Analysis of the Reddit Moderation Network

Group 6

John Hofmann
University of Iowa
john-hofmann@uiowa.edu

Wyatt Wilson
University of Iowa
wyatt-wilson@uiowa.edu

Austin Beschen
University of Iowa
austin-beschen@uiowa.edu

Moria Slabaugh
University of Iowa
moria-slabaugh@uiowa.edu

KEYWORDS

Power Moderator, PRAW, Reddit, Post Removal

1 INTRODUCTION

Reddit is a discussion board website where registered users can submit content such as images, text, links, and videos. While users can upvote or downvote content they like and dislike, there are also moderators that are given extra authority to help uphold the communities' guidelines. The primary focus of our project will be uncovering the moderation network of Reddit. Moderators are given extra authority that grants them the ability to remove posts or comments, ban members who may be spammers or are breaking the community rules, and add other members as moderators. The work done by Reddit moderators is volunteer and as stated by J. Nathan Matias, "Scholarship on volunteer moderation has tended to view this work as digital labor, civic participation, or oligarchy" [8]. We looked through the major subreddits of Reddit and analyzed the information we found to try and understand the moderator network of Reddit.

This type of information can be valuable in determining information regarding Power Mods. "Power Mods" are moderators who moderate a very large amount of subreddits and users [1]. Power Mods are often looked on poorly because of their ability to control such a large number of members and direct the conversations being held. We are not trying to determine whether or not they are negatively affecting Reddit but rather identify the reach that these power mods have. We will also be able to identify who the major players are and how big the network of Reddit is. To build on this, we also investigate two properties of a moderator, their influence, and their power. The influence of a moderator is more shallow-focused and determines how many users a moderator moderates. This is essentially the number of users they can reach through the subreddits they moderate. The power of a moderator focuses instead on how much of their influence they are exerting. Power is defined as the number of posts the moderators are removing from the subreddits they moderate. Due to the availability of certain moderation actions, we focus only on the removal of posts to subreddits they moderate, as banned users are not publicly available in almost all cases.

Our biggest question in this research is determining if there are specific moderators who fall under the category of "Power Mods". Are there moderators who have a significantly large influence compared to the average moderator? The user we found who had the most overall influence, Blank-Cheque, had around 293 million influence and stated "When I was really into it i would do, as i said, literally over a thousand mod actions a day and clear dozens of queues." [3]. This is a motivating idea to see what the influence of the average moderator looks like, as well as identifying other power moderators. Are there moderators who exert a lot of power within the subreddits they moderate? There are some other interesting things to look at when it comes to moderation on Reddit however, such as how many moderators the largest subreddits have and how many subreddits are moderated by those moderators. Understanding moderation on Reddit overall can be extended to other areas such as the average amount of moderators per subreddit and how it changes based on size.

We found that the subreddits with the largest communities have the most moderators on average. This aligns with our expectations as subreddits with more users will need more moderation. We also found that moderators in the top 100 subreddits moderate significantly more communities than those top 1000. This shows that moderators on Reddit have reputations and moderating on one top community increases their chances to moderate on another. When looking at the moderation happening on the top 50 subreddits we found that the subreddits with the most auto-moderation correlated to having the lowest percentages of manual moderation. Although, the opposite correlation was not visible.

2 RELATED RESEARCH

Today, content moderation on any social media platform is a necessity. As we began looking to see how much moderation happens within the top subreddits, we discovered that moderation is not always as transparent as it probably should be. This led us to the study, "Does Transparency in Moderation Really Matter?: User Behavior After Content Removal Explanations on Reddit" by Jhaver, Bruckman, and Gilbert [7]. This paper provides insights into the effectiveness of transparency in content moderation on the social media platform Reddit. Their findings suggest that while transparency in moderation may have some short-term benefits, such as increasing compliance with content rules, it is not a panacea for improving user behavior in the long term. The authors' experiment showed that users who received detailed explanations for why their content was removed were more likely to submit an appeal, indicating that transparency may also increase user challenges to moderation decisions. This study highlights the complex nature of content moderation and the challenges that platforms face in balancing transparency, user satisfaction, and community standards. These discoveries are key to helping us understand why moderation may not be as easy to see as some might want. Reddit,

as a whole, may believe that allowing their moderation decisions to be completely transparent may create more problems and result in the moderators having to work extra which is backed by this paper.

