Analysing The Activities Of Far-Right Extremists On The Parler Social Network

James Stevenson
Matthew Edwards
Awais Rashid

Bristol Cyber Security Group
School of Computer Science
University of Bristol
Bristol, UK
Email: {my19303,matthew.john.edwards,awais.rashid}@bristol.ac.uk

Abstract—A significant gap remains in our understanding of the types of users who utilise online extremist platforms, as well as how their activity on these platforms influences the radicalisation of others and the dissemination of extremist content online. Our research addresses this gap by focusing on the Parler social network, one of the largest social media platforms used by the extreme far-right, boasting a reported 15 million total users as of January 2022. We present an exploration of the Parler social network, specifically reviewing the roles of users in the network and the types of activity and content shared on the platform. Our methodology provides a novel examination of Parler using tools that have previously been tested to understand other extremist groups.

I. INTRODUCTION

In 2019, The Global Terrorism Index, published by the Institute for Economics and Peace (IEP), reported a 320% rise in the total number of far-right terrorism incidents in the West, particularly in Western Europe, North America, and Oceania [16]. This threat is not only present within the US, but also in other western nations such as the UK. In 2021, MI5 Director General Ken McCallum stated that while right-wing terrorism was not at the same scale as Islamic terrorism, it was, however, growing, and of the 29 late-stage attack plots between 2018 and 2021, 10 had been extreme right-wing based [20].

With the rise of internet communication, a large proportion of extremist activities now utilises online communities for communication, including sharing extremist content, radicalising others, and planning terrorist actions. In Q1 of 2023 alone, Facebook took action on approximately 14.5 million pieces of terrorist content on their platform [27]. However, due to the ease of creating such radicalised content on these platforms, these extremist communities continue to be popular and grow.

As of January 6th, 2021, Parler had approximately 15 million total users [14], and the platform was also endorsed by several conservative public figures, encouraging people to migrate from traditional social networks [23]. Due to these factors, many news outlets and researchers debate the platform's importance and role in the January 6th riot [25], and in turn, its significance as a communication hub for far-right extremists. Given these circumstances, our method focuses on the Parler social network as a cornerstone and representative sample of extreme far-right social media usage.

While previous research contributions have broadly explored the types of far-right extremist groups and provided a first-look analysis of platforms such as Parler and Gab, this study provides novel contributions and key insights by answering the following research questions:

- Can methods used to identify Islamic extremism text posts be applied to far-right extremism?
- What are common activities performed by users on the Parler social network?
- Is the Parler social network primarily used to share hate speech and far-right extremism?

II. RELATED WORK

Much research has already been performed by the social science research community into far-right extremism. This has included the effects of far-right extremism on society [22] [29] [31], the differences between far-right extremist groups [7] [9] [12], and how far-right extremism has risen in popularity [1]. Much work has also been performed by the computer science research community in using data driven approaches to identify terrorism online [10]. This paper builds upon these approaches by both developing a machine learning model for the automated identification of far-right extremist content online, and by exploring the types of activities performed by users on the Parler social network, helping explain the various uses of far-right online platforms.

Previous research has attempted to understand the structure and population of far-right extremist platforms, particularly Gab, Parler, and Stormfront [19], [30]. Aliapoulios et al. [2] present a dataset of 183M Parler posts made by 4M users between August 2018 and January 2021, as well as metadata from 13.25M user profiles. In their exploration they provide preliminary information surrounding it, including: bi-grams found in Parler users bios, badges assigned to user profiles, top 20 hashtags in posts and comments, and other related statistics Other work has investigated the Gab social network [19], [32]. Similar to Parler, Gab is a social network boasting free speech and is used in part as an echo chamber for radicalisation among far-right extremists [8]. Zannettou et al. [32] reported general statistics from Gab posts and users as part of their work. In addition to this, research on Gab has also attempted to understand Gab's place in the online misinformation and news ecosystem.

Previous research by Nouh et al. [23] focused on the identification of Islamic extremism in Twitter posts. They used three main feature categories: radical language (TF-IDF scores for each n-gram in the Dabiq extremist corpus), psychological signals (LIWC dictionary scores and the Minoski distance between message scores and the Dabiq dataset), and behavioural features (degree centrality of users in the dataset, follower frequency, and post frequency).

III. DATASETS

Two main datasets were required as part of our methodology. The first is a dataset of far-right extremist posts from the Stormfront [30] neo-Nazi Internet forum[1], and the second is a dataset [2] from the Parler[2] social network.

