Technical Rreport

Yakup Koray Budanaz

Sezin Öztüfek

Abstract—For the computational social science seminar we have collected data from various dating oriented subreddits in the internet forum Reddit and analyzed the posts and comments submitted in these subreddits for hate speech, fanaticism and gatekeeping and their popularity as well as activity trends over time. Our findings show that the hate speech and gatekeeping as well fanaticism does not increase over time but reaches a level of saturation that retains a constant level. The user activity over time shows a certain increase in popularity followed by a decay in popularity.

Index Terms—Reddit, hate speech, gatekeeping, fanaticism

I. INTRODUCTION

Reddit is a social media platform that allows it users to join certain subreddits where each subreddit resembles an internet forum where users can discuss anything related to a certain topic. We have analyzed dating related subreddits, especially male-oriented dating related subreddits: r/TheRedPill[1], r/BlackPillScience[2], r/PurplePillDebate[3], r/MensRights[4] and another controversial female dating subreddit r/FemaleDatingStrategy[5] for hate speech and other toxic behavior like gatekeeping [7] [1] and fanaticism [8] as well as the trends of these subreddits like user activity change over time.

Gatekeeping describes the activity of trying to control who gets particular resources, power or opportunities, and who does not. In our analysis we will assume a more specified definition of gatekeeping by defining it as how particular information and people are allowed into the subreddit, and which ones are discarded. For the fanaticism we have used the following definition: "The pursuit or defence of something in an extreme and passionate way that goes beyond normality" [5] [3]. We are going to assume any act that praises the subreddit's theory as an act of fanaticism and group them with other jargon related to the theory. The analysis for grouping was done manually by identifying terminology and jargon with their most common usage in the subreddit and by categorizing them in separate dictionaries for gatekeeping, hate speech, and fanatic theory.

II. RELATED WORK

Each subreddit has a set of rules and guidelines that determine the behavior inside that subreddit. These rules that govern the subreddits are created by the moderators without any user input. Moderators gatekeep what is deemed acceptable information to them for subreddit users to interact with. Moderator dynamic is what sets reddit apart from other social media platforms, and allow for an explicit gatekeeping of the user's object of view. There is no official built in central leadership on other social media platforms [4]

As gatekeeping evolves, fanaticism follows. George Santayana defines fanaticism as "redoubling your effort when you have forgotten your aim". The fanatic in this case is closed to any idea that is opposing to theirs. This closed environment of a subreddit where all big decisions are made by the moderator and without discussion pushed on to the users create this cult-like environment.

III. DATA

The data was collected from Reddit via The Multithreaded Reddit API Wrapper, PMAW[6], which is a wrapper for the Python Reddit API Wrapper, PRAW[7], as well as using Pushshift Reddit API[8]. The data has been collected from 01.01.2018 to 31.12.2021 and the data has been grouped by month for analysis. All posts, comments and uploaded images for r/FemaleDatingStrategy, r/BlackPillScience and r/TheRedPill were pulled, whereas for r/PurplePillDebate and r/MensRights only posts were pulled. The collected posts, comments and the text extracted from the uploaded messages are available in the code repository pill[9]

IV. METHODOLOGY REGARDING IDENTIFICATION OF GATEKEEPING

Aside from searching gatekeeping related word combinations in posts and comments the following methods for used to identify gatekeeping in subreddits.

A. Percentage of Deleted Posts and Comments

Reddit posts can be removed due to various reasons. It mainly depends on the subreddit's rules. For example in r/FemaleDatingStrategty a rule states "If you are a man, will be banned on sight for disrespecting our female-only space" any post from a man will be deleted. Deletion of posts and comments are not automated and moderators choose which comments will be removed. A higher percentage of removed posts and comments may indicate how closed the community is.

