Social Media as a Lens for Understanding Public Trust in Science

Panagiotis Monachelis
Dept. of Electrical and Electronics Engineering
University of West Attica
Athens, Greece
pmonahelis@uniwa.gr

Emily Maitland
Trilateral Research Ltd
London, United Kingdom
Emily.Maitland@trilateralresearch.com

Kalypso Iordanou
School of Sciences
UCLan Cyprus
Larnaka, Cyprus
KIordanou@uclan.ac.uk

Charalampos Z. Patrikakis
Dept. of Electrical and Electronics Engineering
University of West Attica
Athens, Greece
bpatr@uniwa.gr

Evren Yalaz
Trilateral Research Ltd
Waterford, Ireland
Evren.Yalaz@trilateralresearch.com

Pericles Papadopoulos
Dept. of Electrical and Electronics Engineering
University of West Attica
Athens, Greece
ppapadop@uniwa.gr

***Abstract*— Social media has become a crucial platform for communication and information sharing, but it also facilitates the spread of misinformation, particularly during crises such as the COVID-19 pandemic undermining public's trust in science. Especially, battling misinformation on scientific issues is important to bring closer public to science. The present work is a part of the EU funded project VERITY that aims to address this challenge by using a multidisciplinary approach resulting in in development of strategies to enhance public's trust in science. Among other methodologies, VERITY investigates social media to identify factors that influence trust in science, and this article provides the results of a study on a specific dataset related to COVID-19 vaccination through various methods such as social network analysis, quantitative and qualitative analysis, analysis of the content and the reactions of the messages and the application of deep learning models in order to reveal people's behaviour and the factors that influence it.**

Keywords— Online Social Networks, Trust in Science, Social Network Analysis, Sentiment Analysis, Subjectivity Analysis

I. Introduction

Social media play a significant role in how people communicate, access and share information. The social media applications have become a part of our daily life for entertainment, communication and access to information. Despite the advantages of online social networks, their use is linked to several challenges, and misinformation (an issue which has particularly concerned researchers) is definitely one of them. Misinformation is a phenomenon that is particularly important to deal with in times of crisis, such as the COVID-19 pandemic when many conspiracy theories found fertile ground to grow and create a mindset of scepticism against vaccination. It is therefore more urgent than ever to tackle misinformation especially in scientific topics, since causing public doubt towards science is targeting the foundations over which our modern societies and our way of life (depending on technological evolution) has been based. Considering the findings of the Eurobarometer 516 [1] according to which people choose social networks as their second source for information about science and technology after TV, we can realise the potential risks, since through social networks can affect a large number of users, with the majority of them belonging to the young audience. Misinformation on social media can affect some EU people more and other less, as according to the Eurobarometer 516 social media use to learn about science and technology highly varies among the 27 countries of EU. The countries of southern Europe, including Cyprus (53%), Greece (50%), Malta (44%) and Spain (39%) are above the EU27 average (29%), while countries of norther Europe including Finland (17%), Denmark (22%) and Sweden (23%) use social media for science and technology news much less. An interesting finding of the Eurobarometer is also that the percentage of people who use social media as the most used source for science and technology (28%) tend to believe in conspiracy theories in a double percentage than people who use social media as the least used source (12.7%). The double-edged sword of the use of social media as a source of information is that while science becomes more accessible to everyone, people also become more susceptible to misinformation that spreads through social media widely and continually [2]. Scholars have already documented that misinformation and fake news are prevailing in social media [3] and how social media become the hosts of uncontrolled 'erroneous and misleading information that deviates from scientific consensus' [4]. Thus, a particularly critical question arises, regarding to what extend public trusts science and how this can be studied from the perspective of social media. In the general context of trust, social network analysis does not reveal much about the motivations for trust within these networks [5]. On the other hand, there are studies that approach the topic of trust in science through social media studying specific cases while the topic of COVID-19 has been investigated more than any other topics in the context of conspiracy theories. In general, public health topics are at the top of the list of conspiracy theories, with the most publications, and among them are studies on social media use, conspiracy beliefs and health-protective behaviours [6].

Considering the above, it is extremely important to investigate social media in relation to trust in science examining to what extend trust in science is affected by misinformation studying individuals' online behavior. The present work analyses a dataset related to vaccination against COVID-19 with multiple methodologies, investigating the network of the users, the shared content and the interaction of the users aiming to derive insights about the spread of misinformation and what make them appealing. Deep learning algorithms were also applied to examine people's sentiment

This work was supported by European Union in the context of HORIZON EUROPE Grant agreement ID: 101058623. UK participants in Horizon Europe Project VERITY are supported by UKRI grant number 10039826.

2836-3795/25/$31.00 ©2025 IEEE
DOI 10.1109/COMPSAC65507.2025.00034

and subjectivity regarding the vaccination. The present study responds to emerging calls for more nuanced approaches to examining public trust in science, moving beyond traditional survey methods toward multiple dynamic methodologies that capture the complexity and evolving nature of public's attitudes towards science [7].

