Unmasking Social Media Accountability—an Exploration of Technical Anonymity, Public Expression, and Partisan Media Engagement During COVID-19

ANONYMOUS AUTHOR(S)

Online anonymity has protected individuals from the consequences of inflammatory speech, particularly in the context of the proliferation of partisan news outlets and shock events. During the COVID-19 pandemic, this phenomenon was especially prominent on Twitter, as the platform distinctly enabled contentious conversation among users with varying levels of profile anonymity. While previous literature evaluated anonymity using a singular-factor framework, this study presents a fine-grained, multi-factor framework to assess the anonymity of 9807 Twitter users, the sentiment of 12381 tweets and engagement with 98 partisan news outlets. After normalizing the metric, our analysis indicated that the increase in anonymity further negated sentiment. Engagement with polarized media increased sentiment, particularly with right-leaning news outlets. This presents social media platform developers the opportunity to intentionally improve online accountability by increasing the amount of personal information visible.

CCS Concepts: • **Human-centered computing → Social networking sites**; • **Security and privacy → Human and societal aspects of security and privacy**.

Additional Key Words and Phrases: anonymity, Twitter, coronavirus, sentiment analysis, partisanship, accountability

ACM Reference Format:
Anonymous Author(s). 2023. Unmasking Social Media Accountability—an Exploration of Technical Anonymity, Public Expression, and Partisan Media Engagement During COVID-19. In . ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Shock events prompt the proliferation of social discourse, leading individuals to voice their opinions on social media [17]. As the public transitioned to social media platforms, traditional U.S. news outlets introduced their partisan affiliations to the online community [12]. This created platforms that facilitated open discourse across the political spectrum. Given the sociopolitical polarization within the country, this shift enabled inflammatory speech which became especially prominent during the COVID-19 pandemic [14].

One theory that explains inflammatory speech is the Social Identity Model of Deindividuation Effects (SIDE) [7]. The SIDE finds that with increased perceived anonymity, individuals have a decreased sense of self-identity, referred to as deindividuation, which encourages the deregulation of social behavior. The subset of anonymity that particularly models the absence of an individual's identifying information on the Internet is defined as technical anonymity [18, 25]. This information includes name, location, and other personal details. It is also important to recognize that technical anonymity does not refer to the user's perception of their own anonymity.

While previous research examined inflammatory speech in conjunction with anonymity or the pandemic [20, 23], these studies were largely been based on experiments [23] or surveys [8]. Notably, only a few empirical studies explored the effect of technical anonymity on sentiment using real-world data [4]. In addition, most literature on anonymity

Manuscript submitted to ACM

focused on one to two factors [7, 19]. For this reason, we used a fine-grain scoring framework to represent anonymity and used real-world data from Twitter, a social media platform that allows conversation among users with varying amounts of personal information made public. Twitter was chosen as the platform in question as it played a critical role in disseminating time-critical news and facilitating public discourse in the wake of the COVID-19 pandemic [7].

Although relationships between technical anonymity and factors including controversiality [23], online disinhibition [24], and aggressive language [21] have been explored, there has been a lack of findings related to technical anonymity levels, sentiment, and engagement with partisan media. Given this gap in knowledge, we were motivated to examine the intersection of sentiment, anonymity, and engagement with partisan media during the first 100 days of the COVID pandemic. Our interest was reinforced by the news coverage of the three distinct pandemic waves during the first 100 days [15] and a study that analyzed tweeting patterns related to the pandemic [7]. Our study particularly focuses on interpreting a Twitter user's profile anonymity. Based on existing studies, our anonymity framework was developed using a combination of two main ideas of thought. The first involves several factors used to determine a user's degree of anonymity [10], while the second compares the relationship between the individual anonymity factors [8].

Previous research supported the formulation of our two core hypotheses: 1) higher technical anonymity correlates to more extreme sentiment among users and 2) engaging with more extreme partisan media correlates to more extreme sentiment among users. Both hypotheses were verified by the findings of our paper, allowing creators of social media platforms to understand how to preemptively address inflammatory speech.

