

Cross-platform Analysis of Twitter and Parler during the 2020 U.S. Presidential Election

Jaihyun Park

School of Information Sciences, University of Illinois Urbana-Champaign

JungHwan Yang

Department of Communication, University of Illinois Urbana-Champaign

Katherine Bunsold

Department of Communication, University of Illinois Urbana-Champaign

Amanda Tolbert

Department of Communication, University of Illinois Urbana-Champaign

Abstract

In the recent 2020 Presidential Election, President Trump and his campaign alleged that mail-in ballots were likely to be fraudulent and this claim stood against Twitter's efforts to curb spreading of misinformation (Lima, 2020). This claim resulted in suspending those who participated in voter fraud misinformation (Twitter, 2021), including Trump himself. In response to Twitter's action, Trump and those who supported Trump left Twitter seeking an alternative social media. This migration was a strong collective action by users who felt more than simply constrained (Kiene, Monroy-Hernández & Hill, 2016) by a loss of belonging to the community when users faced increased censorship. Those who left Twitter found Parler as an alternative social networking service, which proclaims that they allow a user to "speak freely and express yourself openly, without fear of being deplatformed for your views" as an asylum. Parler has gained attention from conservatives who are looking for alternative social media, which supposedly accepts them for who they are.

Based on this unique case, this study seeks to understand the impact of echo chambers on people's expressed opinions on social media. Past research efforts on echo chambers, selective exposure, and network homogeneity [CITE] mostly focused on a handful of popular social media, mostly either Facebook or Twitter, while neglecting the unique roles of other niche social media platforms in building online communities [CITE]. We will address this critical gap by leveraging data from two social media platforms: Parler and Twitter as examples that represent distinctive user bases in terms of political ideology. We identify users who have the same account names on both platforms and examine the role of political homogeneity in the online opinion expression and sharing of information.

We rely on the Social Identity Deindividuation Effects (SIDE) model to understand political behaviors of the users who used both Twitter and Parler. The SIDE model explains that deindividuation occurs when group norms are more salient and have a greater effect on individual behaviors than individual processes (Lea & Spears, 1992). The SIDE models focus on anonymity and explicit and implicit norms of online spaces, and supports that anonymity enhances the social influence processes and collective behavior (Spears, 2017). By applying this theoretical model, we are aiming to reveal how Parler’s homogeneous political climate – more conservative than Twitter – helped users to feel more anonymous than Twitter by providing a safe place for them to speak hatred. There are two research questions we wanted to answer. Our focus of interest are the people who used both Twitter and Parler and hereafter, they are called cross-platform users.

- RQ 1. Can we make use of machine learning technique to identify the pattern of increasing or decreasing use of toxic language by cross-platform users in Twitter?
- RQ 2. Can we make use of machine learning technique to identify the pattern of increasing or decreasing use of toxic language by cross-platform users in Parler?

Methods and Data

Username Matching

We collected comparable datasets from Twitter and Parler. For Twitter data, we collected tweets in real-time through Twitter API by using a set of keywords¹ that are related to the U.S.

¹ The keywords are listed here: ‘vote fraud’, ‘voter fraud’, ‘votes fraud’, ‘election fraud’, ‘elections fraud’, ‘rigged elections’, ‘rigged election’, ‘mail-in ballot’, ‘mail-in bal-lots’, ‘mail-in vote’, ‘mail-in votes’, ‘mail in ballot’, ‘mail in balotts’, ‘mail in votes’, ‘mail in vote’, ‘rigged elecion’, ‘rigged’. Incorrectly spelled keywords are included for the comprehensive search.

Presidential Election and voter fraud claims from November 2020 to December 2020. For Parler data, we relied on the data collected and released in Aliapoulios et al. (2021). This public dataset contains 183M Parler posts from 4M users from August 2019 to January 2021. Since Parler stopped service after the January 6th Capitol riot, we used this publicly available data source. To make the Twitter and Parler data comparable, we used the same keywords to filter the data. After this process, the size of Parler data was 539,293 postings from 160,172 unique Parler users and 56,358,632 tweets from 4,297,388 unique Twitter users. As a result of username matching, we identified 9,371 users with case-insensitive matching and their 33,565 Parler postings and 727,636 tweets were used for this study.

