

# It's Complicated

## A Visual Exploration of the Political Landscape of Reddit

Demetrios Fassois  
dimifass@stanford.edu

Tianyi Huang  
thuang97@stanford.edu

Kade Keith  
kade@stanford.edu

### Abstract

We study data of posts and comments on Reddit from 5 months in 2016 and 2017 to explore the nature of political discussions on the platform. We use a flexible graph folding technique to translate user behavior into relationships among subreddits, and also provide a troll detection algorithm to remove potentially distracting contents. We present easy-to-understand visualizations to showcase our results, which suggest that selective exposure exists on Reddit.

## 1. Introduction

### 1.1 Motivation

Following the dot-com boom, the past decade has seen growing interest in the study of online expressions of opinions. As more and more people learned to make their voices heard on the Internet, researchers started paying attention to the nature of discussions that they would experience there, especially in the realm of politics, which invariably comes with much controversy. What is the political landscape like on online platforms? Do these platforms count as “public spheres” where diverse opinions flow freely and reach a broad audience? Or are they in fact “echo chambers” where people are largely insulated from contrary perspectives and only interact with contents that they identify with?

Piqued by these questions, we become particularly curious about what is happening on social media platforms, which host the majority of online discussions. Among them, Reddit stands out due to its high level of user participation. Now with more than 138,000 active communities, 330 million average monthly active users, and higher average user activity than both Facebook and Twitter (Hutchinson), Reddit has in recent years rightly attracted the attention of many network researchers. We as a group are interested in taking a closer look at political discussions take place on such a dynamic platform as Reddit. Does higher overall activity translate into more mobility in the exchange of dissimilar ideas? Or does the tendency for like-minded people to flock together remain strong? We believe that by exploring these questions, we will be able to arrive at a more nuanced understanding of user behaviors on Reddit, as well as to contribute to the expanding and increasingly relevant volume of research on online political discussions.

## 1.2 Problem Statement

In this project, we aim to study the relationship among politically relevant subreddits on Reddit. As a refresher, Reddit is an online discussion platform where users post to boards named “subreddits.” In the context of network analysis, we regard the subreddits as communities.

Specifically, we are interested in learning more about the extent to which communities with dissimilar political opinions are connected. We propose to explore and visualize such connections based on user post or comment behavior, bearing in mind that a pair of communities are more connected if there are more common users who post or comment in both of them. We will also experiment with filtering out troll contents, which do not reflect the general nature of interactions on Reddit, and see if it helps us better understand intercommunity relationships.

## 2. Related Works

### 2.1 Network Analysis Through Graph Folding

Our analysis of the network of Reddit draws inspiration from Goh et al., who propose a novel approach to exploring relationships among diseases by folding a bipartite graph consisting of gene-disease pairs into a “Human Disease Network,” where diseases related to a common gene are connected by an edge. The folded network sheds light on similarities among diseases based on their shared genomic predictors. This, along with an inverse gene network where genes related to a common disease have an edge between them, is able to provide a holistic view of diseases, genes, and their relationships.

While the approach has a drawback of trivializing lesser known diseases due to lack of information on their corresponding genes, we decide that it can translate smoothly to the context of social networks, since it makes sense to focus on communities that are more popular and influential, and therefore more representative of what happens on the network.

### 2.2 Conflicts on Reddit

We are first made aware of trolling (i.e. community harrassing) on Reddit by Kumar et al. (2018), who observe that the vast majority (74%) of negative interactions on the platform are instigated by a small (1%) group of subreddits. This leads us to think more deeply about the role that trolls play in interactions on Reddit, and inspires us to explore whether detecting and removing trolls might help us focus more on the average users on Reddit, and therefore benefit our analysis.

For specific methods of troll detection, we turn to a previous paper with the same lead author (2014), which introduces a novel troll identification algorithm on the “signed” social network (SSN) Slashdot, relying on using features such as post content, user activity, and community response, to capture antisocial or troll-like behavior. While not all features available

for Slashdot are available for Reddit, we are inspired by the general idea that it is possible to detect trolls by a comprehensive examination of both user activity and community response.