Definitely, social media accounts with a large number of followers spread the messages to a wide audience. But is this in itself representative of influence on a broad audience? What are the factors that can influence the audience? How can artificial intelligence contribute to this? There are deep learning algorithms that can detect audience sentiment through messages on social networks and also clarify the objectivity of messages. And in times of crisis, such as that of the COVID-19 period, it is certain that people often expressed themselves with strong emotion. This will be explored below giving more specific results about the extent to which positive or negative sentiment was detected, the extent to which people expressed objectivity or subjectivity and revealing the topics and persons of high interest.

II. Related Studies

Trust is science has been investigated by researchers from different perspectives. Research on the trustworthiness of traditional media use started in the 1980s [8]. Later it has focused on the topic of trust in science in relation to social media usage [9], including a plethora of research papers and surveys focusing on the topic of fake news on social media [10], [11], [12]. Although there is an extended literature about the relation of traditional media exposure and trust in science [13], the studies that examined the connection between trust in science and social media are fewer. According to Huber et. al [9] the question of whether social media supports or hinders trust in science becomes one of the key questions of the literature on public attitudes towards science. On one's hand social media provide the instant access to scientific information on the other hand due to the wide spread of fake news people become more susceptible to misinformation.

The specific issue of misinformation in the scientific domain concerned researchers implementing surveys using different methodological approaches including cross-country [14], [15] and longitudinal [16] surveys revealing that levels of trust in science vary among different countries and change across time respectively. Specifically, the study of Jennings et al. [14] proves that there is a positive relationship between trust in government and vaccine willingness while the longitudinal survey in Germany [16] indicated that trust in science increased substantially after the COVID-19 pandemic began and slightly decreased in the months later. This variation, according to the researchers, is due to expectations about how political institutions should handle the pandemic.

Except of the surveys conducted with questionnaires, the issue of trust in science is being examined using data from the social media platforms where people express their opinion, among other things, on scientific issues. As people in EU use the social media as the second source for scientific information after TV [17], the need for examining the trust in science through social media is imperative. This necessity becomes even more critical as OSNs are not just part of our daily lives, but play a very important role in periods of crises, when it is crucial to filter out trustworthy information from the misinformation messages that may hinder response efforts [18]. Studies examining trust in science using data from OSNs have focused on specific use cases with the COVID-19 pandemic case dominating the research.

A study [19] using data from the platform of Twitter identified that the major factor affecting trust in science in the case of COVID-19 was the political ideology as anti-vaccination was supported by people of specific political ideology. A similar conclusion was drawn from the research of Indiana University [20]. The researchers mined data from Twitter related to COVID-19 and inferred that in USA the states with more Republicans present higher percentages of vaccine refusal than Democratic states. The same study shows that sources with high credibility appear to have greater acceptance, nevertheless sources with low credibility have a significant impact on the public. The COVID-19 case was also studied by Ahmed et al. [21] with data from Twitter related to the hashtag #5GCoronavirus conducting a social network analysis. The biggest group of the network consists of accounts endorsing the connection between 5G and Coronavirus and among them there are accounts of influencers, celebrities, news accounts, and journalists. It is significant that the top list of the accounts didn't include any account combating misinformation. The health sector seems to be often connected to conspiracy theories as in the case of Zika virus as well. Researchers studied data from Reddit platform revealing that the platform conspiracy theories are supported by a critical mass of the network [22]. Also, another study investigating data from three different online social platforms about three different health topics confirms the same finding, while vaccination topic was related to the most misinformation messages [23].

From the perspective of time-based data monitoring, a study conducted using Twitter data from Latin American countries revealed that the spread of COVID-19 in 5 countries was related to the flow of posts in social media. The time flow of the data within the graph coincides with the spread of the virus in these countries [24]. Another time-based analysis of posts from Twitter related to COVID-19 showed that politically motivated actors spread misinformation hindering the communication efforts by the public health agencies [25]. The researchers of this study recommend Social Network Analysis (SNA) as a best practise in risk communication in social media. The same methodology of SNA was conducted to investigate the topic of mandatory vaccination against COVID-19 [26]. This study revealed that the two biggest groups of the network were those with the two opposites opinions. The pro-vaccination group was larger than the anti-vaccination group by 2.7%. The most influential accounts of those groups were political persons or organisations.

According to the literature, the methodologies used produce different conclusions regarding trust in science. This study aims to take a multi-perspective approach and provide answers to the following questions:

- What is the benefit of each methodology?
- To what extent is public trust affected through social networks and what characteristics do misinformation messages seem to have?
- Who seems to influence the public the most? Does the political factor seem to influence more than others?

- What conclusions are drawn by applying deep learning algorithms?
- How do users behave on OSNs regarding misinformation?