A. Stormfront data

The Stormfront dataset contains approximately 5000 unprocessed forum posts originating from the Stormfront platform [30] (originally published between 2002 and 2017). As this dataset is unfiltered, some forum posts are not related to far-right extremism. However, due to the density of far-right extremist content (after a manual review of a random sample), it is hypothesised that this dataset is applicable for the use of setting a baseline for radical language moving forward. Unlike the Parler dataset, this dataset is solely used to develop a radical language corpus that is used as part of identifying far-right messages and posts. This dataset is used for creating a far-right radical language corpus when creating a binary classifier, as discussed in the following sections.

B. Parler data

This dataset includes a total of 183M Parler posts made by 4M users between August 2018 and January 2021, as well as metadata from 13.25M user profiles [2]. Multiple samples

[1]Stormfront is a far-right extremist and neo-Nazi Internet forum, known as the Internet's first major racial hate site [3]

[2]An online social network, self-dubbed as a free-speech and unbiased platform, has a high proportion of conspiracy theorists and far-right extremist users [5]

were drawn from this dataset, as detailed in the relevant portions of Section IV below.

IV. METHOD

We define online-enabled far-right extremism as *the use of social media to spread, incite, demonstrate, or plan activities motivated by an individual's extreme right-wing beliefs*. In particular, this covers:

The use of social media
> Our research focuses on individuals who use social media platforms, such as Parler, Stormfront, and Gab.

Spread, incite, demonstrate, or plan activities
> Our work focuses on the use of social media to share information about previous or current political violence, or to encourage others to perform such acts in the future.

Motivated by an individual's extreme right-wing beliefs
> Our work focuses on actors motivated by extreme far-right beliefs that fall outside of the socio-political norm. These positions may include white supremacism, right-wing fascist ideologies, and various right-wing conspiracy theories.

Our methodology can be broken down into four stages:

1) Building a classifier to identify far-right extremism.
2) Analysing the prevalence of far-right extremism in posts by the most influential Parler users, and in samples of ordinary users.
3) Analysing the overall levels of engagement with extremist content on Parler.
4) Manually identifying distinctive user behaviours visible within Parler.

A. Classifier design

In order to develop a classifier for far-right extremist content [6], we built upon the work of Nouh et al [23], which was developed in the context of Islamic extremism. We first replicated their work on a comparable dataset, and then applied the same methodology to far-right extremism, using Stormfront instead of Dabiq as a reference repository of extremist material. We then deployed our classifier on the Parler social network.

1) Replication of Previous Work: We adopted Nouh et al. [23]'s approach as a baseline due to previous research having highlighted similarities between types of extremism [4], [17], [18]. Our aim was to test whether this similarity would mean methods transfer across the two domains. Replicating Nouh et al.'s method, using a dataset as similar as possible to their original work, we achieved scores within a 0.1 margin in Accuracy, Recall, Precision, and F-Measure of the original published results. These results can be seen in Table I.

2) Transfer to Far-Right Extremism: In order to extend the classifier for far-right extremism, the following steps were taken. Our aim was to closely align with the approach of Nouh et al., while utilizing source data for far-right rather than Islamic extremists:

TABLE I
REPLICATION RESULTS FOR MODEL IDENTIFYING ISLAMIC EXTREMISM

Type	Accuracy	Recall	Precision	F-Measure
All	0.999	0.998	1.0	0.999
Behavioural	1.0	1.0	1.0	1.0
Psychological	0.872	0.740	0.832	0.783
Language	0.827	0.676	0.739	0.706

1) A dataset of known-bad and known-good user posts was created - this dataset was created from a mix of manual labelling and keyword selection using log-liklihood metrics.

2) Parler posts were then represented using features from three main categories: Radical Language, Psychological Signals, and Behavioural Features. These featuresets include word vector embeddings, capital letter and word frequencies, LIWC dictionary scores, Minkowski distance, and user degree centrality scores.

3) A binary classifier was trained using these features to detect far-right messages.

Labelling: Initially, the Parler dataset needed to be divided into a known-good and known-bad dataset (i.e., a dataset of known far-right extremist data and a dataset of known non-far-right extremist data). To accomplish this, 100 posts from the Parler dataset were manually labelled as far-right extremist by a researcher familiar with far-right terminology and rhetoric. Using Log-likelihood [24][21], the frequencies of tokens and ngrams in this corpus were ranked relative to the rest of the text. Subsequently, the top 30 keywords most associated with extremist labelling were selected from this method: *genocidal, fire, destroyers, democraticnazi, fucker, tribunals, invoke, squad, punch, tyrannical, die, treason, traitors, firing, military, death, armed, removed, hung, shoot, proudboys, cowardly, killed, fight, scum, hang, civilwar, executions, whipping,* and *hanged*. Posts from the Parler dataset that contained at least one of these keywords were labelled as extremist. Features were then extracted from the Parler posts as illustrated in Figure 1.

Using the above keywords, posts from the Parler data-set were selected that contained at-least one of these keywords (this was defined as the far-right extremist data-set) and secondly posts were selected that did not include these keywords (these were defined as the baseline data-set).