Toxic Language Detection

Perspective API (Wulczyn, Thain & Dixon, 2017) is the current state-of-the-art natural language processing (NLP) model to identify toxic contents (Rajadesingan, Resnick & Budak, 2020). Perspective API takes languages as input and detects a wide variety of types of toxicity such as rude and disrespectful contents (Toxicity) and negative or hateful comments targeting someone because of their identity (Identity Attack)². Perspective API returns the scores between 0 and 1 to indicate how likely the input comment is to be toxic. If the score is close to 1, it means that the given comment is likely to contain toxic language. After we collected toxic scores for each content, we averaged the score by day in order to observe the patterns of collective development of norm of using toxic language.

Findings

² <https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages>

As we collected Twitter data when the voter fraud claim has started to gain attention, we wanted to examine whether there are possible patterns of hatred statement along with the development of voter fraud claim to answer RQ1 and RQ2. Before we employ time series analysis to determine patterns, we tested Dickey-Fuller analysis to determine whether two situations (the communication environment in Twitter and Parler) of cross-platform users is applicable to time series analysis. The data should be non-stationary in order to further conduct time series analysis. Cross-platform users' toxic language was stationary ($p = 0.0002$) in Twitter, meaning that there was no pattern of increasing or decreasing during the time window of data collection. However, cross-platform users in Parler showed non-stationary ($p = 0.7338$) in Parler, which indicated that Parler contents could be subjected to time series analysis as the average toxic score shows patterns.

We used Auto-Regressive Integrated Moving Average (ARIMA) model to predict whether there is a trend over a given period of data. We trained the differenced first-order autoregressive model to use one lag observation ($p=1$) and differentiated the raw observations once ($q=1$). While other possible independent variables that may impact the increase or decrease in toxicity score, such as the number of users ($p > 0.05$) and the number of tweets ($p > 0.05$) are irrelevant to explain the trend of increasing as shown in Figure 1. We could find that the toxicity score and identity score show a pattern of changes compared to the beginning of the voter fraud claim.

Conclusions and Future Plans

In this study, we revealed that cross-platform users use language differently across Twitter and Parler. While toxic language of cross-platform users in Twitter was stationary, indicating it was

hard to find correlation between maturation of voter fraud claim and toxic language, it was found that cross-platform users in Parler used more toxic language as voter fraud claim develops. Our findings may suggest that SIDE model works differently depending on what social media users are using. Twitter is a place where both right and left-leaning users communicate but Parler created an echo chamber for right-leaning users, which may increase the social norm that they are together through the use of toxic language regarding voter fraud claim. We plan to validate the Toxicity score and Identity Attack score with qualitative coding process to validate the scores we obtained from Perspective API are valid. The next step to understand the different language use by cross-platform users is to compare the toxicity level when cross-platform users talk to cross-platform users and when cross-platform users talk to users who are not found in Parler in Twitter. This may help us understand whether cross-platform users change their code when they talk to those who are not using Parler. This research question can be pertinent to understand code-switching (Gumperz, 1982; Zentella, 1997) behavior of users who use multiple social media platforms.

References

- Aliapoulios, M., Bevensee, E., Blackburn, J., Bradlyn, B., De Cristofaro, E., Stringhini, G., & Zannettou, S. (2021, May). A Large Open Dataset from the Parler Social Network. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 15, pp. 943-951).
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.
- Gumperz, J. J. (1982). *Discourse strategies*. New York: Cambridge University Press.
- Kiene, C., Monroy-Hernández, A., & Hill, B. M. (2016, May). Surviving an "Eternal September" How an Online Community Managed a Surge of Newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1152-1156).
- Lea, M., & Spears, R. (1992). Paralanguage and social perception in computer-mediated communication. *Journal of Organizational Computing and Electronic Commerce*, 2(3-4), 321-341.
- Lima, C. (2020) Twitter plans clampdown on mail-in voting misinformation amid friction with Trump. *Politico*.
- Rajadesingan, A., Resnick, P., & Budak, C. (2020, May). Quick, community-specific learning: How distinctive toxicity norms are maintained in political subreddits. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 557-568).