III. Dataset and Information Collection

The conducted analysis was based on a public available dataset created by DeVerna et al. [20] which contains tweets in English language from January 2021 until February 2023 collected using specific hashtags related to COVID-19 vaccination. The only information related to the origin of the tweet in the dataset is the tweet ID and any other information had to be mined from the Twitter/X[1] API. For this reason, we conducted extraction of the data needed using the free API of Twitter/X until this was available[2]. Using the API we collected data of the tweet IDs such as the text of the message, the number of likes and reshares, the date and the external (to the social media) source. In the rest of the paper, by the term "external source" we will refer to any media or information not created within the context of use of the social media, but rather referenced in a post sharing a web article. Intending to first study people's behaviour in the beginning of the vaccination process we focused on the data collection during the first month of vaccination and then we collected data during the total time. The analysis which follows describes the results of information processing over two periods. The 1st period refers to data mined from January and February of 2021, while the 2nd period refers to data from the period January 2021 – February 2023. Data was collected randomly during these time periods, but we conducted equal requests to the API for each day of the dataset in order to ensure a wide time range of the data and a better time representativeness. In order to investigate the credibility of the messages (using as a criterion the credibility of the external source), all data from the 1st period include tweets with external source, as the criterion of the existence of an external source was set as a necessary prerequisite during the data mining. About 40% of requests to API returned successful results, as many messages were either deleted, or the users no longer existed or suspended, or the messages may have existed but did not contain an external source. The criterion of the existence of external source was removed for the data mining of the 2nd period, given the small amount of data returned with external source from the previous data extraction and in addition to the intention of applying sentiment and objectivity analysis to the wide 2nd period messages independently of the existence of an external source. In the next section follows a numerical analysis of this data.

IV. Assessing the Credibility of External Sources

The credibility of the external sources can be assessed by fact checkers such as Media Bias/Fact Check (MBFC) website[3] that provides an evaluation of the webpage news. In our work presented in this paper, we have relied on the use of fact checkers and in particular the list of unreliable and reliable sources used by DeVerna et al. [20] based on the Iffy index[4]. The Iffy index is a list of untrustworthy sources that was based on the MBFC. The external sources of the dataset were categorised as trustworthy or untrustworthy according to their credibility that MBFC and Iffy index provide. Sources that were not assessed or had an equivalent number of reliable and unreliable articles, were not categorised and thus, they were ignored.

In details, the 1st period of 4,974 tweets (all containing URLs), 3,590 tweets contain trustworthy sources while there are 1,181 tweets containing untrustworthy sources. 203 tweets contain a source that is listed as mixed. In the case of the 2nd period, the collected 9,665 tweets include 526 tweets that contain URLs of which 357 tweets containing trustworthy sources, 141 tweets containing untrustworthy sources and 28 tweets containing mixed sources (Table 1).

Table 1: Total number of tweets with external sources.

Sources	1st period		2nd period	
Tweets containing Trustworthy sources - % of total	3590	72%	357	68%
Tweets containing Untrustworthy sources - % of total	1181	24%	141	27%
Tweets containing Mixed sources - % of total	203	4%	28	5%

Observing the table above, it is clear that the percentages of trustworthy, untrustworthy and mixed sources remain at the same levels without significant divergences over periods. The ratio between trustworthy and untrustworthy sources is 3:1 (especially regarding the 1st period, which contains a significant number of URLs), which means that by average, a Twitter/X user, by reading 4 messages related to COVID-19 vaccination is exposed to one from an untrustworthy source.

V. Social Network Analysis (SNA)

The dataset was examined using the GEPHI[5] software, a powerful user-friendly tool that provides built-in algorithms, important metrics and dynamic layouts for SNA visualisations. The network to be analysed contains hashtags, users and URLs that are represented by nodes while the edges represent the connection between them, indicating the user's reference to another user, hashtag and URL. The node size reflects the degree indicating how often it is mentioned/shared using different colours to distinguish users, URLs and hashtags aiming to identify the differences between the spread of trustworthy and untrustworthy sources. The colour match is purple for accounts, green for URLs, and orange for hashtags.

[1] The platform of Twitter has been renamed to X; the old name is kept in this document as the analysis was conducted before this change.

[2] In March 2023 Twitter platform terminated any free version of the API

[3] https://mediabiasfactcheck.com/

[4] https://iffy.news/index/

[5] GEPHI is a free and open-source software for visualisation and exploration of graphs and networks.

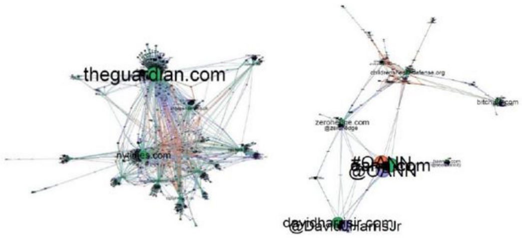

Figure 1: Network graphs for trustworthy (left) and untrustworthy (right) sources of the 1st period data.