---

[1] https://www.reddit.com/r/TheRedPill/
[2] https://www.reddit.com/r/BlackPillScience/
[3] https://www.reddit.com/r/PurplePillDebate/
[4] https://www.reddit.com/r/MensRights/
[5] https://www.reddit.com/r/FemaleDatingStrategy/

[6] https://github.com/mattpodolak/pmaw
[7] https://github.com/praw-dev/praw
[8] https://github.com/pushshift/api
[9] https://github.com/ThrudPrimrose/pill

If a post or comment body is removed by a moderator or deleted by a user then the API returns "[removed]" or "[deleted]". The percentage of comments and posts that solely contain these body texts are counted as deleted comments and posts.
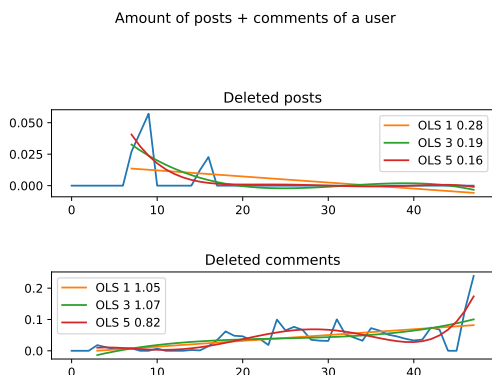
Amount of posts + comments of a user

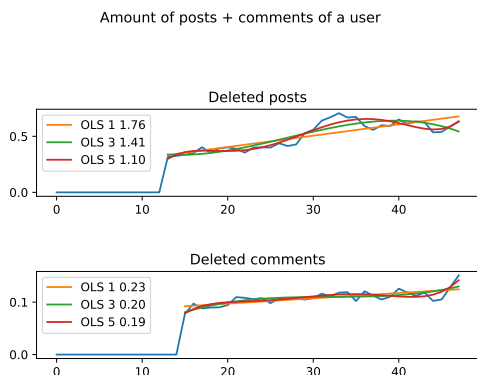Fig. 1. Percentage of posts and comments deleted in r/BlackPillScience from 01.01.2018 to 31.12.2021

Amount of posts + comments of a user

Fig. 2. Percentage of posts and comments deleted in r/TheRedPill from 01.01.2018 to 31.12.2021
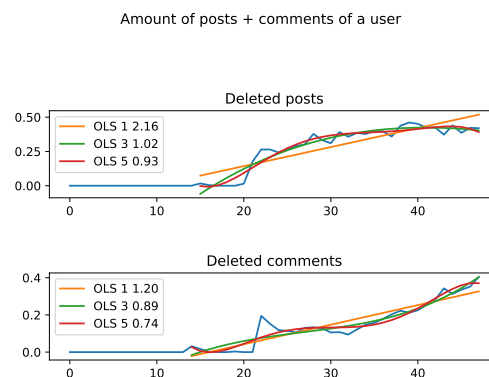
Amount of posts + comments of a user

Fig. 3. Percentage of posts and comments deleted in r/FemaleDatingStrategy from 01.01.2018 to 31.12.2021 (Subreddit was created in February 2019)
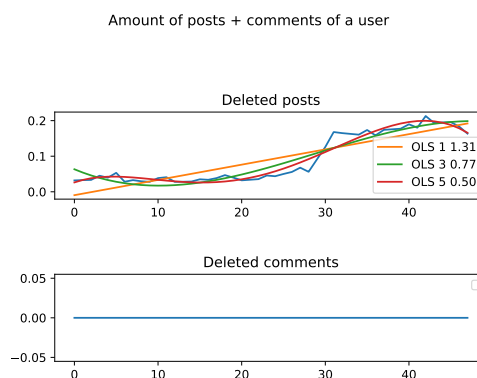
Amount of posts + comments of a user

Fig. 4. Percentage of posts and comments deleted in r/MensRight from 01.01.2018 to 31.12.2021

All subreddits show an increasing trend in the amount of deleted posts and comments over time. Where r/TheRedPill and r/FemaleDatingStrategt indicate more deletion compared to r/BlackPillScience and r/MensRight, indicate a more toxic community that performs regular gatekeeping.

B. Gatekeeping Tagging

Deletion of posts is not the only way gatekeeping is performed in these subreddits. A majority of posts were using the same vocabulary to degrade the non-participants of these subreddits. These posts also had the tendency to have a picture attached to them. For that reason we have determined that a possible gatekeeping post had these two prominent features:

1) Contains a picture: These pictures were usually screenshots that were showcasing either a conversation or it was a post taken from another subreddit in the purpose of mocking the person.

2) Contains gatekeeping vocabulary: For this purpose the vocabulary that was found through manual analysis of the posts were used. If the post had a picture attached, then it was checked if the post had one or more words that were

in our gatekeeping dictionary, if the post did not have any pictures, then the word count that was taken into account was two or more.

