Identifying Misogynistic Rhetoric Within Advice-Seeking Communities

Mid Term Report

Tanvi Bhagwat
tbhagwat6@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

Aiswarya Bhagavatula
abhagavatula8@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

Disha Das
ddas71@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia

ABSTRACT

Reddit as a platform has made it easy for individuals to find communities that support their esoteric interests. Reddit lets the users be active under the cloak of anonymity, thus freeing them to discuss topics that they wouldn't have been able to in real life. Within Reddit, there exist communities that allow posting of personal problems for collaborative solutions. Unfortunately, the same cloak of anonymity that frees up a user to discuss topics without fear of exposure, allows another user to spread about vicious ideologies that are abhorred by the public. These advice-seeking communities are then forced to confront with ideologies shunned by the populace.

These communities tend to be really popular, and often have their posts make it to Reddit's frontpage, or even at times as sources of journalistic articles. Seeing how the language and the opinions expressed by the users tend to be seen by many, we wanted to conclusively see if the discourse on Reddit has grown in misogyny.

Our midterm goals include efficiently scraping the Reddit submissions and then analyzing them. The trend of an increasing negative discourse as seen in our results sets the stage for our final conclusion to this project.

1 DATASET DESCRIPTION

1.1 Data preparation

Source: Of the five subreddits we have chosen to analyze, we honed in on the "AskReddit" community for our midterm baselines. We chose the top 1000 posts from its first year of inception for data mining. To mine the posts we used Pushshift.io [1], which is a Reddit Search Application. After extracting the top 1000 posts, to mine the comments made under each post we used Reddit's API called The Python Reddit Wrapper(PRAW)[2]. Since Reddit only allows 100 requests at a time, PRAW helps break it into multiple API calls with two second delays. Along with the text contained in the posts and the comments, we also have access to other details such as the authors, the time the object was created and the score of the object.

Data preprocessing: Raw reddit data is available in the form of a CSV file. The head of the table is displayed in figure 1. The following preprocessing steps were implemented on the dataset:

- Undesired columns which did not contribute to the analysis were dropped. These columns are given in figure 2.
- Data in the columns 'author_created_utc' (date-time of creation of account) and 'created_utc' (date-time of creation of post) were converted from Unix Epoch Time to datetime object to make analysis easier.
- The dataset was sorted in order of the earliest date of posting to the latest, in accordance with the column 'created_utc'.
- The column 'comments' consists of a list of strings, each string being a top-level comment to the post. These strings are processed by removing non-alphabet letters, converting to lower case and splitting the comments down to individual words.
- 19 columns are added to keep track of misogynist words used in the comments. The 9 categories are *Belittling, Flipping the Narrative, Homophobia, Hostility, Patriarchy, Physical Violence, Racism, Sexual Violence* and *Stoicism.*
  - 'is_misogynist' column gives a 1 for the post if its comments include any misogynist word found in the lexicon.
  - 'num_cat_{number}' column gives the number of misogynist words used in the comment for each of the misogynist categories which are given by 'number' and range from 1 to 9.
  - 'is_cat_{number}' column gives a 1 if the comments contain atleast one word from the misogynist category, which is given by 'number' and ranges from 1 to 9.

The objective of the mid-term report was to analyze the trends in hateful comments among various subreddits over a period of time. This dataset suits the requirements. This is so because it provides us with date-time data of the posting as well as the content of the comments.

For the final report, we want to analyse whether the presence of this negativity drives down a user's involvement in the community. This dataset will suffice for this objective as it gives information about the user making posts. From this, the frequency of the user's posts can be found out over a period of time.

1.2 Raw Data Statistics

Selection of dataset: *r/AskReddit*
The focus of this subreddit is to "ask and answer questions that elicit thought-provoking discussions". It currently consists of 26.3M members.
Available information: The following information can be extracted from the dataset- *Author, post creation time, number of comments, over 18 or not, number of upvotes, title of post* and *comments on post*
Lexicon: We used the Lexicon built by Farrell et al [1]. It consists of 1,300 terms that belong to one of the 9 categories of misogyny,

[1] https://pushshift.io/
[2] https://praw.readthedocs.io/

as mentioned in table 1. This lexicon will help identify if the words used within the submissions are toxic or not, and if they are precisely what category of negative rhetoric they belong to.
Raw data statistics: The data statistics is given in table 2.

