

A Dataset for The Study of Online Radicalization Through Incel Forum Archives

JENNIFER GOLBECK¹

University of Maryland, USA

The incel (involuntary celibate) community is an extremist online community that practices intense misogyny and racism, and that glorifies – and sometimes practices - violence. Work to understand the dynamics within incel communities has been hindered by the fact that these communities are spread over many platforms and many of the more popular forums of the past have been banned and their content deleted. In this paper, we present two main contributions. First, we introduce a carefully reconstructed, nearly complete archive of incel forums dating back to 2016, including millions of posts that can no longer be accessed. Then we illustrate a technique for identifying community-specific language and using that as a marker of extremism to track radicalization over time.

Keywords: radicalization, social media, incels

¹ golbeck@acm.org

Date submitted: 2024-01-10

The incel (involuntary celibate) community is an online subculture. On the surface, incels are men who struggle to form romantic relationships with women. In reality, incel forums are hotbeds for misogyny, racism, homophobia, self-hatred, violent fantasies, and sexual entitlement. A number of self-described incels have committed mass murder, and their forums have hosted videos and livestreams of mass shootings, and encouragement for more violence. Though they have been banned from popular online forums like Reddit, the incel community is active on a number of public platforms. Because their data is public and accessible, they are an ideal community for studying online radicalization.

This paper's goals are twofold: to introduce a dataset for studying the incel community and to understand how radicalization can be tracked over time within this dataset.

We present a dataset with the full archive of five incel communities: incels.is, the most popular active incel forum; the Saidit.net incel forum; the now-defunct Reddit /r/Incels and /r/Braincels forums; and /r/IncelExit, a Reddit support forum for men who are working on leaving the incel community and mindset. The dataset contains 9.2 million posts from 122,079 users that span from 2016 to 2023. We describe each forum and include basic statistics about the data available for each.

Our research questions are:

RQ1: How does an incel community member's pattern of incel-specific language change over time?

RQ2: Is there variation across platforms in users' patterns of incel-specific language?

To answer these RQs, we analyze the language of incel forums against a background corpus to detect incel-specific words that serve as markers of community membership and radicalization. We then track the use of those words over time in incel communities and describe patterns of use.

Background & Summary

Since 9/11, online extremism and radicalization has been a growing concern and active area of research. As extremist groups and online platforms that support them co-evolve, so do the research challenges to understand the process of radicalization and, importantly, deradicalization. In this paper, we present a large dataset focused on one extremist group – the incel community – as a basis for studying many aspects of online extremism and radicalization.

The incel (involuntary celibate) community began as a place of commiseration and discussion among men *and* women who were struggling romantically, but has since morphed into an extreme, violent misogynistic movement. Fantasies of committing rape, murder, and suicide are common. Forums for incels have been repeatedly banned from major platforms because they encourage acts of violence generally and against women specifically. Since the mid 2010s, there has been a rise in mass violence committed by self-described incels.

Notable mass murder events include the 2014 Isla Vista killings, where Elliott Rodger murdered six and injured 14 in an attack that included a sorority house as a deliberate target. He recorded a YouTube video in the middle of his killing spree where he called the violence his "retribution" designed to punish women for sexually rejecting him, and posted it on known incel forums. The 2018 van attack in Toronto's New York City Centre, committed by self-described incel Alex Minassian, killed 11 and injured 15. Beyond these and other large-scale acts of violence, suicide is a common theme in incel communities. Over 80 suicides have been publicly logged on Reddit at /u/incelGraveyard and studied (Daly and Laskovtsov 2021).

The Texas Department of Public Safety has identified incels as an emerging domestic terrorism threat (Texas Department of Public Safety, 2020). Because the community exists almost entirely online, there is a wealth of data that supports analysis of topics like the language of extremism, radicalization, the social structure of extremist groups, community dynamics, and when violent fantasy tips over into violent action.

Related Research

There is extensive research from data science and social science communities into the role that online communities play in the radicalization process and on related topics, including the spread of mis- and dis-information (Benkler et al., 2018), conspiracy theories (Van Raemdonck, 2019), and fake news (Johnson, 2018). Since 2001, there has been extensive research into the role online interactions play in Islamic extremism (Chatfield et al., 2015) and, more recently, domestic extremism like the QAnon movement that led to the January 6 insurrection (Garry et al., 2021). The internet facilitates self-radicalization (Bradbury et al., 2017), where users can discover extreme groups that match their ideals (Hollewell and Longpré, 2022).

There is an extensive body of work in psychology on the social dynamics that lead people to adopt extreme beliefs. We know when people are drawn into fringe movements, cults, and extreme ideologies, many psychological factors are at play. Humans have a need to belong (Baumeister & Leary, 1995; Walton et al., 2012), and interpersonal connections to people in ideologically extreme groups can motivate people to join (Stark & Bainbridge, 1980). Groups with inspiring ideals that strike a chord with an outsider can provide a gateway to joining a group, and commitment to those ideals can override compunctions regarding extreme behavior the person might not otherwise have been willing to engage in (McAdam, 1986; Thrash & Elliott, 2003; Swank & Fahs, 2011). When a person's social identity is tied to an ideal, this can build commitment to a group that connects itself with that ideal (Ysseldyk, 2010). Once a sense of belonging or commitment to a group is created, it can feed deeper commitment and the growth of in-group bias (Efferson et al., 2008)

When stepping into an extremist group, social dynamics are key drivers of the radicalization process. When there are in-group behaviors that strongly diverge from the norm, newer members look to others for social proof, guidance as to what behavior will be rewarded within the new community (Cialdini et al., 1999). Rather than simply establishing an in-group culture, social proof can drive people to extreme behavior, including suicide (Ji et al., 2014). Finally, when belonging to an extremist group comes with incentives – material or psychological – that are tied to performing certain beliefs,

people's genuine beliefs will often shift to match what they are required to express (Solley & Santos, 1958; Festinger & Carlsmith, 1959).

Research into online radicalization leverages these psychological insights (directly or indirectly) and usually applies them through the lens of text analysis, since this is the main body of data available. For example, Grover & Mark (2019) found that alt-right Reddit users clearly presented the same signs of radicalization online that were typical of offline radicalization, including fixation on racial out-groups and increasing negative emotion toward those out groups over time. Habib et al., (2022) analyzed the "manosphere", collection of online misogynistic extremist communities including incels, and found that even a single instance of interacting in these forums could increase the signs of radicalization in men's online activity.

More broadly, there has been extensive work on detecting hate speech (Zhang et al., 2018), harassing language (Marwa et al., 2018), and misogynistic language on social media (Saha et al., 2018). Incel communities use an extensive cryptolite – a set of insider slang and memes used to express their misogynistic ideologies (Gothard, 2021). This lexicon serves to connect the group, confuse outsiders, and is part of the social proof process of radicalization – adoption is required for acceptance, and men are motivated to adopt it when seeing others use it. Some research has used computational linguistics to highlight some important terms to the group (Gothard, 2021), but these approaches include terms that are not incel specific and miss other important elements of the lexicon. Informal lists of incel slang also exist, both online and in print (Sugiura, 2021).