The structure of the two graphs is quite similar, with specific URLs being shared primarily, and with official accounts (such as The Guardian and The New York Times) boosting the sharing of those URLs. Sources from reliable accounts are widely shared and their accounts seem to have large nodes in the network resulting in the exposing of their URLs to more users who are more likely to continue sharing these URLs. Another result we can extract from the visualisation of the Fig. 1 is that hashtags do not play as important a role as would be expected from their intended function. Hashtags are user/added flags that allow topics and ideas to be connected and associated, or to emphasise an event or moment in time, and so we might expect hashtags to play a significant role in the structure of the network, which can be measured by the metrics ‘page rank’ and ‘centrality’ which quantify the importance of a node in information flow across a network. However, the hashtags we see in the reliable source networks do not rank amongst the most important nodes by either of these metrics. Rather, the important nodes are mostly website URLs and the associated accounts of those websites. Contrastingly, in the network of unreliable sources, the hashtag ‘#OANN’ is the 4th most significant node by page rank, and 2nd most significant by centrality, with the other important nodes being primarily accounts of individuals rather than websites or the official accounts of websites/outlets. Table 2 contains the top 10 influential nodes in the network of trustworthy sources while table 3 contains the top 10 influential nodes in the network of untrustworthy sources.

Table 2: The 10 most influentual nodes in the network of trustworthy sources, determined by pagerank.

Node	Type	Pagerank	Centrality
theguardian.com	URL	0.067	1.00
nytimes.com	URL	0.033	0.350
independent.co.uk	URL	0.015	0.152
washingtonpost.com	URL	0.014	0.118
dailymail.co.uk	URL	0.013	0.126
cnn.com	URL	0.012	0.096
@guardian	Account	0.012	0.220
reuters.com	URL	0.012	0.163
huffpost.com	URL	0.009	0.062
npr.org	URL	0.009	0.069

Table 3: The 10 most influential nodes in the network of untrustworthy sources, determined by pagerank.

Node	Type	Pagerank	Centrality
@DavidJHarrisJr	Account	0.047	0.390
davidharrisjr.com	URL	0.046	0.384
oann.com	URL	0.040	1.0
#OANN	Hashtag	0.038	0.982
@OANN	Account	0.038	0.980
zerohedge.com	URL	0.023	0.118
bitchute.com	URL	0.022	0.094
@zerohedge	Account	0.018	0.095
hannity.com	URL	0.016	0.065
childrenshealthdefense.org	URL	0.015	0.086

As seen in Fig. 2, it appears that in the trustworthy source networks, hashtags typically connect to related nodes such as relevant URLs or their associated accounts, and can be used to identify particular themes/topics, whereas, in the untrustworthy sources, the hashtags play more of a bridging role, connecting disparate parts of the graph together that would otherwise not be connected by common themes or accounts. The hashtags in the trustworthy network are not labelled as they are not highly connected. Hashtags in the untrustworthy network are labelled, although the labels overlap slightly because they are often very closely affiliated/connected with corresponding network nodes (URLs or accounts).

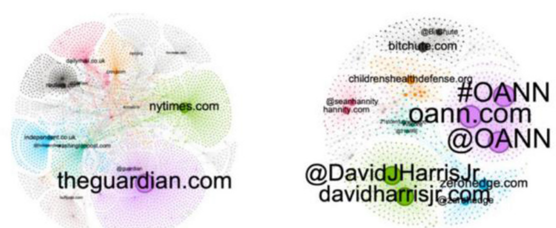

Figure 2: Interactions of Twitter/X for trustworthy (left) and untrustworthy (right) network using Louvain method.

The network analysis was conducted applying the Louvain method as well for the identification of possible communities producing the graphs of the Fig. 2. Nodes have been coloured to indicate modularity classes as inferred by Louvain’s method. It is observed that in the low-credibility network, there are more visible account names as well as URLs.
By using a classification of the URLs shared in the tweets as ‘trustworthy and ‘untrustworthy’ (containing external sources evaluated as trustworthy and untrustworthy respectively) we can build two networks reflecting the topology of social media discussions around trustworthy and untrustworthy information sources. Regarding the first network of ‘trustworthy’, nodes with the highest degree of credibility are highlighted, and these are predominantly news

sites, including ‘The Guardian’, ‘The New York Times’ and ‘Independent’. A metric that measures the structure of a network that can be divided into subgroups is the modularity score. The higher the score, the denser the groups of the networks are. The trustworthy network has a modularity score of 0.791, indicating that the network can be partitioned fairly effectively into non-overlapping communities. For the untrustworthy network the account ‘DavidJHarrisJr’ (conservative speaker, writer and influencer) and ‘OANN’ (OneAmericaNewsNetwork) seem to be the most influential. The modularity score of this network is 0.811, a slightly higher score than the network for lower credibility sources, indicating less connectivity or overlap between subcommunities. From this, we can infer that those users engaging with low-credibility sites appear less likely to interact with or encounter information from multiple sources.

VI. Analysis of the Messages’ Content

Investigating the content of each message is obviously not feasible due to the high volume of messages, but a word cloud can give an overall picture of the content and display terms that appear most frequently in the set of messages. The visualization of the word clouds in this dataset shows what concerns the users of the social network the most. By separating the word clouds between trustworthy and untrustworthy messages we can identify specific terms for these two groups (Fig. 3).