Radical Language Features: Building upon the methodology by Nouh et al., posts published on the far-right forum Stormfront were used to establish a baseline for how far-right extremist messages are constructed. Two feature groups are employed in the radical language category: word vector embeddings and language indicators.

To obtain word vector embeddings, TF-IDF scores for each n-gram (uni-grams, bi-grams, and tri-grams) in the Stormfront corpus were calculated. The top-scoring results were used to train a word2vec model (utilizing the word2vec model implemented in the gensim package). The feature is the
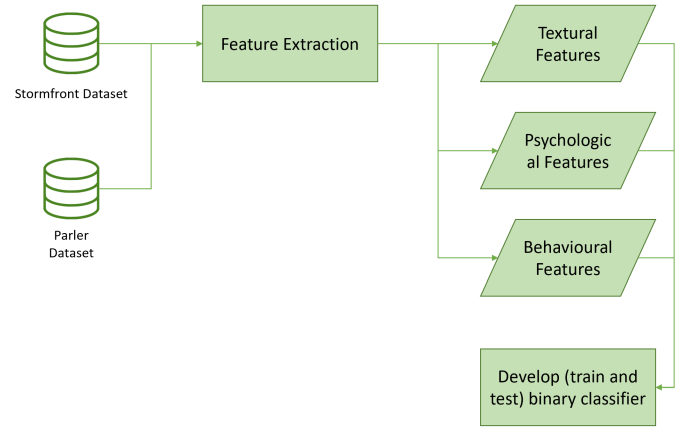

Fig. 1. Creation of far-right extremism binary classifier

outcome of averaging the vectors of each word, combined with the maximum and average for each word. Consequently, each post is represented by a 200-dimension vector.

Our language indicators were inspired by previous work. Previous literature has noted that capital word frequency is helpful in identifying "yelling behaviour" [23]. Furthermore, the use of, radicalized, and terrorist dictionaries can aid in identifying behaviour. For this purpose, several bad-language, hate speech, and word dictionaries were utilized [13], [15].

Psychological Signals: Research in behavioural residue and digital footprinting [11] suggests that individuals leave indicators of their personality online based on their day-to-day choices, including word choice. A considerable amount of past research [26] has highlighted how terrorists and extremists may have different personalities compared to non-extremists. This feature group focuses on these word choices by using LIWC annotation of both the Parler and Stormfront datasets, following Nouh et al. [23]'s methodology.

We used a subset of the scores for LIWC dictionary groups directly as features. This included the categories *Clout, Analytic, Tone, Authentic, Anger, Sadness, Anxiety, Power, Reward, Risk, Achievement, Affiliation, I Pronoun,* and *P Pronoun*. In addition, the Minkowski distance was calculated between a Parler post's LIWC dictionary scores (for the categories above) and the average LIWC scores for Stormfront forum posts.

Behavioural Features: Behavioural features relate to how a specific individual acts and portray themselves online, which includes interactions such as who they follow, what they post about, and how often they post. To capture how users interact with other users, we used a mention interaction graph. Undirected Edges were created between users when a user's post mentioned the other. After all users' posts were added to the graph, the degree of influence each user has over the network is calculated using the degree centrality method.

Training: The sklearn Python package was used to train a Random Forest classifier, configured with 100 estimators, a maximum depth of 50, and out-of-bag samples enabled. To train the dataset, an 80/20 split was implemented between the

training and testing dataset. Accuracy, Recall, Precision, and F-Measure were used as evaluation metrics. In this context, Recall highlights the number of far-right posts identified, while Precision shows the number of posts that can be identified without a false-positive. F-Measure represents the average of Precision and Recall.

B. Application to Analysis of High-Centrality Users

We processed all posts between the date ranges of October 2018 and May 2020 (19 months) - this date range was chosen due to the influx of actvity in later months and the time constraints attached to processing a larger dataset. As part of our methodology, a graph was created for each month. These graphs were formed by adding a node for each user who created a post during each time period, where Un-directed edges were added when a user mentioned another user in their post. Degree centrality was then calculated for each node for each date, and the top 10 users per month with the highest degree centrality were extracted[3].