2 RESEARCH DESIGN AND METHODS

Our research process was split into three parts– data extraction, anonymity framework, and sentiment analysis.

2.1 Data Extraction

We began by extracting the list of 115 Twitter media outlets from Harvard's election report in 2016 [11]. We created a dataset that included the partisanship score as defined in the report and the Twitter handle of each news outlet. As stated in the report, the scores used in the report "highly correlated" with a measure introduced by Bakshy et al in 2015 derived using Facebook data. The correlation between the two scores gave us confidence that the partisanship scores are well-supported by research and should reflect authentic partisan engagement. Accounts that were suspended, private, did not tweet about Coronavirus, or were included more than once in the dataset were removed. The removal of these accounts resulted in a final list of 98 news outlets examined in the study. The distribution of the news outlets are as follows, we partitioned them into 5 clearly defined political groups according to the report [11], such that their partisanship scores ranged (-1.0, -0.5], (-0.5, -0.2], (-0.2, 0.2], (0.2, 0.5] and (0.5 to 1.0]. These values corresponded to left, left-center, center, right-center, and right stances. We have (fill in) of left, (fill in) of left-center, (fill in) of center, (fill in) of right-center), and (fill in) of right new outlets as defined by their scores.

(fill in with frequency distribution in each group?)

Using the Twitter API, we collected 30 COVID-related threads posted by each news outlet within the first 100 days of the pandemic, all replies under each thread, and all publicly available profile information about the individuals who replied. This information included username, bio, external URL, location, profile photo, profile creation date, and verification. A sample size of 30 threads per news outlet was set in compliance to the Twitter API's pull rate limit and the central limit theorem, which states that a sample size of 30 is commonly considered a lower bound for data to produce statistically significant results [6]. We excluded profiles that were suspended or no longer accessible from our dataset. Threads with no replies were also excluded. In total, we collected (fill in later) tweets.

Before performing the data analysis, we wrangled our dataset, modifying it to optimize cohesion for data analysis. As part of this process, Twitter handles and URLs were removed as they tended to disrupt the sentiment analyzer. Special characters, numbers, and punctuation symbols were also removed so that the tweets could be parsed and broken up into individual linguistic units. After the data wrangling process, some replies no longer contained substantial content for sentiment analysis. These tweets were then discarded to avoid empty data analysis subsequently. In total, we had (fill in later) tweets in our final dataset ((fill in later) tweets were discarded).

2.2 Anonymity Framework

2.2.1 Multi-factor Framework. From existing literature, we obtained a set of factors [10] to quantify the technical anonymity of a Twitter profile. It included 1) Name, 2) Bio, 3) GeoPositioning, 4) External URL, 5) Location, 6) Profile Photo, and 7) Cover Photo.

Apart from GeoPositioning (factor 3), we included all public information from Twitter in our framework. This was in accordance with previous studies indicating GeoPositioning to be of lower significance for determining technical anonymity [10].

Anonymity Factors:

- Legitimacy of the username and name
- Presence of a Bio, External URL, Location, and Profile Photo
- User Creation Date
- Verification Status

In computing the profiles' anonymity scores, we first compared handles and account names to a combined database to check if the user provided a legitimate first or last name [5]. This database included surnames that occurred more than 100 times from the 2010 census and baby names during 1880-2021 from the Social Security Administration. We also pre-processed the handles and account names by removing special characters and digits, replacing certain digits that are commonly mapped to letters (e.g. 0 to o) and splitting it to obtain the first and last name. If the profile did not contain a legitimate first name, they received a 1, otherwise, a 0. The same scoring applied for last names.

Since profiles with more information were more likely to be identifiable, for each absence of a factor: bio, URL, or location, the profile received a 1, otherwise, a 0. For the profile photo factor, we compared the URL of their profile picture to the default profile picture, thus checking if the user had uploaded their profile photo. If the user had Twitter's default profile picture, they received a 1, otherwise, a 0. Since profiles created earlier were more likely to be identifiable, we calculated a normalized score that indicated the difference between the tweet date and the account creation date. The tweet date was based on the profile's earliest tweet in our data. Lastly, official verification of a user's profile indicated lower anonymity. In our framework, the profile received a 1 if they were not verified, otherwise, a 0.