Along with looking at human moderators another aspect of moderation is bots. The first of these bots was AutoMod. The paper, "Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator",[6] their findings suggest a need for audit tools to help tune the performance of automated mechanisms, a repository for sharing tools, and improving the division of labor between human and machine decision making. Then came along a new bot in the industry, CrossMod, a much more advanced machine learning bot. The paper, "Crossmod: A Cross-Community Learning-based System to Assist Reddit Moderators", by Chandrasekharan et al. (2021)[5] presents Crossmod, a machine learning system designed to assist Reddit moderators in identifying problematic content. This study addresses the increasing challenges of content moderation in large online communities, where manual moderation by human moderators is often time-consuming and can be prone to errors. Crossmod's cross-community learning approach leverages data from multiple subreddits to improve the accuracy of moderation decisions, making it a promising solution for subreddits with limited moderation resources or expertise. The authors' evaluation of Crossmod's performance using a large dataset of comments from multiple subreddits demonstrates its high precision and recall in identifying problematic content. They deployed Crossmod in a controlled environment, simulating real-time conversations from two large subreddits with over 10 million subscribers each—r/science and r/Futurology. The authors found that Crossmod would significantly extend the capabilities of moderators when deployed on Reddit. This was an important study for our research because it shows why we need to look at how much auto-moderation is done on each subreddit. Due to the large number and increasing number of users on Reddit specifically, perfect manual moderation is almost unachievable for the largest subreddits and thus bots must be put in place to keep communities in line with the rules.

Finally, our research wants to look at which subreddits have the most moderation and the reasons behind the amounts of moderation they have. Iris Birman at the Georgia Institute of Technology looked at a topic very similar to this in that she analyzed subreddits with varying degrees of moderation and the reasons behind those degrees.[2] Birman found that the subreddits which involved more objective opinions such as r/science, r/space, and r/history had the highest percentage of comments removed from the first data gathering to the second. Subreddits that were built around a community sharing a passion for a certain hobby or topic, such as r/pokemon, r/fitness, and r/trees had the lowest percentage of removed comments. Overall, this data helps our understanding of reddit moderation because it gives an indication of what we can likely expect from our own research and thus we can compare our results.

3   METHODS

For our automated data collection, we used PRAW. PRAW is the Python Reddit API Wrapper, which allows data to be scraped off Reddit [4]. PRAW is what we used for all data regarding the moderators of the Top 50, Top 100, and Top 1000 most popular subreddits, as well as the sizes of those subreddits. It is important to note that PRAW's most popular subreddits are based on posts, and not on size. On top of this, our data on the Top 100 and Top 1000 most popular also contains all subreddits that each moderator of those Top 100 or Top 1000 are moderators for, no matter their ranking. This allows us to get a better understanding of how many subreddits the moderators of the most popular subreddits are involved in.

Our analysis also required some manual data collection. Since the deletion of posts by moderators is not publicly available nor is it available via any APIs, we used a site called Reveddit, which allows for a subreddit to be searched and provides information regarding the number of posts made, how many of them were removed by bots, how many were removed by mods, how many were removed by users, as well as another category which cannot distinguish between user removed or mod removed. In our collection of this data, we named these categories "Removed Automatically", "Removed by Mod", and "Removed by Mod and User". However, we decided to exclude "Removed by User", as this information is not valuable to our analysis in any way.

