край на споровете за произхода на COVID-19! Той е изкуствено създавен в САЩ през 2015 г. - Click to see also  thebulgariantimes.com/%d0%9a%d1%80%d...

I just saw End of controversy over the origin of COVID-19! It was artificially created in USA in 2015 - Click to see also  https://t.co/ikFXAmp8bQ

(b) discusses cure

Открито е първото лекарство, което доказано спасява от COVID-19 https://t.co/rUx9EAO5gx

The first drug that has been proven to save from COVID-19 has been found https://t.co/rUx9EAO5gx

Figure 1: Bulgarian tweets with English translation.
Thus, it is important to analyze social media posts in terms of factuality, harmfulness, check-worthiness, etc. It is also useful to understand whether the post is propagandistic, what propaganda techniques are used, and how the issue is framed. While there have been studies focusing on (some of) these issues for high-resource languages such as English and Arabic (Barrón-Cedeño et al., 2020; Hossain et al., 2020; Li et al., 2020; Alam et al., 2021b; Nakov et al., 2021a,c), there has been less work for low-resource languages such as Bulgarian (Dinkov et al., 2019; Alam et al., 2021d; Shaar et al., 2021b,c). Here, we aim to bridge this gap by analyzing tweets and Facebook posts about COVID-19 in Bulgarian, with focus on factuality, harmfulness, propaganda, and framing.

Our contributions can be summarized as follows:

• We create a dataset of tweets and Facebook posts related to COVID-19.
• We perform analysis from various perspectives (factuality, harmfulness, propaganda, and framing), and we discuss some interesting observations from our analysis.

The rest of the paper is organized as follows: Section 2 offers a brief overview of previous work. Section 3 describes the dataset. Section 4 discusses our methodology. Section 5 discusses the findings. Finally, Section 7 concludes and points to possible directions for future work.

2 Related Work

Below, we discuss work relevant to our analysis, focusing on factuality, check-worthiness, propaganda, framing, and fighting the COVID-19 infodemic.

2.1 Factuality

A variety of task formulations have been proposed to address the spread of misinformation and disinformation online, and for each formulation, a number of approaches have been developed. Some good readings on the topic include surveys such as that by Shu et al. (2017), who adopted a data mining perspective on “fake news” and focused on social media. Another survey (Zubiaga et al., 2018) studied rumor detection in social media. The survey by Thorne and Vlachos (2018) took a fact-checking perspective on “fake news” and related problems.

Li et al. (2016) covered truth discovery in general. Lazer et al. (2018) offered an overview and discussion on the science of “fake news”. Vosoughi et al. (2018) focused on the proliferation of true and false news online. Other recent surveys focused on stance detection (Küçük and Can, 2020), propaganda (Nakov et al., 2021b), social bots (Ferrara et al., 2016), false information (Zannettou et al., 2019), and bias on the Web (Baeza-Yates, 2018). Some very recent surveys featured stance for misinformation and disinformation detection (Hardalov et al., 2021), automatic fact-checking to assist human fact-checkers (Nakov et al., 2021b), predicting the factuality and the bias of entire news outlets (Nakov et al., 2021d), and multimodal disinformation detection (Alam et al., 2021a).

A large body of research has focused on developing automatic systems for fact-checking to limit the spread of disinformation and misinformation (Li et al., 2016; Hardalov et al., 2016; Shu et al., 2017; Lazer et al., 2018; Mihaylova et al., 2018; Vosoughi et al., 2018; Nguyen et al., 2020). This includes development of datasets (Wang, 2017; Augenstein et al., 2019), and organizing evaluation campaigns (Derczynski et al., 2017; Nakov et al., 2018; Da San Martino et al., 2019; Elsayed et al., 2019; Gorrell et al., 2019; Mihaylova et al., 2019; Barrón-Cedeño et al., 2020; Nakov et al., 2021c; Shaar et al., 2021b). However, there are credibility issues with automated systems (Arnold, 2020). Hence, another research direction has emerged: building tools to facilitate human fact-checkers (Nakov et al., 2021b).