While the online radicalization literature is extensive, there is very little work on deradicalization. That which exists tends to focus on interventions to counter the messaging of extremist content (Ashour, 2010) or prevent its posting altogether (Habib et al., 2022).

Dataset

We present a large dataset of posts collected from popular incel forums over time. These include the complete accessible archives of the now-banned Reddit forums

/r/Incels and /r/Braincels, and two Reddit-like forums where many in the community have moved to after the Reddit ban, including incels.is and the Saidit.net incel forum. We also have an archive of the active Reddit forum /r/IncelExit which provides support and strategies for men trying to leave the incel community. This latter community is particularly meaningful as it offers ground truth insights into how language and behavior change in men who are deradicalizing. Because the dataset includes a forum for people exiting the incel movement, this opens up new paths to study deradicalization and the detection of those who would consider an off-ramp from extremist communities.

This dataset has the potential to support research on online extremism by easing the barrier to entry that researchers face in collecting data and comparing results. It offers three main benefits:

It collects data into a single, downloadable source, saving researchers the time for creating scrapers and accessing data through rate-limited APIs.

The core data – user id, post text, and threading information – is already parsed into tab-delimited data tables for easy use

Given their history of shutting down, these archives ensure the data will be available for analysis even if existing forums close or go private. For example, the Reddit archives for /r/Incels and /r/Braincels were available in April 2023 when we first collected this data, but have since become unavailable due to changes in Reddit's API.

Data Collection Methods

To collect data from these platforms, we used a combination of available APIs and custom scraping tools written in perl and python. All the data we collected is publicly available to anyone on the web. Neither the forums nor the APIs we used required membership or a login to view the contents. The Terms of Service for all of the platforms and APIs we used allow the type of respectful automated content collection that we performed. Our collection and sharing has been reviewed by the University of Maryland IRB and determined as not human subjects research.

Below, we describe each of the forums present in our dataset.

/r/Incels

Until 2017, Reddit's /r/Incels forum was the most popular online space dedicated to Incels. On October 25 of that year, Reddit introduced a policy banning calls for violence and, on November 7, Reddit banned /r/Incels under that policy. A Reddit spokesperson announced "Communities focused on this content and users who post such content will be banned from the site. On November 7, /r/Incels was banned for violating this policy."² After the ban, the posts became inaccessible on Reddit and community members dispersed to other forums.

To access the pre-ban discussions from /r/Incels, we used the pushshift API and collected all posts and comments available in that archive, spanning May 2016 through November 2017. Pushshift (Baumgartner et al., 2020) is a widely used historical archive of Reddit that, until the Summer 2023 API changes, was updated continuously and in real time.

/r/Braincels

After /r/Incels was shut down, the Incel community spun up a new subreddit, /r/Braincels, which became the most popular incel subreddit. It was quarantined by Reddit in 2018 for rules violations and then reactivated before finally being banned in October 2019.

To collect this full archive, we first accessed a subset of /r/Braincels that covers October 21, 2017 through May 3, 2019 from the Components website³. We processed that data into a standard tab delimited file and stripped some of the media embeddings. The raw data is still available from Components.

The Components dataset was missing all data between September 30, 2018 and February 15, 2019 and after May 3, 2018. During some of this first gap, /r/Braincels was quarantined so the activity was minimal. However, there was some activity. In addition, /r/Braincels existed for around five months after the Components data collection

² <https://www.nytimes.com/2017/11/09/technology/incels-reddit-banned.html>

³ <https://components.one/datasets/braincels-incel-reddit-data>

ended. To fill in the missing data, we collected archived posts and comments from the pushshift API.

incels.is

The most popular active incel forum is incels.is (formerly incels.co). We collected all available posts shared to the Inceldom Discussion forum - the main discussion board on the site. Since we gathered data from the live site, deleted posts and comments are not part of this corpus. We downloaded the HTML for each thread's page using an Apple Automator workflow. We then processed those pages using custom perl scripts.

Saidit.net /s/Incels

Saidit.net is a Reddit-like discussion forum. When /r/Incels was banned, some members moved to Saidit.net and set up an incel forum at /s/Incels. While it is still available, it is largely inactive, averaging only 1-3 posts per month over the last year. Using custom perl scripts, we collected the full history of posts and available user data from this forum. We collected the HTML for each thread's page and parsed it into a data table.

/r/IncelExit

Reddit's /r/IncelExit is a forum for men who want to leave the incel movement. It provides a contrast to the themes of the active incel forums, but with content that comes from men who have been part of that community. Using the pushshift and Reddit APIs, we collected all available posts from the start of the forum, in October 2019, through September 2022.

Table 1. Dataset statistics for each forum

Dataset	Posts	Words	Users	Users ≥5 posts	Start Date	End Date
----------------	--------------	--------------	--------------	-------------------------------	-----------------------	---------------------

Saidit /s/Incels	13,083	654,315	591	257	6/2019	9/2022
/r/iIncelExit	110,989	6,970,071	4,005	1,440	10/ 2019	4/2023
incels.is	5,748,00 9	164,196,095	12,190	9,985	11/ 2017	7/ 2023
/r/Incels	988,244	26,549,906	29,876	9,573	5/ 2016	11/ 2017
/r/Braincels	2,408,17 9	67,449,572	75,417	22,617	10/ 2017	8/2019

Data Records

The dataset is available on the Harvard Dataverse under a CC BY-SA 4.0 license at <https://doi.org/10.7910/DVN/MS9ODP>.

In total, this dataset comprises 9.2 million posts with more than 265 million words. Table 1 shows the statistics for each forum. Users with ≥ 5 posts are shown as an indicator of sustained engagement.

For each forum, we generated a tab-delimited file. Each line represents a comment and contains the following fields:

- user – the username or screen name of the user who posed (anonymized)
- timestamp – the time that the item was posted in Unix time
- text – the text of the post. Most formatting and images have been removed
- link – a URL or page name that can be mapped to the address of the post on each platform
- comment_id – a unique identifier for each post.
- parent – The comment_id of the post being replied to (when available)
- title – The title of the post. This is available for initial posts in threads but not replies, since replies cannot have separate titles.

This dataset follows FAIR principles (Wilkinson, 2016):

- Findable-The data has a persistent digital object identifier, <https://doi.org/10.7910/DVN/MS9ODP>
- Accessible - The dataset is accessible to the public for free.
- Interoperable - All data is available in a standard format.
- Reusable - This data can be reused for many analyses, including studies of online extremism, domestic terrorism, radicalization, and deplatforming.

Data Patterns

To validate that the data we collected for these forums was as close to complete as possible, we compared the statistics for the banned Reddit data to other researchers' work on these forums.

We note that some differences are expected in post and user counts, since that this data is online, changeable, and collected via API. Data availability is impacted by user accounts being deleted, posts being deleted, usernames changing, and API responsiveness. There also may be differences based on researchers' choices about what to include. For example, we kept 72,885 posts from /r/Incels where both the author and post had been removed. We included these because the reply and timestamp information may be useful to some researchers.

For /r/Incels and /r/Braincels, we identified two other papers that use this data and reported numbers against which we could compare. Gothard (2021) studied /r/Incels, /r/Braincels, /r/shortcels. Table 2 shows post and unique user counts from Gothard (2021) and our dataset on /r/Incels and /r/Braincels. We chose not to include /r/shortcels because it is a more specific, and short-lived Reddit incel subreddit that was geared toward shorter incels. Our focus was on general incel forums, and because of its narrower topical role in the community, we did not include it in our data.

