Research Paper

Detection of conspiracy propagators using psycho-linguistic characteristics

Anastasia Giachanou (iD)
Universitat Politècnica de València, Spain; Utrecht University, The Netherlands

Bilal Ghanem
Universitat Politècnica de València, Spain; Symanto Research, Germany

Paolo Rosso
Universitat Politècnica de València, Spain

Abstract

The rise of social media has offered a fast and easy way for the propagation of conspiracy theories and other types of disinformation. Despite the research attention that has received, fake news detection remains an open problem and users keep sharing articles that contain false statements but which they consider real. In this article, we focus on the role of users in the propagation of conspiracy theories that is a specific type of disinformation. First, we compare profile and psycho-linguistic patterns of online users that tend to propagate posts that support conspiracy theories and of those who propagate posts that refute them. To this end, we perform a comparative analysis over various profile, psychological and linguistic characteristics using social media texts of users that share posts about conspiracy theories. Then, we compare the effectiveness of those characteristics for predicting whether a user is a conspiracy propagator or not. In addition, we propose ConspiDetector, a model that is based on a convolutional neural network (CNN) and which combines word embeddings with psycho-linguistic characteristics extracted from the tweets of users to detect conspiracy propagators. The results show that ConspiDetector can improve the performance in detecting conspiracy propagators by 8.82% compared with the CNN baseline with regard to F1-metric.

Keywords

Conspiracy propagators; linguistic analysis; social media analysis

Introduction

Online information disorder[1] has become one of the main threats of our society. Although fake news and conspiracy theories are not a new phenomenon, the exponential growth of social media has offered an easy platform for their propagation that can be faster compared with the one of real news [1]. A great amount of disinformation such as rumours, propaganda and conspiracy theories is propagated in online social media with the aim, usually, to deceive users and formulate specific opinions.

Information disorder can be classified as misinformation, disinformation or malinformation depending on the falseness and intention to harm [2]. Conspiracy theories, which are a type of disinformation (i.e. they contain false information and their intention is to harm), have become a topic of interest for sociologists and psychologists since the 1960s when the assassination of US President John F. Kennedy triggered many conspiracy theories. Different from other types of disinformation, conspiracy theories are alternative explanations offered in opposition to conventional and well-supported explanations of events [3]. For example, well-spread conspiracy theories support that the NASA moon landing was faked, the earth is flat, the climate change is a hoax and vaccines can lead to autism. Conspiracy theorists ignore the

Corresponding author:
Anastasia Giachanou, Utrecht University, Methodology and Statistics, Padualaan 14, 3584 CH, Utrecht, Netherlands.
Email: a.giachanou@uu.nl

overwhelming scientific evidence that has been performed on those fields and tend to explain events as a secret act of powerful forces.

Conspiracy theories are very powerful and can wreak huge amounts of damage in the society. For example, conspiracies about child trafficking caused a gun shooting incident at a pizzeria [4], whereas those about vaccination are to blame for an increase in measles cases in 2019. In addition, Jolley and Douglas [5] found that reading about anti-vaccine conspiracy theories reduced people's intentions to vaccinate, compared with people who read arguments refuting them, or those who did not read any material about vaccination.

Automated detection of information disorder has recently received the increasing attention of researchers. Different types of information have been explored for the detection of disinformation, such as emotions [6,7], user metadata [8], linguistic features [9], visual information [10–12] and temporal features [13]. Apart from the content, users also play a critical role for the detection of inaccurate content. Shu et al. [14] focused on users that share fake news and showed that some features, such as registration time, are different between users that share fake news and those that share real news. Also, they showed that combining user profile features with psycho-linguistic characteristics of the tweets is very effective for fake news detection. Vo and Lee [15] focused on the linguistic characteristics of fact-checking tweets and showed that the fact-checkers tend to use formal language in their tweets.

There are users that use social media platforms to share posts that support conspiracies theories. A conspiracy propagator is a user that tends to share posts that support conspiracy theories, whereas an anti-conspiracy propagator is a user that tends to share posts that refute conspiracies. Understanding the profile of conspiracy propagators is important not only for addressing the problem of conspiracy detection and restrain their negative consequences but also as an additional indicator to be used by intervention techniques and systems. For example, a system that can detect conspiracy propagators could be used by intervention strategies that have been shown to be useful in raising awareness to the users by suggesting them links from trusted news agencies or scientific evidence [16]. These intervention techniques can be particularly useful for the users that unintentionally shared the conspiracy-related posts. Also, such a system could be useful to track new conspiracy posts by tracking what is posted from the propagators' profiles or similar ones.

Several studies from psychology have tried to profile conspiracy theorists [17–20]. For example, Lantian et al. [20] found that belief in conspiracy theories appears to correlate with a need for uniqueness, whereas Goertzel [19] identified a lack of interpersonal trust as a key predictor of conspiratorial belief. In this study, we are interested to profile conspiracy propagators that post about well-known conspiracy theories (e.g. vaccination causes autism, the earth is flat). In particular, we analyse different profile and psycho-linguistic characteristics of conspiracy and anti-conspiracy propagators.[2] To the best of our knowledge, no principled study has been conducted on characterising the profiles of conspiracy and anti-conspiracy propagators who tend to share posts that support/refute widely spread conspiracy theories on social media. In addition, we propose ConspiDetector, a model based on a convolutional neural network (CNN) that can differentiate between conspiracy and anti-conspiracy propagators. In particular, we investigate the following three research questions:

Research Question 1: Which are the profile characteristics of users that are more probably to share posts that tend to support conspiracy theories?

To answer this question, first we build a collection that contains tweets that are about well-spread conspiracy theories using Twitter hashtags. Then, we collect and analyse different profile characteristics of the users (e.g. registration time, number of followers). We perform statistical comparisons of the profile characteristics to better understand the profiles of conspiracy and anti-conspiracy propagators.

Research Question 2: Which are the psycho-linguistic characteristics of users that are more probably to share posts that support/refute conspiracy theories?

Here, we are interested in analysing and understanding the linguistic patterns and personality traits of conspiracy and anti-conspiracy propagators. To this end, we use the state-of-the-art Linguistic Inquiry and Word Count (LIWC) tool [21] to analyse and compare different linguistic patterns (e.g. usage of personal pronouns, swear words) between the tweets that are posted by conspiracy propagators and by anti-conspiracy propagators. In addition, we compare the personality traits of conspiracy and anti-conspiracy propagators that we infer from the users' tweets as well as emotional and sentimental terms that are used in the posts using emotion and sentiment lexicons.

Research Question 3: Can we use the psycho-linguistic characteristics to differentiate between conspiracy and anti-conspiracy propagators?

To answer this question, we propose ConspiDetector, a CNN model to evaluate the effectiveness of linguistic and psychological features which are extracted from the users' tweets in differentiating between conspiracy and anti-conspiracy propagators.

Our analysis reveals that conspiracy propagators tend to have different profile characteristics compared with anti-conspiracy propagators. In addition, we show that there are differences in the linguistic patterns detected in the tweets of conspiracy and anti-conspiracy propagators. More importantly, we found differences in their personality traits. Finally, our experimental results show that the psycho-linguistic features are useful for the effective detection of conspiracy and anti-conspiracy propagators, whereas the profile characteristics are not.

In the remainder of the article, we first discuss related work on disinformation detection with a focus on conspiracy theories. After that, we describe the process we followed to create the dataset of posts extracted from the timelines of conspiracy and anti-conspiracy propagators. Then, we present the analysis of the profile and psycho-linguistic characteristics of conspiracy and anti-conspiracy propagators. Next, we present the ConspiDetector model, the evaluation process and its performance. Finally, we discuss the findings and the limitations of our work followed by the conclusions and future work.

