**Introduction**

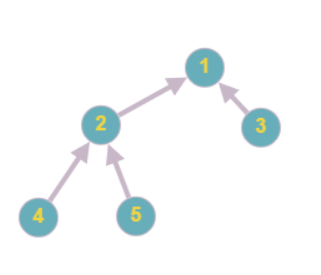
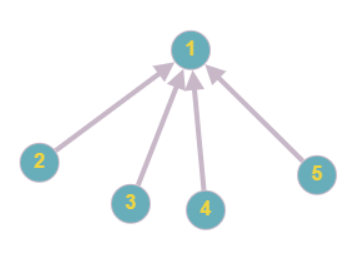
We will use the random walk algorithm to estimate the polarization for each case.

The random walk polarization measure uses the notion of random walks:

Where is the probability of a random walk that starts from partition A and ends in partition B. The more polarized (well-separated) the two partitions and are, the lower the probability of not crossing the two polarized partitions, and therefore, the higher the polarization. So, the higher the the more polarized a graph is.

**Graph Va****riations**

We will use 3 different graph variations to run the random walk algorithm. First, from the comment trees that we created we will make a user multi directed graph, in which each node represents a user, and an edge represents the reply that a user made to another. After this, we take the weakly connected component of the graph and convert it to an undirected graph. Each next step is chosen with uniform probability.

The second variation is that we modify the comment tree structure. Each root node stays normally the root but now all his children, grandchildren and generally all his offspring becomes their children. For example, this tree:

And the third variation is to use weights in the graph as probabilities. If we take the neighbors of a node there are some that the node communicates more often than the others, so we exploit that information by counting in the multidigraph the edges for all the neighbors of each node. Each next step of the random walk now is not uniform but weighted.

**Random Walk Parameters**

*Intra or Inter:* For each type of polarization we partition the graph differently. For the intra polarization we partition the graph with the library metis. Metis is a library that (to be completed). For the inter polarization each partition is the set of users, of each graph before their merge, the users that were a part of both graphs join a partition with ½ probability. So, it should be declared before the execution of the random walk if it is inter or intra polarization.

*Type:* There are 3 different types to run the random walk algorithm. The ‘rr’, ’rp’ and ’pp’. The ‘r’ stands for random and ‘p’ for popular. The ‘rr’ indicates that we randomly select a 10% subpartition of each partition and we start the random walk from a subpartition until we end up to either the subpartition that we started or the other. The ‘rp’ type is the same as ‘rr’ but with different termination criteria. We start, in the same way from random nodes until we end up to a popular one. We take the 10 most popular nodes from each partition. A popular node is defined as the node with the most neighbors. And the ’pp’ means that the random walk starts from popular nodes and ends with the first popular node it finds.

*Weighted:* This should be turned to True when we want to calculate the next step of the random walk based on the weights(probabilities) of the edges of the graph.

*Number of experiments:* The number of experiments that the random walk we want to perform. At the end we calculate the average. In our case we ran 100 experiments for each variation

**The experiments**

**Variation 1**

For the first graph variation, we ran the experiments for each subreddit and for the case of intra polarization we have the following results:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subreddit | RW(rr) | | | RW(rp) | | | RW(pp) | | |
| Top | Contr. | Both | Top | Contr. | Both | Top | Contr. | Both |
| Coronavirus | 0,3038 | 0,5233 | 0,3026 | 0,1378 | 0,5574 | 0,1611 | 0,0041 | 0,0895 | -0,0221 |
| conspiracy | 0,2574 | 0,6181 | 0,2617 | 0,144 | **0,739** | 0,1423 | -0,0492 | 0,2731 | -0,0261 |
| science | 0,5972 | **0,8966** | 0,6034 | 0,4789 | **0,879** | 0,527 | 0,1346 | **0,8482** | 0,0883 |
| MensRights | 0,5601 | **0,7118** | 0,5838 | 0,415 | **0,9151** | 0,4121 | 0,172 | 0,6795 | 0,2283 |
| WitchesVsPatriarchy | 0,3564 | **0,7838** | 0,381 | 0,2411 | **0,9152** | 0,2785 | -0,0142 | 0,5815 | 0,0306 |
| lgbt | 0,369 | 0,5348 | 0,5231 | 0,5047 | **0,711** | 0,5616 | 0,0353 | 0,1685 | 0,0331 |
| Conservative | 0,3565 | **0,8714** | 0,3436 | 0,38 | **0,9901** | 0,3574 | 0,0011 | **0,8284** | 0,0308 |
| conspiracy | 0,56 | 0,2378 | 0,587 | 0,5187 | **0,7578** | 0,5439 | 0,2222 | 0,136 | 0,2498 |
| space | **0,7072** | **0,857** | **0,9408** | 0,586 | **0,9256** | **0,8189** | 0,15 | **0,7772** | 0,6669 |

We can notice that the highest score of polarization is with the type of ‘rp’ then the ‘rr’. The ‘pp’ has the lower scores and there are even negative scores of polarization and this is explained by the high connectivity between the popular nodes. The controversial posts have a tendency of higher polarization and this is expected and the ‘both’ posts have slightly higher score than the ‘top’ ones. The highest score of polarization(0,99) is noted at the ‘Conservative’ subreddit in the controversial posts and it was calculated with the ‘rp’ random walk type, while the lowest score was found in conspiracy subreddit in the set of ‘top’ posts ran by ‘pp’ parameter.