2.2 Check-Worthiness Estimation

Most work on check-worthiness focused on political debates and speeches. This includes the ClaimBuster (Hassan et al., 2015) and the ClaimRank systems (Jaradat et al., 2018), shared tasks at CLEF (Atanasova et al., 2018, 2019; Shaar et al., 2020, 2021c), modeling the context of the claim (Gencheva et al., 2017; Patwari et al., 2017; Shaar et al., 2021a), and multi-task learning from the decisions of multiple fact-checking organizations (Vasileva et al., 2019).

There has been less research on identifying check-worthy claims in social media posts. Previous work in this direction includes check-worthiness estimation of COVID-19 and political tweets (Alam et al., 2021d,b; Shaar et al., 2020, 2021b,c).
More directly related to our work here is the work of Alam et al. (2021d) and Alam et al. (2021b), who developed a multi-question annotation schema to annotate tweets about COVID-19, organized around seven questions that model the perspective of journalists, fact-checkers, social media platforms, policymakers, and the society. In our experiments, we use their schema and data to train classifiers for part of our analysis.

2.3 Propaganda

Propaganda is a communication tool, deliberately designed to influence the opinions and the actions of other people in order to achieve a predetermined goal. Computational propaganda is defined as the use of automated approaches to intentionally disseminate misleading information on social media platforms (Woolley and Howard, 2018).

Most research on propaganda detection has focused on analyzing textual content (Barrón-Cedeno et al., 2019; Rashkin et al., 2017; Da San Martino et al., 2019, 2020). Rashkin et al. (2017) developed the TSHP-17 corpus, which uses document-level annotation and is labeled with four classes: trusted, satire, hoax, and propaganda. They trained a model using word n-gram representation with logistic regression and reported that the model performed well only on articles from sources that the system was trained on. Barrón-Cedeno et al. (2019) developed the QProp corpus with two labels: propaganda vs. non-propaganda. They also experimented on TSHP-17 and QProp corpora, where for the TSHP-17 corpus, they binarized the labels: propaganda vs. any of the other three categories. Similarly, Habernal et al. (2017, 2018) developed a corpus with 1.3k arguments annotated with five fallacies, including ad hominem, red herring, and irrelevant authority, which directly relate to propaganda techniques.

A more fine-grained propaganda analysis was done by Da San Martino et al. (2019), who developed a corpus of news articles annotated with 18 propaganda techniques. Subsequently, the Prta system was released (Da San Martino et al., 2020b), and improved models were proposed, focusing on interpretability (Yu et al., 2021) or addressing the limitations of transformers (Chernyavskiy et al., 2021). Very recently, multimodal content was explored in memes using 22 fine-grained propaganda techniques (Dimitrov et al., 2021a,b).

2.4 Framing

Framing is a strategic device and a central concept in political communication, for representing different salient aspects and perspectives for the purpose of conveying the latent meaning about an issue (Entman, 1993). It is important for news media as the same topics can be discussed from different perspectives, which can influence our understanding, beliefs, and attitudes regarding what is happening in our society. There has been recent work on automatically identifying media frames, which includes developing coding schemes and datasets such as the Media Frames Corpus (Card et al., 2015), developing systems to automatically detect media frames (Liu et al., 2019; Zhang et al., 2019), large scale automatic analysis of New York Times Articles (Kwak et al., 2020), and a semi-supervised approach to detecting frames in online news sources (Cheeks et al., 2020).

2.5 COVID-19 Research

Since the beginning of the COVID-19 pandemic, there has been a large number of work on fighting the COVID-19 infodemic. Most notable work includes developing multi-question annotation schemas of tweets about COVID-19 (Alam et al., 2021d,b), studying credibility (Cinelli et al., 2020; Pulido et al., 2020; Zhou et al., 2020), racial prejudices and fear (Medford et al., 2020; Vidgen et al., 2020), situational information, e.g., caution and advice (Li et al., 2020), as well as on detecting mentions and stance with respect to known misconceptions (Hossain et al., 2020).

Another less relevant research line is on the development of datasets of tweets about COVID-19 (Cinelli et al., 2020; Song et al., 2021; Zhou et al., 2020; Haouari et al., 2021).