Related work

The detection of online disinformation has received a lot of attention during the last years. Researchers have focused among others on the detection of fake news [22], rumours [23], clickbaits [24], bots [25], and fact checking [26]. Rashkin et al. [9] trained a Long Short-Term Memory (LSTM) network to classify credible and non-credible claims. They analysed various linguistic features extracted with the LIWC dictionary [27] such as personal pronouns and swear words. Other researchers focused on emotions expressed in fake news. For example, Vosoughi et al. [28] showed that false rumours triggered fear, disgust and surprise in their replies, whereas the true rumours triggered joy, sadness, trust and anticipation. Giachanou et al. [6] proposed emoCred, an LSTM-based neural network, that leveraged emotions from text to differentiate between credible and non-credible articles and showed that emotions are important for credibility detection. Wang [8] proposed a hybrid CNN to combine user metadata with text for fake news detection.

Some researchers have studied the language that is used in the different types of disinformation. Addawood et al. [29] studied the language that was used from Russian trolls which aimed to manipulate the public opinion during the 2016 US presidential election. In their study, they identified and analysed 49 linguistic cues as potential indicators of deceptive language and showed that the most important indicators of trolls are the number of hashtags and the number of retweets. To detect the linguistic cues, they used standard linguistic dictionaries such as LIWC which has been used in many prediction tasks, such as gender and age prediction [30] or prediction of improvements in mental and physical health [31]. Similar to those works, we also employ LIWC to extract linguistic cues.

Other researchers focused on profiling users that share fake news. Shu and Wang [14] performed an analysis of user profiles that share fake or real news. The analysis showed that there are features (e.g. registration time) that are different between users that share fake news and those that share real news. In addition, they examined the effectiveness of those features on fake news detection and showed that combining user profile features with the psycho-linguistic characteristics of the document can be very effective for fake news detection. Different from their study, we focus on conspiracy propagators and we perform a more thorough analysis of the users' characteristics. Also, in our study, we filter out accounts that are probably trolls or bots with the aim to analyse the characteristics of real users who share posts that tend to support well-known conspiracies (e.g. vaccines lead to autism).

Another related work is the one by Vo and Lee [15] who analysed linguistic characteristics of fact-checking tweets (i.e. tweets that confirm that an article is fake) and also proposed a deep learning framework to generate responses with fact-checking intention. Their analysis showed that the fact-checkers tend to refute fake news and use formal language. In addition, the tweets from fact-checkers emphasised on what happened in the past, whereas random tweets emphasised on present and future. Giachanou et al. [32] proposed a CNN-based system that leveraged a range of psycho-linguistic features and the inferred personality traits of the users to discriminate between potential fake news spreaders and fact-checkers. El Azab et al. [33] explored the effectiveness of different features and indicated the most important ones for fake account detection. In addition, Klein et al. [34] examined users that posted conspiracy theories on Reddit and whether there were any differences in the language and in the social environments used compared with other users. Finally, Rangel et al. [35] organised an evaluation shared task where the participants had to build a system that could determine whether a user is a potential fake news spreader or not.

There are also a lot of studies that have approached the problem of conspiracy theories from a social or psychological perspective. Lantian et al. [20] found that belief in conspiracy theories appears to correlate with a need for uniqueness. Jolley and Douglas [5] demonstrated that beliefs in anti-vaccine conspiracy theories or just the exposure to anti-vaccine

conspiracy theories directly affects vaccination intentions of people. In another study, Douglas and Sutton [36] showed that participants were affected by conspiracies regarding the death of Princess Diana even when they tried to discard the influence. Douglas et al. [18] focused on the psychological factors that make conspiracy theories popular. According to them, the psychological factors can be characterised as epistemic (e.g. desire for understanding), existential (e.g. desire for control) and social (e.g. desire to maintain a positive image of the self or group). In addition, Callaghan et al. [37] found that parents with high levels of conspiratorial thinking and needle sensitivity are more probably to delay the vaccination of their children.

Different from prior work, we approach the conspiracy theories from a computational perspective, and we profile online conspiracy and anti-conspiracy propagators. In particular, we analyse the profile and the psycho-linguistic characteristics of conspiracy and anti-conspiracy propagators based on the tweets that they post. In addition, we propose ConspiDetector that is based on a CNN and which incorporates linguistic characteristics to detect online users that tend to share posts that support or refute the conspiracy theories.

Collection

Several collections have been developed for the task of fake news detection [8,38]. The majority of these datasets contain fake and real articles and can be used for the evaluation of systems developed to detect fake news. However, to the best of our knowledge, there is no available collection that has focused on the conspiracy theories and that could be used for the classification of users into conspiracy and anti-conspiracy propagators. Therefore, we developed our own dataset.[3]

To create our data collection, we used Twitter Application Programming Interface (API) and we collected tweets about some of the most well-known conspiracy theories.[4] Table 1 shows the list of hashtags that we used to collect tweets that refer to conspiracy theories as well as some statistics with regard to the collection. Initially, we collected 25,975 tweets using the hashtags shown in Table 1. We have collected all the different types of posts (e.g. retweets, replies) containing those hashtags. We should mention that we treated replies and retweets independently to the original post. That is because when a user replies or retweets a post, she or he can do it either because she or he agrees or disagrees with the original tweet. In those cases, the new hashtags of the tweet were used for the annotation.

At this stage, we had two groups of hashtags, those that are probably used to support a conspiracy and those that refute them. We should note that these hashtags were used only to collect the initial set of tweets that refer to conspiracy theories and were not used for the final annotation of the tweets as supporting/refuting a conspiracy theory. For every hashtag that supports a conspiracy theory, we tried to have one that probably refutes the conspiracy (e.g. #vaccinesCauseAutism versus #vaccinesWork). However, this was not possible for all the hashtags.

Here, we should note that we collected the tweets that contained the conspiracy-related hashtags in June 2019. However, the final collection contains tweets from previous time periods because we did not put any time limit for collecting the tweets from users' timeline.

Since we focus on the user level, next we take users that have posted between 2 and 10 tweets in any of those hashtags of each group. This step filters out users that are posting a lot of tweets and which are probably bots. Given that the hashtags do not always reflect the content of the tweet, we decided to manually annotate those tweets to reassure if they indeed support or refute the conspiracy theory. In total, we manually annotated 6385 tweets. The manual annotation was

Table 1. Hashtags used to collect the tweets and statistics about the collection.

	Pro-conspiracy	Anti-conspiracy
Hashtags	#vaccinesCauseAutism	#vaccinesWork
	#antiVax	#vaccinessavelives
	#climateChangeIsNotReal	#climateChangeIsReal
	#flatEarth	#earthisnotflat
	#nasaLies	#nasatruth
	#nasaFake	#nasaIRreal
	#spaceIsFake	#spaceIsReal
	#moonLandingFake	#moonlandingisreal
	#bigPharmaFraud	
	#ebolaconspiracy	
	#antiFluoridation	
Users	977	950
Tweets	912,735	992,798

Table 2. Examples of tweets that support and refute a conspiracy theory.

