UNIVERSITÀ DI PADOVA
31/01/2019

IP3 - ABORTION

by Lejla Dzanko, Giulia Rizzoli, Sanja Milijanovic, Sara Shena, Lara Malin Schwarz

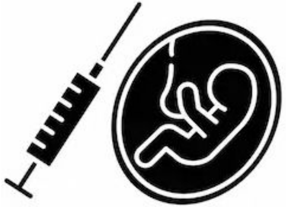
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO DI STUDI LINGUISTICI E LETTERARI

Research Goals:

1. Examine the current abortion sentiment by creating and analyzing semantic networks of tweets, hashtags and words related to abortion posted on Twitter.

2. Investigate if there is a relationship between the gender inequality of different US States and the semantic networks of posts coming from these States.



Theoretical Background:

- Abortion is one of the most controversial topics in social public, political and scientific debates in different disciplines

- Often debates result in reforms of the law → USA 2019

- Two movements:
  - Pro-Life: every human (embryo) has the right to live; abortion is murder → goal to ban it
  - Pro-Choice: every woman should have the right to decide what to do with her body on her own → goal to keep abortion safe and legal

- Gender inequality one aspect of discussions about abortion

Field of Application: Twitter

- Online social media networks are important to acquire mood of the population on a certain topic

- Especially Twitter is used to demonstrate one’s opinion

Network Definition:

- Semantic networks:
  Units of written text connected to each other based on their co-occurrence

- Bipartite networks: hashtag-tweets, hashtag-words and words-tweets related to abortion

- Datasets used to build networks are Pro-Life, Pro-Choice (based on data collected by hashtags) and mixed (samples of collected data put together)

Network Boundaries:

- Only tweets in english language
- Only tweets from year 2019
- Only tweets that contained one or more of pre-handpicked hashtags

Pre-handpicked Hashtags:

Pro-Choice:

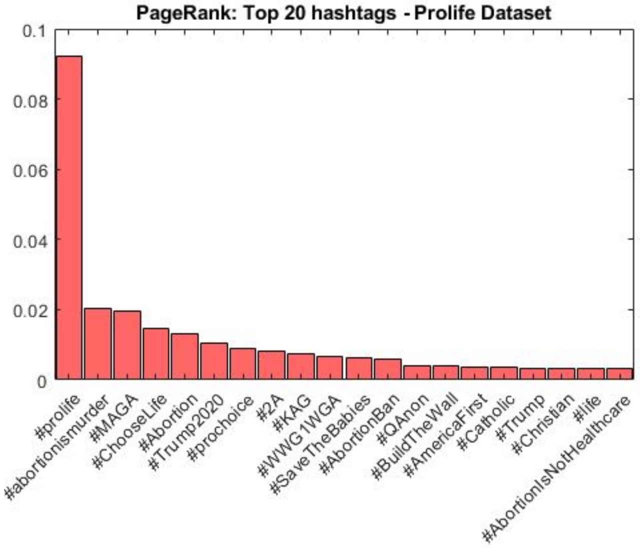
#prochoice
#mybodymychoice
#abortionishealthcare
#abortionisawomansright
#abortionrights
#abortionismurder
#abortionsupportnetwork
#proabortion

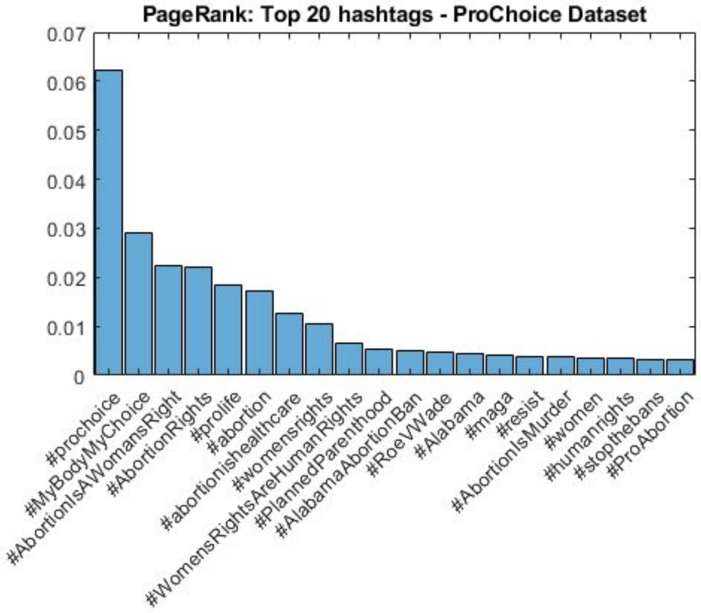
Pro-Life:

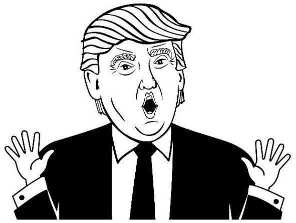
#prolife
#savethebabies
#babiesarehuman
#chooselife
#abortionban
#abortionismurder
#lovethemboth
#whywemarch

Data Collection and Cleaning:

1. Collection of data based on pre-handpicked hashtags: around two hundred thousand Tweets for Pro-Choice and Pro-Life
2. Ranking hashtags in each of the dataset → see if handpicked hashtags were really the most relevant ones or if data needs to be recollected using different, more often appearing hashtags



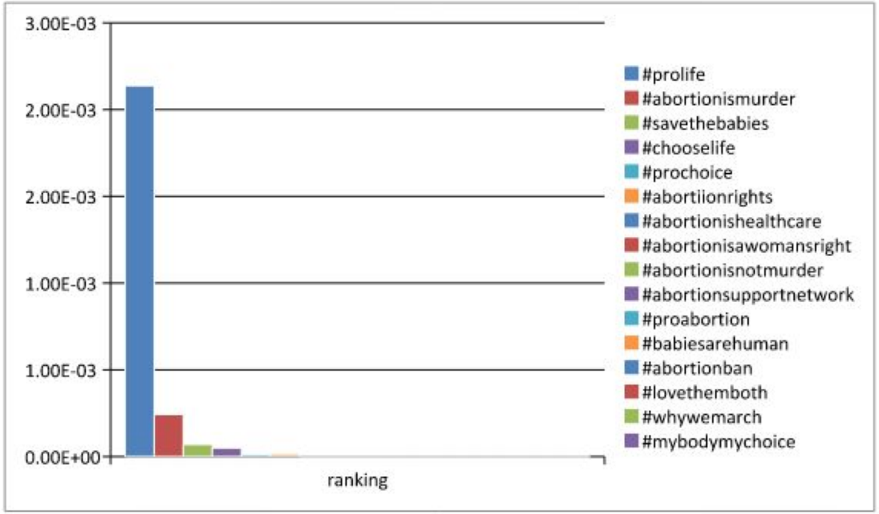




Data Collection and Cleaning:

3. Collection of data using those five hashtags and apply ranking → see if they are relevant for the topic of abortion

Ranking of hashtags related to USA



Data Collection and Cleaning:

4. Recollecting data with final hashtags

5. Collect data from specific states of the USA (depending on their GII)

Gender Inequality Index (GII):
The lower the value, the higher the level of equality.

1. Massachusetts: 0,1579
2. California: 0,1967
3. Maine: 0,1996
4. Connecticut: 0,2068
5. New Hampshire: 0,2113

46. Mississippi: 0,3512
47. Texas: 0,3530
48. Oklahoma: 0,3547
49. Louisiana: 0, 3637
50. Arkansas: 0,3642

List of final Hashtags:

Pro-Choice:

#prochoice

#mybodymychoice

#AbortionIsAWomansRight

#AbortionRights

#prolife

#abortion

#AbortionIsHealthcare

#WomensRights

#WomensRightsAreHumanRights

#PlannedParenthood

Pro-Life:

#prolife

#abortionismurder

#chooselife

#abortion

#prochoice

#savethebabies

#abortionban

#abortionisnothealthcare

#plannedparenthood

#abortionrights

Abortion Network

Final dataset:

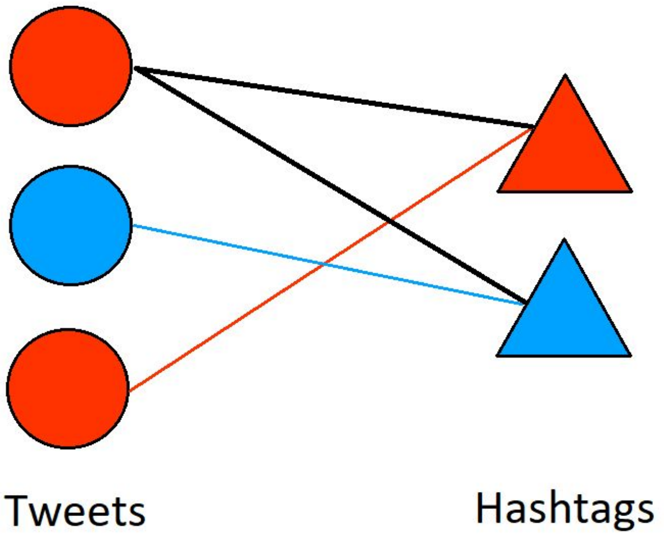
- Pro-Life Dataset: 334744 tweets
- Pro-Choice Dataset: 224294 tweets