We manually collected the removed posts of the top 50 subreddits over two separate time periods a couple of days apart (April 20th, 2023 and April 22nd, 2023). This data collection was originally delayed due to issues causing Reveddit to be unavailable until then. When collecting this data, we only collected the first page of posts that were available for each subreddit on each of those dates. We then split this into three separate categories. Percent Bot Removed shows the percentage of total posts removed by automation. Percent True Mode Removed shows the percentage of total posts removed by mods that were able to be guaranteed as removed by a mod by Reveddit. Percent Mod Removed is the adjusted amount of percentage of total posts removed by mods, which includes all of the posts guaranteed to be removed by mods as well as half of the posts which Reveddit classified as user or mod removed. Since we could not distinguish between user and mod removed posts, we decided to split them in half and give half of those posts as user removed and half as mod removed, as we did not want to completely ignore this data.

After we determined the percentage of removed posts for each of the top 50 subreddits manually, we were able to use it to give a value to the amount of power exerted by a moderator. To do this, we looked through all 999 different moderators of the top 50 subreddits. For each subreddit, we took the total number of posts made * Percent Mod Removed to determine how many posts were removed by moderators. We then divided that by the number of moderators of that subreddit to determine how much power each individual moderator should be given. We divided it equally between all moderators as there is no way to determine which moderators were deleting posts. After this, we went through all 999 moderators, and for each of the top 50 subreddits they moderated for, as some would moderate for multiple, we added that subreddits value to their power. This would be the total power of each moderator exerts.

However, in our manual data collection, we had one more issue to take care of. We also needed to make sure that we removed bots from the equation, as we did not want automated bots in our

moderators dataset to be given power, especially since we separated posts that were deleted automatically out of the power given to moderators. To do this, we went through all of the moderators within the top 50 subreddits and grabbed the group of all moderators that had the word "bot" within them somewhere. We then manually went through the list and whitelisted moderators that were not bots but had "bot" within them. This is especially necessary as there were moderators with names like ex-robot-x, Robotguy39, and immaRealBotacc to name a few. It is important that we did this, as this means that automated moderators would not take away power from the human moderators, as we divided it by the total number of moderators. The whitelist also meant that we would not exclude possibly important moderators or give extra power to moderators of the subreddits they moderated.

4 DATA

As stated above, we collected moderation data from the top 50 subreddits over two separate time periods. The list of all 50 subreddits that we collected data from are shown in Figure 1. However, these lists change over time as subreddits become more or less popular, so it is important to note that these are the top 50 subreddits at the time of data collection. Shown in Figure 2 and 3 are 5 of the top 50 subreddits that we analyzed manually with all of the collected data on the date April 20th, 2023.

Figure 1: The Top 50 Subreddits used in our manual subreddit moderation data collection.

Subreddit	Total Posts	20-Apr	Removed Automatically(Bot)	Removed by Mod	Removed by Mod & User
Home	286		48	38	4
AskReddit	289		62	10	25
funny	296		120	65	54
damnthatsinteresting	288		72	51	23

Figure 2: Example Snippet of 5 of the Top 50 Subreddits showing their total posts that we recorded on that date and their removal breakdown.

Subreddit	Mod + User / 2	Percent Bot Removed	Percent True Mod Removed	Percent Mod Removed
Home	2	0.167832168	0.132867133	0.13986014
AskReddit	12.5	0.214532872	0.034602076	0.077854671
funny	27	0.405405405	0.219594595	0.310810811
damnthatsinteresting	11.5	0.25	0.177083333	0.217013889

Figure 3: Example Snippet of 5 of the Top 50 Subreddits showing the calculation of different removal percentage statistics

We also collected data on the top 100 and top 1000 subreddits and their moderators. This data was collected and stored in Top100Subreddits.xlsx and Top1000Subreddits.xlsx, respectively. Both of these files share the same structure. The first column, column 0, contains the name of the moderator. Every column after that column is one of the subreddits that they moderate for. Some moderators moderate as few as one subreddit, while others moderate hundreds. We also collected the sizes of all of the subreddits. The top 100 sizes and top 1000 sizes are in Top100SubredditSizes.csv and Top1000SubredditSizes.csv, respectively. There is some other data that was gathered and separated using this data. Created using the reddit_mods.py file, there are two sets of spreadsheets. Top100AllModCount and Top1000AllModCount contain all of the subreddit's moderator counts that are present in the respective top number of subreddits. However, they only include the overlap of moderators who moderate the top number of subreddits and are also a moderator of that subreddit. Top100AllModSubCount and Top1000AllModSubCount contain the moderators and the number of subreddits that they moderate that are present from those sets.