As for the inter polarization we have the following table:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2-Subreddits | RW(rr) | | | RW(rp) | | | RW(pp) | | |
| Top | Contr. | Both | Top | Contr. | Both | Top | Contr. | Both |
| Coron. & consp. | 0,7812 | **0,8972** | 0,7809 | 0,5224 | **0,8234** | 0,46 | 0,4701 | **0,8121** | 0,415 |
| science & consp. | 0,7594 | **0,9541** | 0,7633 | 0,568 | **0,9563** | 0,52 | 0,334 | **0,9067** | 0,48 |
| WvP & MR | **0,9811** | **0,9305** | **0,9743** | **0,9465** | 0,7734 | **0,9385** | **0,9442** | 0,7556 | 0,7633 |
| lgbt & Cons. | **0,9956** | no inter | **0,9935** | **0,996** | no inter | **0,987** | **0,993** | no inter | **0,9881** |
| space & consp. | **0,9408** | no inter | **0,9393** | 0,8189 | no inter | **0,854** | 0,6669 | no inter | **0,8068** |

Generally, we can notice that we have higher levels of polarization and this is expected because, we chose opposing subreddits to merge, so it is more unlikely for a user to be part or participate on both subreddits. Users tend to interact with users of similar opinions. For example, the merge of ‘WitchesVsPatriarchy’ and ‘MensRights’ seems to have the highest polarization no matter the parameters of the algorithm, because users have completely opposite opinions about patriarchy, and this is reasonable, while ‘Coronavirus’ and ‘conspiracy’ hit the lowest scores, this happens because they do not necessarily seem to be opposing subreddits. no inter means that there was no intersection in the process of merging the two subreddits. This could also be mathematically translated to polarization score of 1 because, there is no chance of the random walk to end up at a different partition that it started at. The merge of controversial posts tend to have higher scores, while ‘top’ and ‘both’ similar scores.

**Variation 2**

As for the variation that we modify the trees as explained above, in a similar way we perform the experiments and have the following results:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subreddit | RW(rr) | | | RW(rp) | | | RW(pp) | | |
| Top | Contr. | Both | Top | Contr. | Both | Top | Contr. | Both |
| Coronavirus | 0,2842 | 0,504 | 0,2903 | 0,3025 | **0,7214** | 0,3339 | -0,009 | 0,0427 | 0,0172 |
| conspiracy | 0,254 | 0,6109 | 0,2703 | 0,2595 | **0,873** | 0,3223 | -0,027 | 0,2091 | -0,0521 |
| science | 0,5597 | **0,9062** | 0,5173 | **0,7181** | **0,998** | 0,6644 | 0,0483 | **0,8014** | -0,061 |
| MensRights | 0,5565 | **0,7731** | 0,5631 | 0,5783 | **0,9823** | 0,5796 | 0,2735 | 0,6643 | 0,3286 |
| WitchesVsPatriarchy | 0,3517 | **0,8287** | 0,3485 | 0,3734 | **0,983** | 0,3664 | 0,0449 | **0,7051** | 0,0223 |
| lgbt | 0,3741 | 0,5899 | 0,5311 | 0,6318 | **0,8777** | 0,6728 | 0,0311 | 0,2832 | 0,1047 |
| Conservative | 0,3274 | **0,8671** | 0,3227 | 0,6684 | **0,993** | 0,6403 | 0,002 | 0,592 | -0,0775 |
| conspiracy0 | 0,4914 | 0,0195 | 0,5813 | 0,6156 | 0,5003 | 0,6368 | 0,1021 | 0,5118 | 0,2608 |
| space | 0,6958 | **0,8195** | **0,709** | **0,8492** | **0,9831** | **0,7988** | 0,0412 | 0,577 | 0,1104 |

The results are very similar to the previous variation. Controversial posts remain with the highest score, while ‘top’ and ‘controversial’ posts have similar values. The ‘rp’ parameter stays again with the highest score, followed by the ‘rr’ and ‘pp’ with the lowest score with the same reasons as explained above. The highest polarity score again appears at the Conservative subreddit in the set of controversial posts ran by ‘rp’ parameter, whereas the lowest score was found again at the conservative subreddit in the ‘both’ category ran by ‘pp’ algorithm. That was unexpected and contradictory, because from the one hand Conservative Both is a different graph from Conservative Controversial, but from the other Conservative Both was the union between ‘Top’ and ‘Controversial’. It is about the same topic with similar graphs, so we can observe the significance of the parameter type of the algorithm.