Figure 3: Word clouds of trustworthy (left) and untrustworthy (right) messages of the 1st period.

However, the differences are that in the cloud of untrustworthy messages appear terms related to conspiracy theories and accounts that may be related to the spread of fake news. Specifically, the Big Pharma conspiracy theory is one that has caught users as well as specific media such as the ‘OANN’ and ‘The Defender’. The dataset of the 2nd period contains a much smaller number of external sources and the comparison of the two word clouds of trustworthy and untrustworthy messages shows no particular effect which demonstrates the need for a large number of datasets (Fig. 4).

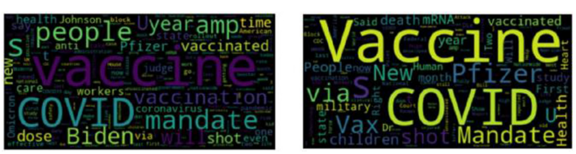

Figure 4: Word clouds of the trustworthy (left) and untrustworthy (right) messages of the 2nd period.

VII. Reactions to Messages

Sorting the data based on the number of likes and the number of shares reveals to us the messages that had the greatest acceptance. These are listed in the table 4 for the data of the 1st period.

Table 4: Sources with the highest reactions of the 1st period.

External Source (Trustworthy)	Likes Retweets	External Source (Untrustworthy)	Likes Retweets
Washingtonpost.com	5,136	OANN.com	11,018
Theguardian.com	1,413	Hannity.com	2,137
Washingtonpost.com	333	Vaccineimpact.com	523
Theguardian.com	307	Zerohedge.com	182
Cnn.com	255	Bitchute.com	161

The table above reveals the messages with the external source that had the highest impact separately for trustworthy and untrustworthy messages. In both cases, there is a specific tweet that has the greatest acceptance, also there is a second message of lower but significant acceptance and then messages with really lower number of likes and reshares. What is clear is that unreliable messages at the beginning of the vaccination period in this dataset seem to have a greater effect. Identifying the content of the posts, the most impactful trustworthy message states ‘Vaccine nationalism is dangerous, only cooperation can stop the virus’ and refers to an article in The Washington Post about the vaccination in poor countries. The most impactful post with untrustworthy source states ‘Biden oversees 40K COVID deaths in 10 days, loses 20M vaccine doses’ is a tweet of the OANN (One America News). The article in the web page of OANN is deleted. The analysis of the 2nd period is summarized in the table 5.

Table 5: Sources with the highest reactions of the 2nd period

External Source (Trustworthy)	Likes & Retweets	External Source (Untrustworthy)	Likes & Retweets
Newyorktimes.com	139,865	childrenshealthdefense.org	105,168
Npr.org	69,666	childrenshealthdefense.org	46,765
newsweek.com	37,950	childrenshealthdefense.org	26,419
Newyorktimes.com	27,168	childrenshealthdefense.org	24,410
Cdc.gov	32,559	childrenshealthdefense.org	24,428

The trustworthy message of NewYorkTimes with the highest impact is a post of a journalist with not so many followers (approx. 20k) containing information about a 66-year-old researcher who contributed in the mRNA vaccines. On the other hand, the message of untrustworthy source with high impact was about a state of Bill Gates that seems to refer to the mRNA vaccine and its complete ineffectiveness. The main difference between these two posts is that the trustworthy source refers to an unknown personality to the general public making clear state in favor of the scientist, while the untrustworthy source is related to a very popular name.

VIII. Applying Sentiment and Subjectivity Analysis

A. Sentiment Analysis

The sentiment analysis was based on a RoBERT a model developed by Laskar et al. [27] trained with a large dataset and achieving the best performance on NLP tasks on six different datasets. This model was used for inference regarding the sentiment analysis since it performed best in a comparative study. The messages from the COVID-19 dataset were processed by the model assigning an extra parameter related to the sentiment analysis. This parameter waives from -1 to 1, where the value 1 represents the positive sentiment, -1 the negative sentiment and 0 the neutral sentiment. Visualising the sentiment analysis score over the time it is noteworthy that the average sentiment waives between 0 and 0.5 revealing that public's opinion had a slightly positive attitude towards the vaccination. The Fig. 5 and 6 are related to the first and the second period respectively noticing that most of the messages have a slightly positive sentiment. The sentiment score of each day is computed as the average value of the scores of the individual messages and the size of the dots represents the number of messages. Looking more closely the first period, it is obvious that there is only one dot with a slight negative value and a significant number of messages. Examining the dataset on this time where the negative value was detected, we found a specific tweet that was widely spread from a source known for conspiracy theories dissemination. The content of the tweet was that a person died hours after being vaccinated. At the time we located this tweet (November 2023) there were 141 comments, 964 shares and 1,200 likes. It is worth mentioning that the tweet was online while the source web article was deleted from the news webpage. Certainly, the existence of a tweet containing a website article is not dependent on the article itself. But as long as the deletion of the articles is not accompanied by the deletion of the tweets, the dissemination of fake news can continue on online social networks.