These factors resulted in a score ranging from 0 to 8, which was then normalized to be within 0 to 1.

2.2.2 Manual Evaluation. To check if our score accurately reflected the anonymity of users, we invited (fill in) 20 students with highschool education level or above to evaluate a random sample of 370 profiles, a sample size determined by the sample size formula with finite population correction (fpc) applied to provide a 95% confidence and a $\leq$ 5% margin of error. Students who helped with evaluating the profiles are compensated at a rate of $10 per hour of work.

In addition to the 8 factors mentioned above, evaluators were asked to score the profiles based on twitter activity e.g. likes and tweets (a score out of 2, higher scores indicates more possible inauthentic activities), and general perception of the profile (a score out of 2, higher scores indicates more possible inauthentic profiles).

The AUC, which reflects the average discrepancy between our framework and our profile of the anonymity scores for the selected profiles is (fill in). The mean and standard deviation of the scores generated by our framework and manually evaluated scores are (fill in) and (fill in) respectively.

2.2.3 Comparison to Single-factor Framework. Previous literature often employed single-factor or double-factor frameworks for defining Twitter profiles' anonymity level [7, 19]. We, researchers of the paper, evaluated a random sample of 370 profiles based on the framework suggested by Peddinti et al [19]: we classified the profiles into four categories- anonymous, partially anonymous, identifiable, and unclassifiable by only looking at the user's handle and display name. We then quantified the categories into a 3-point scale and normalized it so that we can compare it to our multi-factor framework. 0 corresponds to identifiable, 1 corresponds to partially anonymous, and 2 corresponds to anonymous.

We found out that while our framework closely matched (???) evaluators' perceptions of the profiles during manual evaluation and produces a similar distribution of anonymity scores to the manual evaluation, the single-factor framework used in previous literature did not perform as well.

(fill in data to support the claim if the claim is true)

2.3 Sentiment Analysis

After computing anonymity scores for each user, we performed sentiment analysis using Vader (Valence Aware Dictionary and Sentiment Reasoner). Vader is a tool used for sentiment analysis that is lexicon and rule-based [16]. This library was regarded as the primary sentiment tool because it is shown to be an effective sentiment analysis tool for Twitter data [9].

Refer to our GitHub repository (https://github.com/XXXXX) for the source files.

3 RESULTS

In this section, we will examine and interpret the results of our experiments in relation to our hypotheses.

3.1 H1: Higher anonymity facilitates more extreme sentiment

We investigated the relationship between anonymity and sentiment. Although no clear relationship was evident in the raw data ($n = 12,382$), further analysis revealed nuanced patterns. Categorizing users into eight anonymity groups revealed a general decline in sentiment as anonymity increased. Figure 2 displays the correlation between the anonymity score which ranged from (0,1] to (7,8] and the mean sentiment. However, due to substantial standard deviations, definitive confirmation of a correlation was precluded.

3.1.1 Further Analysis of Sentiment Across Anonymity Groups.

Second, we assessed sentiment across the range of smaller anonymity groups. Notably, groups 1-2 demonstrated a decline in positive sentiment as anonymity increased, whereas groups 3–6 exhibited relatively stable sentiment levels. A more distinct shift was observed in groups 7-8, which displayed a marked increase in negative sentiment. To provide empirical support for these observations, user profiles were aggregated into larger anonymity intervals: (0,2], (2,4], (4,6], and (6,8], and an independent samples t-test was performed comparing the extreme groups. The analysis revealed a statistically significant difference in mean sentiment between groups 1-2 ($M = 0.014, SD = 0.46$) and groups

