

Measuring changes in Reddit users' interests in response to coronavirus

Hammas Bin Tanveer

University of Iowa

ABSTRACT

The rampant use of Tor as a tool to attack internet services while maintaining anonymity, has led to severe differential treatment for Tor users by the online service providers who are affected by this abuse. In this study, we aim to detect such abusive traffic in real time while maintaining user anonymity by looking at flow characteristics of traffic which have been approved by the Tor ethics board. We extract these characteristics from traffic logs we collect in order to train a classifier which can detect attack traffic.

First we show that malicious and regular traffic are distinguishable using these characteristics for a baseline case (where Tor is not considered), indicating that malicious traffic has fundamentally different flow characteristics from regular traffic. Then, we show that this same detection is possible for the same traffic sent through Tor.

KEYWORDS

Networks, Privacy and Security, Tor, Abuse Mitigation

ACM Reference Format:

Hammas Bin Tanveer. 2019. Measuring changes in Reddit users' interests in response to coronavirus. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

The world is currently going through a pandemic of an unprecedented scale with around 188 countries and territories reporting cases of the novel Coronavirus [4], known officially as Covid-19. This has led to countries adopting measures which make sure that most of its residents do not come into contact with each other; measures such as state and nation wide quarantine and lock downs. Almost every country that

saw cases of the virus enacted these measures, for amount of time ranging from weeks to months, with India instituting the largest quarantine in the world, closing down every city and village in the country [3].

These lock downs led to the closure of most businesses and restaurants with people resorting to staying inside their homes, even for work, only going out for essential matters like grocery shopping. Epidemiologists and public health experts have suggested that coping with this pandemic would require "a more rapid change of behavior in recent human history" that they have seen before [17]. This sudden change in behavior affected all major walks of human life; more than 50% of all of Americans started working from home [15], which is a stressful experience for many [16]. Furthermore, activities like international and domestic travelling have reduced to about 10-20% of their total activity before the outbreak [18] leading to not only reduced travel for work, but reduced travel for other important activities like going back home for international people working or travelling in foreign countries, and also leisure.

In addition to all this, individuals are forced to let go off all outdoor activities like dining out or even going to games as all restaurants and major sports leagues have been shut down to encourage social distancing [7] [9]. This has led to a shift, in human life, of a degree that was unfathomable before the outbreak and now that it is indeed happening, people are dealing with in various sorts of ways. Some people have said that their productivity has actually gone up during these times as they have much more flexible work hours [15] but others are having a significantly worse experience. Mental health and domestic violence issues are on a rise since the countries went into lock down and in certain parts of Australia, suicide hot-lines have experienced an 800% increase in calls since January [10].

The outbreak has also changed how we live our lives online. There has been a noticeable increase of 15 to 27% on websites like Youtube and Facebook. There has also been a trend in people shifting towards apps like TikTok and houseparty which provide online entertainment, instead of outdoor activities, and interacting more with websites related to online gaming as opposed to sports [8].

In this paper, we will analyze how people reacted to the Coronavirus outbreak on Reddit, which is the self-proclaimed

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

front page of the internet with over 130k+ active communities and 430 million active users [2].

2 BACKGROUND AND RELATED WORK

Reddit

Reddit is currently the sixth most popular site in the USA with over 330M active monthly users [2]. It is often dubbed as the front page of the internet and contains 130k+ active communities.

Reddit is a content-aggregation/social media platform whose users are called *Redditors*. These users interact with other users in groups or communities known as *subreddits*. These *subreddits* can be dedicated to anything from politics to discussion about a very specific video game and *Redditors* subscribing to these *subreddits* can communicate or interact with each other through posting on these communities on which others can comment or commenting on other's posts or even sharing other's posts to a *subreddit* the content was not initially posted in.

Reddit is particularly suited for a study like this because it already has these communities formed around certain general topics and most of the posts within a single *subreddit* belonging to a certain category, for example Politics, will indeed talk mostly about Politics itself. This allows for a better chance to analyze that during situations as pressing as the current pandemic, how people belonging to certain communities react.

