Technology in Society 68 (2022) 101930

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

# People lie, actions Don't! Modeling infodemic proliferation predictors among social media users

Chahat Raj, Priyanka Meel *

*Department of Information Technology, Delhi Technological University, India*

ARTICLE INFO

*Keywords:*
Factor identification
Fake news
COVID-19
Misinformation
Infodemic
Modeling predictors

ABSTRACT

Social media is interactive, and interaction brings misinformation. With the growing amount of user-generated data, fake news on online platforms has become much more frequent since the arrival of social networks. Now and then, an event occurs and becomes the topic of discussion, generating and propagating false information. Existing literature studying fake news elaborates primarily on fake news classification models. Approaches exploring fake news characteristics to distinguish it from real news are minimal. Not much research has focused on statistical testing and generating new factor discoveries. This study assumes fifteen hypotheses to identify factors exhibiting a relationship with fake news. We perform the experiments on two real-world COVID-19 datasets using qualitative and quantitative testing methods. We determine the impact of conditional effects among sentiment, gender, and media usage. This study concludes that sentiment polarity and gender can significantly identify fake news. Dependence on the presence of visual media is, however, inconclusive. Additionally, Twitter-specific user engagement factors like followers count, friends count, favorite count, and retweet count significantly differ in fake and real news. Though, the contribution of status count is currently disputed. This study identifies practical factors to be conjunctly utilized in developing fake news detection algorithms.

## 1. Introduction

COVID-19 spread worldwide faster than a human brain could imagine. People worldwide had hardly heard about it before it became the most fatal. After facing its catastrophic results, many people became aware of the pandemic and started to ponder it. Talks about COVID-19 were everywhere and on everybody's minds and lips. Social media is now an established source to serve people information in one of the easiest ways. Interactions on this hot topic overwhelmed social networking platforms. The pandemic period has demonstrated a higher social media usage than the normal times [1]. Mohammed and Ferraris examine the highly active participation in information sharing among Twitter users during the pandemic [2]. Farooq et al. discuss the impacts of COVID-19 information overload among social media users [3]. The internet is flooded with various types of information. However, not everything that is on the internet is reliable. Information on social media is merely peoples' opinions and has not been validated for credibility. Gradually, most of these talks turned out to be fake news, rumors, conspiracies, misinformation, and disinformation. With the feasibility of posting, sharing, and accessing the information on the web, users can be quickly confounded with fake news. The desks of politicians and public figures made the maiden attempt to spread fake news worldwide, misleading people. As a result of conspiracy theories, 5G towers in the United Kingdom turned into ruins. Fake news exuded harmful political, social, religious, technological, and environmental changes around the globe, generating a sense of distrust among people. Enmity also started grasping its enclosures as people claimed China to be the most causative element in spreading coronavirus. Detection of such malignant talks is one of the greatest needs to prevent society from antisocial online behavior and its impacts.

People are willing participants in the crisis information-sharing process on Twitter [2]. Fake news about the pandemic targets various dimensions of society. One of these fake news is the remedial claims for the coronavirus disease. "A pinch of turmeric or a drop of garlic juice could cure the fatal" was amongst the most prevailing unauthentic fake remedies. Poor perceptions, unproven methods, illogical claims, false figures, and alarming news overwhelmed the global information scenario. Social media platforms are well known for spreading misinformation and denying scientific literature [4]. False social media posts have also tricked users into relying on harmful and poisonous substances

* Corresponding author.
*E-mail addresses:* chahatraj58@gmail.com (C. Raj), priyankameel86@gmail.com (P. Meel).

https://doi.org/10.1016/j.techsoc.2022.101930
Received 4 August 2021; Received in revised form 6 February 2022; Accepted 9 February 2022
Available online 14 February 2022
0160-791X/

like weed, cannabis, and ethanol intake [5]. The rapid evolution of the COVID-19 pandemic has not permitted immediate and specific scientific data [6]. COVID-19 is not the only fake news generating event. Many instancesed to colossal misinformation spread on online social networks, such as the 2016 US presidential elections, Pizzagate, and hurricane Harvey [7]. COVID-19 is one major event generating misinformation on a larger scale than any other event. The term “Infodemic,” coined by David Rothkopf in 2003, refers to the mass propagation of false information that spreads among people like an epidemic. It was first used to denote the widespread misinformation, disinformation, and rumors spreading amidst the SARS epidemic. The term has regained extensive usage in the times of COVID-19.

Previous research has contributed diversely to solving the fake news problem. Researchers from behavioral sciences have covered the factors involved in sharing and accepting fake news [8–10]. Others have investigated several factors like user demographics and background information [11]. Lian et al. proposed real scenario, embedment scenario, and isolation scenario strategies to tackle fake news in similar disaster-like situations [12]. Many studies have developed fake news detection algorithms [13,14]. Such algorithms widely utilize news content, such as linguistic features, visual features, and network features. However, there is an absence of ideal classifiers, while most fake news characteristics are unidentified. In this study, we explore two research questions:

**RQ1**. Which factors significantly explain false news sharing behavior on social media?

**RQ2**. Which user-specific and content-specific characteristics demonstrate biases in real and fake news?

This paper identifies several key factors associated with fake and real news on Twitter. We formulate fifteen hypotheses on the key elements and their direct and mediating relationship with fake news. These hypotheses are evaluated on two real-world datasets which contain tweets related to the COVID-19 pandemic. MediaEval 2020 [15] is a benchmark dataset containing tweets pertaining to coronavirus and 5G conspiracy. CovidHeRA [16] is a collection of tweets associated with spreading health-related misinformation amidst the pandemic. This paper contributes to analyzing characteristics that differentiate between fake and real news. We analyze the following key factors: sentiment polarity, gender, media usage, follower count, friends count, status count, retweet count, and favorite count. Interdependence of factors called “conditional effects” among sentiment polarity, gender, and media usage is studied intensely. We also extend the work of Parikh et al. [17] by demonstrating the relationship between fake news and particular sentiment polarities. This paper comes up with exciting outcomes suggesting important features demonstrating fake news dependence. The research bridges existing gaps in the literature and forms the basis for a new direction in event-specific fake news analysis. Our hypotheses shall help develop efficient fake news detection algorithms covering many fake news components.

The contributions of this article are as follows:

1. We examine the independent bias of sentiment polarities, gender, and media usage towards fake news.
2. We examine the conditional effects among sentiment, gender, and media usage to understand their significance in fake news sharing behavior and highlight some unique insights.
3. We perform a quantitative analysis on five social media engagement attributes, follower count, friends count, status count, retweet count, and favorite count. Through experiments on two real-world coronavirus-specific datasets, we dispute existing research results and compare the insights between existing studies and our findings.

The organization of this paper is as follows: Section 2 studies the existing literature in fake news and the COVID-19 infodemic. The survey provides insights into existing hypotheses and conclusions drawn upon fake news. Section 3 presents the research methodology explaining the datasets used and the motivation for the characteristics assumed in this study. Section 4 covers the results obtained by performing statistical tests on the datasets. Section 5 describes the insights drawn from the results and summarizes the acceptance/rejection of formulated hypotheses. We compare the existing research with the findings of this study. Section 6 concludes the paper by discussing future directions.

## 2. Literature review

The menace of fake news is challenging for information consumers. It has constantly been a topic of concern in the research society. Various studies have proposed identifying and detecting fake news on online social networks [18]. Past studies have focused on a vast dimension of fake news ranging from origin, propagation, consumption, and impact [19]. Several solutions have been proposed to detect fake news with the help of exploiting textual [20], visual [21], and nodal features [22]. In contrast, studies pertaining to hypothesis formulation and testing are very few. There is limited literature discussing the latest trends in online social networks highlighting vulnerabilities in fake news propagation and consumption. It is essential to formulate and discover dependent dimensions of fake news.

Some studies have proposed important insights beneficial for fake news detection. For instance, Parikh et al. proposed hypotheses discussing the origin, proliferation, and tone of fake information [17]. They concluded that such misleading information is published more on lesser-known websites than the popular ones. Unverified users are often shared on social media than by verified accounts in terms of proliferation or sharing. They also demonstrated that fake news has a specific tone or sentiment (positive, negative, or neutral) but did not conclude which type of particular tone is fake news primarily related to. Their study provides ways to form additional hypotheses, which is also a motivation for our work. Zhang et al. [23] demonstrated the importance of dual emotion in fake news detection by establishing that real and fake news emotions are distinctive. They proposed the utilization of both publisher and social emotion, suggesting that these hold emotional resonance or dissonance between each other. Kumari et al. [24] proposed the usage of novelty, emotion, and sentiment features for fake news identification. Their work confirms that novel post content stimulates users to share information, and novel tweets are more likely to be shared than older news. They merge these novelty features with emotional states belonging to negative and positive sentiments into a deep neural network architecture for fake news detection. Their proposed algorithm demonstrates higher classification accuracy with the use of novelty, emotion, and sentiment features compared to the existing state-of-the-art.

As an approach to verify news online, Sethi [25] proposed a social argumentation framework that utilizes crowd-sourcing to perform alternative fact-checking. The community plays a prominent role in this environment by adding arguments and providing trust ratings through critical learning and examination. In another work, Sethi et al. [26] employ crowd-sourced social augmentation with pedagogical agents to build a recommender system for misinformation detection. They extend the principle of analytical and critical thinking in conjunction with users’ emotions and contents’ semantic understanding. Sethi and Rangaraju [27] discuss the ‘backfire effect’ that occurs when users’ opinions counter-intuitively harden upon exposure to facts. Their proposed socially collaborative argumentative system allows users to overcome their subconscious biases against facts by gaining a semantic evidence-based proposition of the statements. Users’ emotional and propositional profiles are modeled together in the framework using sentiments and semantics to counter false user beliefs. Tifferet [28] developed a Verifying Online Information (VOI) self-report scale for users to exploit users’ critical thinking in identifying misinformation.

Linden et al. [29] use the term ‘fake news effect’ to account for the psychological bias in users that discredit unpleasant information sources as fake news. According to their study, such a bias lowers peoples' trust in media and alleviates their belief in conspiracy theories, causing high political bias in users. Demographics and culture form the basis of theories proposed by Rampersad and Althiyabi [11]. They identified the established relationships between age and acceptance of fake news. It was noted that demographics like gender and education played a minor role in fake news acceptance. Another notable hypothesis confirmed that educated people are less likely to accept fake news. It was also observed that culture indirectly impacts the acceptance of fake news significantly. Few works have highlighted the connection between Third Person Effect (TPE) and fake news sharing [30,31]. Brewer et al. have drawn several conclusions towards readers' reactions to consuming fake news [32]. Horne et al. distinguished between real and fake news based on stylistic and psychological features of the text [33]. Bovet and Makse studied the fake news propagation on Twitter during the 2016 US presidential elections and explored its influence [34]. Altay et al. hypothesized the relation between users' reputation and fake news sharing [8]. They studied that very few people were indulged in sharing fake news and identified the causes of such behavior. They concluded that sharing fake news harmed people's reputations and resulted in trust issues, which is a significant reason why very few people indulge in sharing fake news. Osatuyi and Hughes figured that the amount of information available on fake news platforms is lesser than real news [35]. Exploring the role of comments in identifying and rejecting fake news shows that users are less likely to accept fake news if they come across critical comments about the content [36].