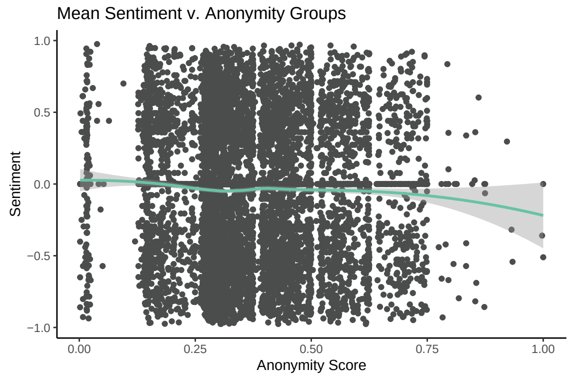

Fig. 1. **Scatter Plot of Sentiment v. Anonymity Score** The anonymity score of Twitter repliers ($n = 12,382$) exhibited no visual relationship with sentiment as displayed by the scatter plot.

7-8 ($M = -0.20, SD = 0.44$); $t(35) = 2.8, p = 0.0094$. These findings corroborate the hypothesis that higher levels of anonymity are associated with more extreme negative sentiment. We observed a general decreasing trend from the grouped data. This indicated the possibility of higher anonymity correlating with more extreme and negative sentiment. However, we found that the standard deviation was large within the groups, so we could not confirm the existence of this correlation.

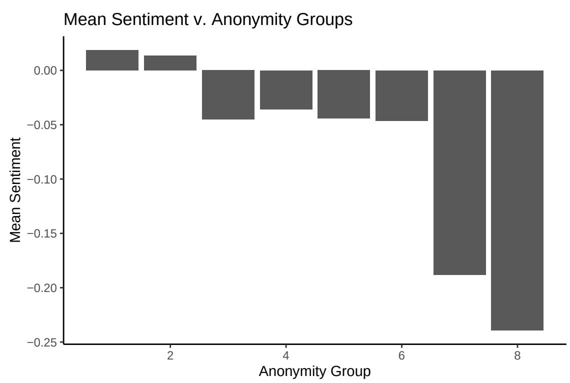

Fig. 2. **Histogram of Mean Sentiment versus Anonymity Groups** - Histogram indicated that as anonymity increased, sentiment further decreased with a sharp decline from group 6 to group 7.

3.2 Correlation between Anonymity Levels and Negative Sentiment

Figure 3 illustrates a correlation between anonymity and negativity, indicating a slight increase in negativity as anonymity levels rise. A t-test was used to compare two sets of negativity scores: the first quartile representing the lowest anonymity ($M = 0.0997, SD = 0.166$) and the fourth quartile representing the highest anonymity ($M = 0.121, SD = 0.185$), t-test: ($t = 4.177, p < 0.05$). These findings reject the null hypothesis, confirming a statistically significant distinction between the least and most anonymous quartiles in influencing negativity.

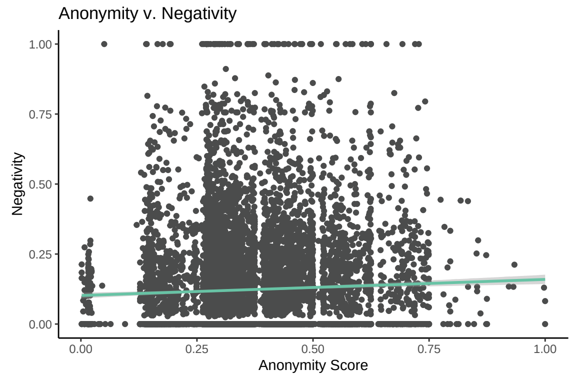

Fig. 3. **Effects of Anonymity on Negativity** - As the mean anonymity score increased ($n = 12,382$), the negativity expressed in the tweets increased.