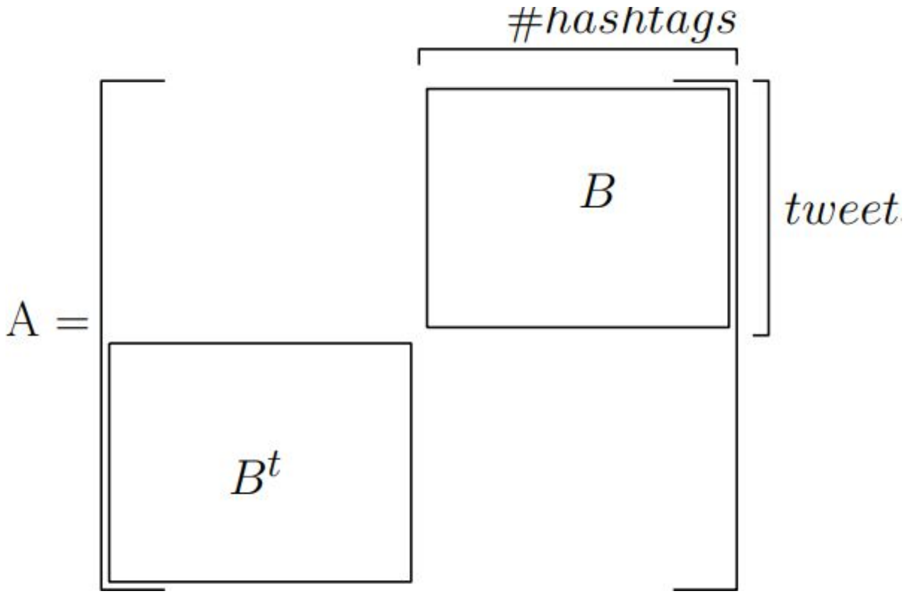
Network building:

- Bipartite network Tweets-Hashtags (Giulia & Sanja)
- Bipartite network Tweets-Words (Sara)
- Bipartite network Words-Hashtags (Lejla)

From .csv to network

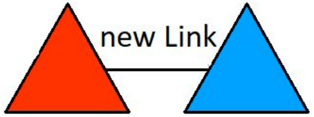
Bipartite Tweets-Hashtags





From bipartite matrix
to hashtags projection:

