Study of various forms polarization on the Reddit network

# Introduction

In this experiment we study various forms of polarization on Reddit.

-First we study ..

-Secondly ..

# **Pipeline preprocessing**

# General

The process selection of the data follows a specific pipeline. For each step, I save the data on the disk, so I can separate the steps of the pipeline. First, I need to select the appropriate subreddits. Secondly, for each subreddit I will divide the data in three categories. *Top, Controversial, Both.* For each category, I select a set of submissions and for each submission I create a comment tree. Then out of these trees, for each category, I create a user graph that represents the communication between users. All of this will be explained in detail below. The tool that has been used for the data collection is a module of python named *PRAW*, a wrapper library that handles the reddit API more conveniently*.*

# Subreddit selection

First of all, I made a general search on reddit with the keywords ‘vaccine’, ‘covid’, ‘coronavirus’, and used the reddit filter ‘top’ for the resulted submissions. Then, I tracked the frequency of the corresponding subreddits for each resulted post. In other words, each post belongs to a subreddit, and I counted how many times each subreddit appeared in our search, so I can continue my research with the most famous ones. More specifically, the results can be shown in the following graph.

With total of 238 submissions the most frequent subreddits that refer to the covid vaccines are the ones that appear on the graph. In our case, I selected ‘science’, ‘worldnews’, ‘Coronavirus’. Also, I manually searched for other subreddits that possibly create a controversy about the topic or some cliques between the users. So, I ended up in the subreddits ‘DebateVaccines’ and ‘conspiracy’. So, we have these subreddits:

* r/Coronavirus
* r/conspiracy
* r/science
* r/worldnews

# Comment Tree Creation

After choosing the subreddits, I crawled on each one and processed them on a specific way. For each subreddit I divide the data in three categories. *Top, Controversial, Both.* For the *Top* case, I put the filter *Top* that the reddit API provides and I collected the top submissions with the keyword ‘*vaccin’.* The same I did with the *Controversial* category. *Both* category is the union of *Top* and *Controversial*.

# User Communication Graph

The next step is taking the trees and out of these create the communication graph.

Subreddit Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Subreddits** | Top | | Controversial | | Both | |
| *# Comms* | *# Posts* | *# Comms* | *# Posts* | *# Comms* | *# Posts* |
| **Coronavirus** | 143116 | 222 | 3395 | 117 | 146511 | 339 |
| **conspiracy** | 118139 | 216 | 14939 | 58 | 132278 | 273 |
| **science** | 68068 | 139 | 1897 | 15 | 69965 | 154 |
| **worldnews** | 94281 | 240 | 155 | 4 | 94436 | 244 |
| **MensRights** | 25051 | 222 | 2024 | 23 | 27075 | 245 |
| **WitchesVsPatriarchy** | 31543 | 241 | 1841 | 49 | 32613 | 285 |

Graph Data

MultiDigraph

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subreddits | Top | | Controversial | | Both | |
| #V | #E | #V | #E | #V | #E |
| **Coronavirus** | 81267 | 264000 | 1923 | 4951 | 82554 | 269073 |
| **conspiracy** | 48592 | 190840 | 8029 | 21938 | 53675 | 211338 |
| **science** | 77872 | 192731 | 604 | 1341 | 78757 | 195166 |
| **worldnews** | 57995 | 86362 | 53 | 110 | 58061 | 158155 |
| **MensRights** | 10359 | 33401 | 827 | 2068 | 10974 | 35467 |
| **WitchesVsPatriarchy** | 15621 | 49602 | 1048 | 2367 | 15957 | 50895 |

Graph

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Subreddits* | Top | | Controversial | | Both | |
| #V | #E | #V | #E | #V | #E |
| **Coronavirus** | 81267 | 132066 | 1923 | 2488 | 82554 | 134614 |
| **conspiracy** | 48592 | 95555 | 8029 | 10987 | 53675 | 105820 |
| **science** | 77872 | 96411 | 604 | 672 | 78757 | 97632 |
| **worldnews** | 57995 | 43258 | 53 | 55 | 58061 | 79120 |
| **MensRights** | 10359 | 16727 | 827 | 1035 | 10974 | 17761 |
| **WitchesVsPatriarchy** | 15621 | 24848 | 1048 | 1186 | 15957 | 25497 |

Average Degree of the graphs

|  |  |  |  |
| --- | --- | --- | --- |
| ***Average Degree*** | **Top** | **Controversial** | **Both** |
| **Coronavirus** | 3,25 | 2,58 | 3,26 |
| **conspiracy** | 3,93 | 2,74 | 3,94 |
| **science** | 2,48 | 2,22 | 2,48 |
| **worldnews** | 2,72 | 2,07 | 2,72 |
| **MensRights** | 3,23 | 2,5 | 3,23 |
| **WitchesVsPatriarchy** | 3,18 | 2,26 | 3,2 |

Modified Graphs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Subreddits* | Top | | Controversial | | Both | |
| #V | #E | #V | #E | #V | #E |
| **Coronavirus** | 81267 | 115981 | 1923 | 2212 | 82554 | 118235 |
| **conspiracy** | 48592 | 79555 | 8029 | 9177 | 53675 | 88119 |
| **science** | 77872 | 85985 | 604 | 609 | 78757 | 87030 |
| **worldnews** | 57995 | 69348 | 53 | 52 | 58061 | 69451 |
| **MensRights** | 10359 | 14033 | 827 | 878 | 10974 | 14911 |
| **WitchesVsPatriarchy** | 15621 | 22156 | 1048 | 1089 | 15957 | 22775 |