1.3 Data Analysis

We had two objectives for the midterm report. First, we wanted to find efficient ways of mining Reddit data with the caveat of being able to only make a certain number of requests or risk being barred from making any ever. Second, we wanted to see the language trends for a subsection of our data set. The data from AskReddit community has been utilized and certain conclusions could be drawn from it.

From Figure 1, we observe a growing trend in the usage of misogynous words across all 9 categories of rhetoric as suggested in the lexicon. The data spans the length of the year 2008, the first year of the subreddit's inception. Among the 9 categories, a steady increase in the usage of words associated with *Physical Violence* is observed. The same can be said for words associated with *Hostility*, which sees a steady increase amidst spikes in usage from time to time. The trends for other categories seem pretty much stable throughout the year. However, we cannot draw an emphatic conclusion about the rise of misogyny in this community from this data alone, given it only spans the length of one year. The usage of misogynous words seems to rise towards the end of the year, from December to January, which is a holiday season for most parts of the world. By taking into account that Reddit will have more user involvement in the holiday seasons, it is difficult to draw a conclusion about the growing trend from this dataset alone.

In Table 3, the frequency of words across all 9 categories are analyzed. The first column shows the percentage of misogynistic posts. A post is tagged misogynous if it's comment section contains atleast one word from the misogyny lexicon. Some of the most common categories seem to be *Physical Violence*, followed by *Hostility, Belittling* and *Racism*. The second column gives the log of the actual percentage of misogynous words used in the comments sections. The higher the number, the greater is the usage of these words. According to this dataset, *Hostility* and *Physical Violence* are the top contenders for the most popular categories of misogyny, followed by *Racism* and *Belittling*. Note that *Racism* has 670 words in the lexicon and is the third most frequently used category. Opposed to that, *Physical Violence* has only 73 words in the lexicon, but it shows a clear dominance in popularity of usage. We can draw a general conclusion that a subreddit user on AskReddit is more likely to come across references to physical violence and hostility than any other category of misogyny.

Our analysis into the categories continues with the results presented in Table 4. This table shows the top 5 most popular words by usage across all categories of misogyny. Upon a closer look at the most popular category, *Physical Violence*, it is observed that the top 5 words paint the picture of a skirmish. The words Hate, Hit, Cut and Kill occupy the top echelons, pointing to violent attitudes in the comments sections. Many terms such as 'Chad' (an attractive white male who gets all the girls), 'Beta' (refers to men who take a passive stance in life) and 'Omega' (refers to a man who chooses not to have an important role in a social or professional situation) are mostly specific to certain communities centered around men's rights. This only goes to show that these communities, that have been shown to be misogynistic, have a measurable influence on the AskReddit community.

Since AskReddit is a community for the general populace, we wanted to find out whether the misogynous content was spread out across all the posts in the dataset or present only in relation to posts that invited such discussions. Figure 2 gives a clear representation of this analysis. It draws a relation between the frequency of misogynous rhetoric found in the title versus the percentage of misogynous rhetoric found in the comments section of that post. There's a clear linear relation between them. A general conclusion about the misogynous content centering around posts that invite such discussions can be made, although this can only be verified with a much larger dataset. In future work, we want to perform sentiment analysis on the comment sections to demarcate hate comments from healthier discussions. This is beyond the scope of this report.

Harsh sentiments towards the posters can be analysed from Figure 3. This shows the relation between the score of the post, the percentage of misogynist rhetoric content in its comments section and the total number of comments. Score is the difference between the number of upvotes (which is the number of likes) and the number of downvotes (which is the number of dislikes). A post that has a low score and a relatively high number of comments must have received plenty of downvotes, implying that it has received negativity. From Figure 3, we observe the second quadrant to find posts that have received plenty of downvotes. First, the outliers were removed. It is observed that most of the points in the second quadrant indeed have 0.2% to 0.5% of misogynous words in their comments section. Thus, it can be safely assumed that a low score with high number of comments is indeed an indication that users using abusive misogynist rhetoric have taken over the comments section.

Lastly, we wanted to know how the young minds are influenced by this negativity in the communities. In Figure 4, the average number of misogynist rhetoric received for users over 18 and under 18 is presented. Thankfully, for the AskReddit community, the hate comments are comparably less frequent for users below the age of 18. Analyzing the effect of this negative influence is beyond the scope of this report.