Cheng et al. [37] thrust upon the importance of causal inference to understand user attributes causing them to share fake news and their susceptibility to such news. They suggest that such attributes depicting user engagement behavior and online activity can potentially characterize users involved in sharing false news. They identified verified accounts, register time, gender, age, organization, number of statuses, favorites, followers, and friends to be statistically significant in characterizing such users. Shu et al. [38] employed Twitter-based user profile features to understand differing characteristics between real and fake tweets, analyzing profile-related, content-related, network-related, and implicit features. Their work identifies specific attributes that make users more likely to trust a piece of false information than real news. Yang et al. [39] exploit user engagement information into a probabilistic graphical method, developing an unsupervised fake news detection framework. They construct a hierarchical user engagement model that extracts tweets from verified Twitter users and social engagement features (likes, retweets, and replies) of unverified users for these tweets. Fake news identification is performed by extracting user opinions on the collected tweets from their engagement behaviors.

With the outbreak of the COVID-19 pandemic, social media communications and interactions rose to a higher level than before. Researchers approached this problem in an early response to the infodemic by analyzing various concerns and suggesting solutions. Moscadelli et al. [40] investigated the topics about the pandemic most polluted with fake news. Calvillo et al. [41] analyzed political associations with the discerning of fake news. Hypotheses linking the fake news belief structure to its acceptance, Kim and Kim [42] proposed that factors like source credibility, quality of information, receiver's ability, perceived benefit, trust, and knowledge decrease people's belief in fake news. Contrastingly, heuristic information, perceived risk, and stigma strengthen the confidence in fake news. Greene and Murphy [43] have discussed the likeliness of people sharing true or false stories on social media, establishing the association with their knowledge concerns. Another study that links conscience and ideology with infodemic sharing behavior is provided by Lawson and Kakkar [44]. Montesi [45] spread light on the nature of infodemic and suggests that the harm caused by fake news is not health-related but more of a moral sort. The dominant infodemic themes are society, politics, and health. Building constructs over the Third Person Effect (TPE), Lui and Huang [46] have facts regarding the susceptibility and perception of fake news in the pandemic era. Similarly, Laato et al. [47] discussed the factors such as information sharing, information overload, and cyberchondria aiding fake news propagation. Sulaiman [43] proposed no relationship between information evaluation and fake news sharing, experimenting on a Nigerian sample. With many hypotheses, Alvi and Saraswat [48] explored connections amongst various heuristic and systematic factors such as Sharing Motivation, Social Media Fatigue, Feel Good Factors. Fear Of Missing Out, News Characteristics, Extraversion, Conscientiousness, Agreeableness, Neuroticism, Trust, and Openness. Balakrishnan et al. [49] investigated user motives of fake news sharing during the pandemic and reported altruism, ignorance, and entertainment as significant factors. Their study exhibits that availability, pass time, and Fear Of Missing Out factors are insignificant in determining fake news sharing behavior. As observed from the existing literature, past studies revolve around identifying psychological and behavioral factors that demonstrate any relationship with fake news. There is a research gap in characterizing features that could aid in distinguishing false information from real and serve as contributing factors to building fake news detection algorithms.

## 3. Research methodology

### 3.1. Data

This study uses two publicly available benchmark datasets, MediaEval 2020 [15] and CovidHeRA [16]. MediaEval 2020 issued a benchmark dataset for its fake news detection task. The dataset consists of 5842 tweets classified into three classes: 5G coronavirus conspiracy, other conspiracy, and non-conspiracy. The tweets contain real and false information revolving around the COVID-19 pandemic. For this study, we classify these tweets into two coarse classes, with non-conspiracy tweets as real and the remaining tweets as fake. CovidHeRA is another benchmark dataset containing false tweets related to coronavirus and health. These tweets are a collection of fake remedies, preventive measures, treatments, and other health-related information spread across Twitter amidst the pandemic. Originally, the datasets consisted of tweet ids. To procure various characteristics of the tweets, the python library Tweepy is used. The scraping results in providing various information of the tweet and user content. This social engagement information forms the basis of this study. To obtain the gender information of Twitter users, a gender predictor algorithm by Sap et al. [50] is utilized. Sentiments on the dataset are extracted using Microsoft's Text Analytics service. Sentiment scores are returned as values in the range of 0.0–1.0. A score between 0.0 and 0.3 signifies negative, 0.3 to 0.7 represents neutral and 0.7 to 1.0 represents positive sentiment. We utilize the 'extended_entities' column from the scraped datasets for media usage. Sizes of both the datasets pertaining to each category are provided in Tables 1–3.

### 3.2. Research hypotheses

To identify characteristics that distinguish fake news and real news and consequently identify fake news based on these characteristics, we have formulated fifteen hypotheses based on the qualitative and quantitative variables present in the dataset. To identify the dependence of social media misinformation, we identify and analyze eight key elements: sentiment polarity, gender, media usage, follower count, friends

**Table 1**
Count of fake and real items with gender as a category.

| Label | CovidHeRA | | | Mediaeval | | |
|---|---|---|---|---|---|---|
| | Male | Female | Total | Male | Female | Total |
| Fake | 1532 | 772 | 2304 | 929 | 837 | 1766 |
| Real | 42,683 | 40,104 | 82,787 | 2011 | 2065 | 4076 |
| Total | 44,215 | 40,876 | 85,091 | 2940 | 2902 | 5842 |

**Table 2**
Count of fake and real items with sentiment polarity as a category.

| Label | CovidHeRA | | | | Mediaeval | | | |
|---|---|---|---|---|---|---|---|---|
| | Negative | Neutral | Positive | Total | Negative | Neutral | Positive | Total |
| Fake | 1292 | 391 | 621 | 2304 | 1042 | 346 | 378 | 1766 |
| Real | 31,004 | 24,638 | 27,145 | 82,787 | 2320 | 690 | 1066 | 4076 |
| Total | 32,296 | 25,029 | 27,766 | 85,091 | 3362 | 1036 | 1444 | 5842 |

**Table 3**
Count of fake and real items with media usage as a category.

| Label | CovidHeRA | | | Mediaeval | | |
|---|---|---|---|---|---|---|
| | With Media | W/o Media | Total | With Media | W/o Media | Total |
| Fake | 150 | 2154 | 2304 | 289 | 1477 | 1766 |
| Real | 17,700 | 65,087 | 82,787 | 791 | 3285 | 4076 |
| Total | 17,850 | 67,241 | 85,091 | 1080 | 4762 | 5842 |

count, status count, retweet count, and favorite count. We assume that fake news characterization, propagation, and acceptance have a relationship with these factors, which can be consequently utilized in fake news detection. To better understand, each tweet labeled as fake/real in the datasets has specific characteristics mentioned above. It is crucial to examine which feasible aspects demonstrate a relationship with false tweets. We also aim to study any significantly different factors between real and fake tweets. We describe certain features useful for real and fake tweet classification by establishing such relationships. Qualitative hypotheses, $H_A$, $H_B$, and $H_C$ are tested to scrutinize the direct relationships between sentiment, gender, and media usage with fake news, respectively. Further, it is vital to analyze the conditional effects, i.e., if the bias of one independent variable influences the bias of another independent variable. We test whether the higher proportion of one categorical variable contributes to the higher proportion of another categorical variable. To do so, we construct six more qualitative hypotheses, $H_D$, $H_E$, $H_F$, $H_G$, $H_H$, and $H_I$. These nine hypotheses are tested using the Chi-square test of independence. The relationship is demonstrated in Fig. 1. To study quantitative variables, we formulate hypotheses $H_J$ to $H_O$ and perform Analysis of Means on each, also calculating their confidence intervals. Fig. 2 demonstrates the quantitative relationships.

#### 3.2.1. Qualitative factors and hypotheses

**Sentiment:** According to Parikh et al. [17], it is widely assumed that most of the news spreading online is negative in terms of its linguistic tone. However, it has not been proven that fake news has a higher negative polarity than neutral or positive polarities. Parikh et al. [17] noted that it was inconclusive to say if fake news is biased towards a particular polarity. Following their assumption, $H_A$ forms the primary hypothesis to test if fake news exhibits a specific sentiment polarity.

**HA0**. There is no bias in the proportion of different sentiments between fake news and real news.

**HA1**. There is a significant bias in the proportion of different sentiments between fake and real news.

**Gender:** Rampersad and Althiyabi [11], examining a sample of Saudi Arabia, observed that gender has a weakly positive effect on people's acceptance of fake news. Their study suggests that it cannot be determined if males or females are more susceptible to fake news. The sample is specific to a particular demographic region. The authors examine the acceptance trend to understand the belief patterns in males and females. However, we intend to examine the proliferation likeliness for both genders. We suggest that if users post false content, they are more opinionated towards that information and tend to be biased towards accepting more of such misinformation. In our research, datasets consist of tweets from Twitter users across the globe, allowing us to examine the assumptions on a universal scale. Instead of focusing on the effect (fake news acceptance), we plan to examine the origination of fake news by determining which gender is more likely to publish fake news. We test this hypothesis by using $H_B$'s statement to verify a significant relationship between gender and false information.

**HB0**. There is no bias of the gender of users involved in fake news with respect to real news.

**HB1**. There is a significant bias of gender of users involved in fake news with respect to real news.

**Media:** Social media platforms are flooded mainly with multimedia posts such as images, videos, URLs, GIFs, and more. Several fake news detection algorithms have been designed that detect whether a visual media in a piece of fake information is credible or not [13,51–55]. Multi-modal fake news detectors demonstrate the critical importance of