Spears, R. (2017). Social identity model of deindividuation effects. *The International Encyclopedia of Media Effects*, 1-9. <https://doi.org/10.1002/9781118783764.wbieme0091>

Twitter (2021) Civil integrity policy. *Twitter Help Center*.

Wulczyn, E., Thain, N., & Dixon, L. (2017, April). Ex machina: Personal attacks seen at scale.

In *Proceedings of the 26th international conference on world wide web* (pp. 1391-1399).

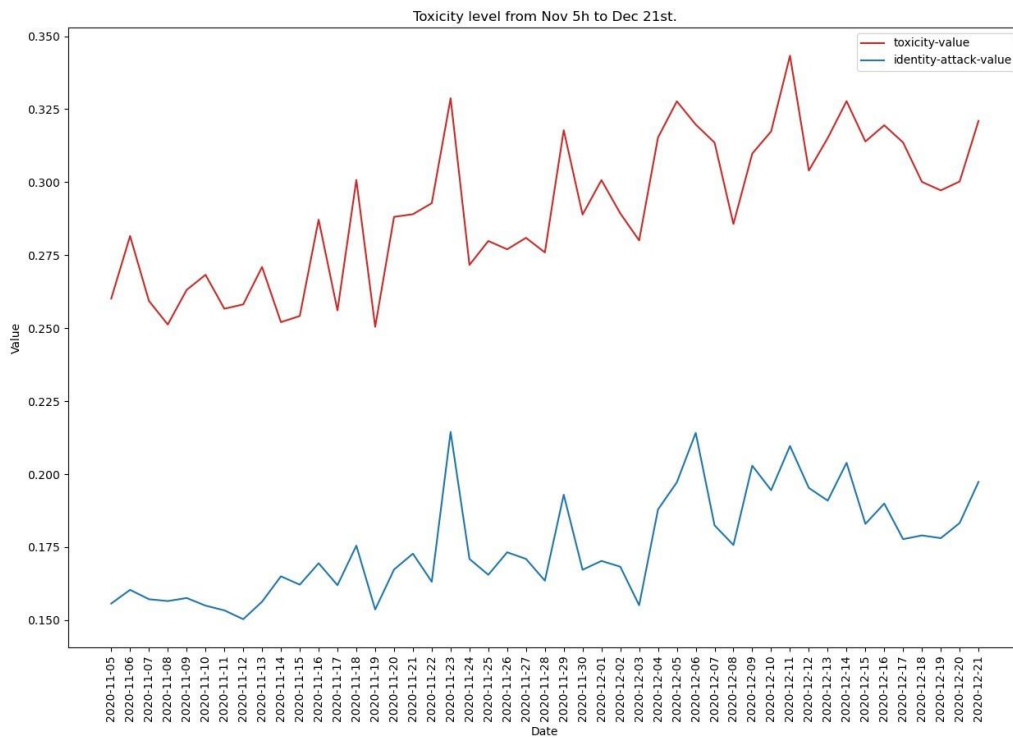
Zentella, A. C. (1997). *Growing up bilingual: Puerto Rican children in New York*. Malden: Blackwell Publishers.

Table 1*ARIMA Results*

	Coef	Std err	z	P> Z	0.025	0.975
constant	0.0010	0.002	0.558	0.577	-0.03	0.005
ar.L1	-0.5497	0.123	-4.484	0.000	-0.790	-0.309

Figure 1

Toxicity score and Identity Attack score of contents created by cross-platform users in Parler



Note. The Toxicity score and Identity Attack score show a similar trend from November 5th to December 21st.