Reddit has been a part of a lot of previous studies which observe patterns in Reddit communities to certain stimuli in a *subreddit* for example [12] which talks about how people respond to certain changes in weight management communities or [13] which examines how different mental health communities react to different mental health problems.

The closest study to what we will be doing in this paper is perhaps [11] which analyzes how the whole of reddit responds to a widespread community unrest. This unrest was caused because of some policy changes enacted by Reddit itself causing people to migrate to other social media/content aggregation platforms; we will analyze how a major social unrest, the global pandemic, caused people to migrate within Reddit.

Coronavirus - Covid-19

Covid-19 is by far one of the deadliest and most contagious viruses that we have come across. Compared to other deadly viruses, Covid-19 has one of the highest R naught values, which roughly translates to the number of people one infectious patient can transmit the disease to [14]. So an R naught value of less than 1 means an infected patient would on average transmit the disease to less than one person and an R naught value of more than 1 means that the person would

spread the disease to more than 1 person. Covid-19 has an R naught value in the range of " 1.4 to about 5. The WHO has estimated the R naught of COVID-19 to be around 1.95" [6] which is more than the R naught for diseases like H1N1, seasonal flu and even the deadly flu pandemics of 1918 and 1957 which killed millions of people.

All this information being available to individuals through the internet coupled with the 2.1 billion media mentions of Covid-19 as compared to only 150 million of HIV, SARS, MERS and Ebola combined [5] has not only convinced people to stay indoors, but to also change their behaviors both, offline and online.

3 METHODOLOGY

Data Collection

For this study, we collected data from the start of May 2019 to the end of April 2020. We analyzed 60 subreddits comprising of 2.5 million different posts during this time frame. We first went through all the 130k+ subreddits and only kept subreddits that had more than 10k subscribers. We then decided the following broad categories we wanted to base our analysis on:

- Politics
- Online Entertainment
- Gaming
- Indoor Activities
- Outdoor Activities
- Travelling
- Sports
- NSFW
- Mental Health
- Random

To decide which subreddits belonged to which category, we first chose known subreddits belonging to each of these categories. We then used [redditstats.com](https://www.redditstats.com/)'s [1], which has Reddit's historical data since 2012, to decide which other subreddits belonged to which category. [redditstats.com](https://www.redditstats.com/) has features called "subreddits with the most common keywords" and "subreddits with the most common users" associated with each of the subreddits. Subreddits with the most number of common users and keywords with the known subreddit of a particular category were chosen as part of the same category. We picked on average 5 subreddits for each of these categories. For the "Random" category, we randomly chose 10 subreddits from our sample of subreddits with 10k+ subscribers.

Online Entertainment					Mental Health				
	May-Aug	Sep-Dec	Jan-April	% Change		May-Aug	Sep-Dec	Jan-April	% Change
netflix	1850 4604	1738 5221	2577 7527	39 63	povertyfinance	850 2639	942 2957	1088 3351	28 27
TrueFilm	421 1021	437 1138	645 1581	53 55	Anxietyhelp	377 1055	441 1297	677 1806	79 71
entertainment	3841 3096	4290 3525	5420 4571	41 47	Needafriend	5130 9970	5689 13850	8050 18136	56 82
podcasts	1449 3588	1207 3596	1388 4141	-4 15	mentalhealth	4672 10784	5644 16183	6293 17556	34 63
ListenToThis	11205 17610	9627 20237	11931 23777	6 35	offmychest	18950 39642	17813 48260	20328 51509	7 30
Indoor Activities					Outdoor Activities				
	May-Aug	Sep-Dec	Jan-April	% Change		May-Aug	Sep-Dec	Jan-April	% Change
cooking	3081 7272	3573 10483	4184 12067	35 66	camping	2005 4998	1107 3198	1083 3026	-45 -40
baking	3168 5725	3971 8693	5849 12741	84 122	hiking	2829 5838	1946 4882	1864 4395	-34 -24
HomeImprovement	5137 11758	4753 13374	5983 16393	16 39	gym	1069 2632	1112 3020	1655 4234	54 61
beauty	1580 2899	1640 3337	2392 3992	51 38	bouldering	553 1301	531 1253	535 1156	-3 -11
homegym	571 1673	604 1770	1402 4097	145 144	festivals	544 1564	379 1075	460 1215	-15 -22