5 TOOLS

As previously stated, we used PRAW, The Python Reddit API Wrapper[4]. PRAW is an API that allows you to grab a Reddit instance which you can probe for information regarding subreddits, which we used for our retrieval of all data about the top 100 and top 1000 subreddits, including the moderator lists and the sizes. We used Python to write our scripts to make use of the data that we retrieved. Within Python, we used Pandas for some of the data manipulation and analysis. Pandas can read from files such as .csv and .xlsx and store that in a data frame to be used. This is what we used for our data analysis regarding the .xlsx files in the supplemental materials. Matplotlib was used for creating the graphs of the data within Python. As well as this, we used the site Reveddit for our manual data collection, as it allows for specific subreddits to be searched for and provides information, with a one-minute delay, on all posts and whether they have been deleted.

6 RESULTS

6.1 Top 100 versus Top 1000

Table 1 and Table 2 show values for the top 100 and top 1000 most popular subreddits. We see that there are on average 18.59 moderators per subreddit in the top 100, and 12.581 moderators per subreddit in the top 1000, which is roughly 1.48 times as many moderators per subreddit in the top 100 as the top 1000. There is an interesting difference in the values of the top 100 and top 1000 as well, as we see that on average, a moderator present in the top 100 moderates 2.59 times as many subreddits as a moderator present in the top 1000, which includes the top 100. Figures 4 and 5 display the number of moderators and how many subreddits they

Table 1: For the top 100 and top 1000 most popular subreddits, the number of moderators, the number of unique subreddits they moderate, and the total number of times those moderators show up in all the unique subreddits.

Top Count	# Moderators	# Unique Subreddits	# Total Unique Subreddit Moderators
Top 100	1859	25774	41719
Top 1000	12581	59533	109060

Table 2: For the top 100 and top 1000 most popular subreddits, lists their number of unique moderators, unique subreddits, average unique subreddit per moderator, and the total number of moderators if you add up the

Top Count	# Moderators	# Unique Subreddit per Moderator	# of Subreddits per Mod
Top 100	1859	13.86	22.44
Top 1000	12581	4.73	8.67

moderate for the top 100 and top 1000, respectively. As well as this, they add 2.93 times as many unique subreddits per moderator on average. There is a likely explanation for why this value is greater than the previous, however. All of the subreddits added by the moderators in the top 100 will have some overlap with subreddits moderated by the moderators of the top 1000, minus the actual top 100 as those moderators are all already included in the top 100 set. This means that there will be fewer overall unique subreddits added per moderator due to the fact that there are already 25774 unique subreddits in the dataset from the top 100. We also looked at the maximum number of moderators across all the subreddits mentioned in both the top 100 and top 1000 data sets. For the top 100 dataset, it was found that "PartyParrot" had the most moderators with a total of 66. However, from the top 1000 set, "science" had the most moderators with a total of 1565. It was surprising to see such a stark difference given the expectation would be that a subreddit from the top 100 would have the most moderators in it.
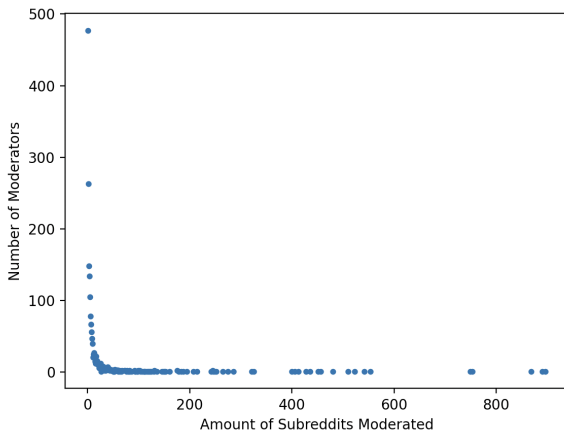