Table 2. User Counts from Gothard and this dataset

	Gothard (2021)	Our Data
--	----------------	----------

/r/Incels comments	881,118	988,244
/r/Incels users	36,301	29,876
/r/Braincels comments	2,486,655	2,408,179
/r/Braincels users	56,698	75,417

Our comment count is larger than Gothard, but, as mentioned above, over 72,000 of the posts we kept in the dataset had both the post text and author information removed. Gothard reports cleaning their data and explicitly removing data with deleted usernames, text, or missing timestamps. This explains most of the differences in the comment count between the two datasets. The differences in unique usernames can also be accounted for by now-deleted users whose information was stripped from the archive in the years between Gothard's and our collections.

Our daily post frequencies closely map to those reported in (Gothard, 2021) as well. The post and comment frequency of our data is shown in Figure 1, and the frequency reported in Gothard is shown in Figure 2.

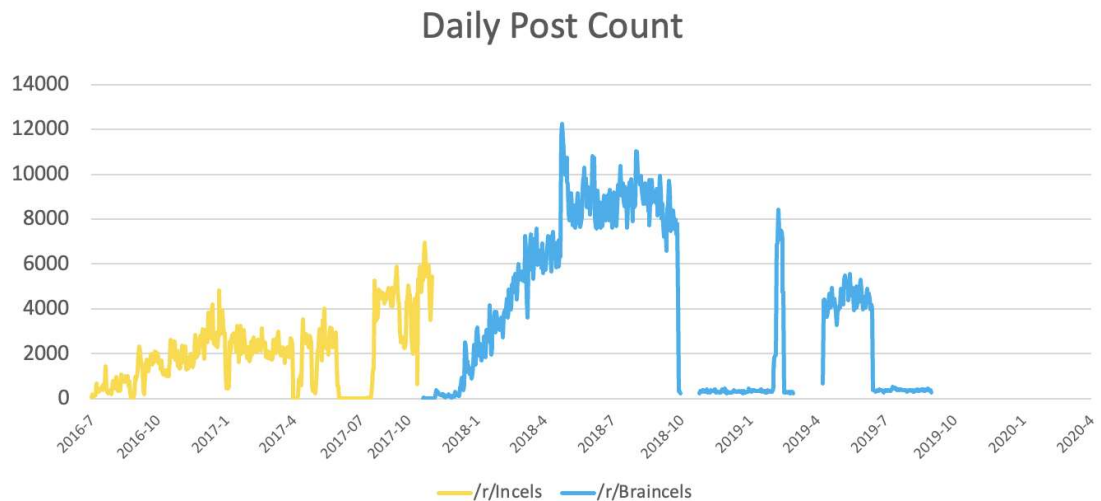


Figure 1. Post frequency for /r/Incels, /r/Braincels, and incels.is in this dataset

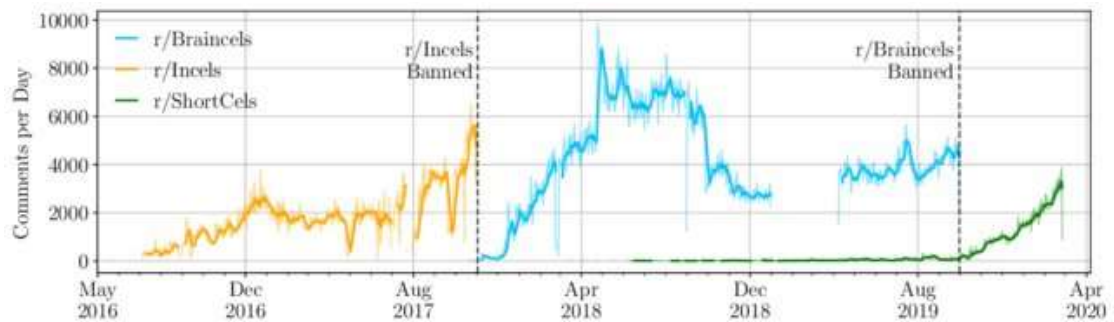


Figure 2. Frequency measures from (Gothard, 2021)

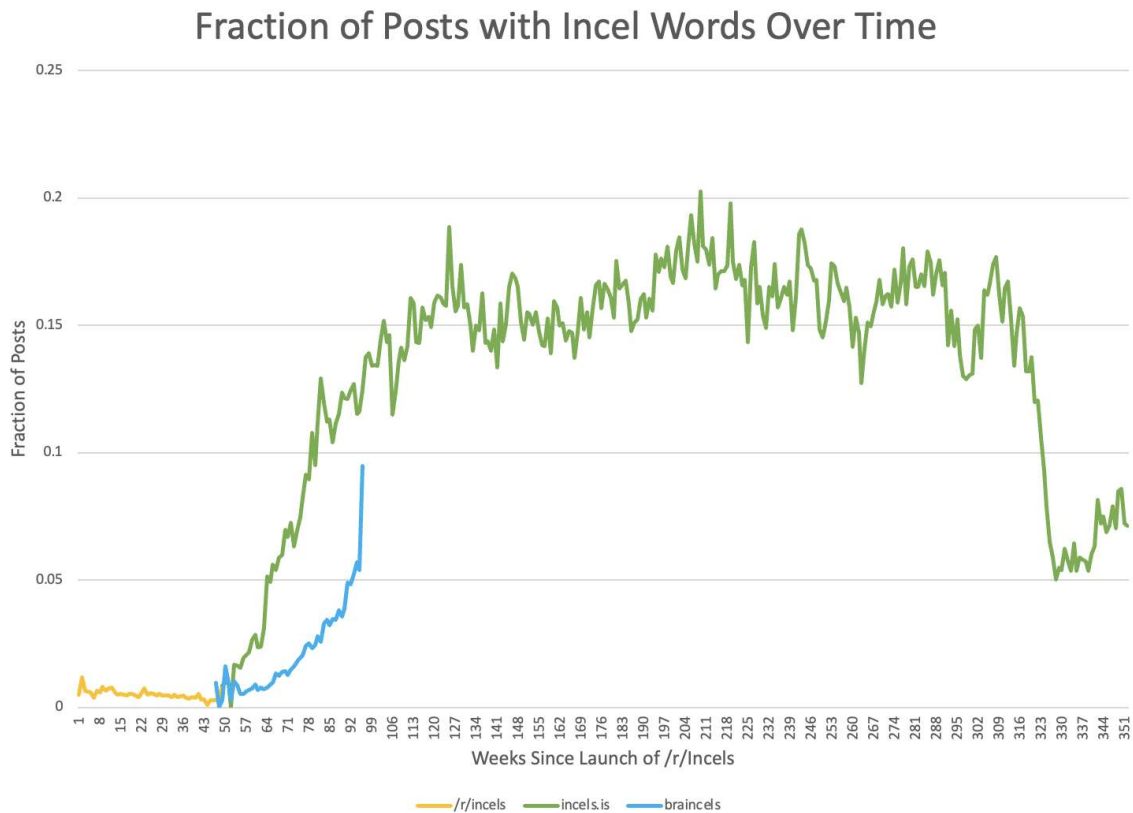


Figure 3. Fraction of Posts with Incel Terms Over Time

Ribeiro et al. (2021) collected Reddit data via pushshift for the last 120 days that /r/Incels was active, from July 10, 2017 to November 7, 2017. Their data contained

340,650 comments. Our data from that same period contains 343,814. Their data had 18,088 users while ours had 14,289. These small variations are likely explained by similar factors as we described for Gothard.