To extract the strings out of the images to analyze them further, we first needed to process the image. The teeseract algorithm that detects text inside the images was not working properly, so we used binary threshold with the python libraries opencv and numpy. Binary threshold allowed us to segment the image into only black and whites, enabling us to ignore any other part that is not related to the text. Then the reliability of the teeseract text detection algorithm increased immensely, we could process the text data that was in the images more efficiently and accurately. Lastly the mentioned features were extracted from the posts, determining the posts that possibly had gatekeeping in them.

V. METHODOLOGY REGARDING FANATICISM AND LIFE CYCLES

In Reddit any user can subscribe and join any subreddit, they are also called as lurkers as they are not active members of the subreddit but only read other peoples' posts and comments. Therefore they do not represent the activity of the subreddit. We define any user that has posted or commented in the given month to be active on that month to analyze how many people join and leave.

A. Last Activity of Active Users

A user is considered to be active on a given month if the user has posted or commented on that month. For every unique active user that has posted or commented between 01.01.2018 and 31.12.2021, the last month that the user has posted or commented is saved. And the graphs below plot the "last active" status of users. Higher values on the right and lower values on the left mean that the users more devoted to their subreddit, and cold imply a more fanatic community.
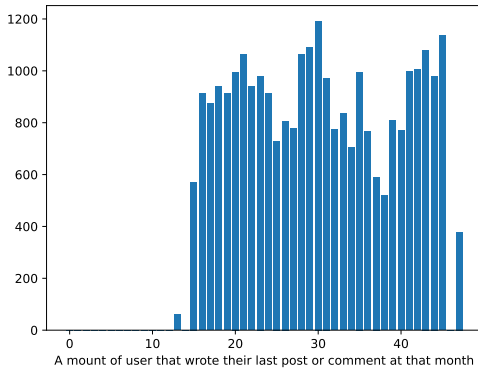
Fig. 5. Amount of users that went inactive on a given month in r/TheRedPill from 01.01.2018 to 31.12.2021
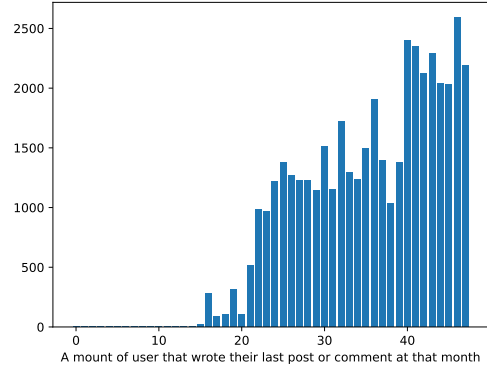
Fig. 6. Amount of users that went inactive on a given month in r/FemaleDatingStrategy from 01.01.2018 to 31.12.2021 (Subreddit was created in February 2019)
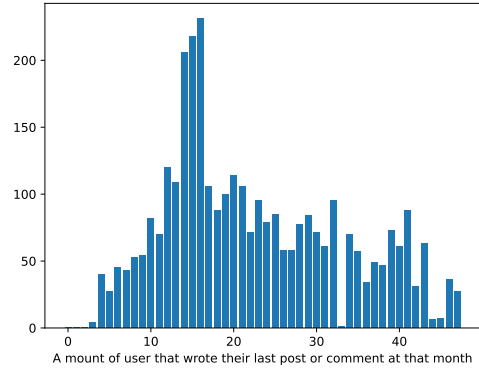
Fig. 7. Amount of users that went inactive on a given month in r/BlackPillScience from 01.01.2018 to 31.12.2021

r/BlackPillScience (BPS) and r/TheRedPill (TRP) show a constant influx and outflux of users where r/FemaleDatingStrategy (FDS) has a higher user base that has posted or commented in recent months, which can indicate that the members of FDS may be more dedicated and thus fanatical to their subreddit. r/BlackPillScience had a maximum of 400 active users in a span of 3 years, r/TheRedPill 3000 and r/FemaleDatingStrategt 6000. All of the 3 above mentioned subreddits show a decay in active user count after attaining a maximum, emphasizing the assumption that users that went inactive are not replaced with enough new users. For the active user change over time see the appendix. Figure 17, Figure 18, Figure 19.

B. Average Activity per User

We assume in case of increasing fanaticism even though there may be less users but the remaining users could post and comment more and result with more activity per user within the remaining user base.