## 2.3 Others

Since the early 2000s, many studies have been published on the nature of political discourse on online news platforms, but more recently the focus has shifted to social networking sites given their tremendous number of users and influence. There was a further spark in academic interest following the 2016 U.S. election, which raised concerns over “fake news” on Facebook (Allcott and Gentzkow). Even before this, research had been done on political homophily on Facebook (e.g. Bakshy et al.) and Twitter (e.g. Colleoni et al.).

On the one hand, we notice that there is a lack of work on political discussions on Reddit, which is part of the motivation for this project. On the other hand, much evidence from other social platform points to selective exposure, which sets the expectation for our analysis of Reddit.

## 3. Approach

### 3.1 Dataset

Our data of Reddit posts and comments come from an online Reddit data dump (<http://files.pushshift.io/reddit/>). We start with 5 months of data, 3 of which are from 2016 (an election year), and 2 of which are from 2017 (a non-election year). We then filter the data to contain only activity in the 211 subreddits in the [Reddit Politosphere](#) (as compiled by the r/Politics subreddit), since we are interested in focusing on political discussions. We exclude comments from deleted accounts since they are all attributed to the user id “[deleted].”

When examining the data, we found a number of suspicious bot accounts that frequently commented similar contents across a wide range of subreddits. Since this does not represent human behavior, we decided to also, for the sake of convenience, exclude hyperactive users with more than 200 comments per year. After filtering, we have 4,229,162 comments from 2016, 2,885,440 comments from 2017, 260,807 posts from 2016, and 299,170 posts from 2017.

### 3.2 Graph Folding

We begin by creating a bipartite graph where the nodes are users and subreddits, and the edges represent either a comment or a post from a user in that subreddit. To account for the disparity in popularity among political subreddits, we impose a cap on the number of comments considered per subreddit. This prevents massively popular subreddits with many posts/comments from forming a giant clique. We also optionally filter out comments and posts that we suspect are from “trolls” (more on that in the subsequent section).

We then use an enhanced version the folding technique described by Goh et al. to build a folded subreddit network based on behavior of Reddit users. Our enhanced graph folding is configurable in the following ways:

1. Whether to consider posts and/or comments as edges in the original bipartite graph
2. How many posts and/or comments are required for an edge to exist in the original bipartite graph
3. How many posts and/or comments to consider per subreddit
4. Whether to consider troll-like contents

This allows us to easily generate a number of folded graphs and compare them. For example, we are able to create a 2017 graph - with subreddits connected by 5 or more shared comments - with 1000 comments per subreddit - with trolls removed. Another example would be a 2016 graph - with subreddits connected by 3 or more shared comments *or* posts - with 2000 comments and 2000 posts per subreddit - with trolls included.

We usually cap at 1000 comments per subreddit, and require at least 3-5 users in common for an edge to exist, since using more comments or requiring fewer users in common resulted in exceedingly dense graphs that were difficult to interpret and visualize.

### 3.3 Troll Detection

Trolling, or community harassing, is generally considered aberrant behavior on online platforms, and does not accurately reflect the nature of discussions that take place, so removing such contents from our data is likely to help make our analysis more realistic.

In order to identify users as trolls and remove them from the data, we use an unsupervised clustering method, namely the Gaussian mixtures model. Every user is assigned a set of features that are calculated from all the comments by that user, and the clusters are computed for each year. The intersection of users who both commented and posted is not big enough as to allow us to combine them and include features from posts as well. This means that only authors of comments were considered for the clusters, and their features derived solely from their comments, i.e. not from their posts. A potential consequence of this is that it could limit the potential of removing users identified as trolls from the contents of their posts.

The features used as inputs to the model are guided by the results from Cheng et al., and grouped in similar categories. We compute comment features including number of words and readability metrics, from the text of all the comments by each user in a given year. We also use activity features such as average comment and link karma (i.e. reward earned for posting popular contents) per user in a year, as well as community features such as average post score, controversiality, and profanity for all their posts in a year. The features that are included in the original data are comment and link karma, and comment score and controversiality, while the rest of the features are computed with text analysis.

In order to determine the number of clusters, we look at the AIC and BIC scores as a function of the number of GMM components. Analyzing the clusters involves manual inspection of the average value of features for each output cluster.