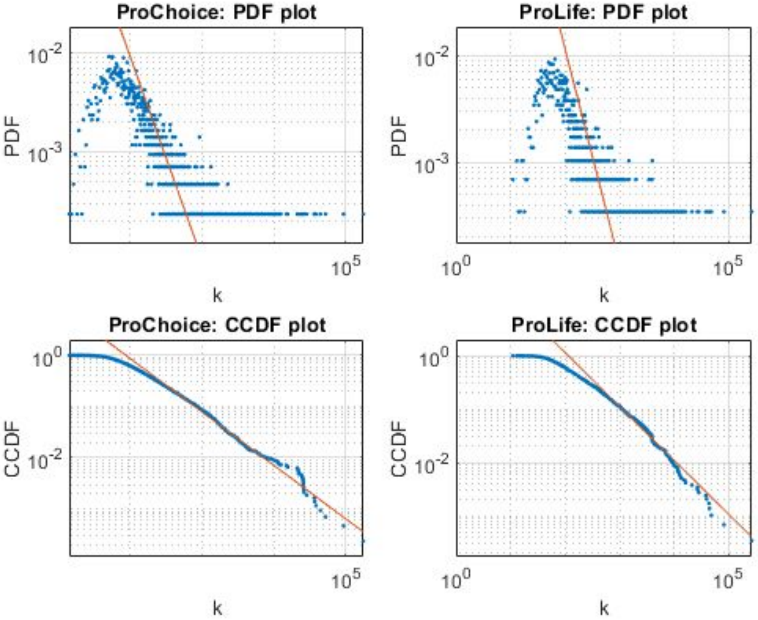
$$A_2 = B^T B$$



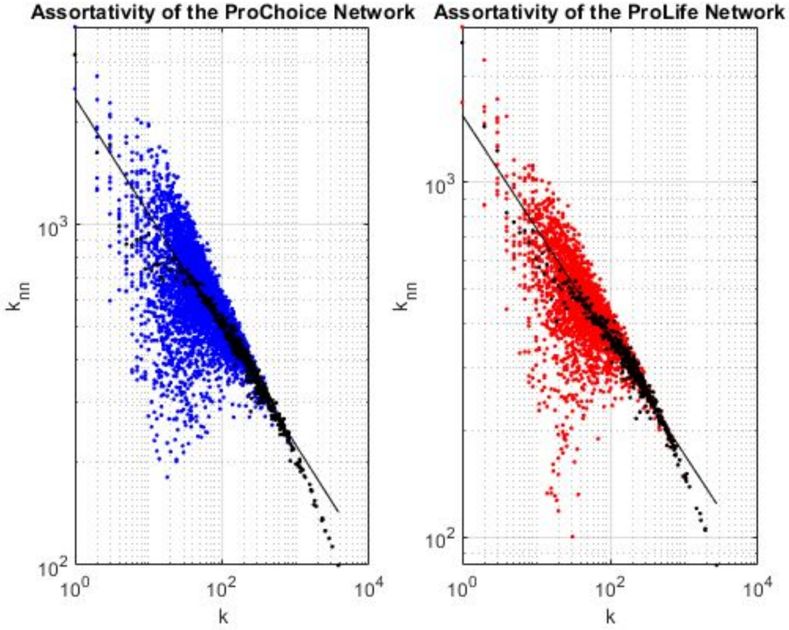
Hashtags Projections: Analysis

- Provide insights for the abortion hashtags projection network.
- Identify more relevant hashtags in abortion network and understand how the network is influenced by them.
- Try to capture the general behaviour in the different dataset of prolife and prochoice. State differences between the two.
- Identify hashtags communities in the abortion dataset.