To test this hypothesis a map was created that mapped every active user within the subreddit to a an array that saved the amount of posts and comments the user authored over months. The plots were generated to show every user's activity as a line to see a general pattern, no pattern could be identified and those plots are left out. Then the average activity on a subreddit on a month is calculated as the amount of posts and comments divided by the count of active users. The average activity plots are below.
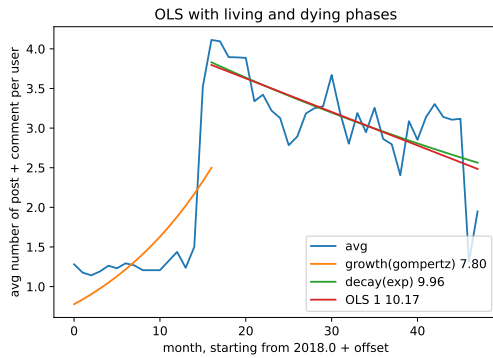

Fig. 10. Average post and comment amount per user in r/BlackPillScience from 01.01.2018 to 31.12.2021

Fig. 8. Average post and comment amount per user in r/TheRedPill from 01.01.2018 to 31.12.2021

Fig. 11. Average post and comment amount per user in r/MensRights from 01.01.2018 to 31.12.2021

Similar to the plots of the active user count per month (see Figure 17, Figure 18, Figure 19), the average activity per user also decreases. Applying a simple heuristics by assuming that the maximum average activity is the peak of the subreddit activity and dividing the life cycle into 2 phases called living phase and dying phase, while applying regression with gompertz function for the living phase and with the exponential function in the dying phase gives the smallest residual compared to other functions used for regression.

The residuals for least squares with polynomials from degree 1 to 7, and with the basis transformation with the gompertz function, exponential function, logarithmic function and square root functions resulted with visibly higher residuals.

The results supports our assumption that subreddits have lifecycles as they get more popular over time and lose popularity afterwards.

VI. VOCABULARY

Manual analysis has shown that r/PurplePillDebate, r/MensRights and r/TheRedPill share similar terminology which are also called subreddits of the Manosphere [2].

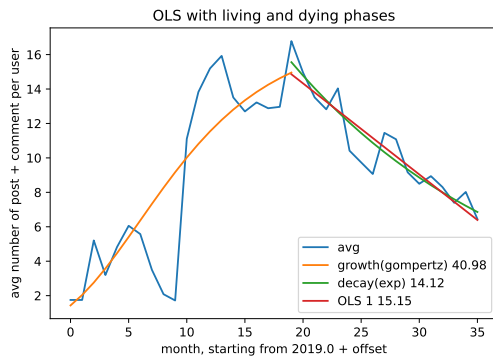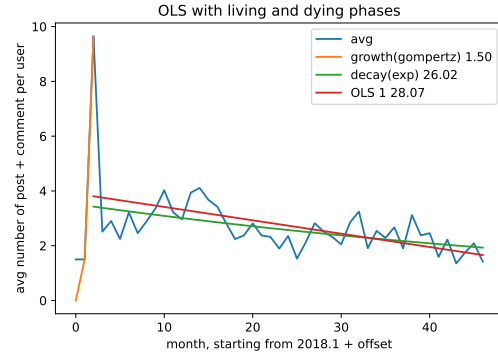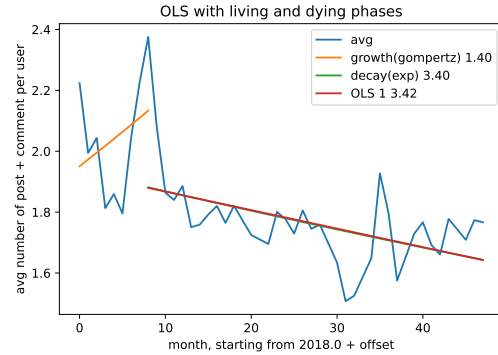Fig. 9. Average post and comment amount per user in r/FemaleDatingStrategy from 01.01.2018 to 31.12.2021 (Subreddit was created in February 2019)

Blackpill is also a part of the Manosphere but manual analysis has shown that they have their unique dialect slightly different from the other Manosphere subreddits, and r/FemaleDatingStrategy has its own jargon.