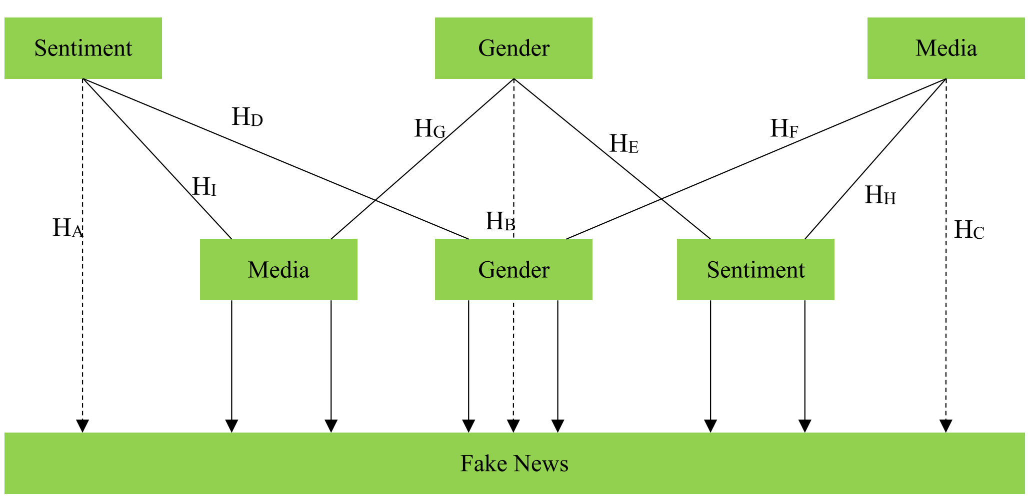


**Fig. 1.** Factors determining fake news (qualitative hypotheses).

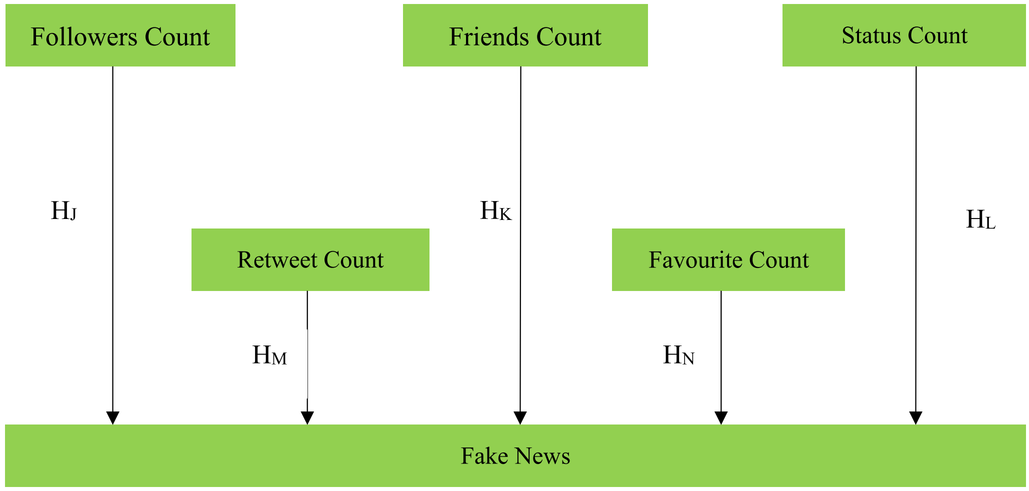


**Fig. 2.** Factors determining fake news (quantitative hypotheses).

visual media consideration for news classification. We analyze whether it can be stated solely based on the presence of visual media that a post/message is false. We aim to discover whether fewer or more users spreading false information utilize visual media to target their victims. We categorize the datasets into two modalities: without and with visual media (pictures/videos). Using the statement $H_C$, we analyze which data modality of social media posts contribute more/demonstrate bias towards misinformation.

**HC0**. There is no bias of visual media usage in fake news with respect to real news.

**HC1**. There is a significant bias of visual media usage in fake news with respect to real news.

Next, we analyze the conditional effects among sentiment, gender, and media to understand the fake news sharing behavior. Based on the above three univariate hypotheses, we decide the mediating relationships among these factors and formulate multivariate hypotheses ($H_D$ to $H_I$) to determine whether bias in one of the above proportions is due to bias in proportions of the other variable.

**HD0**. There is no influence of bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.

**HD1**. There is a significant influence of bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news.

**HE0**. There is no bias in the proportion of a particular sentiment used in fake news between different gender of users.

**HE1**. There is a significant bias in the proportion of a particular sentiment used in fake news between different gender of users.

**HF0**. There is no bias in inducing a particular sentiment with media usage in fake news.

**HF1**. There is a significant bias in inducing a particular sentiment with media usage in fake news.

**HG0**. There is no bias in media usage amongst different sentiments used in fake news.

**HG1**. There is a significant bias in media usage amongst different sentiments used in fake news.

**HH0**. There is no relationship between a particular gender and media usage in fake news.

**HH1**. There is a significant relationship between a particular gender and media usage in fake news.

**HI0**. There is no bias in and media usage in fake news between different gender of users.

**HI1**. There is a significant bias in and usage of media in fake news between different gender of users.

*3.2.2. Quantitative factors and hypotheses*

Using the data scraped from Twitter, we decided on testing our hypotheses on five key social engagement factors, which can be categorized into three user/profile-specific features, i.e., the number of followers, friends, and statuses, and two post-specific features, i.e., retweets count and favorites count. In our approach, we assume that these factors can be utilized to identify tweets' credibility. Moreover, we assume that these factors affect fake news sharing and acceptance.

Followers and friends count determine the reachability of a particular post or message within the user's social network who created it. A retweet is an action of sharing another user's tweet on one's timeline to make it visible to one's followers. Retweet count determines a post's propagation and acceptance behavior by checking the social reach. It is similar to the action "Share" on other social media platforms. It spreads a particular post to the user's social network. The larger the retweet count, the more likely the people reading the post will believe that particular piece of information and further spread it across the web [56]. Status count corresponds to the number of total posts/retweets a specific user has posted since the creation of his account. Favorites are user markings made on a post that other users would like to save for the future.

We determine the relationship between these quantitative social engagement variables and the post's label, i.e., the relationship between the number of retweets and favorites of the post and the followers, friends, and status of the user who posted it, and it being real or fake. Since the source of misinformation can range from a random regular user to a credible account such as commercial news channels, journalists, or celebrities, it becomes difficult to assume any specific range for the count of these quantitative variables. Hence, we test based on a characteristic whether the values of their (quantitative variables) mean and variance within a confidence interval are significantly distinguishable or not. In other words, we determine the probability with which a post or a piece of information under examination can be labeled as fake or real based on the variation of the above-mentioned quantitative variables.

As both the datasets contain an unequal number of fake and real news tweets, we test for 2-variances for each quantitative variable to check for homoscedasticity, that is, if the variables have equal or unequal variances across the two groups, namely, fake and real news. This helps us identify if the two groups are identical with respect to any parameter of these variables. We use Levene's test because the data is highly right skewed with long tails. Further, using $H_O$, we investigate how closely are the user/profile-specific and post-specific quantitative variables associated with each other. We observed a non-linear and/or heteroscedastic linear relationship between pairs of these quantitative variables. Hence, we apply Spearman rank-order correlation to assess the magnitude and direction of association between these variables.

**HJ0.** There is no significant difference in the mean number of followers between the user groups.

**HJ1.** The mean number of followers for the user group posting fake news are significantly different from the user group posting real news.

**HK0.** There is no significant difference in the mean number of friends between the user groups.

**HK1.** The mean number of friends for the user group posting fake news is significantly different from that of the user group.

**HL0.** There is no significant difference in the user groups' mean number of 'Status' posted.

**HL1.** The mean number of 'Status' posted by fake news users is significantly different from those posted real news.

**HM0.** There is no significant difference in the mean number of retweets between real and fake news.

**HM1.** The mean number of retweets on fake news is significantly different from real news.

**HN0.** There is no significant difference in the mean number of favorites between real and fake news tweets.

**HN1.** The mean number of favorite fake news tweets is significantly different from real news tweets.

**HO.** There is a significant effect of fake news sharing behavior on fake news acceptance behavior.

## 4. Results

### 4.1. Qualitative experimental results

To test on the nine hypotheses $H_A$ to $H_I$, which are formed upon the categorical variables, we use the Chi-Square test of independence alongside computing "Cramer's V," "Pearson's r," and "spearman's rho" values. Cramer's V value provides the strength of association between the nominal categorical variables for the conclusion arrived using the Chi-Square test. Its values range between 0 and 1. Pearson's r value signifies both the strength of association and the direction of the association between two continuous variables. Here direction indicates if one variable would increase or decrease with respect to change in another variable. Its values range from −1 to +1, where the value of −1 means that as one variable increases, the other decreases, and +1 means that as one variable increases, the other increases too. A value of 0 indicates no strength of association. Spearman's rho values differ from the outcomes of Pearson's r values by a feature that they can describe the correlation even when the variables do not have a linear association. It is also proof from the long tail of outlier values as it uses the ranks of the values of the variable.

The values in Table 7 include degrees of freedom as df, Chi-Square test value as $\chi^2$, probability value as p-value and Cramer's V value, Pearson's r-value, and Spearman's rho. The first column in this table indicates the hypothesis of the variables and their values. From the first row of the same table, we observe that $\chi^2$ values for testing hypothesis $H_A$ with 2° of freedom (df) for both CovidHeRA and MediaEval datasets are 352.963 and 17.103, respectively, and are more significant than critical value $\chi^2_c = 5.991$ with $p < 0.001$ (Significance level $\alpha = 0.05 = p_c$, critical p-value). This implies a significant difference in proportions of sentiments used between Fake and Real news. However, despite there being a substantial difference in ratios, low values of Cramer's V (less than 0.2), Pearson's r (between −0.20 and + 0.20), and Spearman's rho (between −0.20 and + 0.20) indicate a weak association of label (news being fake or real) and the sentiment (sentiment being negative or neutral or positive). These values (Cramer's V, Person's r, and Spearman's rho) are low for all the hypotheses tested. Therefore, we rely on comparing "actual values" from Tables 1–3 with "expected values" in Tables 4–6, respectively, to determine the association between an independent and a categorical dependent variable, or in other words, the bias of fake news towards a specific or a group of categorical variables. On comparing Tables 2 and 5, we observe that in both CovidHeRA and MediaEval datasets, fake news with negative sentiment has a higher actual count (1292, 1042) with respect to the expected count (874.5, 1016.3), and fake news with positive sentiment has a lower actual count (621, 378) with respect to the expected count (751.8, 436.5). The count of neutral sentiment varies inversely in both the datasets, with CovidHeRA showing a reduced count and MediaEval showing an increase. Similarly, from the same tables, we observe that the actual count of real news with negative sentiment is less than the expected count in both datasets. The actual count of real news with positive sentiment is greater than that of the expected count in both datasets. Therefore, we reject the null hypothesis, $H_{A0,}$ and observe that fake news propagation during CoVID-19 has had a proportional bias towards negative sentiment.

From the second row of Table 7, we observe that $\chi^2$ values for testing hypothesis $H_B$ for both CovidHeRA and MediaEval datasets are 200.321 and 5.261, respectively, and are more significant than the critical value, $\chi^2_c = 3.841$ with $p < 0.001$ and $p = 0.022$, respectively, both less than $\alpha = 0.05$. This implies a significant difference in the proportions of the gender of users between fake and real news. On comparing actual values with expected values from Tables 1 and 4, respectively, we observe that the male gender has a greater actual proportion in fake news than the expected proportion, and the female gender has a higher actual proportion involved in real news than the expected proportion, in both datasets. Therefore, we reject the null hypothesis, $H_{B0,}$ and observe a significant bias in the gender of users involved in CoVID-19 fake news propagation.

To test for the hypothesis $H_C$, from the third row of Table 7, we observe that $\chi^2$ values for both CovidHeRA and MediaEval datasets are 298.995 and 7.765, respectively, and are greater than the critical value, $\chi^2_c = 3.841$ with $p < 0.001$ and $p = 0.006$, respectively, both less than $\alpha = 0.05$. For both the datasets, comparing the values of actual and expected media usage from Tables 3 and 6 shows that the actual values for fake news with media usage is less than the expected values, and the same is more in the case of real news. Therefore, there is a significant difference in the proportion of fake news and real news propagation with media usage than the expected proportion, which leads us to reject the null Hypothesis, $H_{C0}$.