3 Dataset

Tweets: Using the Twitter API, we collected 30k tweets from January 2020 till November 2020. We performed search by specifying the target language to be Bulgarian and asking for the tweet to contain the following keywords and hashtags related to COVID-19 (English translations are shown in bleu color):

#корона, #коронавирус, коронавирус, корона
#corona, #coronavirus, coronavirus, corona
We only selected original tweets (no retweets or replies), we removed duplicates using a similarity-based approach (Alam et al., 2021c), and we filtered out tweets with less than five words. Finally, we selected the most frequently liked and retweeted tweets for annotation. For our analysis, we manually annotated 4k of them using the multi-question annotation schema from (Alam et al., 2021b), with three annotators per tweet (a total of 11k annotations). This Bulgarian data is also used in (Alam et al., 2021d) and for the CLEF 2021 CheckThat! lab task 1 (Shaar et al., 2021c).

Articles in Facebook posts: We further collected articles posted in Bulgarian Facebook groups that discuss COVID-19. We focused on concerned, skeptical, and conspiracy groups; the list is shown in Figure 3. We collected the links to articles posted in these groups, and we manually annotated each article as skeptical or concerned.

4 Method

Figure 2 shows our analysis pipeline. Below, we discuss each element of the pipeline in more detail.

4.1 Manual Annotation

The manual tasks consist of multi-question disinformation annotation of tweets and also of skeptical vs. concerned articles posted on Facebook.

---

**Concerned**

- (10.9k) Covid-19: факти срещу слухове
- (10.9k) Covid-19: facts against rumors
- (2.4k) Коронавирус/COVID-19 - само валидирана информация
- (2.4k) Coronavirus / COVID-19 - validated information only
- (1.0k) Здрав разум за здрава държава
- (1.0k) Common sense for a healthy state

**Skeptical**

- (14.8k) Аз подкрепям доцент Мангъров
- (14.8k) I support Associate Professor Mangarov
- (6.5k) Подкрепа за Атанас Мангъров
- (6.5k) Support for Atanas Mangarov
- (4.0k) С доц. д-р Манг*р*в и Бог обратно в живота
- (4.0k) With Assoc. Prof. Dr. Mang* r * v and God back to life

**Conspiracy**

- (1.4k) ИСТИНСТА: Коронавирус / COVID-19, КОВИД-19/
- (1.4k) THE TRUTH: Coronavirus / COVID-19, COVID-19/
- (1.3k) Изминалият Корона- вирус
- (1.3k) The Corona virus
- (1.2k) Измамата Ковид19
- (1.2k) The Kovid Deception19
- (0.2k) Корона вирус COVID-19 или големата манипулация
- (0.2k) Crown virus COVID-19 or major manipulation

---

![Figure 2: The system architecture of our analysis. The arrows indicate the information flow.](image)

![Figure 3: Facebook groups we collected articles from.](image)

---

4.1.1 Disinformation Annotation for Tweets

For the disinformation analysis, we used the holistic approach in (Alam et al., 2021b). It is formulated into seven questions, asking whether a tweet (Q1) contains a verifiable factual claim, (Q2) is likely to contain false information, (Q3) is of interest to the general public, (Q4) is potentially harmful to a person, a company, a product, or society, (Q5) requires verification by a fact-checker, (Q6) poses harm to society and why, or (Q7) requires the attention of policy makers and why. Three annotators worked on each tweet, following the annotation guidelines in (Alam et al., 2021b).
The annotators were fluent in Bulgarian, two were male and one was female, with qualifications ranging from undergrad students to people with a MSc degree. For disagreed annotations, a final consolidator participated in the discussion to decide the final label. We computed the inter-annotator agreement between the annotators and the final consolidated label using Fleiss Kappa (κ) as shown in Table 1. We can see that there was moderate to substantial agreement between the human annotators across the questions, according to the range of values for κ suggested in (Landis and Koch, 1977).