A random subset of 25 posts for each high-centrality user were classified using our extremism classifier, as well as a toxicity and hate speech classifier [28]. In addition, a qualitative manual review was performed on each of these posts to identify and label behaviours discussed within our findings. This process is detailed in Figure 2.
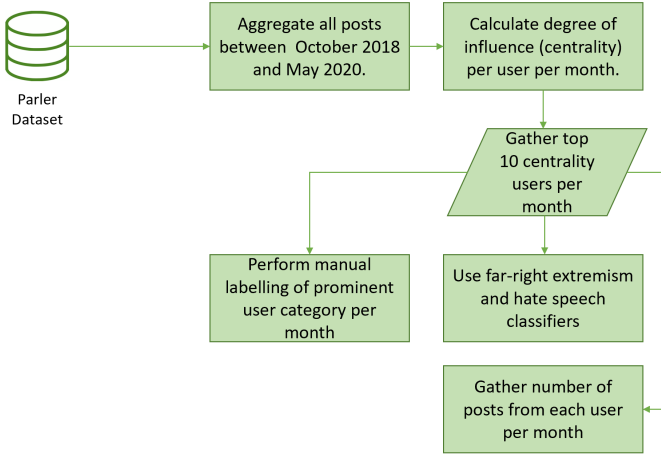

Fig. 2. Analysis of high centrality users

C. Breakdown of Extremist Post Metadata

To further examine the behaviour of Parler users, a random sample of 14,151 user posts were sampled from the Parler dataset. These posts were posted on January 6, 2021, and a year earlier on January 6, 2020. As detailed in Figure 3, all posts were classified using our extremism classifier and labelled for toxicity [28]. Words and hashtags were then counted in relation to their frequency in posts identified as extremist. Metadata was gathered from extremist classified

[3]Barring the months 2018-10 and 2018-11, in which only 5 and 2 users with edges respectively were present in the data.

Type	Accuracy	Recall	Precision	F-Measure
All	0.789	0.697	0.839	0.762
Behavioural	0.693	1.0	0.610	0.758
Psychological	0.742	0.691	0.755	0.722
Language	0.736	0.627	0.784	0.697

posts, including the number of following, followers, upvotes, impressions, and user posts. This data was gathered into ranges of 100 (i.e., 0-100, 101-200, etc.).
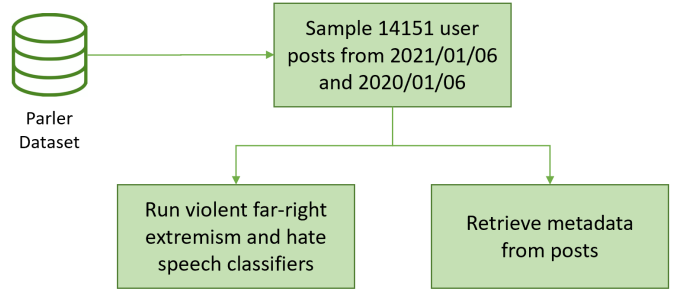

Fig. 3. Breakdown of extremist post metadata

V. FINDINGS

A. Classifier Evalution

Our classifier was trained and evaluated in four separate scenarios: using all available extracted features, using the radical language features, using the behavioral features, and using the psychological features. These results can be seen in Table II. It can be seen that the best performance is achieved by using all feature groups. However, it should be noted that there is no standalone best feature group: language features show the highest precision, psychological features present the best accuracy, and behavioural features have an anomalously high recall. The overall decay in performance relative to our replication results in Table I does demonstrate that the model does not transfer perfectly to this new domain.

Table III shows the individual weights of each feature when all features were in use. Here it can be seen that the message vectors feature holds the highest significance to the model, followed by the degree centrality of the user. The dominance of the textual embedding as a feature is expected, but the relatively high weighting of the behavioural centrality measure is interesting, showing that within Parler centrality contributes to predicting extremist content production.

B. Sub-Groups and Behaviours on Parler

We further analyzed how communities, individuals, and the broader platform communicated on the Parler social network. The details below highlight several key activities observed among high degree centrality users on the Parler network.

TABLE III
FEATURE WEIGHTS FOR MODEL IDENTIFYING FAR-RIGHT EXTREMISM

Type	Feature Weight
anxiety	0.15%
sadness	0.18%
risk	0.29%
reward	0.54%
achievement	0.69%
affiliation	0.83%
cap freq	0.90%
i_pronoun	1.18%
authentic	1.26%
tone	1.27%
p_pronoun	1.38%
clout	1.45%
violent freq	1.46%
analytic	1.64%
power	1.96%
minkowski	2.69%
anger	3.81%
centrality	4.56%
message vectors	73.76%

These activity types appear to be common but are not necessarily shared among all users. These activities were gathered as part of manual observations.

1) Conducting: Conducting[4] is an activity observed on Parler, where users aggregate lists of other users for their followers to follow. The methods for aggregating this list of users to follow differ between conductors; they may use automated scripts, select users who follow or mention them, or simply choose users with whom they politically, socially, or morally align. In addition to sharing these conducting posts, users who follow a Conductor behavior have also been observed having normal conversations with other users, responding to people asking to be added to a conductor post, and responding to posts from other conductors.