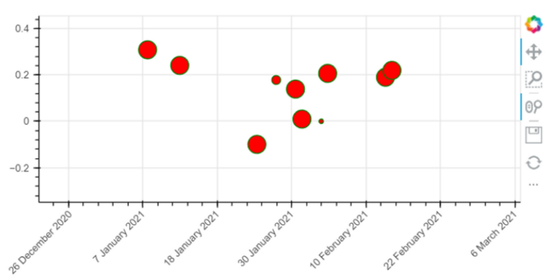

Figure 5: Sentiment analysis of the 1st period.

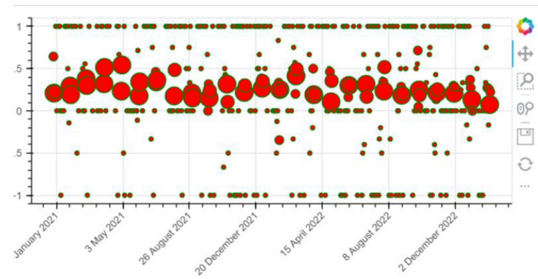

Figure 6: Sentiment analysis of the 2nd period.

B. Subjectivity Analysis

The subjectivity analysis was based on a model developed by Kasnesis et al. [28], that was trained using 5,000 objective and 5,000 subjective sentences achieving the best performance among 11 different models. The messages of the dataset were assessed by the model which provided a value between 0 and 1, where 0 corresponds to an objective opinion and 1 corresponds to a subjective opinion. The visualisation of the two periods provides the plots of the Fig. 7 and 8, where we can notice that the majority of the messages express an objective attitude. Analytically, excluding the messages with a value between 0.40 and 0.60 considering that the attitude is not clear for these messages, the provided results in the table 4 reveal that approximately the 75% of the messages express an objective opinion while approximately 20% of the messages express a more subjective stance.

Table 6: Subjectivity analysis of 1st and 2nd period.

Type of information	1st period data	2nd period data
Objective	77.31%	74.91%
Subjective	18.32%	20.99%
Non-strong subj/obj	4.37%	4.10%

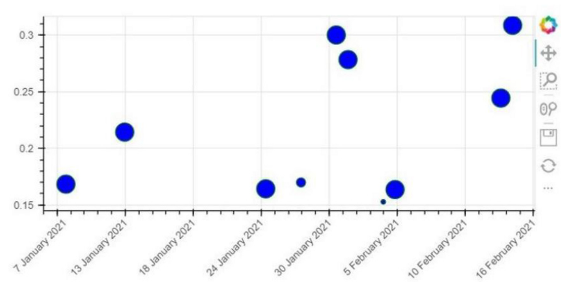

Figure 7: Subjectivity analysis of the 1st period.

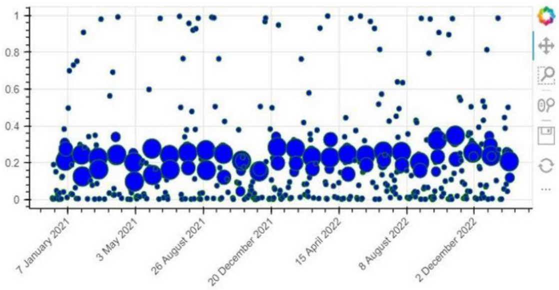

Figure 8: Subjectivity analysis of the 2nd period.

IX. DISCUSSION

The study of data from social networks can be carried out using various methods and produce quite enlightening results. In the case of the public's trust in science, the case of the vaccination against COVID-19 was studied through the lens of social network analysis methods, analysis of the content of messages as well as reactions to them, and deep learning analysis. Using multiple methodologies allows us to examine social media from different perspectives, yielding insights that no single approach could provide. For example, the results of the credibility evaluation of the posts to trustworthy and untrustworthy reveal differences in the users' behaviour in different time periods while the content analysis showed specifically what engaged the public and with which accounts it was associated. Sentiment analysis revealed specific news that negatively affected the world while social network analysis showed a tendency for users who engage with untrustworthy sources to be in more isolated networks as opposed to users who choose trusted sources and choose more news sources. The analysis of the messages' reaction led to a very important finding, that of all posts containing web articles 25% of them were related to untrustworthy sources. From the social network analysis, we can observe that users engaging with untrustworthy sources appear slightly more clustered around single sources and use hashtags that reference specific conspiracy theories or social anxieties to direct traffic towards these sites. Whilst trustworthy sources appear to be politically non-partisan or centre-leaning, the influencers in the untrustworthy network appear to be predominantly conservative and feature more individual personalities such as pundits and commentators.

Our findings show that misinformation and conspiracy theories are widespread in social media, as a quarter of the tweets examined contained web articles from untrustworthy sources. The findings of the subjectivity analysis provide further evidence for this, showing that 20% of the messages in tweets expressed subjective, rather than objective stances. Notably, untrustworthy messages regarding COVID-19 tended to combine unrelated and political information – for example, Iranian appeared as a frequent word in untrustworthy messages as revealed by the content analysis. This finding is consistent with prior research showing that conspiracy theories tend to seek patterns between unrelated issues [29].