<table>
<thead>
<tr>
<th>Agree. Pair</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 - C</td>
<td>0.77</td>
<td>0.44</td>
<td>0.64</td>
<td>0.53</td>
<td>0.49</td>
<td>0.53</td>
<td>0.51</td>
</tr>
<tr>
<td>A2 - C</td>
<td>0.51</td>
<td>0.40</td>
<td>0.59</td>
<td>0.49</td>
<td>0.44</td>
<td>0.56</td>
<td>0.53</td>
</tr>
<tr>
<td>A3 - C</td>
<td>0.47</td>
<td>0.38</td>
<td>0.57</td>
<td>0.49</td>
<td>0.38</td>
<td>0.53</td>
<td>0.40</td>
</tr>
<tr>
<td>Avg</td>
<td>0.58</td>
<td>0.41</td>
<td>0.60</td>
<td>0.50</td>
<td>0.44</td>
<td>0.54</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Table 1: Inter-annotator agreement using Fleiss Kappa for the 7-level annotation for disinformation in tweets.

4.1.2 Skeptical vs. Concerned Annotation for Articles Posted on Facebook

The same annotators further annotated the Facebook articles as skeptical or concerned. This was a fairly straightforward task, with almost no disagreement. Note that we analyzed each article manually to decide whether it is skeptical or concerned (rather than using distant supervision to propagate the label for the group to label articles automatically, even though the vast majority of articles could be labeled with the label of the group).

4.2 Automatic Classification

For the analysis of propaganda and framing, both for tweets and for news articles, we used the automatic models discussed below.

4.2.1 Propaganda Analysis

For this analysis, we used Proppy and Pjeta.

Proppy (Barrón-Cedeño et al., 2019) uses a fragment-level and a sentence-level classifier. They were trained on a corpus of 350K tokens. The performance of the sentence-level classifier is 60.71 in terms of F1 score. The fragment-level classifier identifies the text fragments and the propaganda techniques that occur in them. They consider the following 18 techniques: (i) Loaded language, (ii) Name calling or labeling, (iii) Repetition, (iv) Exaggeration or minimization, (v) Doubt, (vi) Appeal to fear/prejudice, (vii) Flagwaving, (viii) Causal oversimplification, (ix) Slogans, (x) Appeal to authority, (xi) Black-and-white fallacy, dictatorship, (xii) Thought-terminating cliché, (xiii) Whataboutism, (xiv) Reductio ad Hitlerum, (xv) Red herring, (xvi) Bandwagon, (xvii) Obfuscation, intentional vagueness, confusion, and (xviii) Straw man.

Note that both Proppy and Pjeta only support English. To prepare their input, we translated the Bulgarian text to English using Google.

4.2.2 Framing

We used the Tanbih Framing Bias Detection system (Zhang et al., 2019), trained on the Media Frames Corpus (11k training news articles) by fine-tuning BERT to detect topic-agnostic media frames, achieving accuracy of 66.7% on the test set (1,138 news articles). It can predict the following 15 frames: (i) Economy, (ii) Capacity and resources, (iii) Morality, (iv) Fairness and equality, (v) Legality, constitutionality and jurisprudence, (vi) Policy prescription and evaluation, (vii) Crime and punishment, (viii) Security and defense, (ix) Health and safety, (x) Quality of life, (xi) Cultural identity, (xii) Security and defense, (xiii) Politics, (xiv) External regulation and reputation, and (xv) Other.

5 Results and Discussion

Below, we present the results of our analysis.

5.1 Disinformation Analysis

Figure 4 shows a detailed distribution for each question. We can see that (i) most tweets contain a verifiable factual claim, (ii) about half of the tweets are factually true, (iii) most of them are of general interest to the public, (iv) about half of the tweets are not harmful to the society, to a person, a company, or a product, (v) many tweets are trivial to fact-check, (vi) some tweets spread rumors, panic, or make a joke.
5.2 Propaganda Analysis

Propaganda Figure 5 shows the results for the propaganda analysis of tweets associated with check-worthiness and harmfulness. We can see that check-worthy tweets are more propagandistic (right-side bars in Figure 5a). A large portion of them (left-side bars) are neither check-worthy nor propagandistic. On Figure 5b, we can see that harmful tweets (i.e., such spreading rumors, conspiracy, and panic) are (somewhat) more propagandistic than non-harmful ones.

Figure 5c shows the propaganda analysis for articles posted in Facebook groups. We can see that skeptical articles are more propagandistic.

Propaganda Techniques A more fine-grained analysis is important in order to understand the type of content that is shared/posted in social media. Thus, we analyzed tweets by categorizing them using propaganda techniques. Figure 6 shows a propaganda technique analysis for the tweets, which are also labeled for check-worthiness and harmfulness.
(a) Check-worthiness and propaganda in tweets.