NSFW					Gaming				
	May-Aug	Sep-Dec	Jan-April	% Change		May-Aug	Sep-Dec	Jan-April	% Change
porn	5863 2861	5346 3940	6179 6579	5 129	pcgaming	4798 9136	4623 11485	5290 13318	10 45
nsfwvideos	380 381	414 342	519 438	36 14	ps4	1163 2525	9610 23076	13004 31611	1017 1151
gonewild	0 0	14621 14476	100242 62147	- -	xboxone	4712 10216	9329 25174	10749 29486	128 188
AsiansGoneWild	6380 2809	7383 3867	10744 5393	68 91	gaming4gamers	352 871	481 1271	683 1880	93 115
OnOff	1607 2276	1795 2738	2543 3598	58 58	gamernews	421 533	422 597	519 703	23 31
Politics					Sports				
	May-Aug	Sep-Dec	Jan-April	% Change		May-Aug	Sep-Dec	Jan-April	% Change
PoliticalDiscussion	1194 3433	1424 3614	1838 5479	53 59	sports	2797 5234	2652 5405	3050 6065	9 15
worldpolitics	0 0	5563 3065	13972 6088	- -	Ohlympics	177 210	181 198	220 245	24 16
PoliticalHumor	0 0	9048 13565	14651 21106	- -	PremierLeague	506 1105	553 1148	729 1303	44 17
ukpolitics	7097 3794	8219 6948	4398 3851	-38 1	nba	0 0	3747 5913	20031 26343	- -
Libertarian	5115 5447	3915 4526	4477 5365	-12 -1	cricket	6583 7223	3522 4240	3370 4092	-48 -43

Travelling					Random				
	May-Aug	Sep-Dec	Jan-April	% Change		May-Aug	Sep-Dec	Jan-April	% Change
travelhacks	298 697	329 812	273 761	-8 9	philosophy	1325 2872	1206 3375	1383 3880	4 35
solotravel	1861 4555	1657 4906	1487 4403	-20 -3	Futurology	2988 2636	2824 3136	2237 3121	-25 18
vagabond	401 878	306 749	279 714	-30 -18	TropicalWeather	266 504	295 645	49 99	-81 -80
Shoestring	253 772	266 847	272 897	7 16	TrueChristian	1143 1860	1027 1833	1257 2008	10 8
roadtrip	723 2055	588 1579	599 1641	-17 -20	suspiciousquotes	496 1891	276 1026	246 881	-50 -53

Table 1: This table shows the number of average posts and unique users posting in each month in the format "Average post in the phase | Unique users in the post". The % change column in the end shows the percentage change for these metrics in the same format

Data Organization

Now that we have established the categories and the subreddits belonging to each of these categories, we divided the posts from each of these subreddits into 3 time phases of 4 months each:

- May 2019 to August 2019
- September 2019 to December 2019
- January 2020 to April 2020

The first and second phases represent time frames before the pandemic as the first official case of the virus was announced on the 31st of December 2019 by the Chinese

	Online Entertainment	Travelling	Politics	Sports	Mental Health	NSFW	Indoor Activities	Outdoor Activities	Gaming	Random
Online Entertainment	X	36	329	278	187	29	251	54	337	120
Travelling	54	X	54	59	101	10	133	48	55	18
Politics	146	21	X	184	86	8	137	21	145	89
Sports	210	26	312	X	85	21	138	46	261	62
Mental Health	96	64	205	131	X	119	492	74	351	88
NSFW	248	3	13	13	9	X	12	5	26	1
Indoor Activities	248	87	369	221	391	39	X	128	386	118
Outdoor Activities	96	61	122	96	126	30	255	X	125	30
Gaming	192	16	180	200	151	17	159	36	X	37
Random	108	26	249	116	101	7	143	25	106	X
MIGRATION	11,445	-593	26,162	24,950	27,370	66,172	20,000	-2136	51,640	-421