One user we categorized as a conductor was observed to share more than five posts per day; each post contained a similar brief introduction about the posts, the total number of conducting posts they had shared, and then a newline-separated list mentioning a series of users. Due to the nature of these conducting posts, unless the conductor engages in another activity type, they rarely share extremist content.

2) Hate Speech and Inciting Violence: A significant percentage of the top degree centrality users on Parler (although not all) share some form of hate speech or incite violence in some way. This may include anti-minority, anti-Islamic, anti-Jewish, or anti-Chinese hate speech (typically relating to COVID-19).

---

[4]The term conductor comes from the notion of the posts these users share being 'train-like', and sharing a list of 'passengers' for others to follow (though some users may refer to them differently).

This behavior can vary in intensity, ranging from users vaguely inciting others to 'take up arms' or 'stand up and fight' to more extreme posts that single out individuals or groups to be 'hung', 'shot', and 'killed'. This behavior was especially prominent towards January 2021 when many left-wing political figures were targeted with such threats.

3) Parler Account: This type of activity broadly defines accounts managed by people associated with Parler, including the Parler official account, the creators of Parler, administrators, and support accounts. These accounts often appear in the top degree centrality as users mention them for support. Additionally, there was a time when the main Parler account welcomed users to the platform, which significantly increased the account's centrality on the platform.

4) Conspiracy Theorists: A sizable portion of the Parler platform is devoted to the sharing and distribution of conspiracy theories and misinformation. For example, one user was observed sharing a post about the US 2020 election results, claiming they believed the results to be faked. In response to this post, many other users were observed sharing similar or contesting beliefs. Conspiracy rhetoric is also commonplace in regards to other topics including COVID-19 and religions such as Islam.

5) Political, News, or Other Commentary: This type of activity relates to the sharing of news, political opinions, and general commentary about world, national, or public affairs that do not fall into another category of activity. For instance, a user may share their opinion on a by-election without including hate speech, far-right extremist, or violent speech. Some political or news commentators also share conspiracy theories, hate speech, and other types of activity. For example, one user was seen sharing political commentary about the 2020 US presidential election results and also claiming specific members of the Democratic party should be killed.

6) News or External Social Media Aggregation: Some accounts on Parler, which appear to be bots (i.e., users attached to a computer program that automates their activities), but may not necessarily be so, often engage in re-sharing posts from other social networks like Twitter or sharing news articles verbatim. Due to the nature of these accounts, which share verbatim news sources from external media, they rarely shared extremism or hate speech.

7) Anti-Typical-Parler User / Anti-Far-Right: Around January 2021, Parler gained significant attention not only from like-minded right-wing individuals and the media, but also from left-wing individuals looking to challenge right-wing beliefs. As a result, over time, Parler began attracting more self-defined anti-Trump, anti-Republican, and anti-right-wing users. For instance, one user was observed denouncing the 'Trumpublican narrative', calling such individuals brainwashed, and generally challenging right-wing beliefs. While some of these users may spread hate, incite violence, and disseminate misinformation, they are not the focus of this paper due to their political beliefs falling to the left of the spectrum and thus not falling into the far-right extremist definition.

C. Analysis of High-Centrality Users

Figure 4 depicts the observed rates of common Parler activities over time and the changes in such behaviour in the most influential users during each month. It can be seen that there is a steady distribution of activities throughout the date range and most activity types remain relatively stable. Some other categories of content, like non-English posts, emerge in the top-centrality userbase activity only in the months after May 2019.

A moderate positive correlation (0.345) was observed between the number of posts from a user and their degree centrality, indicating that users who post more frequently tend to have higher centrality (degree of influence) in the network. For all high-centrality users, we analysed the rates of extremist posting alongside toxicity measures. Figure 4 shows the comparison of a user's average toxicity rating and average extremist post output across all sampled posts (25 posts were randomly sampled for every month a user was in the top centrality bracket). The correlation coefficient between toxicity and extremism is 0.298. This indicates a weak positive correlation, suggesting that as the average toxicity increases, the rate of extremist post production also tends to slightly increase. However, as the correlation is not strong, there are likely other factors at play influencing these values – extremists are not necessarily posting toxic content.
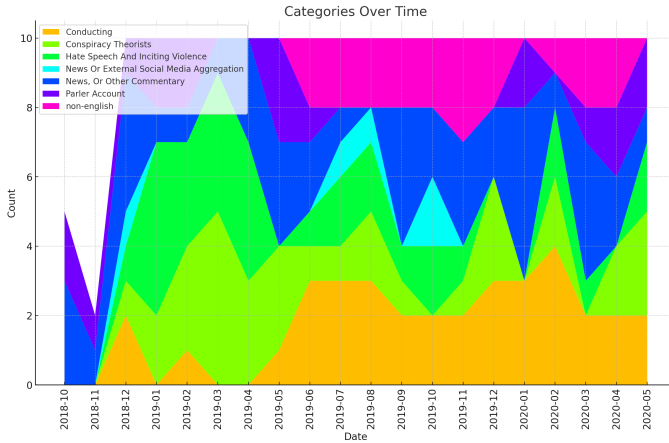