Overall, comparing the size of our datasets to other published work that accessed these archives, we find very similar sizes, indicating we have fully captured the available forum data. We also note that while other research has used the Reddit datasets in the past, those projects did not make the data publicly available and, due to changes in the Reddit API in 2023, this data is no longer available. It had been deleted from Reddit when the forums were banned and pushshift, the API that provided access to archived posts, was shut down in 2023 with the Reddit API changes.

As shown in Figure 3, the incel community as a whole developed and increased their use of these terms over time. After /r/Incels was banned from Reddit, there was a sharp increase in the use of incel terms on alternative forums, including /r/braincels and incels.is. We note that user growth is also generally steady, with coefficient of variation in the number of new users joining per week of around 1 (1.17 for incels.is, 0.99 for /r/Incels, and 0.86 for /r/braincels).

Usage Notes

We anonymized usernames in this dataset. Other social media datasets have chosen not to anonymize (Trujillo et al. 2022), and in this case, existing archives and active forums would make deanonymization relatively straightforward. That said, even though all the data we present was posted publicly and username have been anonymized, there are privacy risks. For example, if people later delete their posts from these forums, they are not deleted from our archived data. These risks are no greater than with any other data archives.

We chose to limit full access to this data because of the ethical issues mentioned above and because the dataset contains deleted posts. The README files are public and describe the data available for each forum. The data files themselves are password protected and access will be granted on request. Researchers who request access will be

asked to first agree to consider issues of privacy, risk, and sensitivity and to take appropriate ethical steps, including anonymization and paraphrasing of quotes in research consistent with ethical principles of working with social media data (Townsend & Wallace, 2016). Researchers who need actual usernames (e.g. to study other posts by incel users on reddit) can contact the author at golbeck@acm.org to discuss collaborations.

Analysis Methods

To measure incel-specific language, we chose to computationally select especially distinctive words. We used the binary log of the ratio of relative frequencies, informally known as the "log ratio". For corpus A which is a subset of corpus B, the log ratio of a word w is given by the base-2 log of its frequency in A divided by the log of its frequency in B. Each time the log ratio increases by 1, it indicates a doubling in frequency of the word in A over its frequency in B (Hardie, 2014). For a word w_i , the log ratio of its relative frequency (RF) in the incel corpus to its relative frequency in a Reddit background corpus is given by:

$$\log_2 \left(\frac{RF_{incel}(w_i)}{RF_{reddit}(w_i)} \right)$$

Where for a word w_i in a corpus c with n words, the relative frequency $RF_c(w_i)$ is given by:

$$RF_c(w_i) = \frac{Freq_c(w_i)}{n}$$

This measure gives the base-2 orders of magnitude difference between the frequency of a word in one corpus vs. another. In this case, it reveals how many orders of

magnitude more often a word appears in the incel forums than in a background Reddit corpus.

We used the combined Reddit data from /r/Incels and /r/Braincels as our incel corpus for detecting incel-specific language. For the base corpus, we used the complete June 2019 archive of Reddit collected via the Pushshift API. We calculated the relative frequency of each word in both corpora and then computed the log ratio.

Looking at 50 words with the highest log ratio, the values ranged from 6.3 to 12.9, meaning these words were $2^{6.3}$ to $2^{12.9}$ (78.8 to 7,643.4) times more common on incel subreddits than on Reddit as a whole. The words were also typical of the Incel cryptolect.

These 50 words are shown in Table 3. Top 50 incel-specific words Table 3. These words show common characteristics of incel language, including blending of words, racist and misogynistic terms, and degrading and dehumanizing language (Bogetić, 2023). They also overlap with many incel lexicon terms identified in previous work (Gothard, 2021). We selected the 50 words with the highest log ratio, i.e. that were the most unique to the incel community, to serve as measures of radicalization. Because these words were frequently used among incels and rarely by the rest of the Reddit community, they serve as a marker for Incel membership. Since incels.is was created as a refuge for the Incel community after it was banned from Reddit, these words discovered on Reddit are effective markers of community membership on that platform as well.

Table 3. Top 50 incel-specific words

foid	coomer	ropefuel	maxx	noodlewhore
ragefuel	soyboi	sexhav	chadfish	lockdown
suicidefuel	tate	dickpil	whitepil	heightpill
jbw	larper	cuckit	betabuxx	dogpill
jfl	sheboon	fd	lifemog	itsover
gtfih	statusmax	gigastaci	ascens	teencel
lifefuel	doofi	brocel	mayocel	wahmen
racepill	inkel	voicepil	moneymaxx	currywheore
itcel	cumskin	japcel	agepil	slavcel

suifuel mogger gigacop femcel sexdoll

As a tangential investigation, we analyzed the impact of using the top-*n* incel-specific words. Using a full lexicon of incel terms (roots, prefixes, and suffixes) developed in (Klein & Golbeck, 2024), we counted the number of posts that would be captured as a base dataset. Then, using only the top *n* most common incel words, we measured what fraction of those posts would be retrieved. As shown in Figure 4, the first nine to ten words alone will capture over 70% of all posts with incel terms. For the top 50 words, over 90% of all posts with incel words will be captured in both incels.is and /r/Incels.