The test for hypothesis $H_D$ involves comparing values from row four and row five of Table 7. From row 4, the $\chi^2$ values of male gender from datasets CovidHeRA and MediaEval are 217.67 and 13.342, respectively, both higher than $\chi^2_c = 5.991$ and p values being $p < 0.001$ and $p = 0.001$, respectively, both less than $\alpha = 0.05$. From row 5, the $\chi^2$ value for female gender from CovidHeRA dataset is 169.979, greater than the critical value, $\chi^2_c = 5.991$ and the value of $p < 0.001$ is less than $\alpha =$

**Table 4**
Expected count of fake and real items with gender as a category.

| Label | CovidHeRA | | | Mediaeval | | |
|---|---|---|---|---|---|---|
| | Male | Female | Total | Male | Female | Total |
| Fake | 1197 | 1107 | 2304 | 888.7 | 877.3 | 1766 |
| Real | 43,018 | 39,769 | 82,787 | 2051.3 | 2024.7 | 4076 |
| Total | 44,215 | 40,876 | 85,091 | 2940 | 2902 | 5842 |

**Table 5**
Expected count of fake and real items with sentiment polarity as a category.

| Label | CovidHeRA | | | | Mediaeval | | | |
|---|---|---|---|---|---|---|---|---|
| | Negative | Neutral | Positive | Total | Negative | Neutral | Positive | Total |
| Fake | 874.5 | 677.7 | 751.8 | 2304 | 1016.3 | 313.2 | 436.5 | 1766 |
| Real | 31421.5 | 24351.3 | 27014.2 | 82,787 | 2345.7 | 722.8 | 1007.5 | 4076 |
| Total | 32,296 | 25,029 | 27,766 | 85,091 | 3362 | 1036 | 1444 | 5842 |

**Table 6**
Expected count of fake and real items with media usage as a category.

| Label | CovidHeRA | | | Mediaeval | | |
|---|---|---|---|---|---|---|
| | With Media | W/o Media | Total | With Media | W/o Media | Total |
| Fake | 483 | 1821 | 2304 | 326.5 | 1439.5 | 1766 |
| Real | 17,367 | 65,420 | 82,787 | 753.5 | 3322.5 | 4076 |
| Total | 17,850 | 67,241 | 85,091 | 1080 | 4762 | 5842 |

0.05. However, for the same gender in the MediaEval dataset, the $\chi^2$ value turns out to be 5.503, which is less than ${\chi^2}_c = 5.991$, and the p-value of $p = 0.064 > \alpha = 0.05$ suggests contradictory inference from these two datasets. Nertheless, since the MediaEval dataset gave both the $\chi^2$ and p values close to their respective critical values for the female gender, we reject the Null Hypothesis, $H_{D0,}$ and conclude that there is a significant bias in the proportion of sentiments used by users of both the gender and the bias in proportion of the user gender has no influence on the bias of proportion of sentiments.

Further, to identify which sentiment is more biased by the users of both genders, we use results from rows six, seven, and eight of Table 7 to test hypothesis $H_E$. For the CovidHeRA dataset, the three rows mentioned above have $\chi^2$ values of 78.005, 13.65, and 146.509 for negative, neutral, and positive sentiment, respectively, which are all greater than ${\chi^2}_c = 3.841$ and their respective p values being $p < 0.001$ for all three, is less than $\alpha = 0.05$. Results from this dataset do not indicate the specific sentiment towards which the bias is more. However, we can infer that there is a significant difference in the proportion of each sentiment when compared to real news. Observing results from these three rows for the MediaEval dataset, we obtain $\chi^2$ values of 1.702, 4.82, and 0.411, for negative, neutral, and positive sentiments, respectively, where $\chi^2$ values for negative and positive sentiments are both less than ${\chi^2}_c = 3.841$ and for neutral sentiment, the $\chi^2$ value is higher than ${\chi^2}_c$. The p values for these corresponding $\chi^2$ values are $p = 0.192$, $p = 0.028$ and $p = 0.521$, respectively. This shows no significant bias of the user gender on negative and positive sentiment as p values (0.192 and 0.521) obtained are greater than $\alpha = 0.05$. However, for the neutral sentiment as the p-value of 0.028 is less than $\alpha = 0.05$. Therefore, we reject the null hypothesis, $H_{E0,}$ and conclude that fake news is more biased towards sentiment neutral, followed by sentiment negative, and show no significant difference in proportions of real news towards sentiment positive.

For testing the hypothesis, $H_F$, the bias of usage of media to induce a particular sentiment in the propagation of COVID-19 fake news, from Table 7, the values from rows nine, ten and eleven for CovidHeRA dataset indicate $\chi^2$ values of 97.382, 59.615, and 124.61 for negative, neutral and positive sentiment, respectively. All of these are greater than ${\chi^2}_c = 5.991$, with a p-value for each of them being $p < 0.001$, less than $\alpha = 0.05$, indicating rejection of the null hypothesis, $H_{F0}$. For the MediaEval dataset, however, the $\chi^2$ values are 0.235, 2.399, and 20.077 for negative, neutral, and positive sentiments, respectively, with the former two being less than ${\chi^2}_c = 3.841$ and the latter being more significant, Their respective p values being $p = 0.628$, $p = 0.121$ and $p < 0.001$ indicate that only for positive sentiment, there is a significant difference of proportion in the usage of media for fake news with respect to real news. From the contradictory results from the two datasets for negative and neutral sentiments, we understand that the usage of media produces a bias for only positive sentiment. Hence, we reject the Null Hypothesis, $H_{F0}$.

From rows twelve and thirteen of Table 7, we test for hypothesis, $H_G$ to observe a bias of the proportion of sentiment caused when media is used and not used, respectively. For CovidHeRA dataset, with media usage (row 12) and without media usage (row 13), $\chi^2$ values are 267.346 and 61.585, respectively, both less than ${\chi^2}_c = 5.991$ and their respective p values of $p < 0.001$ each for both being less than $\alpha = 0.05$, suggest that there is a difference in the proportion of sentiment used in fake news with respect to real news. Similar inference can be obtained from the MediaEval dataset, in which, with media usage (row 12) and without media usage (row 13) have $\chi^2$ values of 7.223 and 25.15, respectively, both less than ${\chi^2}_c = 5.991$ and their respective p values of $p = 0.027$ and $p < 0.001$, both being less than $\alpha = 0.05$. Hence, there is a bias induced in the proportions of sentiment in fake news with respect to real news by

**Table 7**
Chi-square test on qualitative hypotheses.

| (H) | Datasets∏→ | CovidHeRA | | | | | | MediaEval | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | variables↓ | df | $\chi^2$ | p value | $V^2$ | r | rho | df | $\chi^2$ | p value | $V^2$ | r | rho |
| HA | Label vs Sentiment category | 2 | 352.963 | p < 0.001 | 0.004 | 0.047 | 0.048 | 2 | 17.103 | p < 0.001 | 0.003 | 0.037 | 0.032 |
| HB | Label vs Gender | 1 | 200.321 | p < 0.001 | 0.002 | 0.048 | 0.048 | 1 | 5.261 | p = 0.022 | 0.001 | 0.03 | 0.03 |
| HC | Label vs Media usage | 1 | 298.995 | p < 0.001 | 0.003 | 0.059 | 0.059 | 1 | 7.565 | p = 0.006 | 0.001 | 0.035 | 0.035 |
| HD | Label vs Sentiment (Gender - male) | 2 | 217.67 | p < 0.001 | 0.004 | 0.039 | 0.041 | 2 | 13.342 | p = 0.001 | 0.004 | 0.035 | 0.027 |
| HD | Label vs Sentiment (Gender - female) | 2 | 169.979 | p < 0.001 | 0.004 | 0.058 | 0.058 | 2 | 5.503 | p = 0.064 | 0.001 | 0.038 | 0.035 |
| HE | Label vs Gender (Sentiment - Negative) | 1 | 78.005 | p < 0.001 | 0.002 | 0.049 | 0.049 | 1 | 1.702 | p = 0.192 | 0.001 | 0.023 | 0.023 |
| HE | Label vs Gender (Sentiment - Neutral) | 1 | 13.65 | p < 0.001 | 0.001 | 0.023 | 0.023 | 1 | 4.82 | p = 0.028 | 0.004 | 0.068 | 0.068 |
| HE | Label vs Gender (Sentiment - Positive) | 1 | 146.509 | p < 0.001 | 0.005 | 0.072 | 0.072 | 1 | 0.411 | p = 0.521 | 0 | 0.016 | 0.016 |
| HF | Label vs Media usage (Sentiment - Negative) | 1 | 97.382 | p < 0.001 | 0.003 | 0.055 | 0.055 | 1 | 0.235 | p = 0.628 | 0 | −0.008 | −0.008 |
| HF | Label vs Media usage (Sentiment - Netural) | 1 | 59.615 | p < 0.001 | 0.002 | 0.048 | 0.048 | 1 | 2.399 | p = 0.121 | 0.002 | 0.048 | 0.048 |
| HF | Label vs Media usage (Sentiment - Positive) | 1 | 124.61 | p < 0.001 | 0.004 | 0.066 | 0.066 | 1 | 20.077 | p < 0.001 | 0.013 | 0.117 | 0.117 |
| HG | Label vs Sentiment (Media not used) | 2 | 267.346 | p < 0.001 | 0.003 | 0.042 | 0.044 | 2 | 7.223 | p = 0.027 | 0.001 | 0.01 | 0.005 |
| HG | Label vs Sentiment (Media used) | 2 | 61.585 | p < 0.001 | 0.003 | 0.045 | 0.042 | 2 | 25.154 | p < 0.001 | 0.023 | 0.145 | 0.141 |
| HH | Label vs Gender (Media not used) | 1 | 193.333 | p < 0.001 | 0.003 | 0.053 | 0.053 | 1 | 5.561 | p = 0.018 | 0.001 | 0.034 | 0.034 |
| HH | Label vs Gender (media used) | 1 | 5.472 | p = 0.019 | 0 | 0.017 | 0.017 | 1 | 0.345 | p = 0.557 | 0 | 0.017 | 0.017 |
| HI | Label vs Media usage (Gender - male) | 1 | 209.649 | p < 0.001 | 0.005 | 0.069 | 0.069 | 1 | 5.664 | p = 0.017 | 0.001 | 0.043 | 0.043 |
| HI | Label vs Media usage (Gender - female) | 1 | 87.831 | p < 0.001 | 0.002 | 0.046 | 0.046 | 1 | 2.529 | p = 0.112 | 0.001 | 0.03 | 0.03 |

usage and non-usage of media, and therefore we reject the null hypothesis, $H_{G0}$.