A. Gatekeeping related terminology

Any word or any commonly seen combination of words that belittles people outside the group (for example mentioning their ignorance) or sainting group members or degrading opposing ideas is considered to be a word of gatekeeping. Below or some common examples. The full vocabulary is available under the pill code repository (see footnotes). For example in the subreddit r/FemaleDatingStrategy pick-me describes a woman that tries to gain attention of men, is not a part of the subreddit and has low standards whereas a queen is a member of the subreddit that has high standards. For r/TheRedPill blue-pilled people, betas, omegas, zetas and cucks are people outside the group that do not know the redpill theory (also called the game) where as individuals of the red pill try to be an alpha. In the other hand r/BlackPillScience is an incel subreddit [6] describes people who can attain normal relationships as normies and roasties and use misogynistic derogatory terms like foid and femoid the describe women who can't understand their struggles, but also fakcels that should not call themselves incels.

FDS	TRP	BPS
pickme	soyboy	normie
queen	alpha	stacy
bangmaid	beta	bloop
cockholm	niceguy	beta
	bad boy	fakecel
	bloop	

the shoretened names FDS stand for r/FemaleDatingStrategy, TRP for subreddits of the manosphere: r/MensRights, r/TheRedPill, r/PurplePillDebate and BPS for r/BlackPillScience. Similar terms can be used in multiple subreddits udner same context

B. Fanatic Theory

The subreddits are aimed to discuss the theory and the ideology therefore these subreddits use lots of terms and subreddit specific jargon to describe events. Common words and word combinations regarding the ideology in r/RedPill are "alpha fuxx beta buxx" (alpha fucks beta bucks, often shortened as AFBB) that alpha gets the women and beta provides without getting any benefit. And other examples like alpha male of the group (often shortened as AMG), all the women are like that (meaning that women always act the same, often shoretened as AWALT) which is often paired with words like instinct and hypergamy. The incel subreddit r/BlackPillScience is fatalistic and their theory includes betas and alphas like the red pill ideology but also a lot of discussion about looks with height, hypergamy, jawline being common words. Also one has to maximize their status and looks which are described by words like statusmaxxxing, gymmaxxxing and looksmaxxxing. The r/FemaleDatingStrategy groups men

in categories like high value male (HVM), low value male (LVM), zero value male (ZVM) and the members of the group need to weed out, and filter out low or zero value males.

FDS	TRP	BPS
LVM	AWALT	looksmaxxxing
HVM	AFBB	statusmaxxxing
ZVM	AMG	gymmaxxxing
weedout	niceguy	hypergamy
grayrocking	bad boy	chad
gaslighting	heirarchy	dogpill
	game	
	plates	

the shoretened names FDS stand for r/FemaleDatingStrategy, TRP for subreddits of the manosphere: r/MensRights, r/TheRedPill, r/PurplePillDebate and BPS for r/BlackPillScience. Similar terms can be used in multiple subreddits udner same context.

C. Hate Speech

Any term that degrades outsiders, any racist, misandrist, misogynistic remarks are considered in this category. From r/FemaleDatingStrategy we have derogatory terms for men and women with low standarts such as Dickheads, All men are rapists (shortened as AMAR), dickhead, scrote (low value male), where as r/TheRedPill and r/BlackPillScience show more hatred towards women with cock carosel, femcel, alphawidow, hamster, pussy, plate, femoid, foid, landwhale etc, r/BlackPillScience shows hatred towards good looking men (also women) with words like chang, jamal, tryone and intimitated with a combination of looks or a word like chad (used for handsome white male).

FDS	TRP	BPS
AMAR	plate	landwhale
LVM	meek	femoid
scrote	stacy	foid
bitch	landwhale	stacy
	chameleon	
	fiona	

the shoretened names FDS stand for r/FemaleDatingStrategy, TRP for subreddits of the manosphere: r/MensRights, r/TheRedPill, r/PurplePillDebate and BPS for r/BlackPillScience. Similar terms can be used in multiple subreddits udner same context.

D. Fatalism

This is unique to r/BlackPillScience and r/PurplePillDebate where the theory is fatalistic and certain individuals will never be able to obtain love or be in a loving relationship. This can be further divided into subgroups where certain terminology is about coping with this fatalism, some vocabulary it to rant. Rant related vocabulary also contain self hate. Statements like "I hate being short", "I hate how normies have it easy" (to have it easy, HIE is another keyword) and "lie down and rot" (shortened as LDAR) are often used to rant, which also indicates self hate. To cope user often discusses suicide and stories that may give hope. A wizard is a slang term for

someone who had no sex until 30 years old, for example a term that can be derogatory or also a praise for a fellow incel.