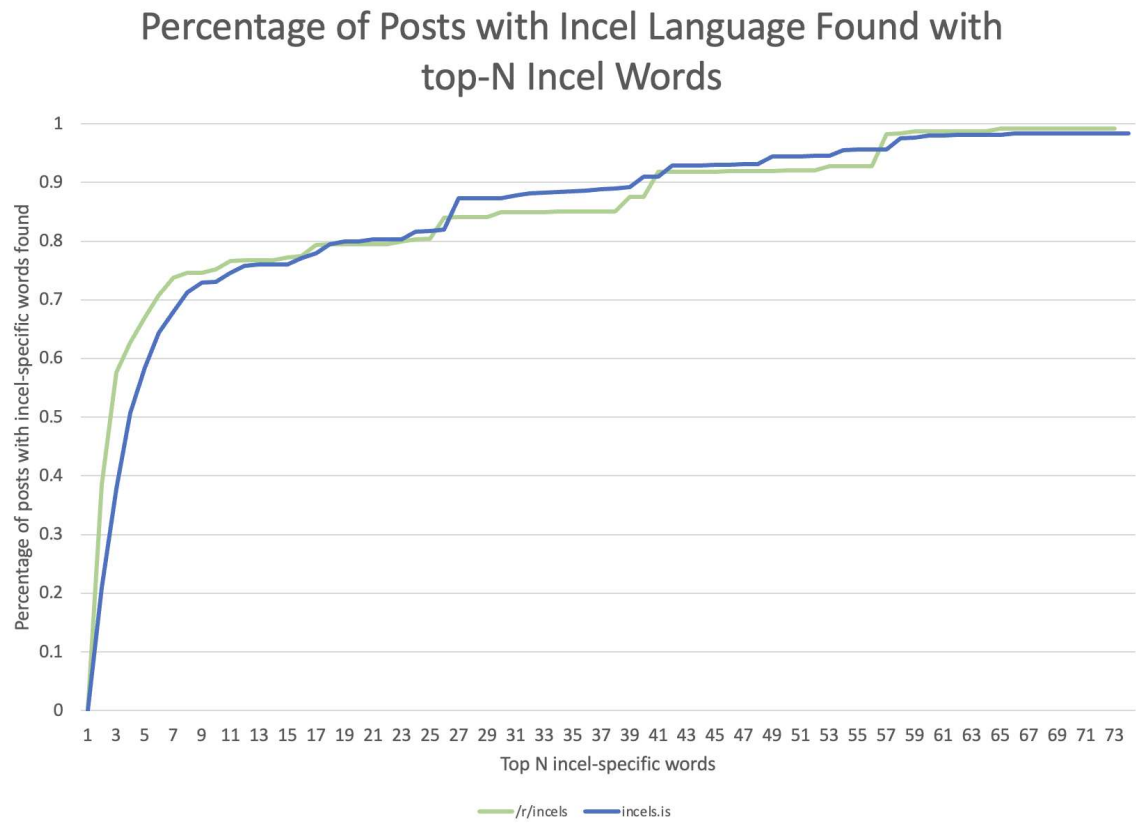


Figure 4. Posts Retrieved with top-n Incel Words

We analyzed the use of these words on Reddit and on incels.is. Because the Saidit dataset is so much smaller and /r/IncelExit is for men attempting to disengage from the incel community, we did not include them in our analysis.

To measure changes in use among individuals, we grouped the text for each user for each week since they joined the platform. For a given user, we have all the text he posted in his first week, second week, third week, and so on. We then aligned these values according to the week-since-joining. Thus, even if User A joined in 2018 and User B joined in 2021, their first week of posting would be aligned. This lets us see if, generally, users follow patterns in their posting and language use over time. To ensure we were measuring users who were fairly active, we limited our analysis to those who had posted at least 5,000 words. This resulted in 2,795 users on Reddit and 3,507 on incels.is.

Using the 50 incel-specific words, we measured their count and frequency by user by week. We applied a snowball stemmer to capture common variations on each word. When users did not post at all in a given week, they were excluded from the calculations around incel-specific word usage, but we did include their total word use count of 0 in the overall word count analysis. Furthermore, because users often disengage and stop posting over time, there is a continual decline in users posting as weeks increase. We only considered weeks where there were at least 100 users who had posted. For incels.is, that point comes at 212 weeks (roughly 4 years) and on Reddit, that happens at 87 weeks.

Analysis Results and Discussion

On both Reddit and incels.is, the number of words posted per week are remarkably similar. As shown in Figure 5, in their first week, users tend to write a lot – over 2,300 words. This drops off sharply but remains over 1,000 words / day on average for the first month. An ANOVA indicates there was no significant difference in the number of words posted per week between incels.is and Reddit ($F(1,172) = 3.65$, $p = 0.57$). Note that these are not weeks on the calendar (e.g. Week 1 does not correspond to a specific date range), but rather the week number of activity for each user. Thus,

similarities in posting frequency are not due to outside events because the week numbers are not coordinated with the calendar.

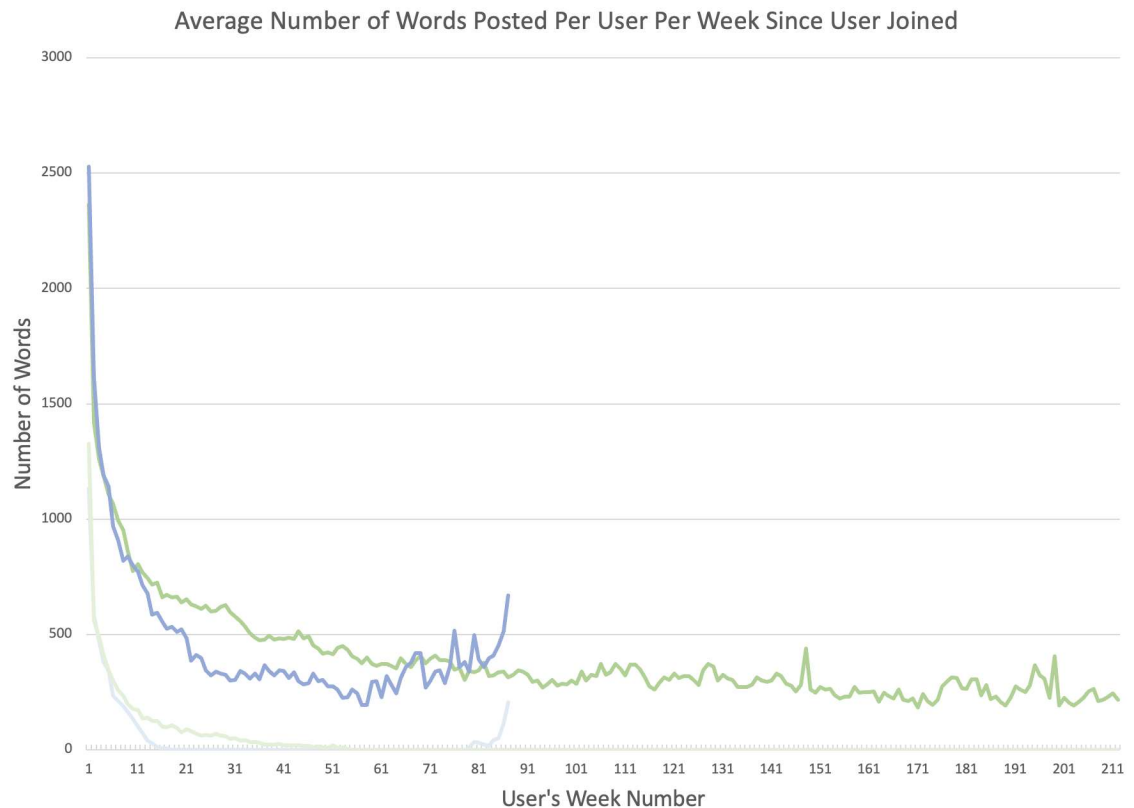


Figure 5. Average (darker) and median (lighter) number of words and posted by users, grouped by the number of weeks since they joined. Incels.is is shown in green and Reddit /r/Incels in blue.

Not only does the average number of words per user decline over time, but the number of users posting on the forums drops off sharply. Looking at this same population of active users with over 5,000 words total posted, we see in Figure 6 that active users each week after joining is exponentially decreasing. Incels.is retains users at a higher rate than Reddit. This may be explained by the fact that incels.is is a forum exclusively for

incels and thus attracts members more committed to the community. Evidence on usage of incel-specific words also supports this hypothesis.