Figure 4: The number of moderators who moderate specific numbers of subreddits in the top 100 most popular subreddits.
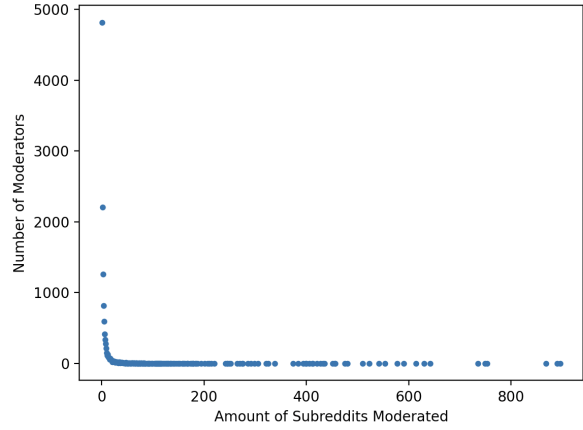

Figure 5: The number of moderators who moderate specific numbers of subreddits in the top 1000 most popular subreddits.

6.2 Moderator Power and Influence

Figures 6 and 7 display moderator power and influence, respectively. Something important to note is that there are clear outliers similar to what we have seen previously when it comes to moderator power. We see a user with utilized power as high as 62.15 and many users sitting with powers in the 0-10 range. We found the average power of moderators across all 50 top subreddits to be 7.01, significantly less than some of the moderators on the higher end. It is important to note that the values of utilized power are completely dependent on the size of the dataset. Our dataset is based off of the two separate time-frames of data collection, specifically the first page of Reveddit for those subreddits. The addition of more time-frames would increase the numbers, however, the trends of the data should stay the same. When we take a look at moderator influence, it also follows a similar trend. We see users as high as 293 million, as Blank-cheque is, and a large majority of users sitting below 1 million. However, there can be an explanation for these similarities as we distributed power equally across moderators, which means that the more subreddits a moderator moderates, with the more users, the more posts will most likely be deleted and the more power they will be credited for.

7 ANALYSIS

When we compare the difference between the top 100 most popular subreddits and the top 1000 most popular subreddits, there is a clear distinction between the moderators of the top 100 and top 1000 most popular subreddits. Not only do the subreddits have more moderators in general, but the moderators themselves moderate more subreddits on average and also introduce more unique subreddits to the overall dataset. This would imply that moderators of the top 100 are much more involved in the Reddit space in general. From Figure 4 and Figure 5, we can see that when we compare the top 100 and top 1000 subreddits, the y-axis increases by nearly 10x and keeps a similar pattern while the x-axis stays the same,
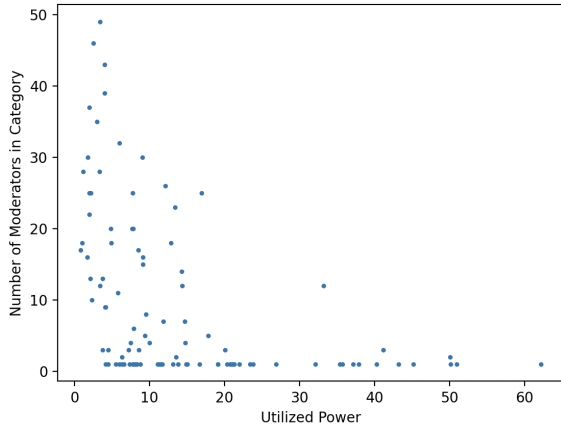
Figure 6: The number of moderators who have utilized specific amounts of power, defined as the number of posts they were given credit for deleting.
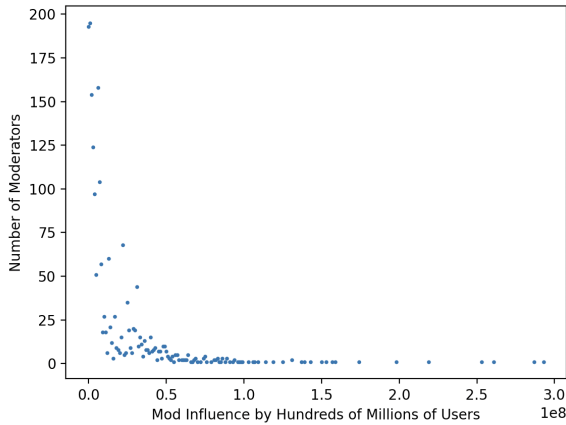