Tweets that support a conspiracy theory	Tweets that refute a conspiracy theory
My babies will grow up to be healthier than all of you because I'm not injecting that poison into them **#antivax**	DID YOU KNOW? being **#antivax** #antivaxx decreases the chances of your child surviving adulthood.
Spread my newborn with peanut butter, better than vaxxing that little thing! **#antivax** #vaxxer #antivaxxer #antivacccine_movement	**#Antivax** groups spreading lies to immigrant communities & harming children are despicable. #vaccineswork
I don't vaccinate my children but one of them caught measles and the school sent him home! Really rude of them to try and compromise his education!! #DoctorsUnderOpression #antivaccine_movement **#antivax** #angry #EducationFest	#Measles Outbreak in Minnesota, US, Caused by **#antivax** Campaign, Officials Say http://www.livescience.com/59105-measles-outbreak-minnesota.html #antivaxx #vaccineswork
Hoping to learn more about the dangers of vaccines and ways to keep my babies healthy. I'll never let some doctor inject poison into them! But this is all new for me, so any advice from other parents on ways to keep them healthy would be appreciated! Thank you! **#antivax**	You can't get autism from a vaccine. Antivaxxers clearly don't understand how the brain works. #vaccinate #vaccines **#antivax** #vaccinateyourkids #antivaxxersareidiots #vaccineswork #measles #measlesoutbreak
I mean really, NASA lost the original tapes, telemetry data AND specs for how to get back to the moon?? I guess [they] figured if we fell for this contraption,we would fall for anything. **#SpaceIsFake** #MoonLandingHoax	If space is fake, how can we experience night and day? Solar and lunar eclipses? The northern lights? Stars and planets? Solar storms? Sunburns? Comets and asteroids. The moon? **#spaceisfake**
Still think the images brought to you by NASA and the other space agencies are real? Many are waking up to the deception. **#FlatEarth** #SpaceIsFake #EarthIsNotAGlobe	Then why is it so hard to produce a working map? Or to find inaccuracies in the globe map! That alone already proves the earth is a globe! **#flatearth** fail!

necessary since some of the hashtags are used to support as well as to refute a conspiracy theory. During the manual annotation, we noticed that the *#flatEarth* and *#antiVax* hashtags were the most controversial ones in the means that they were used to support as well as to refute the conspiracy. Table 2 shows some examples of tweets that support and refute a conspiracy theory. Some of those examples refer to vaccination and others to the flat earth conspiracy theory. For example, we observe that in the first tweet that supports the antiVax theory, the user is referring to the vaccines as a poison that do not want to give it to his or her children. It is clear that the user believes that vaccination can be harmful. On the contrary, the first tweet in the refuting column is against the theory by saying that if you believe in this theory, you 'decrease the chances of your child surviving adulthood'.

After the annotation of the tweets as *supporting, refuting* or *uncertain* to a conspiracy theory, we proceed with the annotation of the users. Let $u_{support}$ and u_{refute} be the number of tweets that a user u posted and which support or refute any of the conspiracy theories, respectively. Then for every user, we calculated the ratio of tweets that support a conspiracy as $u_{ratio} = u_{support}/(u_{support} + u_{refute})$. In case u_{ratio} was larger than 0.5, the user was annotated as a *conspiracy propagator*; otherwise, the user was annotated as *anti-conspiracy propagator*. Here, we should note that the majority of the users had a u_{ratio} of either 0 or 1, indicating that all their tweets belonged to one class.

After the annotation at a user level, we randomly selected 977 conspiracy propagators and 950 anti-conspiracy propagators. Finally, we collected the 1000 most recent tweets of those users. Our analysis is based on those tweets that refer to 912,735 and 992,798 tweets for conspiracy and anti-conspiracy propagators, respectively. Here, we should mention that we did not do any further manual annotation of the tweets collected for each user since these tweets are only used to infer and calculate the profile and the psycho-linguistic characteristics of the users.

Bots in the collection

As already mentioned, we have filtered out accounts that post a large amount of tweets because those accounts are more probably to be bots. This step is important since our intention is to focus on real users that share posts regarding these conspiracy theories. To check if there are any bots remained, we used the Botometer API.[5] Botometer uses a set of

features, such as sentiment, network and content, to detect whether a Twitter account is a bot or not. Botometer gives as a result a score on a scale of 0–5. Scores that are close to 0 are probably human accounts, whereas scores closer to 5 are more probably bots. In our experiments, we consider users whose Botometer score was larger than 2.5 as bots. The analysis showed that the ratio of bots in conspiracy propagators is 14.4%, whereas in anti-conspiracy propagators is 9%. Although the ratio is higher in the users that support conspiracies compared with those that refute them, this difference is not big. Also, the ratio is rather low, and this is important for the comparative analysis of our study.

Analysis of conspiracy propagators

In this section, first we analyse and compare the profile characteristics between conspiracy and anti-conspiracy propagators. Then we analyse the linguistic patterns, personality traits, emotions and sentiment expressed in the posts.
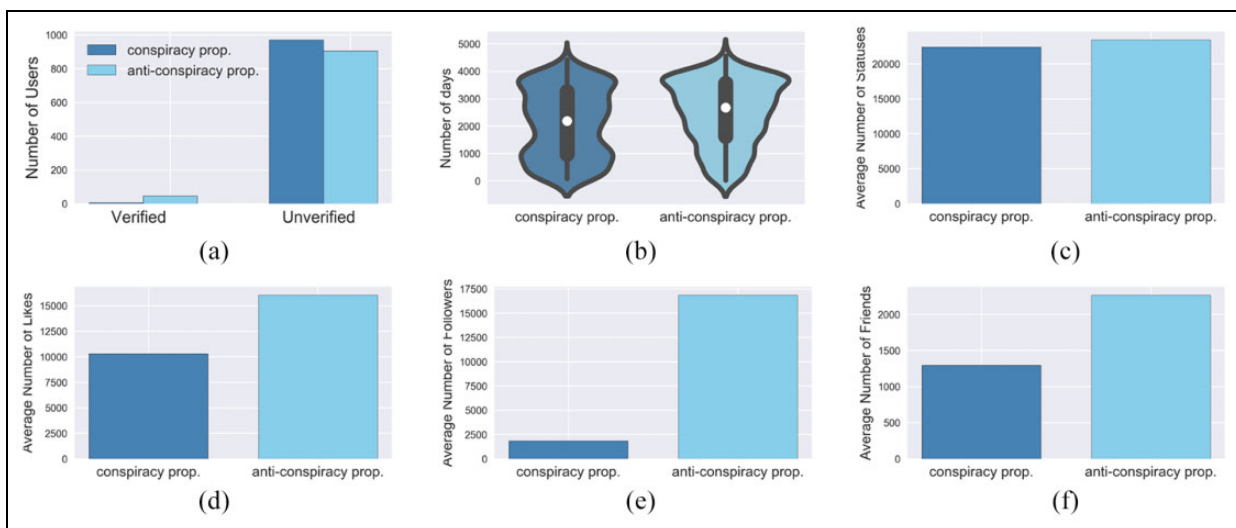
Profile characteristics

In this section, we try to answer the first research question: *Which are the profile characteristics of users that are more probably to share posts that tend to support conspiracy theories?* To highlight the profile characteristics of conspiracy propagators, we compare various profile characteristics of their accounts with those of anti-conspiracy propagators. In particular, we calculate the average scores of the different characteristics for the conspiracy and anti-conspiracy propagators. To measure if the differences are statistically significant, we use the Mann–Whitney U test [39].