2 EXPERIMENTAL SETTINGS

The experiment was run on Google Colab
Runtime Type: Python3
Hardware Accelerator: GPU
RAM: 12 GB

3 BASELINE RESULTS AND DISCUSSIONS

3.1 Baseline description

The approach used by Farrell et al. in their paper [1] was used as a model for our midterm report. They chose six Reddit communities that have identified themselves as "incels". They gathered all the posts and comments of each community from their creation all

Table 1: Lexicon of Misogyny

Category of Misogyny	Num terms	Examples
Belittling	58	femoid, titties, stupid cow
Flipping the narrative	7	beta, normie, men's rights
Homophobia	126	dyke, fistfucker, faggot
Hostility	303	bitch, cunt, whore
Patriarchy	8	alpha male, subjugate, suppress
Physical Violence	73	hit, punch, choke
Racism	670	nigger, raghead, pikey
Sexual Violence	22	rape, sodomise, gangbang
Stoicism	33	blackpill, cuck, hypergamy

Table 2: Raw data statistics

Community	Num Posts	minDate	maxDate	Avg num of comments	Max num of comments	Min num of comments	Comment Vocab size	Avg num of words per comments section	Avg score	Max score	Min score
AskReddit	1000	2008-02-08	2009-01-24	137.30	3467	3	84594	4916.06	64.12	681	22

Table 3: Misogynous posts across categories

Community	% of Mis. Posts	Log(% of mis. words used)
Belittling	48	-4.10
FlippingNarr	4	-7.04
Homophobia	11	-5.94
Hostility	78	-2.60
Patriarchy	6	-6.74
P. Violence	80	-2.63
Racism	47	-4.15
S. Violence	18	-5.47
Stoicism	11	-5.97

Figure 1: Frequency of misogynous words over time

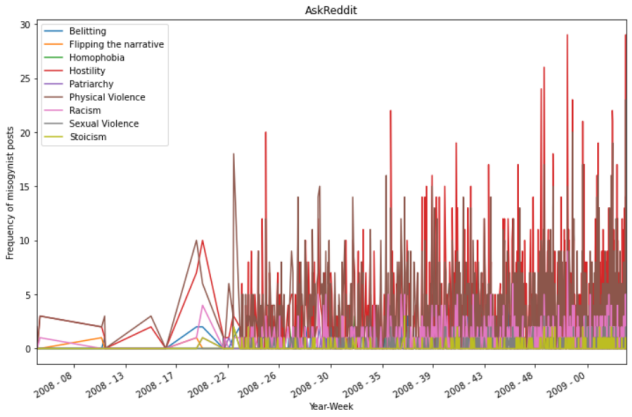
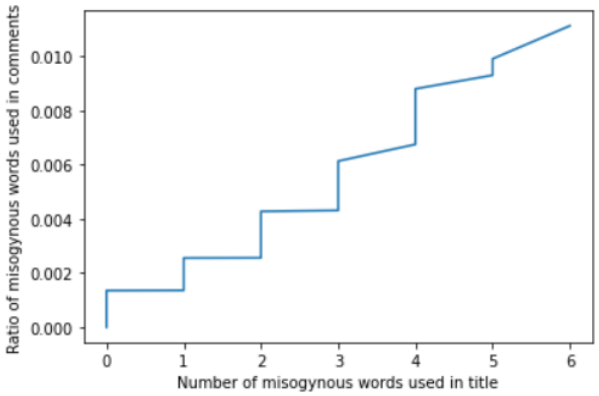


Figure 2: Comparison of misogynous words present in title vs comments section



the way through to 2019. They used the Pushshift API to crawl for all the posts and the comment submissions. Then these comments were checked for the presence of words from the hate lexicon and for misogynistic content.

For our data collection, we utilized the Pushshift API to scrape for the top 1000 submissions and all the comments posted under it. To have a better understanding of the data set, we only looked at the top 1000 submissions for one year of the "AskReddit" subreddit.

This paper was chosen as a baseline as it archived the growing trend of using negative discourse within Incel-identifying communities. We plan to showcase similar trends but for more general advice-seeking communities. The paper guided us in the right direction on choosing the correct lexicons, and to highlight the correct results.