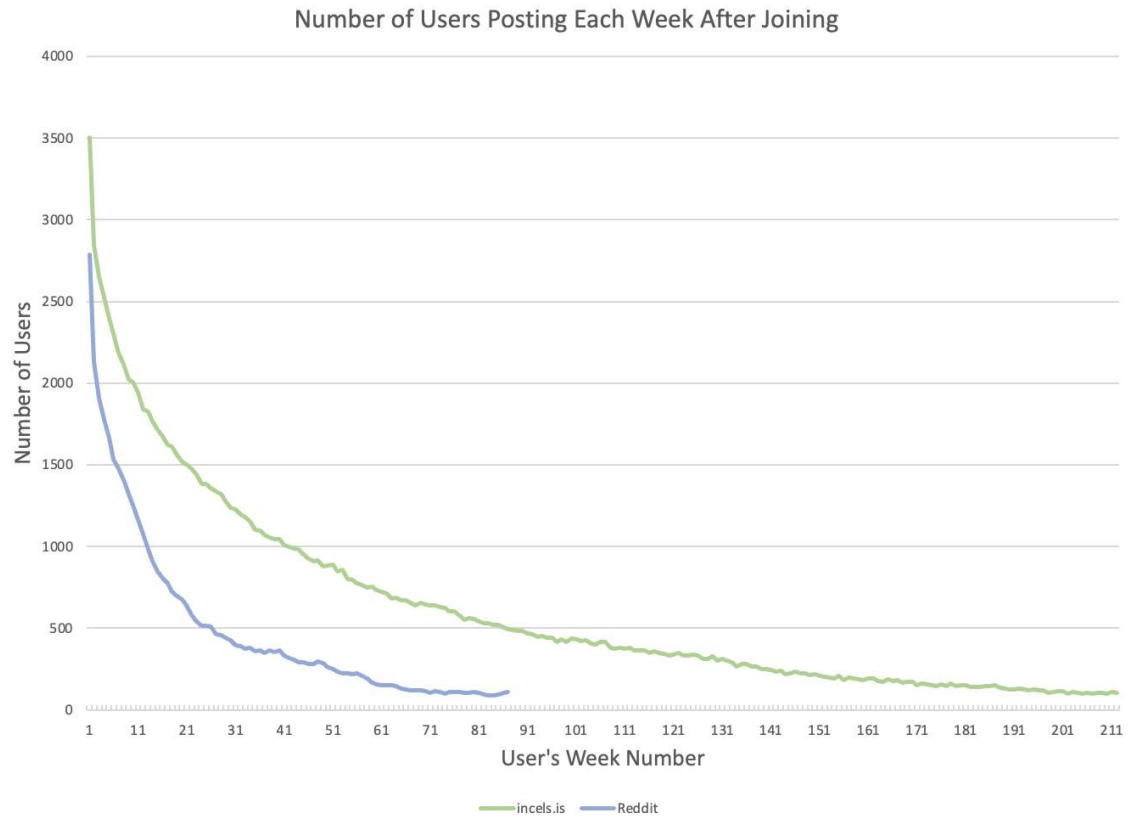


Figure 6. Number of Active Users by Week Since Joining

The average frequency of incel-specific words used per user per week is shown in Figure 7. A 12-week moving average is shown in red. There are two interesting insights from this data.

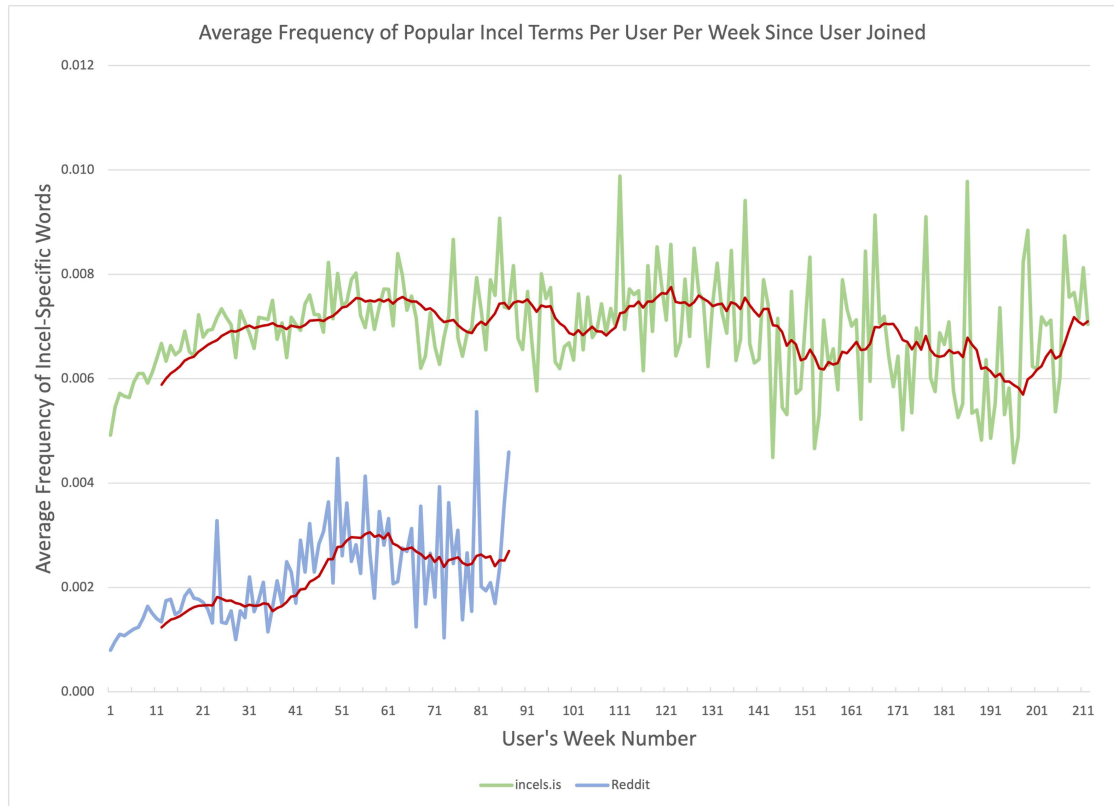


Figure 7. Average Frequency of Incel-Specific Terms Per User

First, users on incels.is are using incel-specific language at roughly 3 times the rate of users on Reddit. If we consider these words, which are racist, misogynistic, and dehumanizing, as markers of incel extremism, this indicates that incels.is is a much more extreme community than the subreddits. This echoes the previous insight that incels.is users disengage at a slower rate than Reddit users.

Second, we see that on both Reddit and incels.is, the average frequency of incel-specific words increases over time, roughly doubling. Two possible explanations for this are (1) Users become more radicalized over time and this is reflected in their increased use of radicalized language or (2) since less committed users drop out over time, later weeks include only the more extreme incels who are consistently using extreme language at a higher rate.

To test this, recall that we considered weeks until there were fewer than 100 users posting in that week. We analyzed the 110 Reddit users and 103 incels.is users who were still active in the final week of consideration. As Figure 8 shows, it appears that even among these longest-participating users, radicalization increases over time. A spot-check statistical comparison of the longest-participating users and those who dropped out earlier in weeks 50-59 also shows that the groups use incel specific language at similar rates. There was no significant difference between the two user groups on incels.is (Student's t-test, $p = 0.12$) or on Reddit (Student's t-test, $p = 0.570$).