We compare the following attributes:

- Verified: if the user has a verified account or not
- Registration time: time of registration of the user account
- Number of statuses: number of statuses posted by the user
- Number of likes: number of favourites of the user
- Number of followers: number of followers of the user
- Number of friends: number of friends of the user

Figure 1 shows the profile characteristics of conspiracy and anti-conspiracy propagators. From Figure 1(a), we observe that verified users are more probably to refute conspiracies, whereas conspiracy propagators are more probably to be unverified. This observation is in compliance with previous studies [14] that showed that users who tend to share fake news are not verified. Figure 1(b) shows the amount of accounts that were registered with regard to time for the conspiracy and anti-conspiracy propagators. We observe that there are conspiracy propagators that have new as well as old

Figure 1. Account, content and network-related characteristics of conspiracy and anti-conspiracy propagators: (a) Verified and unverified users. (b) Registration time. (c) Statuses. (d) Favourites. (e) Followers. (f) Friends.
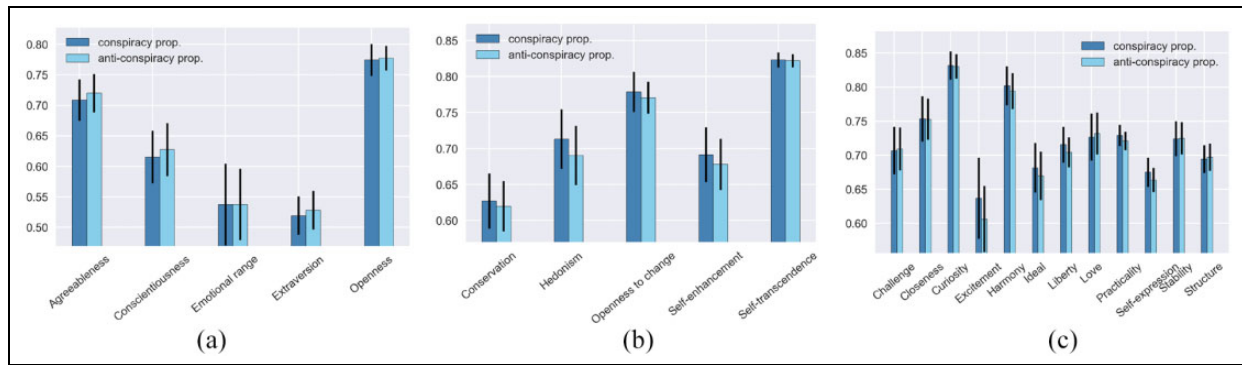
Figure 2. Personality-related characteristics for conspiracy and anti-conspiracy propagators: (a) Big Five. (b) Values. (c) Needs.

accounts. However, those that refute the conspiracy theories tend to have old accounts. This observation can be explained by the fact that some of the new accounts are deliberately created to support conspiracy theories and propagate false information. This finding is consistent with previous studies that have analysed profile characteristics for users that spread fake news [14,40].

Regarding the number of statuses shown in Figure 1(c), we observe that the anti-conspiracy propagators have a larger number of statuses compared with conspiracy propagators. However, it is interesting that the difference in the number of statuses is not large, although users that refute conspiracies have accounts that were created more recently. This happens because the accounts that are deliberately created to spread the conspiracy theories tend to post a large number of tweets.

In addition, we observe that conspiracy propagators have nine times less *followers* (μ = 1861.57) compared with anti-conspiracy propagators (μ = 16,871.21, p < 0.001). Similarly, conspiracy propagators have a lower number of statuses (μ = 22,441.7, p < 0.001), favourites (μ = 10,316.31, p < 0.001) and friends (μ = 1297.01, p < 0.001) compared with anti-conspiracy propagators.

Psycho-linguistic characteristics

In this section, we focus on the second research question: *Which are the psycho-linguistic characteristics of users that are more probably to share posts that support/refute conspiracy theories?* by analysing and comparing various psycho-linguistic patterns extracted from the tweets of the users. In particular, the *psycho-linguistic characteristics* include the following:

- Personality traits: the inferred personality traits of the user
- Sentiment: the sentiment polarity expressed in the user's tweets (i.e. positive, negative)
- Emotions: the amount of emotions expressed by the user in the tweets
- Linguistic patterns: the amount of different linguistic patterns expressed in the tweets

Personality traits: For the personality traits, we use the IBM Personality Insights API[6] to analyse users from multiple aspects based on their tweets. Figure 2 shows the personality traits regarding different aspects between conspiracy and anti-conspiracy propagators. In particular, we show the personality traits regarding the *Big Five, Values* and *Needs*. The Big Five model [41] is one of the most well-studied and well-known models and identifies five dimensions of personality traits of people: agreeableness, conscientiousness, emotional range (also known as neuroticism), extroversion and openness. *Values* refers to motivating factors that influence the user's decision-making process and includes conservation, hedonism, openness to change, self-enhancement and self-transcendence. *Needs* includes 12 categories (i.e. challenge, closeness, curiosity, excitement, harmony, ideal, liberty, love, practicality, self-expression, stability and structure).

We observe that users that share posts that refute conspiracy theories have a higher score in *agreeableness* (μ = 0.72) compared with the conspiracy propagators (μ = 0.708, p < 0.001) as well as in *conscientiousness* (p < 0.001) and *extroversion* (p < 0.001). On the other hand, regarding *Values*, conspiracy propagators have higher scores in *conservation* (μ = 0.627, p < 0.001), *hedonism* (μ = 0.713, p < 0.001), *openness to change* (μ = 0.779, p < 0.001) and *self-enhancement* (μ = 0.692, p < 0.001) compared with anti-conspiracy propagators. With regard to *Needs*, conspiracy
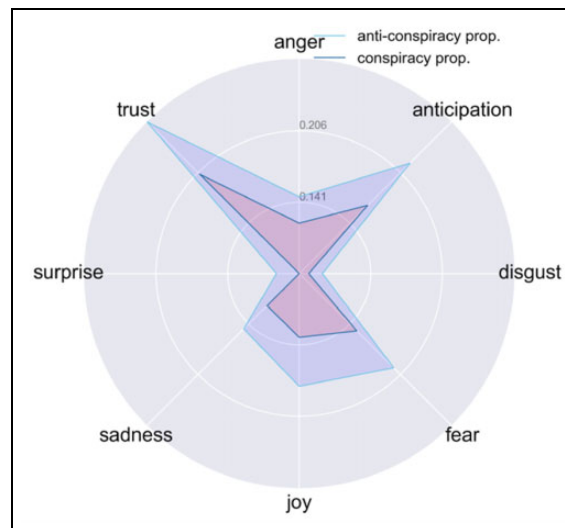
Figure 3. Average emotion scores for conspiracy and anti-conspiracy propagators.
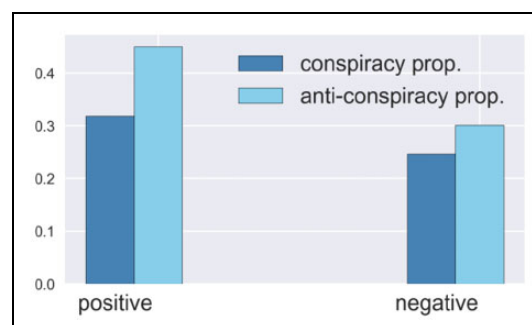
Figure 4. Average sentiment scores for conspiracy and anti-conspiracy propagators.

propagators have higher scores in *excitement* ($\mu = 0.637$, $p < 0.001$), *harmony* ($\mu = 0.802$, $p < 0.001$), *ideal* ($\mu = 0.682$, $p < 0.001$) and *liberty* ($\mu = 0.716$, $p < 0.001$) compared with anti-conspiracy propagators.