3.3 H2: Engaging with more extreme partisan media correlates to more extreme sentiment among users

To verify our second hypothesis, we plotted sentiment score versus partisanship. From the plot of the raw data ($n = 12,382$) in Figure 5, individuals appeared to display more neutral sentiment when engaging with centrist political outlets. To further explore engagement with partisan media, news outlets were partitioned into 5 clearly defined

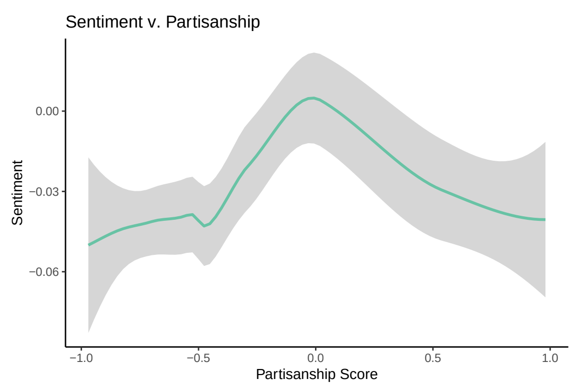

Fig. 4. **Plot of Sentiment & Partisanship** - More neutral sentiment scores were observed in tweets ($n = 12,382$) engaging with centrist political outlets.

political groups such that their partisanship scores [11] ranged (-1.0, -0.5], (-0.5, -0.2], (-0.2, 0.2], (0.2, 0.5] and (0.5 to 1.0]. These values corresponded to left, left-center, center, right-center, and right stances. While there was no linear relationship between the two variables, the graph in Figure 6 displayed a relationship between the sentiment rising to a neutral score when moving from the left to the center. As individuals interacted with more right-leaning outlets, they displayed more negative sentiment. Since the standard deviations within the groups were large, a t-test analysis further provided context to the findings. A t-test was performed to compare the mean sentiment of tweets engaging with right and left partisan media. There was no significant difference between the mean sentiment of the tweets by the least ($M = -0.051, SD = 0.45$) and most partisan groups ($M = -0.046, SD = 0.45$); $t(5847) = -0.35$, $p = 0.72$. Given

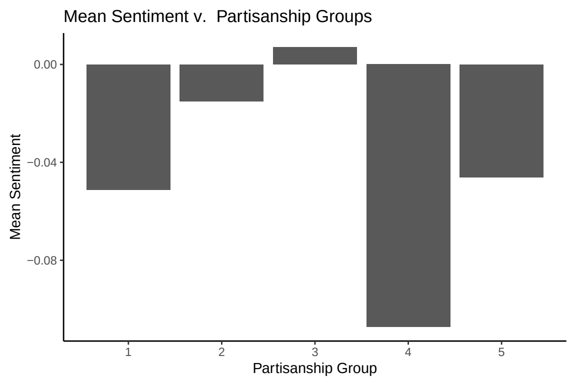

Fig. 5. **Histogram of Mean Sentiment versus Partisanship Group** - Sentiment increased to a neutral score then decreased past centrist new outlets with a final slight increase in the last group.

that $p > 0.05$, we cannot reject the null hypothesis of the absolute value comparison of highly polarized groups. We explored whether sentiment when engaging with the outer and center groups was correlated by conducting two t-tests. The first indicated that it was statistically significant with the right group ($M = -0.046, SD = 0.45$) and the center group ($M = 0.0071$, $SD = 0.46$); $t(2640) = 3.0$, $p = 0.0029 < 0.05$. The second t-test further verified that there was a statistically significant difference between the two groups with the center group ($M = 0.0071$, $SD = 0.46$) and the left group ($M = -0.051, SD = 0.45$); $t(5949) = -4.2$, $p = 0.000031 < 0.05$, indicating that engaging with more extreme partisan media correlated to more extreme sentiment among users.

3.4 Engagement with Partisan Media & Anonymity

On top of the two main hypotheses, we also observed a relationship between engagement with partisan media and anonymity. We plotted anonymity scores versus partisanship scores ($n = 12,382$) in Figure 7. An increasing trend was observed as partisanship increased. This indicated that users who engaged with extreme right-leaning (more positive partisan) content tended to masquerade themselves on Twitter to propagate their viewpoints (likely to avoid any stigma or ostracism). This difference observed, however, is quite subtle.