Hashtags Projection: Insights

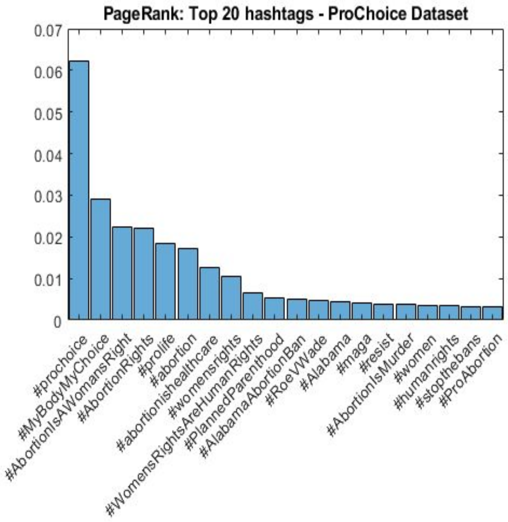


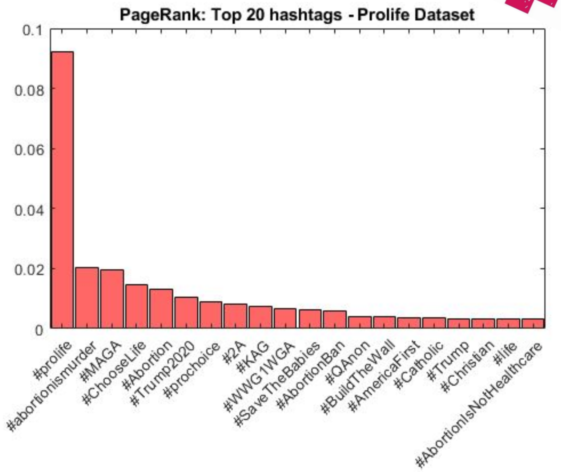
Scale-free

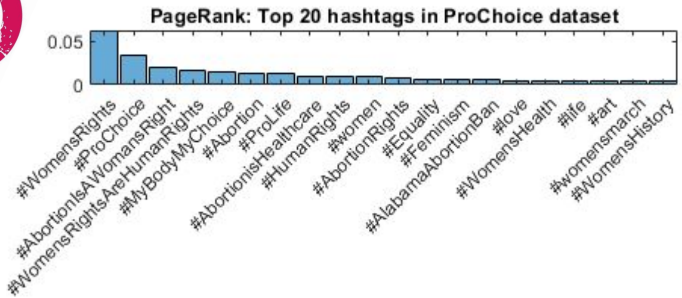


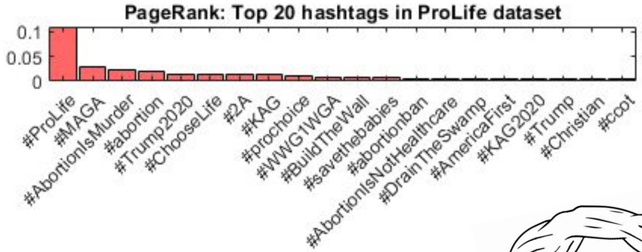
Disassortative

Hashtags Projection: PageRank









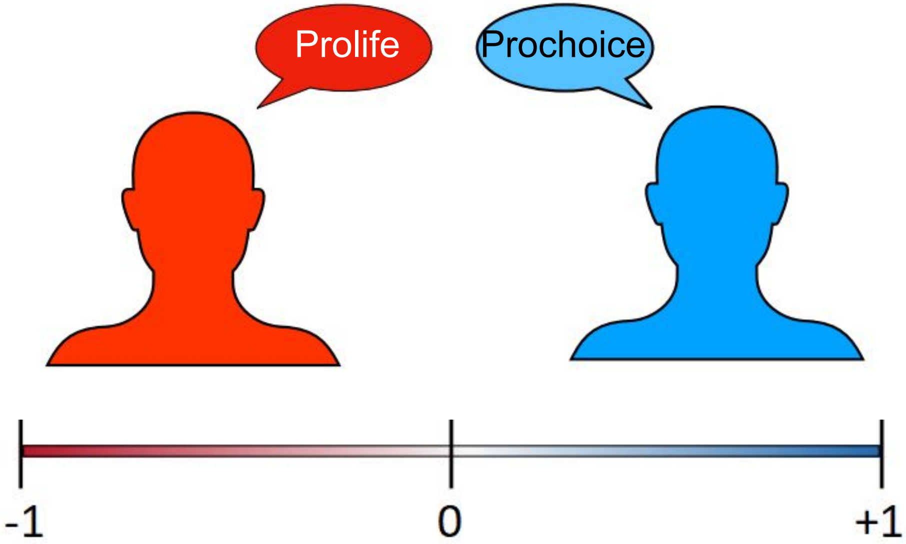