Emotions: We extract the emotions expressed in the tweets of a user. We follow Plutchik's model [42] and focus on the following eight emotions: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. To extract the emotions, we use National Research Council (NRC) emotions lexicon [43] that contains around 14K words labelled with regard to these emotions.

Figure 3 shows an analysis of the emotions that are expressed in the posts shared by conspiracy and anti-conspiracy propagators. We calculate the scores for eight different emotions (i.e. anger, anticipation, disgust, fear, joy, sadness, surprise, trust). We observe that the prevalent emotion expressed in the posts of both is trust. Also, anti-conspiracy propagators express more emotions in their tweets compared with conspiracy propagators. Studies showed that emotions are very important for similar tasks such as author profiling [44], credibility detection [6] and other types of fake news detection [7]. However, a key difference in our study is that we analyse the emotions that are expressed by the users in their posts and not only in the posts that support/refute a conspiracy theory.

Sentiment: To extract the sentiment, we use NRC emotions lexicon [43] that in addition to emotions provides also annotations for sentiment. Figure 4 shows that the tweets of users that tend to refute conspiracies express a larger amount of sentiment compared with the conspiracy propagators' tweets. This difference is smaller for the negative sentiment compared with the positive. In particular, anti-conspiracy propagators show a high usage regarding positive sentiment with an average score of $\mu = 0.45$ in comparison with conspiracy propagators ($\mu = 0.319$, $p < 0.001$) as well as regarding negative sentiment ($p < 0.001$).
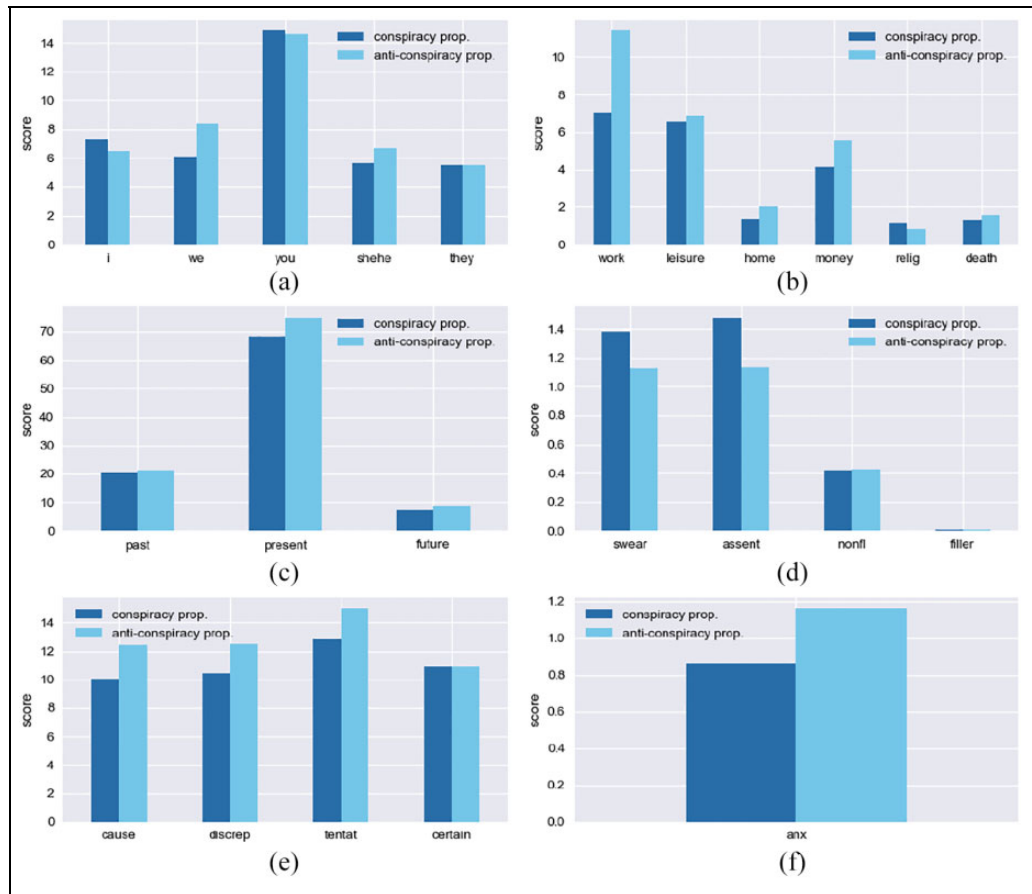
Figure 5. LIWC categories for conspiracy and anti-conspiracy propagators. (a) Pronouns. (b) Personal concerns. (c) Time. (d) Informal. (e) Cognitive. (f) Affective processes.

Linguistic patterns: For the linguistic patterns, we employed LIWC [21], a standard approach for mapping text to 73 psychologically meaningful categories, for comparing the psychological characteristics between conspiracy and anti-conspiracy propagators. In particular, we extract pronouns (I, we, you, she or he, they), personal concerns (work, leisure, home, money, religion, death), time focus (past, present, future), informal language (swear, assent, non-fluencies, fillers), cognitive processes (causation, discrepancy, tentative, certainty) and affective processes (anxiety).

For the linguistic analysis, we apply the following process. Given one tweet of a user, first we count how many words of the LIWC category appear in that tweet. We do this for all the tweets of the user, and then we calculate the average score for that user by dividing the total count with the number of tweets. Finally, we calculate the average of the scores for the conspiracy and anti-conspiracy propagators list.

Figure 5 shows the scores of the different psycho-linguistic characteristics between the conspiracy and anti-conspiracy propagators. Figure 5(a) shows that the users that refute conspiracy theories exhibit a higher usage of the third singular (i.e. she or he) and the first plural person (i.e. we) in comparison with the users that tend to support the conspiracy theories.

In Figure 5(b), we can see that users who share posts refuting conspiracy theories exhibit higher usage of personal concerns in comparison with users that tend to support conspiracies. In particular, anti-conspiracy propagators show a high usage regarding *work* (e.g. work, class, boss) with an average score of $\mu = 11.411$ in comparison with conspiracy propagators ($\mu = 7.058$, $p < 0.001$). An example of a tweet that refers to work is *In the office today after being gone for a week. Greeting me was a pile of cards, letters and flowers*. In addition, anti-conspiracy propagators exhibit a statistically significant higher usage regarding *leisure* (e.g. house, TV, music), *money* (audit, cash, owe), *home* (e.g. house, kitchen, lawn) and *death*. On the other hand, we observe that the conspiracy propagators have more concerns regarding *religion* ($\mu = 1.175$, $p < 0.001$). For example, the tweet *RT @TIME: Pope Francis opens the door for future female deacons* was posted by a conspiracy propagator.
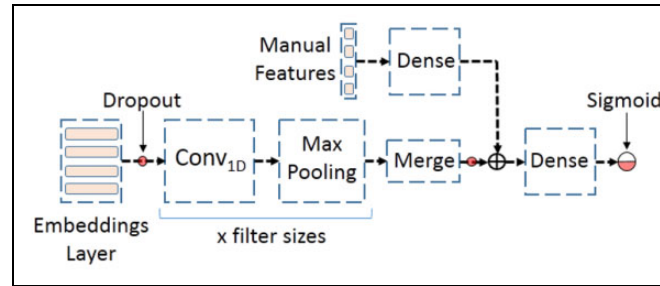
Figure 6. Architecture of ConspiDetector.