Average Degree of the modified graphs

|  |  |  |  |
| --- | --- | --- | --- |
| ***Average Degree*** | **Top** | **Controversial** | **Both** |
| **Coronavirus** | 2,85 | 2,31 | 2,86 |
| **conspiracy** | 3,27 | 2,28 | 3,28 |
| **science** | 2,21 | 2,02 | 2,21 |
| **worldnews** | 2,4 | 1,96 | 2,39 |
| **MensRights** | 2,7 | 2,12 | 2,71 |
| **WitchesVsPatriarchy** | 2,84 | 2,08 | 2,85 |

2-subreddits

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 2-subreddits | Top | | Controversial | | Both | |
| #V | #E | #V | #E | #V | #E |
| Coron. & consp. | 128088 | 227621 | 9943 | 13475 | 134345 | 240434 |
| science & consp. | 95150 | 152372 |  |  | 130699 | 203451 |
| wrldns & consp. | 105277 | 174561 | 8078 | 11042 | 110347 | 184939 |
| WvP & MR | 25957 | 41575 | 1874 | 2221 | 26907 | 43258 |

Modified 2-subreddits

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 2-subreddits | Top | | Controversial | | Both | |
| #V | #E | #V | #E | #V | #E |
| Coron. & consp. | 128088 | 195536 | 9943 | 11389 | 134345 | 206354 |
| science & consp. | 95150 |  |  |  | 130699 |  |
| wrldns & consp. | 105277 | 148903 | 8078 | 9229 | 110347 | 157570 |
| WvP & MR | 25957 | 36189 | 1874 | 1967 | 26907 | 37686 |

**Random Walk Results**

INTRA POLARIZATION

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subreddit | RW(rr) | | | RW(rp) | | | RW(pp) | | |
| Top | Contr. | Both | Top | Contr. | Both | Top | Contr. | Both |
| Coronavirus | 0,30 | 0,52 | 0,30 | 0,14 | 0,56 | 0,16 | 0,00 | 0,09 | -0,02 |
| conspiracy | 0,26 | 0,62 | 0,26 | 0,14 | 0,74 | 0,14 | -0,05 | 0,27 | -0,03 |
| science | 0,60 | **0,90** | 0,60 | 0,48 | **0,88** | 0,53 | 0,13 | **0,85** | 0,09 |
| worldnews | 0,42 | -0,30 | 0,44 | 0,17 | **0,92** | 0,20 | -0,05 | 0,72 | -0,05 |
| MensRights | 0,56 | **0,71** | 0,58 | 0,42 | **0,92** | 0,41 | 0,17 | 0,68 | 0,23 |
| WitchesVsPatriarchy | 0,36 | **0,78** | 0,38 | 0,24 | **0,92** | 0,28 | -0,01 | 0,58 | 0,03 |

Modified

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subreddit | RW(rr) | | | RW(rp) | | | RW(pp) | | |
| Top | Contr. | Both | Top | Contr. | Both | Top | Contr. | Both |
| Coronavirus | 0,28 | 0,50 | 0,29 | 0,30 | **0,72** | 0,33 | -0,01 | 0,04 | 0,02 |
| conspiracy | 0,25 | 0,61 | 0,27 | 0,26 | 0,87 | 0,32 | -0,03 | 0,21 | -0,05 |
| science | 0,56 | **0,91** | 0,52 | 0,72 | **1,00** | 0,66 | 0,05 | **0,80** | -0,06 |
| worldnews | 0,41 | -0,36 | 0,42 | 0,36 | -0,50 | 0,36 | 0,00 | -0,51 | -0,01 |
| MensRights | 0,56 | **0,77** | 0,56 | 0,58 | **0,98** | 0,58 | 0,27 | 0,66 | 0,33 |
| WitchesVsPatriarchy | 0,35 | **0,83** | 0,35 | 0,37 | **0,98** | 0,37 | 0,04 | 0,71 | 0,02 |

INTER POLARIZATION

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2-Subreddits | RW(rr) | | | RW(rp) | | | RW(pp) | | |
| Top | Contr. | Both | Top | Contr. | Both | Top | Contr. | Both |
| Coron. & consp. | 0,78 | 0,90 | 0,78 | 0,52 | 0,82 | 0,46 | 0,47 | 0,81 | 0,42 |
| science & consp. |  |  | 0,76 |  |  | 0,52 |  |  | 0,48 |
| wn & consp. | 0,81 | 0,64 | 0,81 | 0,48 | 0,95 | 0,52 | 0,40 | 0,83 | 0,40 |
| WvP & MR | **0,98** | **0,93** | **0,97** | **0,95** | 0,77 | **0,94** | **0,94** | 0,76 | **0,76** |

Modified

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2-Subreddits | RW(rr) | | | RW(rp) | | | RW(pp) | | |
| Top | Contr. | Both | Top | Contr. | Both | Top | Contr. | Both |
| Coron. & consp. | **0,80** | **0,79** | **0,80** | 0,72 | **0,84** | 0,70 | 0,65 | **0,81** | 0,65 |
| science & consp. |  |  | 0,77 |  |  | 0,72 |  |  | 0,49 |
| wn & consp. | **0,82** | 0,61 | **0,82** | 0,70 | **0,98** | 0,67 | 0,58 | **0,88** | 0,57 |
| WvP & MR | **0,99** | **0,89** | **0,98** | **0,97** | **0,97** | **0,98** | **0,97** | **0,73** | **0,94** |