Further, from rows fourteen and fifteen of Table 7, we test for the hypothesis $H_H$ to check if the bias in the proportion of gender of fake news with respect to real news is influenced by a bias in media usage. For CovidHeRA, we obtain $\chi^2$ values of 193.333 and 5.472 for “media used” and “media not used”, respectively, both greater than ${\chi^2}_c = 3.84$ with their respective p values being $p < 0.001$ and $p = 0.019$, both less than $\alpha = 0.05$. For the MediaEval dataset, for the same rows, we obtain $\chi^2$ values of 5.561 and 0.345 and p values of $p = 0.018$ and $p = 0.557$ for “media used” and “media not used,” respectively. We observe that for “media not used,” the test shows the opposite result with that compared from the CovidHeRA dataset, meaning that there is no difference in the proportion of users' gender when media is not used in fake with respect to real news propagation. These contradictory results make hypothesis $H_H$ inconclusive.

From the values in rows 16 and 17 in Table 7, for the CovidHeRA dataset, both genders show a difference in the proportion of media used for fake news propagation with respect to real news. This can be observed as the $\chi^2$ values of 209.649 and 87.831 for the male and female gender, respectively, are greater than ${\chi^2}_c = 3.84$, and their respective p values, both $p < 0.001$ is more diminutive than $\alpha = 0.05$. In the MediaEval dataset, we observe from rows 16 and 17 of Table 7 that while male users with $\chi^2$ value of 5.664 and $p = 0.017$ show difference in the proportion of media used for fake news with respect to real news, but for the female gender, indifference in proportions of usage of media in fake news with respect to real news is observed as the $\chi^2$ value of 2.529 is less than ${\chi^2}_c = 3.84$ and its p-value of $p = 0.112$ is more remarkable than $\alpha = 0.05$. Therefore, for hypothesis $H_I$, we cannot come to any conclusive decision.

### 4.2. Quantitative experimental results

In Levene's test for homoscedasticity, the common null hypothesis ($H_0$) is such that $H_0$: $\frac{\sigma_1}{\sigma_2} = 1$ and alternate hypothesis ($H_1$) is such that, $H_1$: $\frac{\sigma_1}{\sigma_2} \neq 1$, where σ is the standard deviation of that particular independent variable on the two groups, fake (1) and real (2). From Table 9, we observe that only for the friends' count, the p-value of Levene's test is greater than 0.05 for both datasets. Therefore, we fail to reject the null hypothesis for friends count and conclude that variances from the mean are the same for both fake and real news. All other quantitative variables have a p-value less than 0.05, and we, therefore, reject the null hypotheses and conclude that the variance from the mean value is significantly different for fake and real news for these variables. As results from Levene's test indicate unequal variances for most variables, this eliminates the necessity to perform an independent sample *t*-test. Therefore, we plot the data distribution around the mean with a 95% confidence interval. This will distinguish the central values of the variables and help us determine the strength of the distinguishment, i.e., the smaller the upper and lower bound distance from the mean, the more the reliance on these values representing the true mean value of the population. Table 8 provides descriptive statistics of both datasets. From Fig. 3 and Fig. 4, for CovidHeRA and MediaEval datasets, we observe that users who propagated fake news have smaller followers than users posting real news. The mean values for fake news for these datasets are 5421.65 and 23255.37 in the order mentioned above. These are distinct from the mean number of followers of real news, 63656.21 and 99511.34 for the two datasets. We also observe a significant bias in the number of followers of fake news and real news proliferators as the range of 95% CI for mean does not overlap for fake and real news. Therefore, attributing a label to a piece of information on Twitter by comparing the number of followers of the user who shared it with the mean range of these plots can be done more accurately.

From Table 10 and Table 11, it becomes evident that post-specific variables (retweet and favorite count) are more strongly associated with each other and have a mutual positive influence, i.e., an increase in one also leads to an increase in the other. We also observe that user-specific variables (followers, friends, and status count) are either strongly (followers-friends, followers-status) or moderately (friends-

**Table 9**
Levene's Test results on Quantitative variables for both datasets.

| Quantitative Variable | Dataset | Levene's test statistic | Significance of Levene's test (p-value) | Decision |
|---|---|---|---|---|
| Followers Count | CovidHeRA<br>MediaEval | 15.84<br>60.80 | 0.000<br>0.000 | Reject Null Hypothesis, Unequal Variances |
| Friends Count | CovidHeRA<br>MediaEval | 1.20<br>0.31 | 0.194<br>0.579 | Fail to Reject Null Hypothesis, Equal Variances |
| Status Count | CovidHeRA<br>MediaEval | 8.84<br>9.20 | 0.003<br>0.002 | Reject Null Hypothesis, Unequal Variances |
| Retweet Count | CovidHeRA<br>MediaEval | 2.52<br>35.36 | 0.113<br>0.000 | Reject Null Hypothesis, Unequal Variances |
| Favorite Count | CovidHeRA<br>MediaEval | 8.03<br>36.98 | 0.005<br>0.000 | Reject Null Hypothesis, Unequal Variances |

**Table 8**
Descriptive statistics of CovidHeRA(C) and MediaEval dataset(M).

| Statistics | Followers | | Friends | | Retweets | | Status | | Favorites | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Fake | Real | Fake | Real | Fake | Real | Fake | Real | Fake | Real |
| Mean (C) | 5421.657 | 63656.21 | 3181.374 | 2293.652 | 154.132 | 628.718 | 56262.98 | 46189.4 | 2.238 | 7.766 |
| Standard Error (C) | 445.735 | 3847.024 | 149.150 | 35.658 | 38.207 | 17.616 | 2269.618 | 497.010 | 0.255 | 0.490 |
| Median (C) | 1742.5 | 21,733 | 953 | 605 | 52 | 173 | 17,180 | 9238 | 1 | 2 |
| Mode (C) | 706 | 7810 | 775 | 209 | 18 | 28 | 3145 | 1760 | 0 | 0 |
| Standard Deviation (C) | 21395.32 | 1,106,894 | 7159.233 | 10259.94 | 1833.94 | 5068.56 | 108941.7 | 143003.4 | 12.261 | 141.126 |
| Sample Variance (C) | 4.58 E+08 | 1.23 E+12 | 51,254,619 | 1.05 E+08 | 3.36 E+06 | 2.57 E+07 | 1.19 E+10 | 2.04 E+10 | 150.344 | 19916.57 |
| Count (C) | 2304 | 82,787 | 2304 | 82,787 | 2304 | 82,787 | 2304 | 82,787 | 2304 | 82,787 |
| Confidence Level (95.0%) (C) | 874.085 | 7540.138 | 292.483 | 69.890 | 74.886 | 34.527 | 4450.709 | 974.136 | 0.500 | 0.961 |
| Mean (M) | 23255.37 | 99511.34 | 3012.989 | 1999.394 | 260.701 | 644.781 | 38369.96 | 55846.16 | 679.669 | 2244.092 |
| Standard Error (M) | 9979.619 | 11302.57 | 302.646 | 159.182 | 71.410 | 61.498 | 1787.307 | 1744.372 | 201.229 | 224.780 |
| Median (M) | 4711.5 | 37160.5 | 733 | 609 | 60 | 155 | 12,955 | 18305.5 | 48 | 292 |
| Mode (M) | 1180 | 12,548 | 650 | 138 | 77 | 92 | 547 | 446 | 48 | 177 |
| Standard Deviation (M) | 419381.5 | 721596.4 | 12718.35 | 10162.77 | 3000.955 | 3926.28 | 75109.43 | 111366.9 | 8456.42 | 14350.77 |
| Sample Variance (M) | 1.76 E+11 | 5.21 E+11 | 1.62 E+08 | 1.03 E+08 | 9,005,729 | 15,415,676 | 5.64 E+09 | 1.24 E+10 | 71,511,039 | 2.06 E+08 |
| Count (M) | 1766 | 4076 | 1766 | 4076 | 1766 | 4076 | 1766 | 4076 | 1766 | 4076 |
| Confidence Level (95.0%) (M) | 19560.05 | 22153.04 | 593.583 | 311.998 | 140.058 | 120.570 | 3505.461 | 3419.921 | 394.672 | 440.692 |

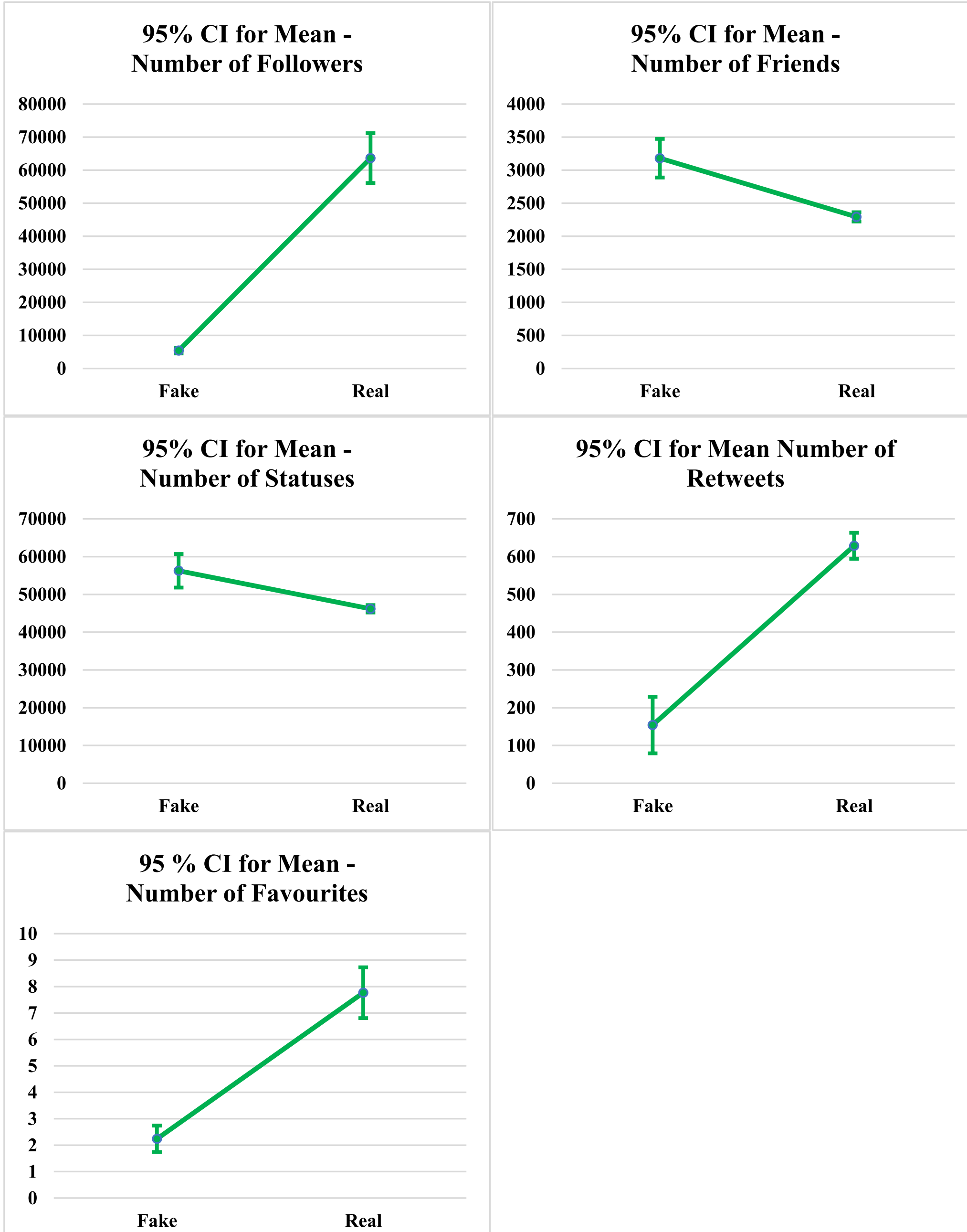


**Fig. 3.** 95% Confidence Interval on Mean for quantitative factors on CovidHeRA dataset.

status) associated with each other and an increase in one of these leads to an increase in the other. Further, there is a weak influence of user-specific features (followers, friends, and status count) on post-specific features (favorite and retweet count) as the correlation values are significantly low. From the above-discussed scenario, we interpret that a particular user's activity or engagement level to keep a stand on social media (friends, followers, status counts) has no significant impact on the post (favorite and retweet counts).