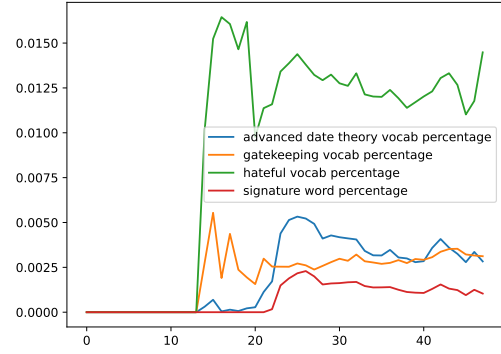| Rant	Cope
kisless	rope
touchless	gymcel
suifuel	wizard
shortcel	suifuel
LDAR	lifefuel

Fig. 12. Jargon and hate speech usage in posts and comments r/FemaleDatingStrategy from 01.01.2018 to 31.12.2021 (Subreddit was created in February 2019). The signature word is "Low Value Male"

VII. METHODOLOGY REGARDING HATE SPEECH

From posts, their titles and body texts whereas from posts their body texts were tokenized, lemmetized; cleared from punctuation, stop words, demonstrative determiners, pronouns and any word that was not defined in our dictionary that had a length less than 3. The remaining tokens were saved in a separate file. The amount of words for a given for specific subreddit group. r/TheRedPill, r/MensRights, r/PurplePillDebate belong the TRP group, r/BlackPillScience to BPS and r/FemaleDatingStrategy to FDS group. The words and abbreviations that were contained in the vocabulary list were searched within saved array of tokens and their percentage were plotted for analysis.

The plotted word groups correspond to the dictionaries defined above in the vocabulary section. The plotted groups are defined as following:

- Signature word: Defined only for FDS "Low Value Male" and for BPS "Chad"
- Gatekeeping vocab percentage: Vocabulary used within the group for gatekeeping.
- Advanced Date Theory Vocab: Vocabulary used within the group for fanaticism and the theory of the subreddit.
- Fatalism: Special the r/PurplePillDebate and r/BlackPillScience where their ideology is a fatalistic way of viewing romantic relationship where they will never obtain love or a loving relationship.
- Hateful vocab percentage: This is the umbrella dictionary that combines every category (Fatalism, Gatekeeping, Fanatic Theory, Signature Words) of every group (TRP, FDS, BPS) and checks for toxic and hateful vocabulary usage.

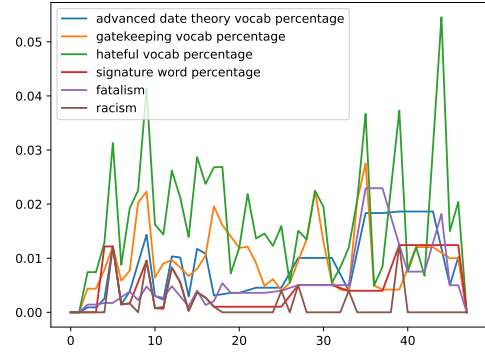The vocabulary mentioned in the vocabulary section

Fig. 13. Jargon and hate speech usage in posts and comments in r/BlackPillScience from 01.01.2018 to 31.12.2021, The signature word is "Chad"

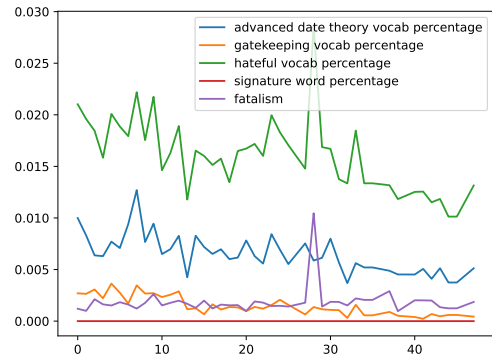Fig. 14. Jargon and hate speech usage in posts in r/PurplePill from 01.01.2018 to 31.12.2021, No signature word, in this subreddit
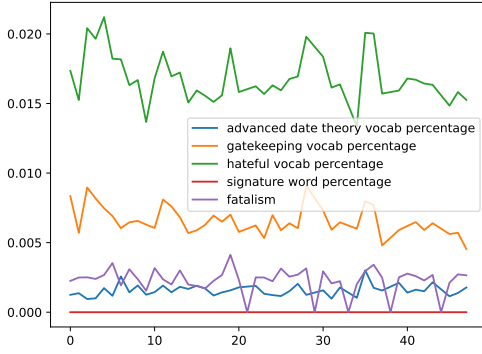
Fig. 15. Jargon and hate speech usage in posts in r/MensRights from 01.01.2018 to 31.12.2021, No signature word in this subreddit
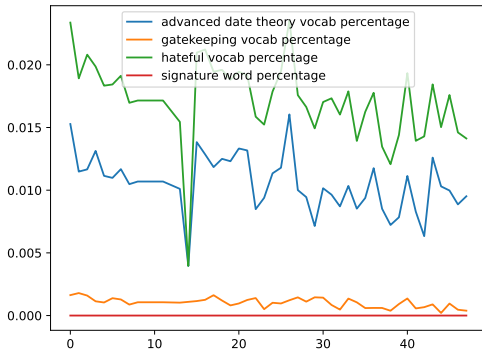