Initial PageRank

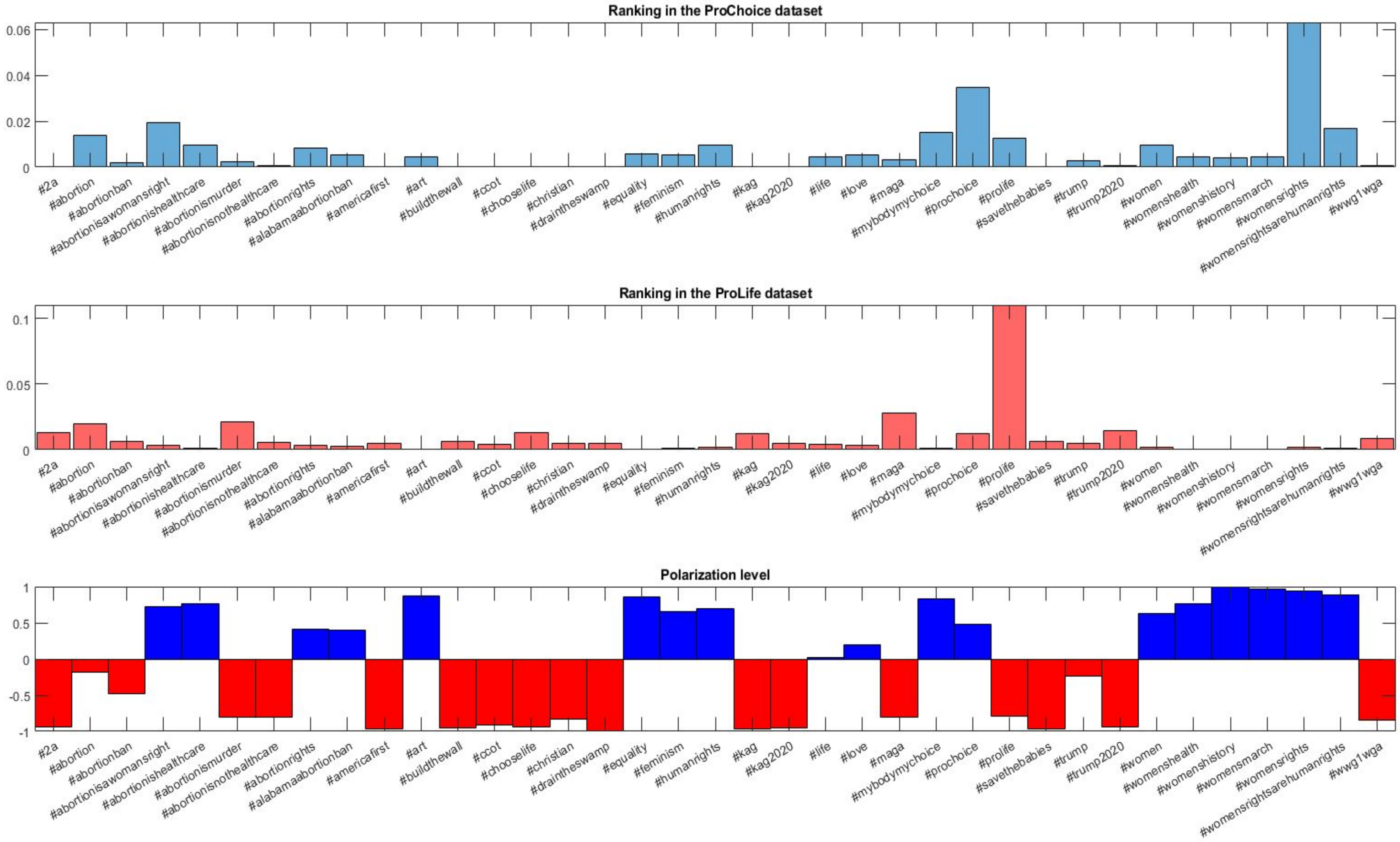
Projection PageRank

Hashtags Projection: Polarization

- Extract which opinion an hashtag holds
- Measure of hashtags centralities among the two dataset

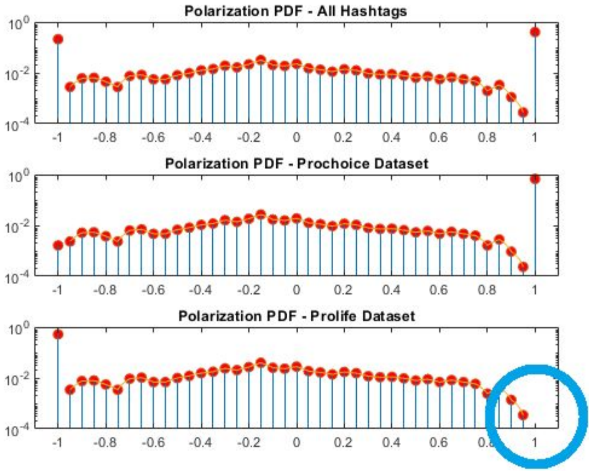
$$P_i = \frac{W_{pc_i} - W_{pl_i}}{W_{pc_i} + W_{pl_i}}$$