From the plots of the number of friends in Figs. 3 and 4 for CovidHeRA and MediaEval datasets, respectively, the previously mentioned inference becomes much more robust as not only the 95% CI bounds remain distinct for fake news and real news, but also the closer proximity of the value of mean for a particular label in both datasets shows the repeatability of the trend. The mean values for fake news in CovidHeRA and MediaEval dataset are 3181.374 and 3012.989, respectively, and the same for real news in these datasets are 2293.652 and 1999.394, respectively. There is a significant bias in the mean number of friends for users who propagated fake news compared to those of users who propagated real news.

The plots from Figs. 3 and 4 for the number of retweets have similar mean values for fake and real news. For fake news, the mean values are 154.132 and 260.701 for the two datasets, and for real news, the mean values are 628.718 and 644.781, which show the closeness within the label and distinction between the labels. Therefore, this bias can help label a piece of information based on its proximity to one of the 95% CI interval mean values.

For status count, the 95% CI interval for mean and the mean value for fake and real news alternate between the two datasets. Therefore, we cannot come to any specific conclusion using the information of this variable of a particular information sample despite there being a bias in the mean values between the labels. The same conclusion can be drawn for the number of favorites as the datasets' ranges are significantly

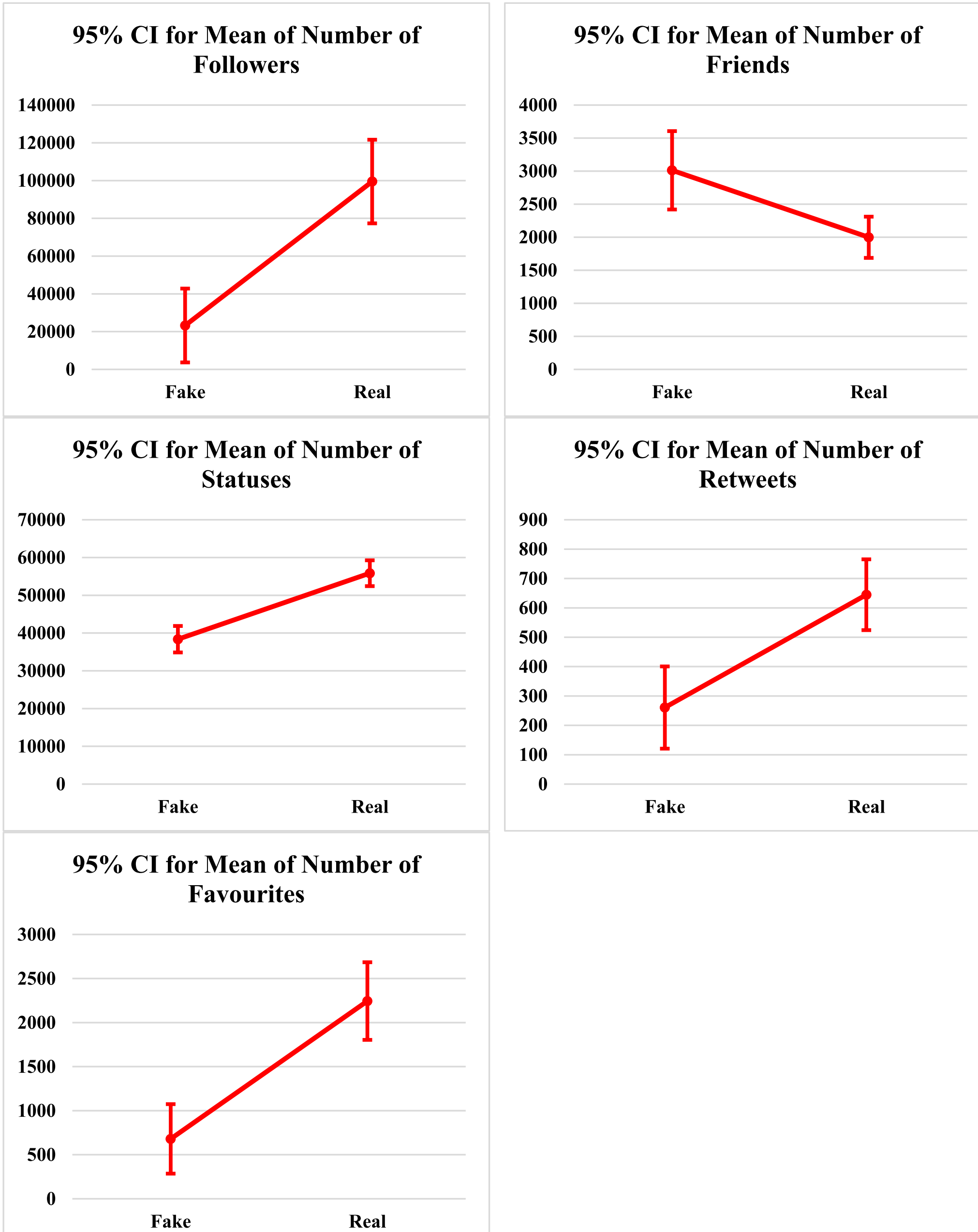


**Fig. 4.** 95% Confidence Interval on Mean for quantitative factors on MediaEval dataset.

**Table 10**
Spearman correlation for CovidHeRA dataset.

| | Retweet Count | Favorite Count | Followers Count | Friends Count | Status Count |
|---|---|---|---|---|---|
| Retweet Count | 1 | | | | |
| Favorite Count | 0.647 | 1 | | | |
| Followers Count | 0.396 | 0.367 | 1 | | |
| Friends Count | 0.156 | 0.141 | 0.674 | 1 | |
| Status Count | 0.077 | 0.000 | 0.603 | 0.480 | 1 |

**Table 11**
Spearman correlation for MediaEval dataset.

| | Retweet Count | Favorite Count | Followers Count | Friends Count | Status Count |
|---|---|---|---|---|---|
| Retweet Count | 1 | | | | |
| Favorite Count | 0.837 | 1 | | | |
| Followers Count | 0.689 | 0.666 | 1 | | |
| Friends Count | 0.236 | 0.229 | 0.538 | 1 | |
| Status Count | 0.274 | 0.245 | 0.605 | 0.511 | 1 |

different amongst the same variable. Hence, any information about this variable in sample information under test cannot be determined as fake or real.

## 5. Discussion

Fake news on social media is a menace hard to identify and characterize. It is unclear which factors help distinguish between real and fake news. Past literature has identified several psychological and behavioral features associated with fake news propagation and acceptance [57]. Little research has been done in identifying key factors characterizing fake news. This study delves deep into factor analysis and their interdependence, examining how certain factors influence fake news detection and propagation on Twitter. Table 12 summarizes the results of all the hypotheses considered in this article.

Table 13 depicts the inferences/insights gained from each of the hypotheses. In the qualitative hypothesis $H_A$, it is assumed that there is a bias in the proportions of sentiment (linguistic tone) in fake news. Although, the central polarity of bias was unclear. Our study on two COVID-19 specific datasets found a strong bias of fake news towards neutral sentiment followed by negative sentiment with respect to real news, which is proved by the results of our first hypothesis. In the second hypothesis, $H_B$, we tested the bias in the proportion of gender in fake news. The results predicted a strong bias of the male gender towards fake news propagation with respect to real news. The influence of the gender ratio of Twitter users is not taken into account as the test is performed to distinguish the characteristics of real and fake news. This influence is assumed to affect both types of news equally and nullify its effect. In other words, the speculated gender ratio of 6.85:3.15 should be observed in any random sample collection of tweets. Hence, we directly compare the dataset's actual ratio without considering the speculated ratio's deviation. In our datasets, the proportion of tweets (both real and fake) with media is lesser than tweets without media. From the chi-square test results on hypothesis $H_C$, we find that the proportion of fake news with media is significantly less than expected and substantially more than anticipated for real news with media. Further, we explore if the bias in proportions of one category amongst sentiment, gender, and media usage, is significantly influenced by the bias in proportions of these categories. From the test for Hypothesis $H_D$, we find that fake news shared by both male and female gender show bias in proportion of sentiment. The result for hypothesis $H_E$ indicates that this bias is towards fake news being sentiment neutral, followed by sentiment negative, with respect to real news. This supports our Hypothesis $H_A$. Further, from the results of testing Hypothesis $H_F$ and $H_G$, $H_G$ concludes that sentiment is biased in both "with" and "without" media usage. From $H_F$, we conclude that this bias in fake news propagation is proportional to using positive sentiment. For the remaining combination of gender and media usage, from the results of hypotheses $H_H$ and $H_I$, it cannot be concluded if there is a mutual influence of media usage and the gender of the user in the bias observed in hypotheses $H_B$ and $H_C$ due to the contradictory results from the two datasets. In hypothesis $H_H$, the results are contradictory for "media used," and for $H_I$, the results are contradictory for the "female" gender.