With regard to time focus, in Figure 5(c), we observe that both conspiracy and anti-conspiracy propagators focus more on *present* than *past* or *future*. This can be explained by the type of medium that is used to post what is happening at a specific time (i.e. tweeting). Also, users who tend to support conspiracies focus less on *present* (μ = 68.161, $p < 0.001$) and *future* (μ = 7.382, $p < 0.001$) in comparison with those that tend to refute the conspiracies.

From Figure 5(d), we observe that the conspiracy propagators tend to use more *swear* words (μ = 1.381) compared with anti-conspiracy propagators (μ = 1.126, $p < 0.001$). Also, conspiracy propagators tend to use more *assent* words (e.g. agree, yup, okey) (μ = 1.481) compared with anti-conspiracy propagators (μ = 1.131, $p < 0.001$).

Finally, regarding the cognitive processes, Figure 5(e) shows that anti-conspiracy propagators exhibit a higher usage in *causation* (because, effect, hence) in comparison with conspiracy propagators (μ = 10.079, $p < 0.001$). That is explained by the fact that users that refute the conspiracy theories use more explanations and arguments in their posts. Similarly, anti-conspiracy propagators show a statistically significant higher usage regarding *discrepancy* (should, would, could) and *tentative* (e.g. maybe, perhaps).

ConspiDetector

To reply to the third research question *Can we use the psycho-linguistic characteristics to differentiate between conspiracy and anti-conspiracy propagators?* we present ConspiDetector. ConspiDetector is based on a CNN and psycho-linguistic features that are extracted from the users' tweets with the aim to classify a user as a conspiracy or anti-conspiracy propagator. The model consists of two branches, a content-based and a lexicon-based. The content-based consists of an embedding layer followed by convolutional, max pooling and dense layers as shown in Figure 6.

Since the classification task is binary (conspiracy propagators versus anti-conspiracy propagators), as output we use a sigmoid layer. In our implementation, we use dropout both after the embedding layer and before the dense layer. With regard to multiple sizes of convolutional filters, we concatenate their outputs in one vector after the max pooling layer. To feed user's tweets to this branch, we concatenate all her or his tweets into a single document. It is worth to mention that we choose CNN rather than LSTM network since the process of concatenating all the tweets discards the sequential nature of the input document. This was also confirmed from the fact that when we applied LSTM on our data, we obtained a lower performance compared with CNN.

The second branch is based on the four groups of psycho-linguistic features (i.e. personality traits, emotions, sentiment and linguistic patterns) extracted from the user's tweets. Given a tweet of a user, first we count how many words from the categories/lexicons appear in that tweet (count frequency feature vector). We do this for all the tweets of the user, and then we calculate an average vector for that user by summing the tweets vectors and dividing them by the number of tweets. The final averaged vector is fed into the second branch of ConspiDetector.

Evaluation of ConspiDetector

In this section, we describe the evaluation process of ConspiDetector. We start by presenting the experimental settings and then we present and discuss the performance results with regard to conspiracy and anti-conspiracy propagator classification.

Experimental settings

For our experiments, we use the collection that we created and which contains tweets regarding conspiracy and anti-conspiracy propagators. We use 25% of the users from our corpus for validation, 15% for test and the rest for the

Table 3. Hyperparameters of the different combinations of features and ConspiDetector.

	Epochs	Filter sizes	No. of filters	Activation	Optimiser
CNN	8	3, 4	16	tanh	rmsprop
CNN + Profile	5	3, 5	64	tanh	sgd
CNN + IBM	2	4	8	tanh	adadelta
CNN + LIWC	12	3, 4, 5	8	tanh	adadelta
NN + Sentiment	4	6	8	relu	adadelta
CNN + Emotion	7	6	16	selu	adam
ConspiDetector (psycho-linguistic)	5	2, 3, 4	8	relu	adadelta
CNN + Psycho-linguistic + Profile	5	5	64	selu	adadelta

CNN: convolutional neural network; LIWC: Linguistic Inquiry and Word Count.

Table 4. Performance of the different combinations on the conspiracy and anti-conspiracy propagator detection.

	Precision	Recall	F1
Majority class	0.51	1.00	0.34
Random	0.50	0.47	0.50
USE	0.70	0.69	0.69
CNN	0.68	0.82	0.68
CNN + Profile	0.61	0.78	0.58
CNN + Personality	0.75	0.79	0.73
CNN + LIWC	0.73	0.78	0.71
CNN + Sentiment	0.67	0.76	0.66
CNN + Emotion	0.77	0.58	0.67
ConspiDetector (psycho-linguistic)	0.77	0.76	**0.74**
CNN + Psycho-linguistic + Profile	0.72	0.70	0.68

USE: Universal Sentence Encoder; CNN: convolutional neural network; LIWC: Linguistic Inquiry and Word Count.

training. We initialise our embedding layer with the 300-dimensional pre-trained GloVe embeddings [45]. In addition, we evaluate the performance of majority class, random classifier, CNN with only embeddings (CNN), and Universal Sentence Encoder (USE) [46]. We also evaluated the performance of the Bidirectional Encoder Representations from Transformers (BERT) [47]. However, we decided not to present the results of BERT because it achieved a very low performance that can be explained from the fact that BERT has been trained on contextual sentences, whereas in our experiments, the different tweets of a user are semantically unrelated. Also, when we concatenate all the tweets of a user, the final document length becomes very large,[7] where BERT input length is limited to 512 tokens.

For the USE baseline, we represent the final concatenated documents using USE embeddings[8] and then we feed them to a logistic regression classifier, which achieved the highest performance among the other tested classifiers (Random Forest, Support Vector Machine and Naïve Bayes). ConspiDetector is based on the combination of word embeddings and the psycho-linguistic characteristics (i.e. personality traits, emotions, sentiment, linguistic patterns). Also, we evaluate the performance of CNN using word embeddings and one group of features each time.

We use the *hyperopt* library to search for the best parameters on the validation set for each combination.[9] Table 3 shows the parameters we have used in our experiments. For all the systems, the batch size is 4. For the evaluation, we report precision, recall and macro-averaged F1 score.

Performance results

Table 4 shows the results of our experiments. We observe that ConspiDetector (CNN + psycho-linguistic) achieves the best performance. In particular, ConspiDetector manages to improve the performance by 8.82% compared with the CNN baseline.

Regarding using individual groups of features, the most effective is the IBM personality traits with a performance of 0.73 with regard to F1. The lowest performance is achieved with the profile characteristics (CNN + Profile) that are lower than the CNN baseline. Also, we observe that sentiment and emotion are not helpful and similar to the profile characteristics, and they obtain a lower performance compared with the CNN baseline. This is an interesting observation

since sentiment and emotion have been shown to be important in the detection of false information [6,7]. However, in the previous studies, the emotions were extracted from the false claims, whereas in our study, we analyse the emotional and sentimental language used by the users and therefore we use all the available published tweets of the users.

Finally, we observe that the result of the USE shows comparable performance to the CNN.

Discussion and limitations

In this study, we focused on conspiracy theories which are a specific type of disinformation and we compared various profile and psycho-linguistic characteristics between users that share posts that tend to support conspiracy theories and users that share posts that tend to refute them. We collected tweets about well-known conspiracy theories and we developed a collection of tweets posted by conspiracy and anti-conspiracy propagators. Our analysis showed that anti-conspiracy propagators have a higher number of followers, friends and favourites compared with conspiracy propagators. Also, the accounts of anti-conspiracy propagators are created earlier than the ones of the conspiracy propagators. Regarding the psycho-linguistic characteristics, we found that conspiracy propagators use a larger number of swear words. On the other hand, it seems that the anti-conspiracy propagators have more personal concerns (e.g. work), more cognitive processes and a higher score in anxiety. This analysis is very helpful for understanding linguistic differences between conspiracy and anti-conspiracy propagators and provides further insights regarding the research on conspiracy theories.