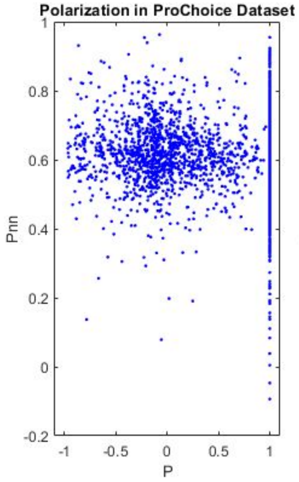


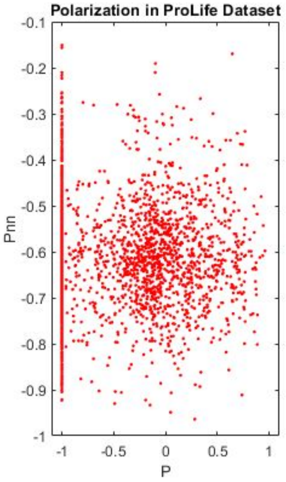


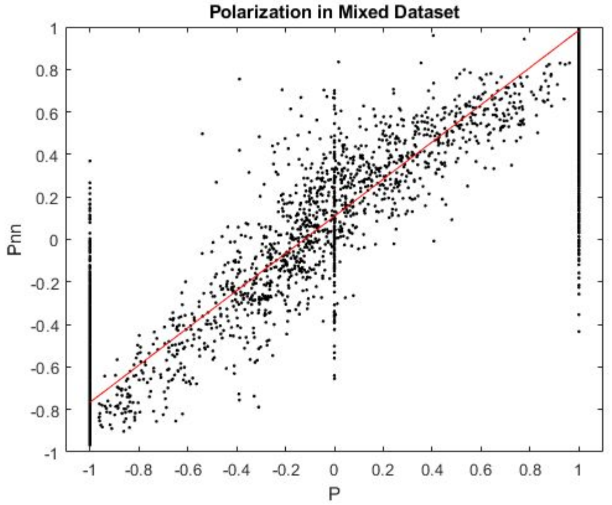
Ranking in the ProChoice dataset
0.06
0.04
0.02
0
#2a
#abortion
#abortionban
#abortionisawomansright
#abortionishealthcare
#abortionismurder
#abortionisnothealthcare
#abortionrights
#alabamaabortionban
#americafirst
#art
#buildthewall
#ccot
#chooselife
#christian
#draintheswamp
#equality
#feminism
#humanrights
#kag
#kag2020
#life
#love
#maga
#mybodymychoice
#prochoice
#prolife
#savethebabies
#trump
#trump2020
#women
#womenshealth
#womenshistory
#womensmarch
#womensrights
#womensrightsarehumanrights
#wwg1wga
Ranking in the ProLife dataset
0.1
0.05
0
#2a
#abortion
#abortionban
#abortionisawomansright
#abortionishealthcare
#abortionismurder
#abortionisnothealthcare
#abortionrights
#alabamaabortionban
#americafirst
#art
#buildthewall
#ccot
#chooselife
#christian
#draintheswamp
#equality
#feminism
#humanrights
#kag
#kag2020
#life
#love
#maga
#mybodymychoice
#prochoice
#prolife
#savethebabies
#trump
#trump2020
#women
#womenshealth
#womenshistory
#womensmarch
#womensrights
#womensrightsarehumanrights
#wwg1wga
Polarization level
1
0.5
0
-0.5
-1
#2a
#abortion
#abortionban
#abortionisawomansright
#abortionishealthcare
#abortionismurder
#abortionisnothealthcare
#abortionrights
#alabamaabortionban
#americafirst
#art
#buildthewall
#ccot
#chooselife
#christian
#draintheswamp
#equality
#feminism
#humanrights
#kag
#kag2020
#life
#love
#maga
#mybodymychoice
#prochoice
#prolife
#savethebabies
#trump
#trump2020
#women
#womenshealth
#womenshistory
#womensmarch
#womensrights
#womensrightsarehumanrights
#wwg1wga