We observe a significant distinguishable difference in the mean number of followers, friends, and retweets for fake and real news from the quantitative variables. The smaller value of mean for followers can be attributed to the case that most real news proliferators are official media channels and celebrity users who share information on Twitter. In contrast, fake news comes mostly from regular Twitter users who do not have a huge following. Similar reasons can be attributed to a smaller mean value for retweets of fake news. For the larger value of mean for the number of friends, we understand that the users who propagate fake news are involved in more mutual social connections. Understandably, when compared to regular active Twitter users, celebrities and official media sources do not have many mutual connections that Twitter classifies as "friends" and, therefore, the resulting smaller value of the mean. The confidence interval for the mean for each of these plots acts as a range for true mean for fake and real news. These can be used to identify any sample of data by comparing its mean to the 95% CI for the mean of these plots. The non-distinguishable mean value and reverse in the plotted trend for the number of statuses posted by the users who propagated fake and real news and the difference of range for the mean of the number of users who favorited the tweet between the two datasets make

**Table 12**
Summary table.

| Hypotheses | Results |
|---|---|
| $\mathbf{H_A}$: Bias of sentiment in fake news with respect to real news. | Reject Null Hypothesis |
| $\mathbf{H_B}$: Bias of the gender of users involved in fake news with respect to real news. | Reject Null Hypothesis |
| $\mathbf{H_C}$: Bias of media usage in fake news with respect to real news. | Reject Null Hypothesis |
| $\mathbf{H_D}$: Bias in the proportion of a particular gender of the user on the bias in the proportion of sentiments in fake news with respect to real news. | Reject Null Hypothesis |
| $\mathbf{H_E}$: Bias in the proportion of a particular sentiment used in fake news between different gender of users. | Reject Null Hypothesis |
| $\mathbf{H_F}$: Bias of inducing a particular sentiment with the usage of media in fake news. | Reject Null Hypothesis |
| $\mathbf{H_G}$: Bias in the usage of media amongst different sentiments used in fake news. | Reject Null Hypothesis |
| $\mathbf{H_H}$: Relationship between a particular gender and media usage in fake news. | Inconclusive |
| $\mathbf{H_I}$: Bias in and usage of media in fake news between different gender of users. | Inconclusive |
| $\mathbf{H_J}$: Significantly distinguishable bias of "follower" count in fake news. | Reject Null Hypothesis |
| $\mathbf{H_K}$: Significantly distinguishable bias of "friends" count in fake news. | Reject Null Hypothesis |
| $\mathbf{H_L}$: Significantly distinguishable bias of "status" count in fake news. | Fail to Reject Null Hypothesis |
| $\mathbf{H_M}$: Significantly distinguishable bias of "retweet" count in fake news. | Reject Null Hypothesis |
| $\mathbf{H_N}$: Significantly distinguishable bias of "favorite" count in fake news. | Fail to Reject Null Hypothesis |
| $\mathbf{H_O}$: Significant effect of fake news sharing behavior on its acceptance behavior. | Hypothesis verified |

**Table 13**
Inferences/insights gained from all the proposed hypotheses.

| Hypotheses | Insights |
|---|---|
| $H_A$ | Fake news is more often written in a negative or neutral tone. |
| $H_B$ | Male users are more likely to tweet/post fake news than female users. |
| $H_C$ | Fake news is less likely to contain visual media than real news. |
| $H_D$ | Both genders show bias in sentiments upon sharing fake news. |
| $H_E$ | Male users are more likely to share fake news with a negative or neutral linguistic tone.<br>Female users are likely to share fake news with a positive linguistic tone. |
| $H_F$, $H_G$ | Fake news is more likely to have a negative sentiment/linguistic tone when visual media is absent. |
| $H_H$, $H_I$ | It is inconclusive to specify if a gender type is biased towards media usage. |
| $H_J$ | Users spreading fake news have fewer followers than users spreading real news. |
| $H_K$ | Users spreading fake news have a larger number of friends than real news. |
| $H_L$ | It is inconclusive to state whether fake news proliferators have a greater or lesser status count. Users who spread fake news can display either high engagement through tweeting or a less frequent tweet. |
| $H_M$ | Fake news proliferators are less likely to retweet posts from other users. |
| $H_N$ | Fake news is less likely to be marked as a favorite by other users. |
| $H_O$ | Higher sharing (retweet) leads to trust and a higher acceptance (favorite) of fake news. |
| – | User-specific features and post-specific features have weak association. Hence, fake news sharing is impervious to the social media presence/stand/popularity of the user. |

these variables unsuitable for classification of the tweets' labels.

Table 14 describes the results presented by the existing studies for the user engagement attributes and compares/contrasts them with the findings of this study. Sentiments and emotions are of great importance in understanding the polarity of publishers and respondents. Our findings indicate that fake news is written with a negative publisher emotion. We also suggest that real news largely constitutes neutral sentiment polarity than being positive or negative. Our conditional experiments demonstrate that males and females equally spread negative fake news, and there is no significant bias towards a particular gender demonstrating negative sentiment while sharing fake news. However, fake news sharing demonstrates a strong positive bias towards gender, suggesting that male users are more involved in sharing fake news. Since acceptance or susceptibility to false information leads to sharing such news, the above finding contradicts the study by Rampersad and Althiyabi [11] that states that gender has a weakly positive effect on the acceptance of fake news. Our findings contradict the result by Shu et al. [38] that states that females are more likely to trust fake news. Since we consider original posts/tweets and disregard retweets, our experiments demonstrate that false news is more likely to originate from male users than female users. This result can be potentially used to trace user accounts spreading misinformation.

Existing studies argue that status count (total number of posts shared by an account) is crucial in characterizing false information. The study by Shu et al. [38] indicates that users spreading fake news generally have a lower status count than the users sharing real news. They attribute this finding to the idea that authentic users are more likely to engage on social media. In contrast, Cheng et al. [37] suggest that fake news proliferators have a higher status count. However, our findings on two different coronavirus-specific datasets indicate contrasting trends between both. In CovidHeRA, it is seen that users with a considerably large number of statuses are more involved in sharing false information. This can be attributed to the nature of the dataset, implying that highly engaging users are more likely to spread health-related misinformation such as precautional and remedial measures to curb coronavirus. Whereas users posting authentic health information such as official notices or advisories comparatively have a lower status count suggesting that these are trustworthy user profiles belonging to users applying cognitive sense and critical thinking before sharing a piece of information. In the MediaEval dataset, a reverse trend is observed with more posts by users sharing real news, again attributed to the dataset type. Results suggest that users spreading conspiracy theories are more likely to have a lower status count than the users avoiding conspiracies. Considering the varying results in existing studies and our research, we conclude that status count is an arguable factor, and it is inconclusive to determine that false news sharing profiles have a higher or lower status count.

Shu et al. [38] discover that users sharing false information express more 'favor' actions in order to engage and interact with their network. In contrast, we examine the number of favorites marked on a given real or fake tweet. Our results demonstrate that fake news receives a lot fewer favorites than fake news. This highlights the unpopularity of false news among social networks, which implies that users are less oriented towards saving a false claim for their future reference. We observe that real news receives a high number of favorites demonstrating higher user interaction around authentic information. Users are more likely to 'favorite' or save trustworthy information.

We confirm the results obtained for follower count by Shu et al. [26], indicating that users spreading false news have fewer followers than the users sharing real news. This finding suggests that false news proliferators might not have a well-regarded reputation on the platform or have a considerably smaller social network. This strengthens the belief that most Twitter users do not follow malicious accounts spreading misinformation or conspiracies.

Our findings contradict Cheng et al. [25] regarding friends count, stating that users sharing false information have fewer friends. Our experiments demonstrate that such users can have a larger number of friends, and these users follow a large number of social media profiles to potentially build a larger friend network. This highlights their intention to increase or gain more social attention by exploiting the 'follow for follow' user behavior. The 'follow for follow' social media engagement is a highly used behavior where users simply follow other user accounts to gain more followers. Malicious false news proliferators might intend to gain more visibility to their false content by increasing their friends' count.

Another observation from $H_O$ proposes that false news is less often shared or retweeted in a social network. We can observe a psychological bias that users are more concerned about their online reputation and think critically before sharing any information. Whereas real news obtains a much higher number of retweets, which confirms that social media users are more inclined towards sharing authentic and verified information, and most users are restrained from sharing false information. Through spearman correlation among the five social media engagement attributes, we observe that higher number of retweets correlates with higher number of favorites on that post. This implies that users sharing false news demonstrate trust and acceptance of that information.

**Table 14**
Comparison between results from previous literature and our findings.

| Variable | Existing Results | Our Findings | Confirm/ Dispute |
|---|---|---|---|
| Sentiment | The fake news stories are written in a specific linguistic tone, though it is inconclusive to say which one (negative, positive, or neutral) [17] | Irrespective of the gender, most of the fake news is written in negative sentiment (negative publisher emotion) in proportion than positive sentiment. Fake news with negative and neutral sentiment contains fewer visual media than real news. | Findings dispute existing results |
| Gender | Gender has a weakly positive effect on people's acceptance of fake news [11]. Female users are more likely to trust fake news than male users [38]. | Gender has a strong positive effect on the sharing of fake news. Male users are more likely to post fake news | Findings dispute existing results |
| Status Count | The users spreading fake news generally publish fewer posts than users spreading real news, which indicates those users trusting more real news are more likely to be active and express themselves [38]. Users with a higher status count are less susceptible to sharing fake news [37]. | Though there is a difference in the status count of real and fake news proliferators, it is inconclusive to state if fake news sharing profiles have a higher or lower status count than real news sharing profiles. | Findings dispute existing results |
| Followers Count | Users sharing fake news have fewer followers [38]. | Users sharing fake news have a lower follower count, as established through experiments on both datasets. | Findings confirm existing results |
| Friends Count | Users with a higher friends count (the number of users an account follows) are less susceptible to sharing fake news. (Users with more friends share less fake news) [37]. | Users spreading false information have a higher friends' count. | Findings dispute existing results |

## 6. Conclusion

Fake information on social platforms has constantly been increasing. In the state of the COVID-19 pandemic, this problem has grown at an exponential rate globally. The pandemic is one major event generating misinformation and promoting its consumption through social networks worldwide. In the absence of a holistic fake news detection model, it is unclear what factors can be used to identify misinformation. Very few past works are dedicated to identifying such factors. This work examines several factors from two Twitter datasets, MediaEval 2020 and Covid-HeRA, using fifteen hypotheses $H_A$ to $H_O$. The study uses Chi-square tests for nine qualitative theories ($H_A$ to $H_I$), whereas for six quantitative tests ($H_J$ to $H_O$), we have calculated variances from mean, confidence intervals using Analysis of Means and correlated tweet-specific variables. Observations from this study unravel specific characteristics to distinguish fake news from real news, especially the conditional effects of categorical variables that are one amongst the unique findings of this work. These new findings pave the way for future research and the development of robust fake news detection algorithms. We motivate fellow researchers to design algorithms that utilize the discovered dependencies using their combined decisions. Also, we encourage to discover more identifiers that can characterize false information present online ubiquitously. This study provides a new dimension to the existing literature in the fake news and infodemic domains.

## Author statement

**Priyanka Meel**: Conceptualization, Methodology, Formal Analysis, Resources, Supervision, Validation.

**Chahat Raj**: Data Curation, Implementation, Software, Writing-Original Draft, Reviewing and Editing, Revision.

## Acknowledgments

We thank Mihir P Mehta for his support and feedback in the experimentation and analysis.

## References


[1] T. Kaya, The changes in the effects of social media use of Cypriots due to COVID-19 pandemic, Technol. Soc. 63 (2020), https://doi.org/10.1016/j.techsoc.2020.101380.

[2] A. Mohammed, A. Ferraris, Factors influencing user participation in social media: evidence from twitter usage during COVID-19 pandemic in Saudi Arabia, Technol. Soc. 66 (2021), https://doi.org/10.1016/j.techsoc.2021.101651.

[3] A. Farooq, S. Laato, A.K.M.N. Islam, J. Isoaho, Understanding the impact of information sources on COVID-19 related preventive measures in Finland, Technol. Soc. 65 (2021), https://doi.org/10.1016/j.techsoc.2021.101573.

[4] H. Rosenberg, S. Syed, S. Rezaie, The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic, Can. J. Emerg. Med. 22 (4) (2020), https://doi.org/10.1017/cem.2020.361.