Table 2: This table shows the number of users who have migrated from one category to the other. The number in each cell represents the number of users who were not present in the category in the horizontal axis but were present in the category in the vertical axis during phase 1 but ended up posting in the horizontal category in phase 3. The last MIGRATION row represents the total number of new users seen in phase 3 as compared to phase 1 for each of the horizontal categories.

Government. The period from May to December were divided into two phases to serve as a control for increased activity; if activity increased from the first time frame to the second, we would know it was not because of the pandemic itself but some other external factor. This would allow us to compare and contrast, the changes in each of these subreddits as a direct result of the pandemic itself or simply because of unknown random external factors.

Data Analysis

We based our analysis on the basis of number of posts a *redditor* makes on a subreddit. A post is content in any form - text, images, videos, GIFs etc. - that a user posts and others can interact with it by "upvoting" or "downvoting" it or even by commenting on it. We measure activity in a subreddit in terms of the number of posts made in that particular subreddit and not the comments made under each of these posts, or the number of "upvotes" or "downvotes".

From our posts data, we got each post of each author made in each of these subreddits along with the UNIX timestamp associated with it. We then divided the activity - the number of posts - into individual months and averaged these number of posts over each of the time phases. For example, for a subreddit belonging to the "Politics" category, we first isolated posts made in each of the months - May 2019 to April 2020 - and then averaged posts made in each phases by dividing the total number of posts in this phase by 4. This helps us gauge the activity in these subreddits and analyze trends from one phase to the next as shown in Figure 1.

We then analyzed how many users shifted from one category to another. To perform this analysis, we first calculated how

many users were present in a certain subreddit category **A** and not in subreddit category **B** in the first phase. We then calculated how many of these users which were exclusive to subreddit category **A** (relative to subreddit category **B**). Once we have this set of unique users, we then calculated how many of these users were present in subreddit category **B** in the third phase. This would tell us how many users who were initially not part of a certain category of subreddits migrated to this category because of the pandemic. We calculated this for each category to the other, giving us the 2D matrix in Table 2.

4 EVALUATION AND RESULTS

Increase in user Activity

Figure 1 shows the increase in user activity in the individual subreddits belonging to each of the 10 categories. We calculated the percentage increase of posts from the first time phase to the last.

We observe that subreddits belonging to categories "Online Entertainment", "Mental Health", "Indoor Activities" and "NSFW" have a clear increase in the number of posts and individual users posting from the first phase to last. We observe subreddits like **r/netflix** and **r/mentalhealth** see a higher increase in the number of individual users posting than the increase in average number of posts over the three phases. On the other hand, we see subreddits like **r/nsfwvideos** and **r/beauty** see a larger increase in the number of average posts over three months as compared to the increase in individual users. We also observe that certain subreddits in categories belonging to *Outdoor Activities* and *Travelling* have a negative change in percentage in both the number of

average posts and the number of individual users posting in these subreddits.

User Migration between categories

Next we observed, how many users, who were initially not a part of a certain category of subreddits ended up being a part of the subreddit in question over the course of the three phases.

Table 2 shows a clear trend in the *Mental Health* and *Indoor Activities* categories getting the most number of users from other subreddit categories; they also have similar numbers for total number of users that migrated to them over the course of these 3 phases. Note that the total Migration number shows the total number of users that ended up posting in a subreddit of the category in question and not necessarily from the subreddits we are analyzing.

We also see low numbers for categories like *Travelling* and *Outdoor Activities* which is expected given that most people had to stay indoors due to the quarantines and lock downs enacted by governments during the third phase.