Furthermore, our findings suggest that inaccurate or conspiratorial messages have the potential to spread concerns online. Our analysis showed that even a single tweet with conspiratorial content was widely spread online – with 964 shares and 1,200 likes – and succeeded in propagating negative feelings online. Of particular concern is the finding that the detrimental effect of the conspiratorial message continued even after the retraction of the web paper which originally contained these claims, as the tweets mentioning it were not retracted.

Our finding that users engaging with low-credibility sites appear less likely to interact with multiple sources, is consistent with other research findings using different methodology. For example, Wineburg and McGrew [30] found that professional fact checkers interact with multiple sources to verify the credibility of an information, what they called lateral reading, rather than interacting with the same source as novices do.

Overall, the present study showing that a quarter of the messages in social media involve conspiratorial or misinformation highlight the need for citizens to develop the skills to discern misinformation from accurate information. The different pattern of interaction between those who interact with reliable vs. unreliable sources, indicates that particular strategies are related to trustworthy messages showing the potential role of education in helping individuals develop these strategies – such as lateral reading [30] – which are related with discerning reliable from unreliable messages. Our findings provide further support to different stakeholder voices who view education and regulation as the key to having more ethical social networks and emerging technologies [31].

Ultimately, the multiple methodology followed in this work seems to address the questions raised regarding the individuals' online behavior and trust in science. The widespread dissemination of information does not depend solely on the number of followers but on the content of the message. The example of the widespread dissemination of a message with scientific content from an account that does not have a particularly large number of followers is typical. It is also noteworthy that the content of messages from unreliable sources involves names widely known in the political and business world. Deep learning algorithms can indeed detect messages that evoke strong emotion, as was seen in the case of a message from an unreliable source with the news of a man's death hours after his vaccination while subjectivity analysis revealed the general stance of users who expressed themselves primarily objectively rather than subjectively.

The above shows how important it is to research social network data as it produces results that can be forwarded to policy makers to adopt policies that can strengthen trust in science as project VERITY aims to do. A very critical recommendation to policy makers is the adoption of fact-checking mechanisms and regulations from social media to address misinformation, but also it is crucial to empower individuals with digital literacy and critical evaluation skills to counter misinformation with intervention strategies to the society.

X. References

[1] European Commission. Directorate General for Communication. (2021). Citizens' knowledge, perceptions, values and expectations of science: Report. Publications Office. https://data.europa.eu/doi/10.2775/071577.

[2] Mousoulidou, M., Christodoulou, A., Argyrides, M., Siakalli, M., & Constantinou, L. (2022). Trust in Science and COVID-19. Encyclopedia, 2(1), Article 1. https://doi.org/10.3390/encyclopedia2010040.

[3] Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211.

[4] Allgaier, J. (2016). Science on YouTube: What users find when they search for climate science and climate manipulation. https://doi.org/10.48550/arXiv.1602.02692.

[5] Adam, M.B., & Donelson, A., (2022), Trust is the engine of change: A conceptual model for trust building in health systems, System Research and Behavioral Science, Vol.39, Issue 1, pp.116-127, 2022, https://doi.org/10.1002/sres.2766.

[6] Mahl, D., Schäfer, M.S., & Zeng, J., (2023). Conspiracy theories in online environments: An interdisciplinary literature review and agenda for future research, New Media & Society, 25(7), 1781–1801. https://doi.org/10.1177/14614448221075759.

[7] Iordanou , K., Ravn, T., & Zwart, H. (Eds.) (2025). Trust in Science. Springer Nature. Springer Briefs in Research and Innovation Governance.

[8] Durant, J., Evans, G., & Thomas, G. (1989). The Public Understanding of Science. Nature. 340. 11–14. 10.1038/340011a0.

[9] Huber, B., Barnidge, M., Gil de Zúñiga, H., & Liu, J. (2019). Fostering public trust in science: The role of social media. Public Understanding of Science, 28(7), 759–777. https://doi.org/10.1177/0963662519869097.

[10] Pierri, F., & Ceri, S., (2017). False News On Social Media: A Data-Driven Survey, ACM SIGMOD Record, 48(2), pp 18–27. https://doi.org/10.1145/3377330.3377334. "Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009." U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2.

[11] Kondamudi, M.R., Sahoo, S.R., Chouhan, L., & Yadav, N., (2023). A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches, Journal of King Saud University - Computer and Information Sciences, 35 (6), 101571. https://doi.org/10.1016/j.jksuci.2023.101571.

[12] Meel P., & D.K. Vishwakarma, D.K., (2020), Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities, Expert Systems with Applications, 153. https://doi.org/10.1016/j.eswa.2019.112986.

[13] Anderson, A., Scheufele, D., Brossard, D., & Corley, E. (2011). The Role of Media and Deference to Scientific Authority in Cultivating Trust in Sources of Information about Emerging Technologies. International Journal of Public Opinion Research (Advance Online Publication), 24. https://doi.org/10.1093/ijpor/edr032.