### 3.4 Visualization

To visualize our folded graphs, we use node2vec (Grover and Leskovec) to create embedding representations of the subreddits in the main connected component of the folded graph. The embedding size used is 128, and the return and in-out parameters used are both 1, resulting in the deepwalk equivalent model. We subsequently divide the embedding vectors into 2 clusters using K-means. In order to plot the clusters, we project the embeddings using PCA as well as t-SNE to produce two scatterplots.

We also use the Louvain algorithm mentioned in class as an alternative method to extract clusters. This ends up producing a more intuitive visual representation of our embeddings, which is further discussed in section 4.2.

## 4. Findings

### 4.1 Trolls

The average features per cluster for 2016 is shown below. A similar analysis was performed using the comments from 2017.

Cluster	Comment karma	Link karma	Read ability index	Readability score	Reading ease	Difficult words	Text standard	Number of words	Score	Controversiality	Profanity
1	4114	572	8.29	6.49	64	2.98	6.03	18.79	2.82	0	0.01
2	21444	6626	93.15	15.81	-112.97	31.25	8.27	184.67	9.25	0.07	0.18
3	16338	1637	43.68	10.8	-7.24	15.37	8.94	88	3.74	0.02	0.12
4	1350168	6395765	2.35	4.4	69.31	1	2.8	11.3	-2	0.4	0
5	22878	3178	13.34	8.6	34.48	8.2	7.91	48.11	4.25	0.04	0.08
6	130045	122813	24.89	8.53	29.55	8.72	7.35	50.88	5.68	0.03	0.07
7	41886	9992	11.97	7.12	56.79	4.18	6.66	25.98	5.48	0.04	0.07
8	17890	4001	5.05	5.64	71.65	1.91	4.78	13.62	3.8	0.08	0.15

We interpret these results according to the analysis of troll characteristics presented in Cheng et al. Cluster 4 is not considered because of its small sample size, but it is interesting to note

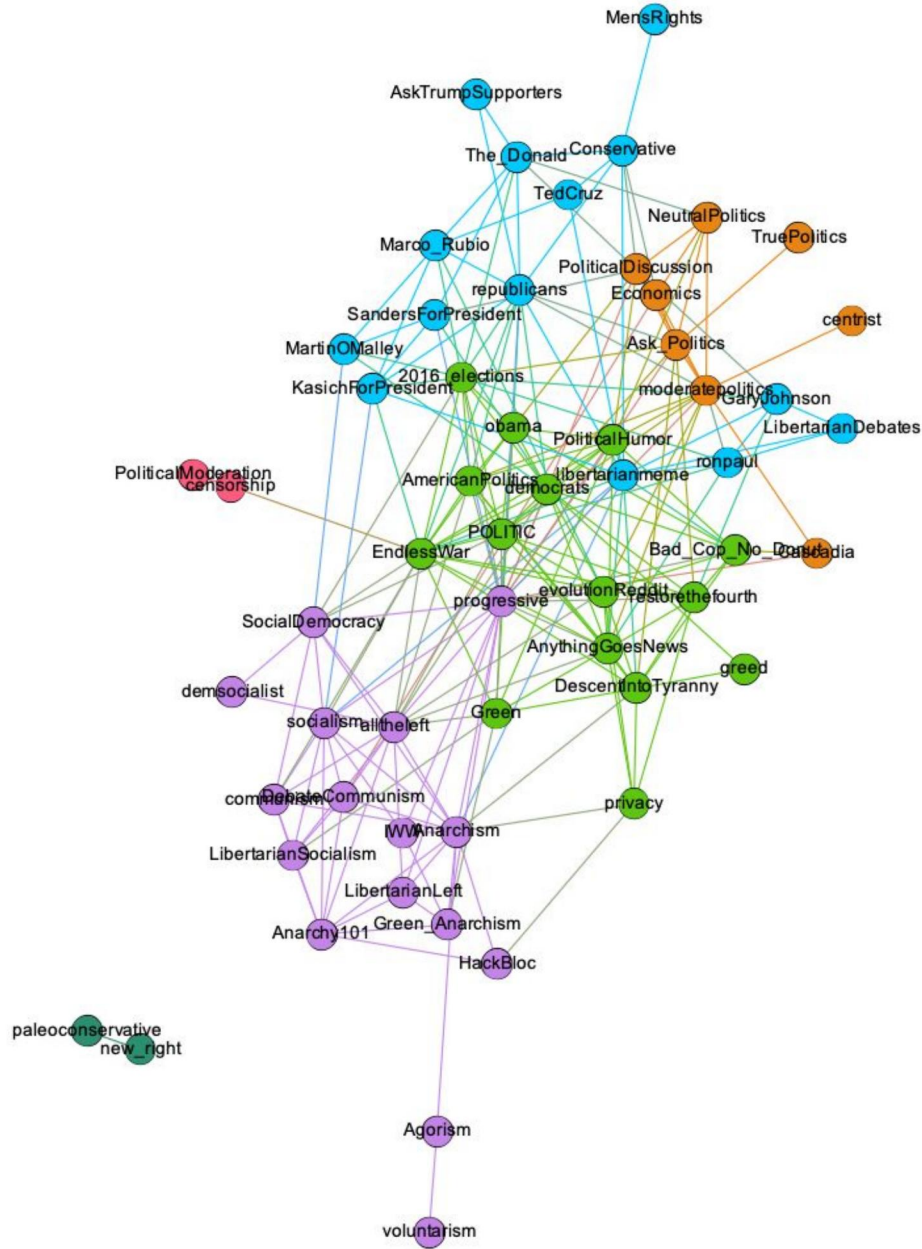
that it includes very few users with very high karma scores and very controversial comments. Cluster 1 represents a large majority of users with low karma and comment scores but no controversial content. Another interesting result that exists for both years is cluster 2, which has the smallest reading ease score, high readability scores (which represent the grade level needed to comprehend the text), many difficult words, high text standard, but also many profanity words. Those are lengthy, higher quality comments that can include profanity as well. At the end, we identify users in cluster 8 to be troll users. This cluster includes users with low karma, that write short comments that receive low scores and that are easily comprehensible without difficult words but with low text standard, and also include profanity words and are controversial.