5 CONCLUSION AND FUTURE WORK

In this work, we looked at how people are reacting to the global pandemic online, specifically on Reddit. We observe that people are posting much more, generally in most subreddits, since the countries went into lock down. The most increase in posts and unique users posting was seen in categories like *Indoor Activities*, *NSFW*, *Mental Health* and *Gaming*, whereas categories like *Outdoor Activities* and *Travelling* saw a drop in the activity and unique users posting. We also observe that people migrated the most to subreddit categories that saw the highest increase in user activity.

For future work, we could look at what kind of sentiments people belonging to these categories have. We can also look at the kind of content that was being posted before and after the pandemic in these different categories and how that has changed.

REFERENCES

- [1] [n.d.]. . <https://subredditstats.com/>
- [2] 2019. *Reddit by the Numbers*. <https://www.redditinc.com/>
- [3] 2020. *Coronavirus: India extends world's biggest lockdown*. <https://www.dw.com/en/coronavirus-india-extends-worlds-biggest-lockdown/a-53300137>
- [4] 2020. *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*. Retrieved May 13, 2020 from <https://coronavirus.jhu.edu/map.html>
- [5] Fabio Bergamaschi Dr Stephanie Starling Unvers Labs David McCandless, Omid Kashan. 2020. *Coronavirus Infection Trajectories*. <https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/>
- [6] Berkeley Lovelace Jr. 2020. *The coronavirus may be deadlier than the 1918 flu: Here's how it stacks up to other pandemics*. <https://www.cnn.com/2020/03/26/coronavirus-may-be-deadlier-than-1918-flu-heres-how-it-stacks-up-to-other-pandemics.html>
- [7] Amanda Kludt and Brenna Houck. 2020. *Restaurants and Bars Shuttered Across the U.S. in Light of Coronavirus Pandemic*. <https://www.eater.com/2020/3/15/21180761/coronavirus-restaurants-bars-closed-new-york-la-chicago>
- [8] Ella Koeze and Nathaniel Popper. 2020. *The Virus Changed the Way We Internet*. <https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html>
- [9] Victor Mather. 2020. *Updates on Sports Canceled by Coronavirus*. <https://www.nytimes.com/2020/03/13/sports/sports-canceled-coronavirus.html>
- [10] Chuck Morris. [n.d.]. .
- [11] Edward Newell, David Jurgens, Haji Mohammad Saleem, Hardik Vala, Jad Sassine, Caitrin Armstrong, and Derek Ruths. 2016. User migration in online social networks: A case study on Reddit during a period of community unrest. In *Tenth International AAAI Conference on Web and Social Media*.
- [12] Gisele Lobo Pappa, Tiago Oliveira Cunha, Paulo Viana Bicalho, Antonio Ribeiro, Ana Paula Couto Silva, Wagner Meira Jr, and Alline Maria Rezende Beleigoli. 2017. Factors associated with weight change in online weight management communities: a case study in the Loselt Reddit community. *Journal of medical Internet research* 19, 1 (2017), e17.
- [13] Albert Park, Mike Conway, and Annie T Chen. 2018. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in human behavior* 78 (2018), 98–112.
- [14] Vanessa Bates Ramirez. 2020. *What Is R0? Gauging Contagious Infections*. <https://www.healthline.com/health/r-nought-reproduction-number>
- [15] Zapier Editorial Team. 2020. *Half of America just started working from home. So, how's it going?* <https://zapier.com/blog/wfh-report/>
- [16] Derek Thompson. 2020. *The Coronavirus Is Creating a Huge, Stressful Experiment in Working From Home*. <https://www.theatlantic.com/ideas/archive/2020/03/coronavirus-creating-huge-stressful-experiment-working-home/607945/>
- [17] Jay J Van Bavel, Katherine Baicker, Paulo S Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J Crockett, Alia J Crum, Karen M Douglas, James N Druckman, et al. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* (2020), 1–12.
- [18] Gregory Wallace. 2020. *Thousands of grounded planes. Nearly empty flights. How the coronavirus is affecting US airlines*. <https://www.cnn.com/2020/03/28/politics/coronavirus-impact-airlines/index.html>