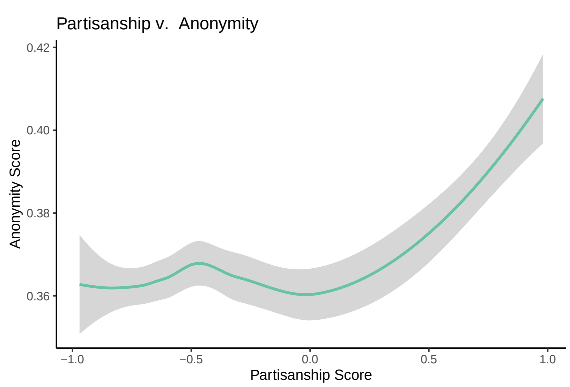

Fig. 6. **Plot of Anonymity versus Partisanship** - Anonymity is observed to increase as partisanship scores increased

Table 1. T-Test Summary

Comparison Groups	μ Sentiment	Standard Deviation	t-value	df	p-value
Groups 1-2 vs. 7-8	0.014 vs. -0.20	0.46 vs. 0.44	2.8	35	0.0094
Group 1 vs. Group 8	0.0997 vs. 0.121	0.166 vs. 0.185	4.177	-	<0.05
Right vs. Center	-0.046 vs. 0.0071	0.45 vs. 0.46	3.0	2640	0.0029
Center vs. Left	0.0071 vs. -0.051	0.46 vs. 0.45	-4.2	5949	0.000031

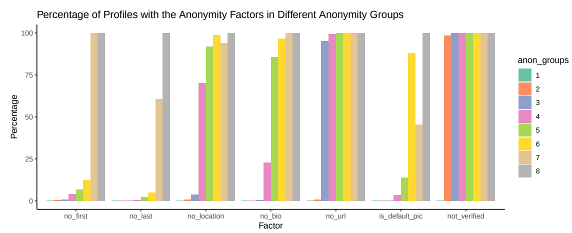

Fig. 7. **Histogram of percentage of profiles with the Anonymity Factors in Different Anonymity Groups** - External URL is the last field of user information added to a profile.

Based on Figure 8, an account was often only verified if all the other profile details were filled in. This demonstrated that Twitter's mechanism of verifying its accounts was accurate based on our anonymity framework since most verified accounts had an anonymity score of 0. In other words, the account was identified as least anonymous by our framework.

We can also gauge the general order in which users likely enter profile information using Figure 7. The order is as follows: 1) Last Name 2) First Name 3) Profile Picture 4) Bio 5) Location 6) URL.

The external URL was identified to be the last piece of information that was added to the profile, which implies that profiles with external URLs are less anonymous. External URLs are a common phenomenon among other social media platforms like Facebook, Instagram and Dating Apps as shown in Figure 9.

4 DISCUSSION

4.1 Anonymity Factor Analysis

As per Figure 8, the presence of External URLs correlated with less anonymous accounts. Thus, Twitter could reduce the presence of fake accounts by requiring this factor as method of verification. Facebook allows users to connect Instagram to their user Profile. Some dating apps also make the external URL mandatory in order to verify users when making a profile. All these forms of external links reveal additional trackable information about the user that likely lower their anonymity score.

To reduce inflammatory content on their platform, the platform managers of Twitter could thus discourage individuals from creating fake accounts by making the external URL mandatory. As a result, users would be held more accountable of their online behavior.

4.2 Implications of our Findings

Past research has linked lower levels of accountability to hateful speech [20, 23]. To address hateful speech, social media platforms have introduced filtering mechanisms [1]. However, we propose that social media developers can

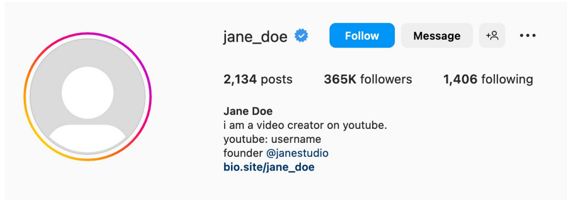

Fig. 8. **Example of external links on a Instagram Bio** - Instagram allows users to enter links, like Linktree, in their bio. This screenshot was an Instagram profile modified using HTML by the authors.

preemptively address these issues. Based on our results, platform developers can make users more accountable by potentially increasing public user information, reducing the effects of deindividuation.