Interestingly, looking at the longest-participating users and the remaining users in weeks 10-19, we do see significant differences in the frequency of incel-specific term use both in incels.is ($p < 0.001$) and Reddit ($p < 0.001$), but the longest-participating users are using significantly *fewer* incel-specific terms. We suspect there are several factors at work that are impacting use of incel specific language. While the patterns are consistent that use of radicalized language increases with longer participation, an important area of future work will be to understand the different types of users and the reasons for their changing rates of use.

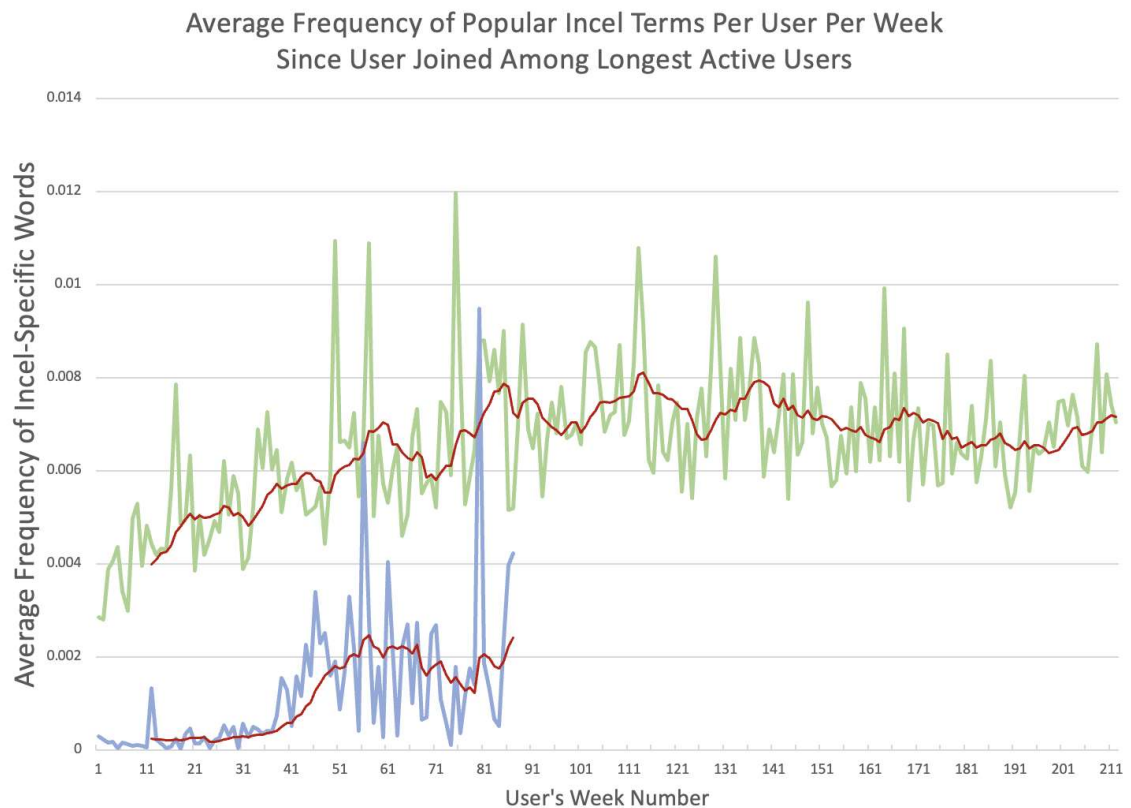


Figure 8. Average weekly frequency of Incel specific term use among the longest active members of both platforms

Deradicalization

Part of leveraging insights from this analysis is being able to identify men who are on the precipice of deradicalization. Our hypothesis is that there are men in active incel communities whose posts look similar to those of men who have disengaged. Since active forums can see hundreds or even thousands of comments every day, there are benefits from being able to get insights automatically and at scale.

A pilot analysis shows that there are differences in language use between active incel communities and /r/IncelExit. We applied a text-based Bayesian machine learning algorithm, Naïve Bayes Multinomial Text, to our data. It could accurately classify which

community a post came from (active incel vs. /r/IncelExit) with 82% accuracy (ROC AUC 0.932). This shows that the language differences between communities are clear and raises the questions about whether we could detect men in active incel groups who look like those who are exiting. If so, it would highlight people who could be targeted with supportive messaging. Positive results would also offer a starting place to study what events trigger disengagement from the community.

Discussion and Limitations

The data collected as part of this research is from forums that were publicly available and whose terms of use allowed (or did not prohibit) the automated collection of data. While these represent the main places incels gather, they are not the only forums.

The most prominent forum that is not represented here is 4Chan's /r9k board. R9K was not established as an incel forum and does not require users to be incels (unlike, for example, incels.is), but it is still a popular destination for incel discussions and hosts roughly 20,000 threads per month. We chose to adhere to the terms of use in our data collection so did not include it in this dataset release.

There are other smaller discussion groups for incels, including various discord servers and message boards. None of these had the prominence of the forums included in the dataset. In addition, most that we could identify had terms of service that prohibited automatic data collection. Many were also non-exclusive to incels, making them less useful for our goal of measuring radicalization among members.

These exclusions do impose some limitations on understanding incel behavior, particularly the exclusion of the r9k forum data. Norms may differ across platforms and language may evolve in smaller communities in interesting ways.

IncelExit has the potential to illustrate how language and behavior shifts in the deradicalization process. This forum supports research into when men decide to leave, to potentially detect men in active forums who are thinking about leaving, and to highlight the types of questions and struggles these men face when trying to deradicalize themselves.

In addition, we have used incel lexicon words as a proxy for radicalization into the incel mindset in this paper and measured that by the number of words used per week. However, this may mask more complexities. Is a person who writes on long, angry post each week that is full of incel terminology differently radicalized than a person who writes 100 milder posts that each use one or two of these largely dehumanizing terms? Our results do not account for this and probing such nuances is an interesting area for future study.

Conclusions

In this paper, we presented a dataset that contains extensive archives of five online incel communities: Reddit's /r/Incels and /r/Braincels, both of which were banned, incels.is, Saidit's /s/Incels, and the /r/IncelExit subreddit for men leaving the incel community. We have made processed versions of these data available on the Harvard Dataverse.

We analyzed the archives of the two main platforms, incels.is and the combined /r/Incels and /r/Braincels subreddits, to detect incel-specific words that serve as markers of incel radicalization. We then tracked the use of these words over time by user.

We found that users post similar numbers of words on both platforms, and that the average number of words posted weekly per user declines over time. The number of active users declines over time as well. However, the frequency of use of incel-specific language increases overtime, especially among the longest-participating users, indicating some process of radicalization takes place.

Our planned future work includes a deep dive into the language of incels, the patterns of use of that language, and whether it is an accurate proxy to track radicalization in this extremist community.

Funding Information

This work was supported by a University of Maryland Catalyst Fund New Directions Award and NSF Award #2331257.