Fig. 16. Jargon and hate speech usage in posts and comments in r/TheRedPill from 01.01.2018 to 31.12.2021, No signature word in this subreddit. The sudden spike downwards in February of 2019 is caused by the quarantine of reddit.

Fatalism and hateful vocabulary shows a high correlation in the r/BlackPillScience. Since the amount of posts and comments in r/BlackPillScience are not enough to create a smooth plot therefore the plot is spiky but in other subreddits the plots are much smoother. Fanatic theory related vocabulary and hateful vocabulary find a level of saturation. They stay around a constant level in subreddits r/FemaleDatingStrategy, r/TheRedPill, r/MensRights. The percentage of vocabulary usage in the subreddits of the Manosphere is higher than in the r/FemaleDatingStrategy. This could also be caused by the fact that Manosphere has been existing for a longer time and has developed a bigger jargon, in our vocabulary the TRP and BPS group has more entries than in the FDS group. Even though a constant saturation level can be seen, there are sudden spikes in increase of hateful vocabulary and fanatic vocabulary but every spike results with a swift decay. In many subreddits gatekeeping related and fanatic vocabulary slightly decreases over time, with the expection of r/FemaleDatingStrategy where gatekeeping vocabulary also increases over time.

VIII. CONCLUSION

The percentage of hate speech, theory related and gatekeeping related terminology stay at a relatively constant level. With the exception of r/FemaleDatingStrategy where gatekeeping related terminology increases. The percentage of tagged gatekeeping posts also increase over time in this subreddit, where in others the percentage of gatekeeping related terminology and tagged gatekeeping posts remain either constant or decrease over time. Even though the most popular subreddit, out of all the others mentioned, the average user activity in the r/FemaleDatingStrategy is decreasing with a faster rate than every other subreddit. Even with the declining active user count it is still the most active subreddit in comparison to the all of the analyzed subreddits. From this data we can assume that more gatekeeping results faster decay of popularity of the subreddit.

Actions against these subreddits might result with spikes in popularity. For example quarantining the r/TheRedPill corresponds to a spike in hate speech and in popularity (Feburary 2019). Therefore one could conclude that leaving these subreddits (and other hate subreddits) to themselves and not taking any action against them can be a better action as the graphs of their life cycles has shown that they will lose popularity over time naturally.

IX. FUTURE WORK

The gatekeeping tagging can be improved to use more indicators and to make use of machine learning to tag such posts better. Semi-supervised learning method can be used to first determine what makes a gatekeeping post, to then further analyze them better. The methodology used in this research was only simplified to two determinants, a post having a picture and using gatekeeping vocabulary. These determining factors can be broadened further.

The sudden spikes in the data regarding hate speech must be further analyzed as there still obvious spikes in hate speech that are yet unexplained. Our assumption is that spikes are caused by real world events. Another assumption could be that an outlier (or multiple) is introduced that singlehandedly changes the plots, checking for outliers and whether these outliers leave the community must be calculated and scripts have to be improve to detect such outliers.

The vocabulary used for detection can be further expanded and improved.

To further analyze the incel and blackpill communities, the fatalistic vocabulary can be expanded and divided into two groups for rant related vocabulary and cope related vocabulary. This can be used to analyze self hate further.

Reddit has banned many incel subreddits in the past and banning incel subreddits has caused incels communities to migrate away from Reddit and to create their own websites and forums. r/TheRedPill and r/FemaleDatingStrategy has their own sites too, and it is especially stated that the websites are created in case they are banned. Banning these subreddits will also make it much harder to analyze these groups further as it will require writing web-scrappers specially for their

websites which makes these groups much less open to further research. The website that belongs to the Manosphere, the websites for the r/TheRedPill and r/FemaleDatingStrategy can be analyzed if these websites will be more hateful than their Reddit counterparts. To conclude whether allowing less toxic groups to exist to prevent users from migrating to more radical and more toxic communities.