Figure 7: The number of moderators with a specific amount of influence, defined as reachable users.

meaning that we can most likely expect the majority of power mods to be present within the top 100 subreddits. Even if we were to go to greater limits of subreddits we would likely see little change, especially once we take into consideration that the average moderator within the top 100 moderated roughly 2.59 times as many subreddits as the average moderator within the top 1000.

Looking at the influence and power of moderators within the top 50, we see that there is the existence of power mods. There are both moderators who have a significantly larger amount of influence than the average moderator and those who utilize a significantly larger amount of power than the average moderator. Throughout all of the connecting graphs of moderators and the subreddits they moderate, their influence, and their power, we see similar trends that all point in the same direction of a large number of lower-valued

moderators and a small number of higher-valued moderators, the power mods.

8 LIMITATIONS

This study has three main limitations. The first of these limitations is when we looked at the influence of each moderator represented by totaling the number of people they moderated. This may not be a true reflection of their influence because for one, not all users of a given subreddit are necessarily active on that subreddit. Secondly, Reddit used to subscribe new users to certain subreddits automatically, thus on these communities there are some users who are unaware that they are subscribed to certain subreddits. These two factors would increase the influence number of moderators although it would not accurately reflect their true influence over users.

The second limitation of our study is the inability to determine which moderators are actually using their moderation power. As stated before, power was distributed equally between all moderators of a subreddit, minus bots, so someone who is inactive as a moderator will receive the same power as someone who is active. While this serves the purpose of our general idea, it does not allow us to focus in on specific moderators as we would like. Given Reddit's current moderation policies, however, this is not possible for us to achieve.

Another limitation is the amount of manual data collected. We collected manual data from the top 50 subreddits over two separate time frames, acquiring roughly 600 posts on average, with some as high as 1200, for each subreddit. While this is not a super small number, it is capable of being influenced by outliers which could happen due to current events. For example, r/politics will most likely have different amounts of removed posts depending on the current political landscape. If there are highly contentious topics in public discussion, more conflict will likely result in more posts being removed, which would give those moderators a higher amount of power. In a perfect world, consistent data collection over a longer period, such as a year, would provide a much more consistent outlook.

9 CONCLUSIONS

Moderation is important in online communities like Reddit because it helps to create a safe and productive environment for users to share and engage with content. We hoped to uncover if there are certain moderators that do more moderation than others to keep their subreddits safe.

Through our data, we have discovered that there are clearly power moderators within the Reddit moderation network. We found moderators who exerted far more power than the average moderator, as well as moderators who had far more influence than the average moderator. We are not capable of deciding whether power moderators are a good or a bad thing for Reddit as a whole, or if they are abusing their power, but we do know that they exist. We also discovered some stark differences between the top 100 and top 1000 most popular subreddits, determining that moderators of the top 100 were far more likely to be involved in a large number of subreddits than the top 1000.

9.1 Future Expansion

There are multiple routes this data could be taken in the future to provide more insight overall. We have determined whether there are power moderators within the moderator network, but are power mods inherently a bad thing? While this question is not a simple yes or no, it could be useful to further analyze power mods to determine if they are abusing their power or not. They have access to a large number of users or perform a large amount of moderation, but if this power is being abused then there could be a very clear issue with such major figures. This could even be taken a step further to see if we can understand more specific traits of moderators, such as their views and if there are specific topics or views that they moderate more heavily than others, showing a significant bias in their moderation.