References

- Ashour, O. (2010). Online de-radicalization? Countering violent extremist narratives: Message, messenger and media strategy. *Perspectives on terrorism*, 4(6), 15-19.
- Baker, P. 2016a. What Happens to Men Who Can't Have Sex. *ELLE*, Feb, 1: 2016.
- Baker, P. 2016b. The woman who accidentally started the incel movement. *ELLE*, March, 1: 2016.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497- 529.
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- Bestvater, S. E., & Monroe, B. L. (2023). Sentiment is not stance: target-aware opinion classification for political text analysis. *Political Analysis*, 31(2), 235-256.
- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020, May). The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media* (Vol. 14, pp. 830-839).
- Bogetić, K. (2023). Race and The Language Of Incels: Figurative Neologisms In An Emerging English Cryptoelect. *English Today*, 39(2), 89-99.
- Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The Development and Psychometric Properties Of LIWC-22. Austin, TX: University of Texas At Austin, 1-47.
- Bradbury, R., Bossomaier, T., & Kernot, D. (2017). Predicting the emergence of self-

- radicalisation through social media: a complex systems approach. *Terrorists' Use of the Internet: Assessment and Response*, IOS Press, Amsterdam, 379-89.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Buntain, C., & Golbeck, J. (2014, April). Identifying social roles in reddit using network structure. In *Proceedings of the 23rd international conference on world wide web* (pp. 615-620).
- Chatfield, A. T., Reddick, C. G., & Brajawidagda, U. (2015, May). Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. In *Proceedings of the 16th annual international conference on digital government research* (pp. 239-249).
- Cialdini, R. B., Wosinska, W., Barrett, D. W., Butner, J., & Gornik-Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin*, 25(10), 1242-1253.
- Daly, Sarah E., And Albina Laskovtsov. (2021). "Goodbye, My Friendcels": An Analysis of Incel Suicide Posts. *Journal of Qualitative Criminal Justice & Criminology* 11, No. 1: 1-33.
- Efferson, C., Lalive, R., & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science*, 321(5897), 1844-1849.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, 58(2), 203-210.
- Gothard, K. C. (2021). The Incel Lexicon: Deciphering the Emergent Cryptolect Of A Global Misogynistic Community. The University of Vermont And State Agricultural College.
- Hardie, A. (2014). Statistical Identification of Keywords, Lockwords And Collocations as

- A Two-Step Procedure. *Proceedings of the ICAME 35 Conference*, University of Nottingham.
- Horta Ribeiro, M., Jhaver, S., Zannettou, S., Blackburn, J., Stringhini, G., De Cristofaro, E., & West, R. (2021). Do Platform Migrations Compromise Content Moderation? Evidence From R/The_Donald and R/Incels. *Proceedings of The ACM on Human-Computer Interaction*, 5 (CSCW2), 1-24.
- Jaki, S., De Smedt, T., Gwózdź, M., Panchal, R., Rossa, A., & De Pauw, G. (2019). Online hatred of women in the Incels. me forum: Linguistic analysis and automatic detection. *Journal of Language Aggression and Conflict*, 7(2), 240-268.
- Jensen, M., James, P., LaFree, G., Safer-Lichtenstein, A., & Yates, E. (2018). The use of social media by United States extremists. *National Consortium for the Study of Terrorism and Responses to Terrorism*.
- Ji, N. J., Lee, W. Y., Noh, M. S., & Yip, P. S. (2014). The impact of indiscriminate media coverage of a celebrity suicide on a society with a high suicide rate: epidemiological findings on copycat suicides from South Korea. *Journal of affective disorders*, 156, 56- 61.
- Johnson, J. (2018). The self-radicalization of white men: “Fake news” and the affective networking of paranoia. *Communication Culture & Critique*, 11(1), 100-115.
- Klein, E. & Golbeck, J. (2024). A Lexicon for Studying Radicalization in Incel Communities. *Proceedings of the ACM Web Science Conference*, 2024.
- Marwa, T., Salima, O., & Souham, M. (2018, October). Deep learning for online harassment detection in tweets. In *2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS)* (pp. 1-5). IEEE.
- McAdam, D. (1986). Recruitment to high-risk activism: The case of the freedom summer. *American Journal of Sociology*, 92(1), 64-90.
- Necaise, A., Williams, A., Vrzakova, H., & Amon, M. J. (2021, August). Regularity

- versus novelty of users' multimodal comment patterns and dynamics as markers of social media radicalization. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media* (pp. 237-243).
- Rodríguez-Sánchez, F., Carrillo-de-Albornoz, J., & Plaza, L. (2020). Automatic classification of sexism in social networks: An empirical study on twitter data. *IEEE Access*, 8, 219563-219576.
- Saha, P., Mathew, B., Goyal, P., & Mukherjee, A. (2018). Hateminers: Detecting hate speech against women. *arXiv preprint arXiv:1812.06700*.
- Shugars, Joseph, K., S. Gallagher, R. Green, J. Mathé, A. Q. An, Z. & Lazer, D. (2021). (Mis) alignment between stance expressed in social media data and public opinion surveys. *arXiv preprint arXiv:2109.01762*.
- Solley, C. M., & Santos, J. F. (1958). Perceptual learning with partial verbal reinforcement. *Perceptual and Motor Skills*, 183-193.
- Stark, R., & Bainbridge, W. S. (1980). Networks of faith: Interpersonal bonds and recruitment to cults and sects. *American Journal of Sociology*, 85(6), 1376-1395.
- Sugiura, L. (2021). *The incel rebellion: The rise of the manosphere and the virtual war against women*. Emerald Group Publishing.
- Swank, E., & Fahs, B. (2011). Students for peace: Contextual and framing motivations of antiwar activism. *Journal of Sociology & Social Welfare*, 38, 111.
- Texas Department of Public Safety. (2020). Texas Domestic Terrorism Threat Assessment.
- Thrash, T. M., & Elliot, A. J. (2003). Inspiration as a psychological construct. *Journal of Personality and Social Psychology*, 84(4), 871-889.
- Trujillo, M. Z.; Gruppi, M.; Buntain, C.; and Horne, B. D. 2022. The MeLa BitChute Dataset. In *Proceedings of the International AAAI Conference on Web and Social*

Media, volume 16, 1342–1351.

Trujillo, Amaury, And Stefano Cresci. (2022). "Make Reddit Great Again: Assessing Community Effects of Moderation Interventions On R/The_Donald." *Proceedings of The ACM On Human-Computer Interaction*. Vol 6, No. CSCW2, 1-28.

Van Raemdonck, N. (2019). The echo chamber of anti-vaccination conspiracies: mechanisms of radicalization on Facebook and Reddit. *Institute for Policy, Advocacy and Governance (IPAG) Knowledge Series, Forthcoming*.

Walton, G. M., Cohen, G. L., Cwir, D., & Spencer, S. J. (2012). Mere belonging: The power of social connections. *Journal of Personality and Social Psychology*, 102(3), 513- 532.

Weill, K. (2022). *Off the edge: Flat earthers, conspiracy culture, and why people will believe anything*. Algonquin Books.

Ysseldyk, R., Matheson, K., & Anisman, H. (2010). Religiosity as identity: Toward an understanding of religion from a social identity perspective. *Personality and Social Psychology Review*, 14(1), 60-71.

Zhang, Z., Robinson, D., & Tepper, J. (2018). Detecting hate speech on twitter using a convolution-gru based deep neural network. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15* (pp. 745-760). Springer International Publishing.