X. CONTRIBUTIONS

- Data Collection with PMAW API - Yakup
- Tokenization and Lemmetization - Yakup
- Analysis For Percentage of Deleted Posts and Comments - Yakup
- Manual analysis for Vocabulary - Yakup
- Analysis for Last Activity - Yakup
- Analysis for Average Activity per User - Yakup
- Image parsing and Analysis gatekeeping detection - Sezin
- Chapter for Related Works - Sezin
- Chapter for Future Work - Joint
- Chapter for image parsing and gatekeeping detection - Sezin
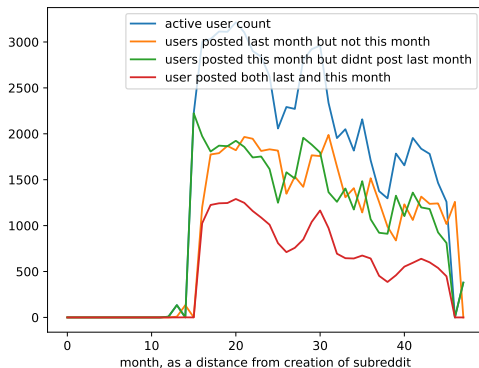- Chapters for the rest - Yakup

APPENDIX

Fig. 17. Amount of active users on a given month in r/TheRedPill from 01.01.2018 to 31.12.2021
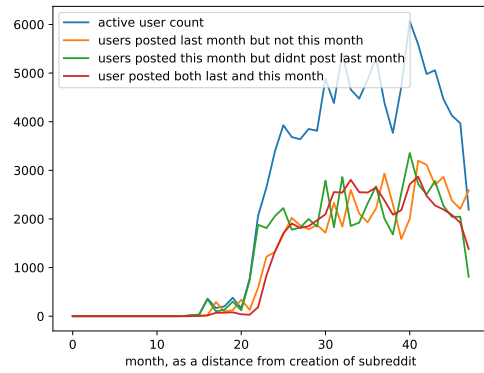

Fig. 18. Amount of active users on a given month in r/FemaleDatingStrategy from 01.01.2018 to 31.12.2021 (Subreddit was created in February 2019)
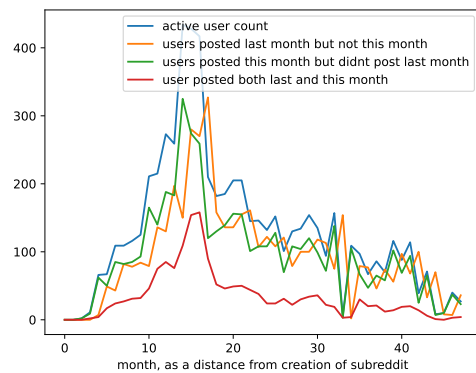

Fig. 19. Amount of active users on a given month in r/BlackPillScience from 01.01.2018 to 31.12.2021

REFERENCES

[1] Karine Barzilai-Nahon. Gatekeeping: A critical review. *Annual review of information science and technology*, 43(1):1–79, 2009.

[2] Tracie Farrell, Miriam Fernandez, Jakub Novotny, and Harith Alani. Exploring misogyny across the manosphere in reddit. In *Proceedings of the 10th ACM Conference on Web Science*, pages 87–96, 2019.

[3] Tõnu Lehtsaar. *Äärmusliku usugrupi psühholoogia*. T. Lehtsaar, 1997.

[4] Kimery S. Lynch. Fans as transcultural gatekeepers: The hierarchy of bts' anglophone reddit fandom and the digital east-west media flow. 2020.

[5] Kalmer Marimaa. The many faces of fanaticism. 2011.

[6] Eero Salojärvi, Matti Rantanen, Emilia Nieminen, Alina Juote, and Heidi Hanhela. The'incel'phenomenon in the digital era–how echo chambers have fueled the incel movement. 2020.

[7] Pamela J Shoemaker and Timothy Vos. *Gatekeeping theory*. Routledge, 2009.

[8] Alberto Toscano. Fanaticism: a brief history of the concept. *Eurozine*, 2006.