Another area for future research would be a heavier focus on the specific subreddits' moderation. We mainly focus on moderators themselves within this research but it is clear from the data we collected that certain subreddits experience more moderation than others. This could be a mix of reasons, such as due to the nature of the topics within the subreddit, the size of the subreddit, or the existence of power abuse as mentioned above. It would be interesting to compare the amount of moderation power mods do on specific subreddits versus other ones and their topic. Is the cause of their moderation linked to the subreddit itself, or the moderator themselves?

Another possible topic would be the comparison of bot and human moderation over large time frames. We collected automatically removed posts through the use of bots with our manual dataset. This data could be useful in some of the other areas of research, such as determining specific subreddits and volatile the content or regulation of it, or both, is, but also could lead into another question. Has the use of automated moderation increased as time has moved on, or will it increase in the future? Does this correlate to a decrease in human moderation, or is there no effect, possibly implying heavier content moderation? All of the aforementioned areas for future research are possible to expand upon but were out of the scope of this project.

9.2 Challenges

One of the major challenges of this project was finding ways to measure the power of moderators. There were a lot of things that could pose possible roadblocks that we needed to solve. First, there was no API available that would allow us to see the total moderation of a subreddit, as it is not made public. For this, we used Reveddit to allow us to grab the data manually. Next, we faced the issue of not being able to tell which moderator deleted each post, so we decided on distributing deletions equally among all moderators. We also faced issues with undecided deletions categorized as user and moderator deleted as well as the existence of bots as moderators, both of which needed to be handled.

Another major challenge we faced was working with the current Reddit API to gather a multitude of data. This data included: the sizes of the subreddits, the moderators of each subreddit, and the top 1000 moderators by number of users they moderate. The Reddit API was extremely slow in gathering these data types which lead to large quantities of time needing to be allocated to letting our script run to gather this information. For example, when finding the top 1000 moderators by influence, the script had to run for approximately 18 hours. This was an unavoidable problem and we had to let our computers run overnight to finish the data collection.

REFERENCES

[1] Auror Andrachrome. Power mod. *Reddit Wiki*, 2022.
[2] Iris Birman. Moderation in different communities on reddit – a qualitative analysis study. 3(CSCW), may 2018.
[3] Blank-Cheque. Overiew for blank-cheque. *Reddit*, sep 2019.
[4] Boe Bryce. Praw. 2016.
[5] Eshwar Chandrasekharan, Chaitrali Gandhi, Matthew Wortley Mustelier, and Eric Gilbert. Crossmod: A cross-community learning-based system to assist reddit moderators. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019.
[6] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. Human-machine collaboration for content regulation: The case of reddit automoderator. 26(5), jul 2019.
[7] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. Does transparency in moderation really matter? user behavior after content removal explanations on reddit. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019.
[8] Nathan Matias. The civic labor of online moderators. 5, 2016.

9.3 Supplemental Material

Code and Data: https://github.com/jhofmann33/RedditPowermodAnalysis

10 CONTRIBUTIONS

Austin Beschen worked on creating reddit_mods.py and compared the top 100 and top 1000 most popular subreddit datasets to each other. He created the associated tables and the figures of the data. He also worked on the manual data collection of the moderation done on the top 50 subreddits. Finally, he worked on writing, editing, and reviewing the final paper.

Wyatt Wilson worked on editing and reviewing the final paper. He worked on the manual data collection of the total amount of moderation done on the top 50 subreddits. Analyzed and generated graphs based on the manually gathered data. He gathered many of the sources used as references throughout the paper and wrote the related research portion. Also worked on creating the bibliography.

John Hofmann worked on data collection and analysis, totaling, and graphing/figuring. Data collection was done using the PRAW API. Code base work and graphing in python3. Generation of output and csv files for use in analysis and further code based work. Code and files uploaded to github repository (see Supplemental Material). Manual data collection for totaling utilized moderator power. Helped in paper writing and formatting for ACM format.

11 SIGNATURES