A challenge with detecting trolls in our case is that this is an unsupervised clustering problem that involves manual inspection of the results. The reason behind this is that data about comment deletions and user bans are not available for us to develop a supervised learning model as in Cheng et al. Direct user to user upvote/downvote data are not available either, so neither are we able to follow an approach using the troll identification algorithm on the “signed” social network developed in Kumar et al. (2014).

## 4.2 Graph Folding: An Example

Below is a 2016 graph - with subreddits connected by 5 or more shared comments - with 1000 comments per subreddit - with the trolls included, whose clusters are generated by the Louvain algorithm, and the details fine-tuned in Gephi. We see that the majority of politically relevant subreddits are connected with one another, with ones that represent similar political inclinations (such as *r/The\_Donald* and *r/republicans*, or *r/obama* and *r/democrats*) largely in the same cluster, and ones with dissimilar political inclinations largely in separate clusters, indicating possible selective exposure, which will be further analyzed in the subsequent section.

The graphs generated with PCA and t-SNE are less intuitive, but are included, along with all the other graphs generated with various configurations, in our GitHub repository (link included in section 7).



### 4.3 Evaluation

Most of the preliminary evaluation of our results is empirical. Specifically, we pay attention to whether subreddits that are close to one another by common sense end up as neighbors or in the same grouping, given that social networks in general have been shown to promote selective exposure. We realize that we may encounter some surprises along the way, e.g. subreddits that are generally thought to have opposing views could end up with high similarity. This has so far not occurred, presumably given that we have not run our program on a particularly large dataset.

We can also compare our results (from the Louvain algorithm) against human-curated lists of subreddits, such as the one of the Reddit Politosphere mentioned in section 3.1. Specifically, we choose to do so with the completeness score which can be computed through the sklearn package in Python. This score measures the extent to which all members of a given class (in the Reddit Politosphere list) are assigned to the same cluster (by our algorithm). The results are as follows.