Inflamed language poses a danger to society, since it has the potential to incite violence and hateful conduct [3]. Creators of social media platforms have a responsibility to regulate activity on their platform. From our findings, we theorized that the restructuring of social media platforms by manipulating the technical anonymity of users will address the fundamental causes of inflammatory behavior, creating a more amicable environment.

For developers who want to increase user's emotional expression on their platform, it is beneficial to allow higher levels of anonymity for users. This can make users feel more comfortable expressing their true thoughts without the risk of being identified. For example, Sidechat is a anonymous platform for unfiltered discourse and topical debate. In order to create such an environment, Sidechat's developers lower the amount of data input required to create a profile on the platform (only an email address is required to create an account) [2].

4.3 Limitations

Our data collection relied on U.S. Census data for name authentication, which may not recognize foreign names. On a similar note, we only extracted threads from English-Medium US-based news outlets; thus the study is U.S. centric and our findings may not be global truths. Additionally, anonymity scores of organizations may be high due to a lack of identifiable names.

During the manual verification process, we encountered a few accounts that had scores that may not be reflective of the user's true identity. This included pet accounts, fan accounts, and other examples of impersonation and fictitious creations [13]. While these accounts existed, they reflected the 0.1 % of accounts that were verifiably fake [13].

4.4 Future Work

Further research could expand on the criteria used to score anonymity. Facial recognition and analysis could be conducted on user profiles to classify the validity of the profile picture. AI Models can be used to analyze Twitter bios, differentiating vague bios from ones that reveal identifying information [22]. Through further analysis, we can get more granular data on user profile anonymity that can help us differentiate between different types of impersonation accounts (interest accounts, fan accounts, and character impersonation accounts).

Since our study was focused on the U.S., there is an opportunity to expand it to other regions. By looking at user profiles and tweets from other countries, we can understand their user polarization and behavior. This study could also be conducted on shock events other than the COVID-19 pandemic.

5 CONCLUSION

Our study provides empirical evidence supporting the validity of the SIDE Theory by analyzing real-life data. We have demonstrated a correlation between the SIDE Theory and user behavior within online platforms, employing a multi-factor framework that extends beyond previous single-framework approaches to anonymity research. Specifically, our findings indicate that heightened technical anonymity correlates with increased instances of negative and extreme sentiments among platform users [7]. Moreover, interaction with polarized content exacerbates these sentiments. Additionally, we observed that users accessing right-leaning news outlets tend to exhibit higher levels of anonymity, and partisan affiliation contributes to heightened sentiment scores, particularly in discussions characterized by elevated anonymity levels. These insights may provide the basis for platform design practices to address inflammatory speech effectively. By using this multi-factor framework and considering variables such as external links and location requirements, platform developers can foster greater user accountability and reduce the prevalence of hate speech. Our study thus provides actionable guidance for creating digital spaces that prioritize safety, inclusivity, and positivity.

REFERENCES

[1] 2021. Media Settings. https://help.twitter.com/en/rules-and-policies/media-settings. Twitter Help Center.

[2] 2022. Community Guidelines. https://www.sidechat.lol/community-guidelines

[3] 2022. Enhancing the Forensic Analysis of Human Remains. https://www.ojp.gov/pdffiles1/nij/grants/304532.pdf

[4] M. Alonzo and M. Aiken. 2004. Flaming in electronic communication. *Decision Support Systems* 36, 3 (2004), 205–213.

[5] Census.gov. 2018. 2010 Surnames. [Online] Available: https://www.census.gov/topics/population/genealogy/data/2010_surnames.html..

[6] H. J. Chang, K. Huang, and C. Wu. 2006. Determination of sample size in using central limit theorem for Weibull distribution. *International Journal of Information and Management Sciences* 17, 3 (2006), 153–174.