In addition, we showed that the psycho-linguistic characteristics (personality, emotions, sentiment, linguistic patterns) are effective for the detection of conspiracy and anti-conspiracy propagators, whereas the profile characteristics are not useful. Given the damage that conspiracy theories can cause on the society, the automated detection of conspiracy propagators can be a very useful tool. Users that are probably to share posts that support conspiracy theories could be identified and tracked regarding the content of their posts, or even blocked in cases of sharing harmful information.

Even if our study can provide valuable insights regarding the profile of conspiracy propagators and their automated detection, there are some limitations. The first limitation of our study is the use of Twitter data for the analysis. Our decision is based on the fact that Twitter provides easy and fast access to data and it was feasible to collect our data on conspiracy theories and make the analysis. However, Twitter users are not a representative sample of the population and is possible that there is some bias in the data. In addition, it is very hard to estimate if our data are representative since we do not have access to the demographics of those users.

Another limitation of our study is the use of IBM personality insight tool to infer the personality traits of the users based on the tweets that they posted. All the automated tools that are used for predictions are subject to errors. That means that some of the predictions regarding the personality traits that were inferred from the text are not correct. However, it is not possible to evaluate the performance of IBM personality insights since we do not have ground truth data regarding the users' personality traits. An alternative way would be to contact those users and ask them to fill in one of the standard questionnaires (e.g. Myers-Briggs Type Indicator (MBTI) questionnaire) that have been evaluated based on several psychological studies and tend to have more precise results. However, the feasibility of this approach depends on willingness of the users to fill the questionnaire.

Our study has also some ethical concerns. We should mention that the aim of a system that can differentiate between conspiracy and anti-conspiracy propagators should be used by no means to stigmatise any users. On the contrary, such a tool should be used only for the benefit of the users. For example, it could be used as a supportive tool to prevent propagation of conspiracy theories that lack scientific evidence and to raise awareness to users by improving the effectiveness of intervention techniques and strategies.

This study has also some ethical concerns regarding the collection and the release of the data. First, we plan to make this collection available only for research purposes. To protect the privacy of users, we plan to publish the data anonymously. Also, we plan to use neutral annotation labels regarding the two classes (i.e. 0 and 1 instead of conspiracy propagator or not). Future researchers that want to use the collection will not have access to the information of which class each label refers to. Finally, we will not make available the labels at a post level since this information can reveal the information regarding the annotation labels at a user level.

Conclusions and future work

In this article, we focused on conspiracy theories and we analysed the profile and psycho-linguistic differences between users that share posts that tend to support and those that share posts that tend to refute conspiracy theories. We developed a collection that contains posts of conspiracy and anti-conspiracy propagators that can be used for further research on the area of computational conspiracy detection. In addition, we proposed ConspiDetector, a CNN-based model that leverages different psycho-linguistic features to differentiate between conspiracy and anti-conspiracy propagators. We exploited

users' tweets to extract information that represents the personality traits, emotion, sentiment and linguistic patterns (pronouns, personal concerns, time focus, informal language, cognitive processes and affective processes).

Our analysis showed that anti-conspiracy propagators have a higher number of followers, friends and favourites compared with conspiracy propagators and tend to have older accounts compared with the ones of the conspiracy propagators. Also, we found that conspiracy propagators use a larger number of swear words, whereas anti-conspiracy propagators have more personal concerns (e.g. work), more cognitive processes and a higher score in anxiety.

In addition, we showed that the personality traits and the linguistic patterns are the most effective when the different groups are used individually. We showed that ConspiDetector that combines word embeddings with the psycho-linguistic characteristics achieved the highest performance. The results showed that incorporating information from the users' profile did not improve the performance; even the analysis showed that there are differences in the number of followers, friends and favourites. Finally, we believe that our study and detection approach could be very useful for developing policies to limit the spread of mis/disinformation or enhancing fact-checking platforms, combining them with theoretical studies that model mis/disinformation spreading and especially the role of some targeted nodes in the network [48–51].

In future, we plan to explore ways to overcome the current limitations of the study by contacting the users and ask them to fill in personality questionnaires. Also, we think it would be interesting to analyse the profile of conspiracy propagators across different countries using the geographical location information available in tweets. Finally, we plan to investigate how our findings can be used to improve the effectiveness of fake news detection systems and of intervention techniques.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Notes

1. Following Wardle and Derakhshan [2], we use the term information disorder to refer to all the different types of false and inaccurate information
2. We should note that a user who shares a post that supports a conspiracy theory does not mean that the user is a conspiracy believer/theorist since a user can post different things compared with what he believes.
3. The dataset and the code are available for research purposes upon request.
4. https://en.wikipedia.org/wiki/List_of_conspiracy_theories
5. https://botometer.iuni.iu.edu
6. https://personality-insights-demo.ng.bluemix.net/
7. The average length of the final documents in our collection is larger than 4000 tokens.
8. https://tfhub.dev/google/universal-sentence-encoder-large/3
9. https://github.com/hyperopt/hyperopt

ORCID iD

Anastasia Giachanou 🆔 https://orcid.org/0000-0002-7601-8667

References

[1] Lazer DMJ, Baum MA, Benkler Y et al. The science of fake news. *Science* 2018; 359(6380): 1094–1096.
[2] Wardle C and Derakhshan H. Information disorder: toward an interdisciplinary framework for research and policy making. *Council of Europe Report*, 2017, https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c
[3] McCaffrey P. *The reference shelf: conspiracy theories*. New York: HW Wilson, 2012.
[4] Kang C and Goldman A. In Washington Pizzeria attack, fake news brought real guns. *New York Times*, 2016, p. 5, https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

[5]   Jolley D and Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. *PLoS ONE* 2014; 9(2): 1–9.

[6]   Giachanou A, Rosso P and Crestani F. Leveraging emotional signals for credibility detection. In: *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, SIGIR '19*, Paris, 21–25 July 2019, p. 877–880. New York: ACM.

[7]   Ghanem B, Rosso P and Rangel F. An emotional analysis of false information in social media and news articles. *ACM Trans Intern Tech* 2020; 20(2): 1–18.

[8]   Wang WY. Liar, liar pants on fire: a new benchmark dataset for fake news detection. In: *Proceedings of the 55th annual meeting of the association for computational linguistics (Volume 2: Short Papers), ACL '17*, 2017, pp. 422–426, https://arxiv.org/abs/1705.00648

[9]   Rashkin H, Choi E, Jang JY et al. Truth of varying shades: analyzing language in fake news and political fact-checking. In: *Proceedings of the 2017 conference on empirical methods in natural language processing, EMNLP '17*, Copenhagen, 7–11 September 2017.

[10] Singhal S, Shah RR, Chakraborty T et al. SpotFake: a multi-modal framework for fake news detection. *In: 2019 IEEE 5th international conference on multimedia big data. BigMM '19*, Singapore, 11–13 September 2019, pp. 39–47. New York: IEEE.

[11] Giachanou A, Zhang G and Rosso P. Multimodal fake news detection with textual, visual and semantic information. In: *23rd international conference, TSD 2020*, Brno, 8–11 September 2020. New York: Springer.