[14] Jennings, W., Valgarðsson, V., McKay, L., Stoker, G., Mello, E., & Baniamin, H. M. (2023). Trust and vaccine hesitancy during the COVID-19 pandemic: A cross-national analysis. Vaccine: X, 14, 100299. https://doi.org/10.1016/j.jvacx.2023.100299.

[15] Lindholt, M. F., Jørgensen, F., Bor, A., & Petersen, M. B. (2021). Public acceptance of COVID-19 vaccines: Cross-national evidence on levels and individual-level predictors using observational data. BMJ Open, 11(6), e048172. https://doi.org/10.1136/bmjopen-2020-048172

[16] Bromme, R., Mede, N. G., Thomm, E., Kremer, B., & Ziegler, R. (2022). An anchor in troubled times: Trust in science before and within the COVID-19 pandemic. PLOS ONE, 17(2), e0262823. https://doi.org/10.1371/journal.pone.0262823.

[17] European Commission. Directorate General for Communication. (2021). Citizens' knowledge, perceptions, values and expectations of science: Report. Publications Office. https://data.europa.eu/doi/10.2775/071577.

[18] Hagar, C., (2013). Crisis Informatics: Perspectives of Trust – Is social media a Mixed Blessing?, School of Information Student Research Journal, Vol.2, Issue 2, 2013, https://doi.org/10.31979/2575-2499.020202.

[19] Thelwall, M., Kousha, K., & Thelwall, S., (2021). COVID-19 vaccine hesitancy on English-language Twitter, Information Professional , 30 (2). https://doi.org/10.3145/epi.2021.mar.12.

[20] DeVerna, M.R., Pierri, F., Truong, B.T., Bollenbacher, J., Axelrod, D., Loynes, N., Torres-Lugo, C., Yang, K.-C., Menczer, F., & Bryden, J., (2021). CoVaxxy: A Collection of English-Language Twitter Posts About COVID-19 Vaccines, Proceedings of the International AAAI Conference on Web and social media, 15(1), 992-999. https://doi.org/10.1609/icwsm.v15i1.18122.

[21] Ahmed, W., Vidal-Alaball, J., Downing, J., & López Seguí, F., (2020). COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data, Journal of medical Internet research, 22(5), e19458. https://doi.org/10.2196/19458.

[22] Kou, Y., Gui, X., Chen, Y., & Pine, K., (2017). Conspiracy Talk on Social Media: Collective Sensemaking during a Public Health Crisis, Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), pp 1–21. https://doi.org/10.1145/3134696.

[23] Pulido, C., Ruiz-Eugenio, L., Redondo-Sama, G., & Villarejo-Carballido, B., (2020). A New Application of Social Impact in Social Media for Overcoming Fake News in Health, International Journal of Environmental Research and Public Health, 17(7), p. 2430. doi: 10.3390/ijerph17072430.

[24] Ceron, W., Gruszynski Sanseverino, G., de-Lima-Santos, M. F., & Quiles, M. G. (2021). COVID-19 fake news diffusion across Latin America. Social network analysis and mining, 11(1), 47. https://doi.org/10.1007/s13278-021-00753-z.

[25] Pascual-Ferrá, P., Alperstein, N., & Barnett, D.J., (2022). Social Network Analysis of COVID-19 Public Discourse on Twitter: Implications for Risk Communication, Disaster Med Public Health Prep. 16(2):561-569. doi:10.1017/dmp.2020.347.

[26] Olszowski, R., Zabdyr-Jamróz, M., Baran, S., Pięta, P., & Ahmed, W., (2022). A Social Network Analysis of Tweets Related to Mandatory COVID-19 Vaccination in Poland, Vaccines, 10(5), p. 750, doi: 10.3390/vaccines10050750.

[27] Laskar, M.T.R., Huang, J.X., & Hoque, E., (2020). Contextualized Embeddings based Transformer Encoder for Sentence Similarity Modeling in Answer Selection Task. In Proceedings of the Twelfth Language Resources and Evaluation Conference, pages 5505–5514, Marseille, France. European Language Resources Association.

[28] Kasnesis, P., Toumanidis, L., & Patrikakis, C. Z. (2021). Combating Fake News with Transformers: A Comparative Analysis of Stance Detection and Subjectivity Analysis. Information, 12(10), 409. https://doi.org/10.3390/info12100409.

[29] van Prooijen, J. W., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. European Journal of Social Psychology, 48(7), 897-908. https://doi.org/10.1002/ejsp.2530

[30] Wineburg, S., & McGrew, S. (2019). Lateral reading: Reading less and learning more when evaluating digital information. Teachers College Record, 121(1), 1-40.

[31] Iordanou, K., & Antoniou, J. (2023). Tackling Bias in AI and Promoting Responsible Research and Innovation: Insights from Discussions with Different Stakeholders. In AI and Society (pp. 141-157). Chapman and Hall/CRC.