(b) Harmfulness and propaganda in tweets.

(c) Propaganda for skeptical vs. concerned articles posted in Facebook groups.

Figure 5: Propaganda.

Figure 6a shows that a higher proportion of check-worthy tweets are associated with fear, doubt, and loaded language, whereas non-check-worthy tweets are associated with exaggeration, flag-waving, name-calling, slogans, causal oversimplification, and repetition.

Figure 6b further shows that check-worthy tweets are associated with name-calling; rumor and conspiracy tweets are associated with doubt; and panic tweets are associated with fear, exaggeration, loaded language, and flag-waving.

Figure 5c shows the distribution of the propaganda techniques in skeptical vs. concerned articles posted in Facebook groups. We can see that skeptical articles are associated with doubt, flag-waving, and slogans, whereas concerned articles use loaded language, appeal to fear, appeal to authority, and causal oversimplification.

5.3 Framing

Our analysis of framing in tweets shows that economy is the dominant perspective, health and safety come second, and legality is third. Figure 7 reports the distribution of tweets manually annotated for check-worthiness and harmfulness and automatically analyzed for framing. Figure 7a shows that the most frequent check-worthy tweets are associated with health, legality, crime and punishment, whereas non-check-worthy are associated with economy, politics, and quality of life. Figure 7b reports the distribution of the framing and the harmfulness labels. Frames labeled as economy are non-harmful; cultural identity, crime and punishment are associated with rumor/conspiracy, while health and safety frames show panic.

Figure 7c reports the distribution of articles manually categorized as skeptical vs. concerned and automatically analyzed for framing. The plot shows that skeptical articles are associated with quality of life, policy, legality, economy, and politics, whereas concerned articles are associated with health and safety, and cultural identity.

6 Limitations

Manual annotations Our manual annotation for disinformation in tweets shows moderate to substantial agreement across the questions. We believe that this is reasonable given the complexity of the task.

Automatic analysis The performance of the automatic analysis varies across the different tasks (e.g., for propaganda analysis vs. framing), which can introduce noise in the results.

Translation We needed to translate the text from Bulgarian to English, which can add noise in case of translation errors. Although we performed a qualitative analysis on a sample of propaganda annotations and we found a good quality for our model’s predictions, in future work, we would like to train a model directly for Bulgarian.

7 Conclusion and Future Work

We presented our analysis of COVID-19 in Bulgarian social media with focus on tweets and on news articles posted in Facebook groups, which we collected in different time frames starting from January till November 2020. Then, we manually and automatically analyzed them using different aspects of disinformation, propaganda, and framing.
We believe that the kind of analysis we perform here would help in better understanding various trends in social media about COVID-19. See also a related study about COVID-19 and vaccines in Qatar (Nakov et al., 2021a).

There are a number of interesting research directions that could be pursued using the approaches we used in this study. While we only focused on Twitter and Facebook, similar analysis can be done on other platforms e.g., WhatsApp, Gab, Reddit.
(a) Check-worthiness and framing in tweets.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>check-worthy</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>non-check-worthy</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

(b) Harmfulness and framing in tweets.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>non-harmful</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>rumor/conspiracy</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>panic</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

(c) Framing for skeptical vs. concerned articles posted in Facebook groups.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Skeptical articles</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Concerned articles</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

Figure 7: Framing.

Acknowledgments

We would like to thank the Atlantic Club in Bulgaria and DataBee for helping us with the manual analysis of Bulgarian tweets and of news articles posted on Facebook. We further thank the anonymous reviewers for their constructive comments.

This research is part of the Tanbih mega-project (http://tanbih.qcri.org), developed at the Qatar Computing Research Institute, HBKU, which aims to limit the impact of “fake news”, propaganda, and media bias by making users aware of what they are reading, thus promoting media literacy and critical thinking.
References


Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI ’20, pages 4826–4832.


Yaliang Li, Jing Gao, Chushu Meng, Qi Li, Lu Su, Bo Zhao, Wei Fan, and Jiawei Han. 2016. A survey on truth discovery. SIGKDD Explor. NewsL., 17(2):1–16.