Fig. 5. Cross reference of the average extremism and toxicity (in all top 10 centrality users) over time

Fig. 4. Spread of categories present in the top 10 centrality users per month, over time

D. Extremist Post Metadata

In total, 4,164 posts were identified as extremist (29.46%) by our extremism classifier. Furthermore, there was an average toxicity of 0.207% across all posts identified as extremist, compared to an average toxicity of 0.164% across posts not identified as extremist and an average toxicity of 0.177% across all posts.

The list of the top 20 most frequent hashtags in extremist posts was compared to the overall most frequent hashtags in Parler [2]. 55% of the overall top 20 hashtags overlapped with our list of the topmost 20 hashtags, again highlighting
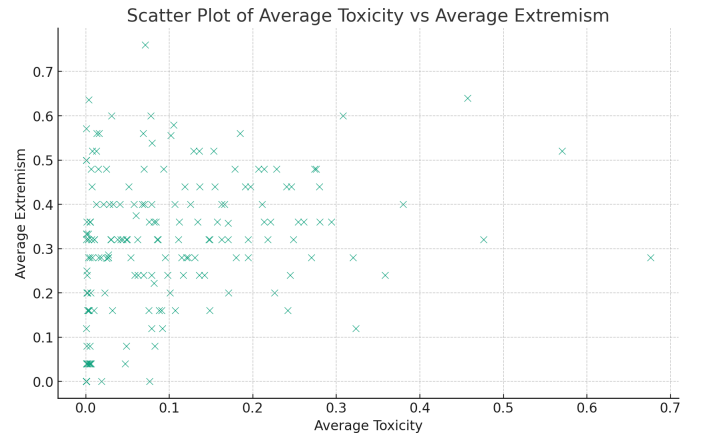
the degree centrality of this extremist behaviour on Parler. Hashtags that were present in the extremist post list and not present in the overall list include: '#Patriots', '#WeThePeople', and '#FightBack'.

Table IV presents the frequency of impressions, followers, following, and author posts for users that produce extremist content. We found that the highest individual ranges for all threshold categories are 0-100, 101-200, and 201-300. However, collectively, the 301+ range outweighs all of these other ranges. This indicates the significant amount of interaction that extremist posts received on Parler.

TABLE IV
FOLLOWER, FOLLOWING, IMPRESSION, AND AUTHOR THRESHOLDS FOR
EXTREMIST POSTS

	# 0-100	# 101-200	# 201-300	# 301+
Followers	1453 (10.267%)	378 (2.671%)	264 (1.865%)	12056 (85.195%)
Following	4064 (28.718%)	23 (0.162%)	11 (0.077%)	10053 (71.040%)
Impressions	3018 (21.327%)	309 (2.755%)	168 (1.187%)	10657 (75.309%)
Posts	825 (5.829%)	343 (2.423%)	222 (1.568%)	12761 (90.177%)

VI. DISCUSSION

This study provides a detailed examination of the Parler social network, investigating common activities and identifying far-right extremism on the platform.

The results from our analysis reveal several findings that address the research questions outlined in this paper. It is evident that the network is not solely used for sharing or practicing far-right extremism but also serves as a regular social network where like-minded users can share and discuss information within a community of individuals who hold similar beliefs. However, it is apparent that Parler's lack of moderation has resulted in hate speech and extremist rhetoric becoming common across the platform, with at the very least users with the highest degree of influence commonly sharing such violent and extreme rhetoric. The focus on violent extremism is further reflected in the use of popular violent and extreme hashtags among extremist and non-extremist posts.

The analysis of popular words and hashtags in extremist posts on the Parler platform reveals that many of these terms, such as '#StopTheSteal', 'Election', 'Trump', and 'KAG' (Keep America Great), are associated with American right-wing terminology. However, it is worth noting that only 29.46% of the sampled posts were identified as extremist, indicating that the use of Parler is multifaceted. The platform serves as a meeting place for like-minded right-wing users, a venue for sharing right-wing views and messaging alongside, consequently due to the lack of moderation, a space for sharing violent and hate speech messages. The identification of this last behavior is particularly concerning and influential for this research.

This study also presents a range of tools and techniques for identifying extremist user groups, such as those on the Parler social network, that pose a risk to society with their extremist views. Some of these identification techniques are derived from previous tools and methods designed for other forms of extremist groups, such as Islamic extremism. The study demonstrates that methods used for identifying one group of extremist users can be applied to other forms of extremism, although there are differences in performance between the two domains.