[7] C. C. Chen, H. Yuan, and M. Z. Yao. 2021. Identifiability as an "Antidote": Exploring Emotional Contagion and the Role of Anonymity in Twitter Discussions on Misinformation. In *International Conference on Human-Computer Interaction*. Springer, Cham, 240–252.

[8] Daegon Cho, Soodong Kim, and Alessandro Acquisti. 2012. Empirical analysis of online anonymity and user behaviors: the impact of real name policy. In *2012 45th Hawaii international conference on system sciences*. IEEE.

[9] Shihab Elbagir and Jing Yang. 2019. Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists*, Vol. 122.

[10] Zoraida Esteve, Asier Moneva, and Fernando Miró-Llinares. 2019. Can metadata be used to measure the anonymity of Twitter users? Results of a Confirmatory Factor Analysis. *International E-journal of Criminal Sciences* 13 (2019).

[11] R. Faris, H. Roberts, B. Etling, N. Bourassa, E. Zuckerman, and Y. Benkler. 2017. Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. *Berkman Klein Center Research Publication* 6 (2017).

[12] Stephanie Gleason. 2010. Harnessing social media: News outlets are assigning staffers to focus on networking. *American Journalism Review* 32, 1 (2010), 6–8.

[13] Supraja Gurajala and et al. 2016. Profile characteristics of fake Twitter accounts. *Big Data & Society* 3, 2 (2016), 2053951716674236.

[14] P. S. Hart, S. Chinn, and S. Soroka. 2020. Politicization and Polarization in COVID-19 News Coverage. *Sci. Commun.* 42, 5 (2020), 679–697. https://doi.org/10.1177/1075547020950735

[15] Alison J. Head, Steven Braun, Margy MacMillan, Jessica Yurkofsky, and Alaina C. Bull. 2020. Shape of Coronavirus Story. https://projectinfolit.org/pubs/covid19-first-100-days/shape-of-coronavirus-story/index.html.

[16] C.J. Hutto and E. Gilbert. 2014. VADER Sentiment Analysis. https://github.com/cjhutto/vaderSentiment.

[17] J. Kalyanam, M. Quezada, B. Poblete, and G. Lanckriet. 2016. Prediction and characterization of high-activity events in social media triggered by real-world news. *PloS one* 11, 12 (2016), e0166694.

[18] N. Lapidot-Lefler and A. Barak. 2012. Effects of anonymity, invisibility, and lack of eye contact on toxic online disinhibition. *Computers in human behavior* 28, 2 (2012), 434–443.

[19] Srinivas T Peddinti, Kevin W Ross, and Justin Cappos. 2017. User anonymity on twitter. *IEEE Security & Privacy* 15, 3 (2017), 84–87.

[20] K. Sharma, S. Seo, C. Meng, S. Rambhatla, A. Dua, and Y. Liu. 2020. Coronavirus on social media: Analyzing misinformation in Twitter conversations. (2020).

[21] L. Sproull. 1986. Reducing Social Context Cues: Electronic Mail in Organizational Communication. https://www.researchgate.net/publication/227445694_Reducing_Social_Context_Cues_Electronic_Mail_in_Organizational_Communication

[22] J. Vainio and K. Holmberg. 2017. Highly tweeted science articles: who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics* 112, 1 (2017), 345–366.

[23] K. Zhang and R. F. Kizilcec. 2014. Anonymity in social media: Effects of content controversiality and social endorsement on sharing behavior. In *Eighth International AAAI Conference on Weblogs and Social Media*.

[24] P. G. Zimbardo. 1969. The human choice: Individuation, reason and order, versus deindividuation, impulse and chaos. In *Nebraska Symposium on Motivation: 1969*, W. Arnold and D. Levine (Eds.). University of Nebraska Press, Lincoln.

[25] A. G. Zimmerman and G. J. Ybarra. 2016. Online aggression: The influences of anonymity and social modeling. *Psychology of Popular Media Culture* 5, 2 (2016), 181.

Received 15 January 2023