[5] K. Soltaninejad, Methanol mass poisoning outbreak, a consequence of COVID-19 pandemic and misleading messages on social media, Int. J. Occup. Environ. Med. 11 (3) (2020), https://doi.org/10.34172/ijoem.2020.1983.

[6] F. Tagliabue, L. Galassi, P. Mariani, The 'pandemic' of disinformation in COVID-19, SN Compr. Clin. Med. 2 (9) (2020), https://doi.org/10.1007/s42399-020-00439-1.

[7] S. Zannettou, et al., What Is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber, 2018, https://doi.org/10.1145/3184558.3191531.

[8] S. Altay, A.S. Hacquin, H. Mercier, Why do so few people share fake news? It hurts their reputation, New Media Soc. (2020), https://doi.org/10.1177/1461444820969893.

[9] A. Ardèvol-Abreu, P. Delponti, C. Rodríguez-Wangüemert, Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content, Prof. de la Inform. 29 (5) (2020), https://doi.org/10.3145/epi.2020.sep.07.

[10] S. Talwar, A. Dhir, P. Kaur, N. Zafar, M. Alrasheedy, Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior, J. Retailing Consum. Serv. 51 (2019), https://doi.org/10.1016/j.jretconser.2019.05.026.

[11] G. Rampersad, T. Althiyabi, Fake news: acceptance by demographics and culture on social media, J. Inf. Technol. Polit. 17 (1) (2020), https://doi.org/10.1080/19331681.2019.1686676.

[12] Y. Lian, Y. Liu, X. Dong, Strategies for controlling false online information during natural disasters: the case of Typhoon Mangkhut in China, Technol. Soc. 62 (2020), https://doi.org/10.1016/j.techsoc.2020.101265.

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

[14] D.K. Vishwakarma, C. Jain, Recent State-Of-The-Art of Fake News Detection: A Review, 2020, pp. 1–6, https://doi.org/10.1109/incet49848.2020.9153985.

[15] K. Pogorelov, et al., FakeNews: corona virus and 5G conspiracy task at MediaEval 2020, in: CEUR Workshop Proceedings, vol. 2882, 2020.

[16] A. Dharawat, I. Lourentzou, A. Morales, C. Zhai, Drink bleach or do what now? Covid-HeRA: a dataset for risk-informed health decision making in the presence of COVID19 misinformation [Online]. Available: http://arxiv.org/abs/2010.08743, Oct. 2020.

[17] S.B. Parikh, V. Patil, P.K. Atrey, On the Origin, Proliferation and Tone of Fake News, 2019, https://doi.org/10.1109/MIPR.2019.00031.

[18] M. García Lozano, et al., Veracity assessment of online data, Decis. Support Syst. 129 (2020), https://doi.org/10.1016/j.dss.2019.113132.

[19] B. Narwal, Fake news in digital media, in: Proceedings - IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, vol. 2018, ICACCCN, Oct. 2018, pp. 977–981, https://doi.org/10.1109/ICACCCN.2018.8748586.

[20] R. Oshikawa, J. Qian, W.Y. Wang, A Survey on Natural Language Processing for Fake News Detection, 2020.

[21] S.B. Parikh, P.K. Atrey, Media-rich fake news detection: a survey, in: Proceedings - IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018, 2018, pp. 436–441, https://doi.org/10.1109/MIPR.2018.00093. April.

[22] X. Zhou, R. Zafarani, Network-based fake news detection, ACM SIGKDD Explorations Newsletter 21 (2) (2019), https://doi.org/10.1145/3373464.3373473.

[23] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, K. Shu, Mining Dual Emotion for Fake News Detection, 2021, https://doi.org/10.1145/3442381.3450004.

[24] R. Kumari, N. Ashok, T. Ghosal, A. Ekbal, A multitask learning approach for fake news detection: novelty, emotion, and sentiment lend a helping hand, in: Proceedings of the International Joint Conference on Neural Networks, 2021-July, 2021, https://doi.org/10.1109/IJCNN52387.2021.9534218.

[25] R.J. Sethi, Spotting Fake News: A Social Argumentation Framework for Scrutinizing Alternative Facts, 2017, https://doi.org/10.1109/ICWS.2017.108.

[26] R.J. Sethi, R. Rangaraju, B. Shurts, Fact checking misinformation using recommendations from emotional pedagogical agents, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11528, LNCS, 2019, https://doi.org/10.1007/978-3-030-22244-4_13.

[27] R. Sethi, R. Rangaraju, Extinguishing the Backfire Effect: Using Emotions in Online Social Collaborative Argumentation for Fact Checking, 2018, https://doi.org/10.1109/ICWS.2018.00062.

[28] S. Tifferet, Verifying online information: development and validation of a self-report scale, Technol. Soc. 67 (2021), https://doi.org/10.1016/j.techsoc.2021.101788.

[29] S. van der Linden, C. Panagopoulos, J. Roozenbeek, You are fake news: political bias in perceptions of fake news, Media Cult. Soc. 42 (3) (2020), https://doi.org/10.1177/0163443720906992.

[30] S.M. Jang, J.K. Kim, Third person effects of fake news: fake news regulation and media literacy interventions, Comput. Hum. Behav. 80 (2018), https://doi.org/10.1016/j.chb.2017.11.034.

[31] S. Talwar, A. Dhir, D. Singh, G.S. Virk, J. Salo, Sharing of fake news on social media: application of the honeycomb framework and the third-person effect hypothesis, J. Retailing Consum. Serv. 57 (2020), https://doi.org/10.1016/j.jretconser.2020.102197.

[32] P.R. Brewer, D.G. Young, M. Morreale, The impact of real news about 'fake news': intertextual processes and political satire, Int. J. Publ. Opin. Res. 25 (3) (2013), https://doi.org/10.1093/ijpor/edt015.

[33] B. D. Horne and S. Adalı, "This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News." [Online]. Available: https://github.com/rpitrust/.

[34] A. Bovet, H.A. Makse, Influence of fake news in Twitter during the 2016 US presidential election, Nat. Commun. 10 (1) (2019), https://doi.org/10.1038/s41467-018-07761-2.

[35] B. Osatuyi, J. Hughes, A tale of two internet news platforms-real vs. Fake: an elaboration likelihood model perspective, in: Proceedings of the Annual Hawaii International Conference on System Sciences, 2018-January, 2018, https://doi.org/10.24251/hicss.2018.500.

[36] J. Colliander, 'This is fake news': investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media, Comput. Hum. Behav. 97 (2019), https://doi.org/10.1016/j.chb.2019.03.032.

[37] L. Cheng, R. Guo, K. Shu, H. Liu, Causal Understanding of Fake News Dissemination on Social Media, 2021, https://doi.org/10.1145/3447548.3467321.

[38] K. Shu, S. Wang, H. Liu, Understanding User Profiles on Social Media for Fake News Detection, 2018, https://doi.org/10.1109/MIPR.2018.00092.

[39] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, H. Liu, Unsupervised Fake News Detection on Social Media, A generative approach, 2019, https://doi.org/10.1609/aaai.v33i01.33015644.

[40] A. Moscadelli, et al., Fake news and covid-19 in Italy: results of a quantitative observational study, Int. J. Environ. Res. Publ. Health 17 (16) (2020), https://doi.org/10.3390/ijerph17165850.

[41] D.P. Calvillo, B.J. Ross, R.J.B. Garcia, T.J. Smelter, A.M. Rutchick, Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it), Social Psychological and Personality Science 11 (8) (2020), https://doi.org/10.1177/1948550620940539.
[42] S. Kim, S. Kim, The crisis of public health and infodemic: analyzing belief structure of fake news about covid-19 pandemic, Sustainability 12 (23) (2020), https://doi.org/10.3390/su12239904.
[43] C.M. Greene, G. Murphy, Individual differences in susceptibility to false memories for COVID-19 fake news, Cognitive Research: Principles and Implications 5 (1) (2020), https://doi.org/10.1186/s41235-020-00262-1.
[44] A. Lawson, H. Kakkar, Of pandemics, politics, and personality: the role of conscientiousness and political ideology in sharing of fake news, J. Exp. Psychol. Gen. (2021), https://doi.org/10.1037/xge0001120.
[45] M. Montesi, Understanding fake news during the Covid-19 health crisis from the perspective of information behaviour: the case of Spain, J. Librarian. Inf. Sci. 53 (3) (2021), https://doi.org/10.1177/0961000620949653.
[46] P.L. Liu, L.V. Huang, Digital disinformation about COVID-19 and the third-person effect: examining the channel differences and negative emotional outcomes, Cyberpsychol., Behav. Soc. Netw. 23 (11) (2020), https://doi.org/10.1089/cyber.2020.0363.
[47] S. Laato, A.K.M.N. Islam, M.N. Islam, E. Whelan, What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? Eur. J. Inf. Syst. 29 (3) (2020) https://doi.org/10.1080/0960085X.2020.1770632.
[48] I. Alvi, N. Saraswat, Information processing - heuristic VS systematic and susceptibility OF sharing covid 19-RELATED fake news ON social media, Journal of Content, Community and Communication 12 (2020), https://doi.org/10.31620/JCCC.12.20/06.
[49] V. Balakrishnan, K.S. Ng, H.A. Rahim, To share or not to share – the underlying motives of sharing fake news amidst the COVID-19 pandemic in Malaysia, Technol. Soc. 66 (2021), https://doi.org/10.1016/j.techsoc.2021.101676.
[50] M. Sap, et al., Developing Age and Gender Predictive Lexica over Social Media, 2014, https://doi.org/10.3115/v1/d14-1121.
[51] C. Raj, P. Meel, ARCNN framework for multimodal infodemic detection, Neural Network. 146 (2022), https://doi.org/10.1016/j.neunet.2021.11.006.
[52] C. Raj, P. Meel, ConvNet frameworks for multi-modal fake news detection, Appl. Intell. 51 (11) (2021), https://doi.org/10.1007/s10489-021-02345-y.
[53] S. Singhal, R.R. Shah, T. Chakraborty, P. Kumaraguru, S. Satoh, SpotFake: a multi-modal framework for fake news detection, in: Proceedings - 2019 IEEE 5th International Conference on Multimedia Big Data, BigMM 2019, 2019, pp. 39–47, https://doi.org/10.1109/BigMM.2019.00-44.
[54] P. Meel, D.K. Vishwakarma, HAN, Image Captioning, and Forensics Ensemble Multimodal Fake News Detection, vol. 567, Information Sciences, 2021, https://doi.org/10.1016/j.ins.2021.03.037.
[55] N. Saini, M. Singhal, M. Tanwar, P. Meel, Multimodal, Semi-supervised and Unsupervised web content credibility analysis Frameworks, in: Procedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020, Iciccs, 2020, pp. 948–955, https://doi.org/10.1109/ICICCS48265.2020.9121005.
[56] K. Shu, et al., Combating disinformation in a social media age, Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 10 (6) (2020), https://doi.org/10.1002/widm.1385.
[57] M. Karami, T.H. Nazer, H. Liu, Profiling Fake News Spreaders on Social Media through Psychological and Motivational Factors, 2021, https://doi.org/10.1145/3465336.3475097.