Furthermore, it is observed that several users appear in the top 10 degree centrality users over multiple months (with the average top user appearing 1.85 times and the most central figure appearing 12/19 times). Many of the most influential users being present over several months appear to be either 'personality' figures or engaging in the "conductor" activity. This suggests that these conductors serve as influential figures within the community, connecting users to other extreme and radical individuals, which further propagates incitement to violence on the platform. Future research could delve deeper into the role of seemingly innocuous conductors (or their analogues on other platforms) as amplifiers of unwanted behaviour.

Additionally, it is evident that Parler follows a quantity-over-quality mindset, similar to platforms like Twitter, where users post consistently in short intervals, often without having a large following or significant impressions or upvotes on their posts. This is highlighted by the correlation between number of posts and degree centrality. This poses challenges for manual identification of extremism and highlights the need for data-driven, machine learning, and automated approaches.

Our research also indicates a weak correlation between extremist posts and higher levels of hate speech/toxicity. This raises questions about the relationship between extremism and toxicity or hate speech: incitement to violence may not always be related to toxic forms of expression, and using one as a proxy for identifying the other could be misleading moderation attempts in other platforms.

VII. CONCLUSION AND FUTURE WORK

The goal of this research was to review the Parler userbase and the prevalence of hate speech and extremism shared on the platform. In doing so, we have presented several tools and techniques which have been applied on other social networks or extremist user groups and demonstrated their applicability to far-right extremism on Parler.

Our work has replicated the results of Nouh et al. [23] within Islamic extremist content identification, and shown that a classifier designed along the same lines can be deployed for identifying other forms of extremism on different platforms, albeit at a performance penalty. We have also identified several interesting forms of behaviour present on the Parler social network, and the prevalence of hate speech and extremism both among the most popular users on the platform and across a sample of all users. Our analysis shows that extremist content, while not the exclusive focus of Parler users, was highly central to Parler activity.

There are several topics which could aptly be explored by future research. First, various approaches could be taken to improve model performance in this shifted domain. For example, in creating our training data, a keyword-based methodology was applied, creating a risk of a high textual dependence. Alternative labelling approaches could be fruitful both for the Parler dataset in particular and application to other extremist social media platforms. Nouh et al. also compared performance to a counterpoise dataset, where the classifier is tested against data where extremism is discussed but no extremist ideology is shared (for example news on terror attacks) instead of a baseline non-extremist dataset. Secondly, we found only a weak correlation between a post being extremist and having a high hate speech or toxicity rating. Further research may be required to explore the nuanced relationship between extremism and toxicity. In addition to this, a limitation is present in the way that the known-good and known-bad datasets were created. This is due to how keywords were used to automatically label the known-good dataset, not accommodating for the word's context (such as disavowing a word), future research may seek to extend on the work hear by using manually labelled datasets for a stronger ground-truth.

REFERENCES

[1] Adamczyk, A., Gruenewald, J., Chermak, S.M., Freilich, J.D.: The relationship between hate groups and far-right ideological violence. Journal of Contemporary Criminal Justice 30(3), 310–332 (2014). https://doi.org/10.1177/1043986214536659

[2] Aliapoulios, M., Bevensee, E., Blackburn, J., Bradlyn, B., De Cristofaro, E., Stringhini, G., Zannettou, S.: An early look at the Parler online social network. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 15, pp. 943–951 (2021)

[3] Backover, A.: Hate sets up shop on internet. The Denver Post Online (2013), https://extras.denverpost.com/enduser/digital/digital1108c.htm

[4] Basit, A.: Conspiracy theories and violent extremism. Counter Terrorist Trends and Analyses 13(3), 1–9 (2021)

[5] BBC: Parler 'free speech' app tops charts in wake of Trump defeat. BBC News (Nov 2020), https://www.bbc.co.uk/news/technology-54873800

[6] Cartographer Labs: CartographerLabs/Pinpoint (Apr 2022), https://github.com/CartographerLabs/Pinpoint

[7] Chermak, S., Freilich, J., Suttmoeller, M.: The organizational dynamics of far-right hate groups in the United States: Comparing violent to non-violent organizations. Studies in Conflict & Terrorism 36(3), 193–218 (2013). https://doi.org/10.1080/1057610x.2013.755912

[8] Corera, G.: Is there a growing far-right threat online? BBC News (Jul 2019), https://www.bbc.co.uk/news/world-europe-48830980

[9] Ellinas, A.A.: Far-right parties in Europe. Oxford Bibliographies Online: Political Science (2015). https://doi.org/10.1093/obo/9780199756223-0165