3.2 Baseline result

The original paper calculated the frequencies of the misogynistic content across categories such as misogyny, belittling, flipping

Table 4: Top 5 Misogynous words used across categories

Belittling		FlippingNarr		Homophobia		Hostility		Patriarchy		P. Violence		Racism		S. Violence		Stoicism	
Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq
Female	203	Beta	32	Fag	35	Hate	479	Oblige	13	Hit	406	Black	328	Rape	101	Fuel	47
Dumb	196	Overthrow	5	Faggot	22	Hurt	264	Suppress	12	Cut	350	Bigger	189	Pound	47	Cope	37
Failure	108	Prevail	5	Homo	17	Bitch	238	Overwhelm	11	Kill	322	Nigger	19	Conquer	15	Rope	33
Bird	80	Vanquish	1	Queer	16	Beat	229	Omega	11	Force	242	Arab	17	Spank	10	Chad	7
Birds	56			Fags	15	Asshole	201	Compel	7	Kick	178	Crackers	17	Lolita	9	Friendless	1

Figure 3: Score vs ratio of misogynous words in the comments

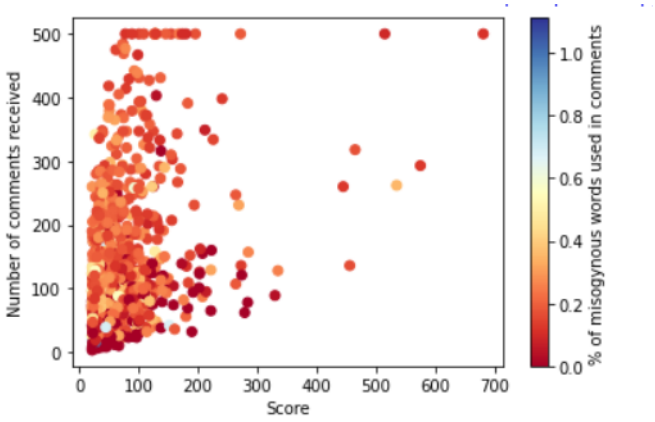
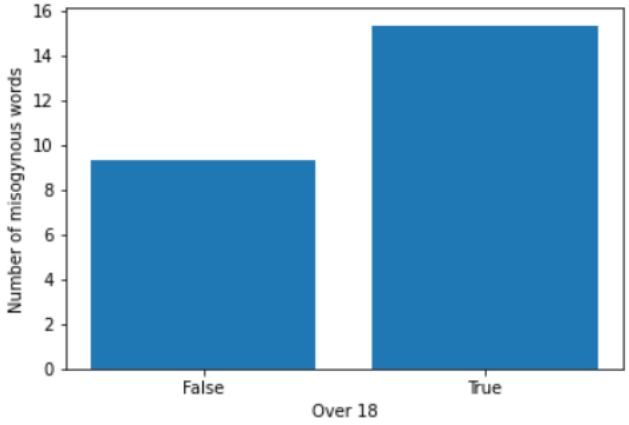


Figure 4: Average number of misogynous words used per age group



narrative, homophobia, hostility, patriarchy, violence, racism and stoicism. It also looked at the users using such words across the similar categories. It then mapped the usage of these words over time for all the communities.

We tried to emulate similar results. In comparison to the baseline results, our results look promising as they produce similar trends. For example, the words that top Table 4 in each category is consistent in both results. Similarly, the highest percentage of the negativity from the posts that are analysed fall under Hostility. Additionally, our results show some new words that were not covered by the paper. For example, the words 'asshole' and 'beat' are amongst the top five misogynous words under hostility, that were not a part of the results shown by the baseline paper. This shows that the limited data that has been chosen produces statistically similar results but identifies different words.

3.3 Result discussion

The paper by Farrell et al. [1] has scraped a wider dataset than the one we have currently available. This, coupled with the already toxic communities they are looking at, the frequencies of the negative words utilized is higher. The results also show that the vocabulary used in a general community is less specific than the vocabulary used by the Incel communities.

4 NEXT STEPS

With a limit placed on the amount of requests that can be made and the amount of engagement within the Reddit community, crawling to create a Reddit Data set takes time. Having seen promising results on a subsection of the data, for the remaining part of the project, we will continue gathering more data. We plan to conduct a similar analysis such as the one conducted in this report on the new dataset. Then we plan to reference what communities the users with the higher frequency of the negative rhetoric in their submissions post in more frequently. This information will aid the subreddit moderators as well as the users posting if they want their posts to be visible to certain users and or subreddits. We will also analyze the frequency of the users' posts over a period of time to verify our assumption that negativity drives down the users' involvement in the communities.

5 CONTRIBUTION

All the team members have contributed a similar amount of effort.

REFERENCES

[1] Tracie Farrell, Miriam Fernandez, Jakub Novotny, and Harith Alani. 2019. Exploring Misogyny across the Manosphere in Reddit. In *Proceedings of the 10th ACM Conference on Web Science* (Boston, Massachusetts, USA) *(WebSci '19)*. Association for Computing Machinery, New York, NY, USA, 87–96. https://doi.org/10.1145/3292522.3326045