Graph	Score	Null Model Score	Difference
2016 Comments - Only trolls	0.579	0.369	0.210
2016 Comments - All users	0.521	0.242	0.278
2016 Comments - Exclude Trolls	0.652	0.353	0.298
2016 Posts	0.554	0.368	0.185
2017 Comments - Only trolls	0.514	0.477	0.037
2017 Comments - All users	0.428	0.368	0.059
2017 Comments - Exclude Trolls	0.413	0.309	0.104
2017 Posts	0.506	0.266	0.240

Note that for the 2016 comments graph, removing trolls increases the completeness score significantly, which implies that trolls post across community, or in this particular case, partisan, boundaries, further justifying that it is necessary to remove trolls to obtain an accurate, realistic representation of interactions on Reddit, which is intuitive. There is, however, a decrease in the case of the 2017 comments graph, but only a very slight one, and likely due to the fact that our troll detection algorithm is far from perfect (note the difficulties mentioned at the end of section 4.1) For graphs based on posts, including or excluding trolls does not make a visible difference (presumably because trolling takes place predominantly in comments rather than posts), so the corresponding results are omitted.

## 5. Conclusion

Going forward from our discussion in section 4.3, it is important to bear in mind that our “ground truth” given by the Reddit Politosphere list is merely a reference, and the fact that a grouping that we find deviates a lot from it does not necessarily mean that the grouping is not accurate; it simply means that the political landscape that that grouping reflects is different from the one reflected by the ground truth. But insofar as our current results show (with completeness scores largely over 0.5), selective exposure, i.e. the phenomenon of people only



interacting with people with similar opinions (political opinions, in this particular case), forming something of an “echo chamber”, is far from non-existent on Reddit. And hopefully, our graphs provide an easy-to-understand and informative visualization of such phenomenon.

There is a lot that can still be improved in our project. For example, we can experiment more with our troll detection algorithm to improve its accuracy, and we can include more data to expand the scope of our project (we decided to keep to our current scope due to concerns in running time and feasibility). In addition, as we mentioned at the beginning of section 4.3, we have not yet seen any “surprises,” but should any of them come along during future explorations, it would likely be beneficial to examine the relevant subreddits more closely on a case-by-case basis.

We are grateful to Alex Haigh for helping us brainstorm ideas for the project, to Srijan Kumar, whose works provided us with many inspirations, even though we did not get to meet him in person, and to Jure and the rest of the course team for showing us how interesting networks are.

Last but not least, our respective contributions are as follows:

- Dimitris: Report, data pre-processing, troll detection, community detection, deconvolution
- Kade: Report, data pre-processing, building bipartite graph and folding, evaluation metrics, visualization
- Tianyi: Problem formulation, writing up and coordinating the report, poster session

## 6. References

- Hunt Allcott, Matthew Gentzkow. Social Media and Fake News in the 2016 Election. *Journal of Economics Perspectives*, Vol. 31, No. 2, Spring 2017, 211-36.
- Eytan Bakshy, Solomon Messing, Lada A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science* 05 Jun 2015: Vol. 348, Issue 6239, 1130-32.
- Justin Cheng, Cristian Danescu-Niculescu-Mizil, Jure Leskovec. Antisocial Behavior in Online Discussion Communities. *ICWSM*, 2015.
- Elanor Colleoni, Alessandro Rozza, Adam Arvidsson. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication*, Vol. 64, Issue 2, 1 April 2014, 317-32.
- Kwang-Il Goh, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences* 104.21 (2007): 8685-90.
- Aditya Grover, Jure Leskovec. node2vec: Scalable Feature Learning for Networks. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016.
- Andrew Hutchinson. Reddit Now Has as Many Users as Twitter, and Far Higher Engagement Rates. *Social Media Today*, Apr 20th 2018.

- Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, Mathieu Bastian. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS One*, 2014; 9(6): e98679.
- Srijan Kumar, William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Community Interaction and Conflict on the Web. The Web Conference (WWW), 2018.
- Srijan Kumar, Francesca Spezzano, V.S. Subrahmanian. Accurately Detecting Trolls in Slashdot Zoo via Decluttering. *ASONAM*, 2014.
- Our GitHub repository is available at <https://github.com/keithkade/224w-project>.
- The data of Reddit posts and comments that we worked with are available at <https://drive.google.com/drive/folders/1yJrqn5QhTWUoqiPlp4DwkGHlknUp3B>