[10] Fernandez, M., Alani, H.: Artificial intelligence and online extremism. In: McDaniel, J., Pease, K. (eds.) Predictive Policing and Artificial Intelligence, pp. 132–162. Routledge Frontiers of Criminal Justice, Routledge (2021). https://doi.org/10.4324/9780429265365-7

[11] Gosling, S.D., Augustine, A.A., Vazire, S., Holtzman, N., Gaddis, S.: Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. Cyberpsychology, Behavior, and Social Networking **14**(9), 483–488 (2011). https://doi.org/10.1089/cyber.2010.0087

[12] Gruenewald, J., Chermak, S., Freilich, J.D.: Distinguishing "loner" attacks from other domestic extremist violence. Criminology & Public Policy **12**(1), 65–91 (2013). https://doi.org/10.1111/1745-9133.12008

[13] Guille, A.: Swear words (Jan 2018), https://www.kaggle.com/datasets/highflyingbird/swear-words

[14] Hagey, K., Horowitz, J.: Parler, a platform favored by Trump fans, struggles for survival. The Wall Street Journal (Jan 2021), https://www.wsj.com/articles/parler-struggles-survival-amazon-lawsuit-trump-fans-11610414745

[15] Hansindu: Inappropriate words classification (Aug 2022), https://www.kaggle.com/datasets/hansindu/inappropriate-words-classification

[16] Institute For Economics And Peace: Global Terrorism Index 2019 (Nov 2019), https://www.visionofhumanity.org/wp-content/uploads/2020/11/GTI-2020-Briefing.pdf

[17] Knight, S., Keatley, D., Woodward, K.: Comparing the different behavioral outcomes of extremism: A comparison of violent and non-violent extremists, acting alone or as part of a group. Studies in Conflict & Terrorism **45**(8), 682–703 (2022)

[18] Knight, S., Woodward, K., Lancaster, G.L.: Violent versus nonviolent actors: An empirical study of different types of extremism. Journal of Threat Assessment and Management **4**(4), 230 (2017)

[19] Lima, L., Reis, J.C., Melo, P., Murai, F., Araujo, L., Vikatos, P., Benevenuto, F.: Inside the right-leaning echo chambers: Characterizing Gab, an unmoderated social system. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2018). https://doi.org/10.1109/asonam.2018.8508809

[20] MI5: Director General Ken McCallum gives annual threat update 2021 (2021), https://www.mi5.gov.uk/news/director-general-ken-mccallum-gives-annual-threat-update-2021

[21] Mikesuhan: Mikesuhan/keyness: Rank orders log likelihood values for every token or ngram in a corpus. (2019), https://github.com/mikesuhan/keyness

[22] Mills, M.F.: Obamageddon: Fear, the far right, and the rise of "Doomsday" prepping in Obama's America. Journal of American Studies **55**(2), 336–365 (2019). https://doi.org/10.1017/s0021875819000501

[23] Nouh, M., Nurse, J.R., Goldsmith, M.: Understanding the radical mind: Identifying signals to detect extremist content on Twitter. In: Proceedings of the 2019 IEEE International Conference on Intelligence and Security Informatics (ISI). pp. 98–103 (2019)

[24] Rayson, P.: Log-likelihood and effect size calculator (2016), https://ucrel.lancs.ac.uk/llwizard.html

[25] Rondeaux, C., Dalton, B.: What role did the far-right platform Parler play in the Jan. 6 insurrection? Slate Magazine (Jan 2022), https://slate.com/technology/2022/01/parler-jan-6-capitol-facebook-twitter.html

[26] Sahito, F., Slany, W., Zaman, S., Sahito, I., Qaiser, R.: Terrorist profiling as a counter-insurgency strategy: Applying the concept to law enforcement. International Journal of Innovation and Applied Studies (2013)

[27] Statista Research Department: Facebook terrorist content deletion per quarter 2023 (Jul 2023), https://www.statista.com/statistics/1013864/facebook-terrorist-propaganda-removal-quarter/

[28] Unitaryai: UNITARYAI/detoxify (2020), https://github.com/unitaryai/detoxify

[29] Vertigans, S.: Beyond the fringe? Radicalisation within the American far-right. Totalitarian Movements and Political Religions **8**(3-4), 641–659 (2007). https://doi.org/10.1080/14690760701571254

[30] Vicomtech: Hate-speech-dataset: Hate speech dataset from Stormfront forum manually labelled at sentence level. (2020), https://github.com/Vicomtech/hate-speech-dataset

[31] Wojcieszak, M.: 'Don't talk to me': Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. New Media & Society **12**(4), 637–655 (2010). https://doi.org/10.1177/1461444809342775

[32] Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., Blackburn, J.: What is Gab? A bastion of free speech or an alt-right echo chamber? In: Companion Proceedings of the The Web Conference 2018. pp. 1007–1014 (2018). https://doi.org/10.1145/3184558.3191531