[12] Giachanou A, Zhang G and Rosso P. Multimodal multi-image fake news detection. In: *2020 IEEE 7Th international conference on data science and advanced analytics, DSAA '20*, Sydney, NSW, Australia, 6–9 October 2020, pp. 647–654. New York: IEEE.

[13] Ma J, Gao W, Wei Z et al. Detect rumors using time series of social context information on microblogging websites. In: *Proceedings of the 24th acm international on conference on information and knowledge management, CIKM '15*, Melbourne, VIC, Australia, 19–23 November 2015, pp. 1751–1754. New York: ACM.

[14] Shu K, Wang S and Liu H. Understanding user profiles on social media for fake news detection. In: *Proceedings of the 2018 IEEE conference on multimedia information processing and retrieval, MIPR '18*, Miami, FL, 10–12 April 2018, pp. 430–435. New York: IEEE.

[15] Vo N and Lee K. Learning from fact-checkers: analysis and generation of fact-checking language. In: *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval. SIGIR '19*, Paris, 21–25 July 2019. New York: ACM.

[16] Lutzke L, Drummond C, Slovic P et al. Priming critical thinking: simple interventions limit the influence of fake news about climate change on Facebook. *Global Environ Change* 2019; 58: 101964.

[17] Bensley DA, Lilienfeld SO, Rowan KA et al. The generality of belief in unsubstantiated claims. *Appl Cognit Psychol* 2019; 34: 16–28.

[18] Douglas KM, Sutton RM and Cichocka A. The psychology of conspiracy theories. *Curr Direct Psychol Sci* 2017; 26(6): 538–542.

[19] Goertzel T. Belief in conspiracy theories. *Political Psychol* 1994; 1994: 731–742.

[20] Lantian A, Muller D, Nurra C et al. I know things they don't know! *Social Psychol* 2017; 48: 160–173.

[21] Pennebaker JW, Boyd RL, Jordan K et al. The development and psychometric properties of LIWC 2015. *The University of Texas at Austin*, Austin, TX, 2015.

[22] Ruchansky N, Seo S and Liu Y. CSI: a hybrid deep model for fake news detection. In: *Proceedings of the 2017 ACM on conference on information and knowledge management CIKM '17*, Singapore, 6–10 November 2017, pp. 797–806.

[23] Derczynski L, Bontcheva K, Liakata M et al. SemEval-2017 Task 8 RumourEval: determining rumour veracity and support for rumours. In: *Proceedings of the 11Th International Workshop on Semantic Evaluation. Semeval '17*, Vancouver, BC, Canada, 3–4 August 2017, pp. 69–76, https://dblp.org/rec/conf/semeval/2017.html

[24] Anand A, Chakraborty T and Park N. We used neural networks to detect clickbaits: you won't believe what happened Next! In: *Proceedings of the 2017 European conference on information retrieval*. ECIR '17, Aberdeen, 8–13 April 2017.

[25] Rangel F and Rosso P. Overview of the 7th author profiling task at PAN 2019: bots and gender profiling in Twitter. In: *CLEF 2019 Labs and Workshops, Notebook Papers*, 2019, https://dblp.org/rec/conf/clef/RangelR19.html

[26] Vlachos A and Riedel S. Fact checking: task definition and dataset construction. In: *Proceedings of the ACL 2014 workshop on language technologies and computational social science*, Baltimore, MD, 26 June 2014.

[27] Tausczik YR and Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2010; 29(1): 24–54.

[28] Vosoughi S, Roy D and Aral S. The spread of true and false news online. *Science* 2018; 359(6380): 1146–1151.

[29] Addawood A, Badawy A, Lerman K et al. Linguistic cues to deception: identifying political trolls on social media. In: *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, Palo Alto, CA, 25–28 June 2019, p.15–25. Reston, VA: AAAI.

[30] Pennebaker JW and King LA. Linguistic styles: language use as an individual difference. *J Personal Soc Psychol* 1999; 77(6): 1296.

[31] Pennebaker JW, Mayne TJ and Francis ME. Linguistic predictors of adaptive bereavement. *J Personal Soc Psychol* 1997; 72(4): 863.

[32] Giachanou A, Ríssola EA, Ghanem B et al. The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In: Métais E, Meziane F, Horacek H et al. (eds) *Natural language processing and information systems*. New York: Springer, 2020, p.181–192.

[33] El Azab A, Idrees AM, Mahmoud MA et al. Fake account detection in twitter based on minimum weighted feature set. *Int J Comput Inform Eng* 2015; 10(1): 13–18.

[34] Klein C, Clutton P and Dunn AG. Pathways to conspiracy: the social and linguistic precursors of involvement in reddit's conspiracy theory forum. *PLoS ONE* 2019; 14(11): e0225098.

[35] Rangel F, Giachanou A, Ghanem B et al. Overview of the 8th Author Profiling Task at PAN 2020: profiling Fake News Spreaders on Twitter. In: *CLEF 2020 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings*, 2020, http://ceur-ws.org/Vol-2696/paper_267.pdf

[36] Douglas KM and Sutton RM. The hidden impact of conspiracy theories: perceived and actual influence of theories surrounding the death of Princess Diana. *J Social Psych* 2008; 148(2): 210–222.

[37] Callaghan T, Motta M, Sylvester S et al. Parent psychology and the decision to delay childhood vaccination. *Soc Sci Med* 2019; 2019: 238.

[38] Shu K, Mahudeswaran D, Wang S et al. FakeNewsNet: a data repository with news content, social context and dynamic information for studying fake news on social media, 2018, https://arxiv.org/abs/1809.01286

[39] Mann HB and Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 1947; 18(1): 50–60.

[40] Castillo C, Mendoza M and Poblete B. Information credibility on Twitter. In: *Proceedings of the 20th international conference on World Wide Web WWW '11*, 28 March–1 April 2011, p.675–684. New York: ACM.

[41] Digman JM. Personality structure: emergence of the five-factor model. *Ann Rev Psychol* 1990; 41(1): 417–440.

[42] Plutchik R. A psychoevolutionary theory of emotions. *Soc Sci Inform* 1982; 21: 529–553.

[43] Mohammad SM and Turney PD. Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In: *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, Los Angeles, CA, June 2010, p.26–34.

[44] Rangel F and Rosso P. On the impact of emotions on author profiling. *Inform Process Manag* 2016; 52(1): 73–92.

[45] Pennington J, Socher R and Manning C. Glove: global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing EMNLP '14*, Doha, 25–29 October 2014, p.1532–1543.

[46] Cer D, Yang Y, Kong S et al. Universal sentence encoder, 2018, https://arxiv.org/abs/1803.11175

[47] Devlin J, Chang MW, Lee K et al. BERT: pretraining of deep bidirectional transformers for language understanding, 2018, https://arxiv.org/abs/1810.04805

[48] Kitsak M, Gallos LK, Havlin S et al. Identification of influential spreaders in complex networks. *Nature Physics* 2010; 6(11): 888–893.

[49] Tambuscio M and Ruffo G. Fact-checking strategies to limit urban legends spreading in a segregated society. *Appl Netw Sci* 2019; 4(1): 116.

[50] Borge-Holthoefer J, Meloni S, GoncÇalves B et al. Emergence of influential spreaders in modified rumor models. *J Stat Phys* 2013; 151(1–2): 383–393.

[51] Ghosh R and Lerman K. Predicting influential users in online social networks. In: *Proceedings of the KDD workshop on social network analysis*, 2010, https://arxiv.org/abs/1005.4882