Hashtags Projection: Polarization





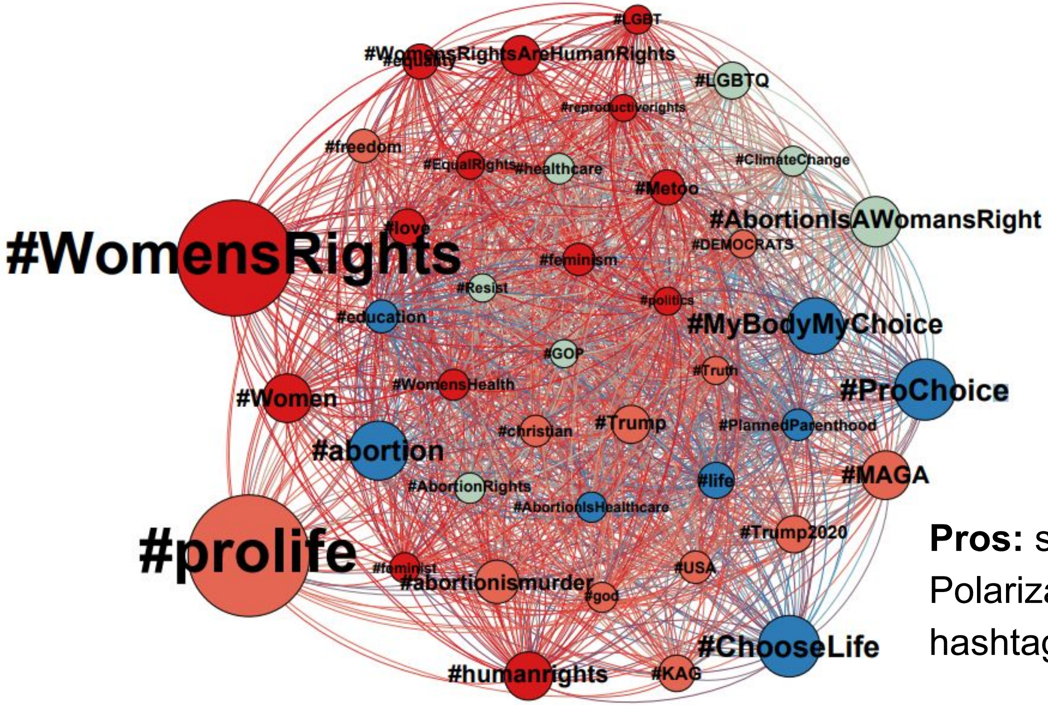




Absence of a debate in the prolife dataset

Linear trend among hashtags induced by polarization

Hashtags Projection: Finals



- Ranking based on betweenness centralities (avg. path)
- Community division on modularity

Main findings:

1. Womens'rights biased.
2. #Prolife has higher centrality.
3. Communities are not perfect. Projections lead to loss of infos. There could be other parameters, different from hashtags, to state user-opinion.

Pros: strong opinion hashtags can be identified. Polarization could be a good way to account hashtags opinion to perform community division.

Con: No real information about tweets content.

Bipartite Network Hashtags-Tweets

The goal is creating bipartite network of tweets and hashtags and showing the characteristics of networks for three different sets of data, prolife, prochoice and prolife and prochoice together.

Dataset used for creating network contains 334769 tweets in the prolife and 224302 tweets in prochoice dataset. Besides prolife and prochoice datasets there is a network based on mixed dataset which contains 50% of samples from both datasets.

Bipartite Network Hashtags-Tweets

Network analysis

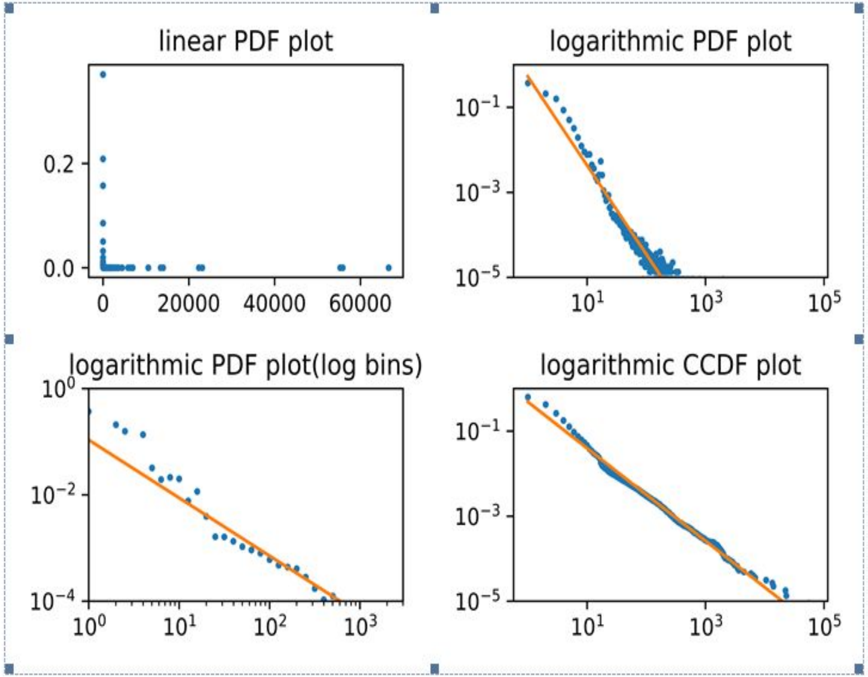
For network analysis are considered degree distribution, asortativity for datasets of tweets from all datasets.

Assortativity represents a preference for nodes to be attached to other nodes with similar characteristics. Analysing results we can conclude that networks showed diassortativity as the nodes with low degree have only high degree neighbours and vice versa.

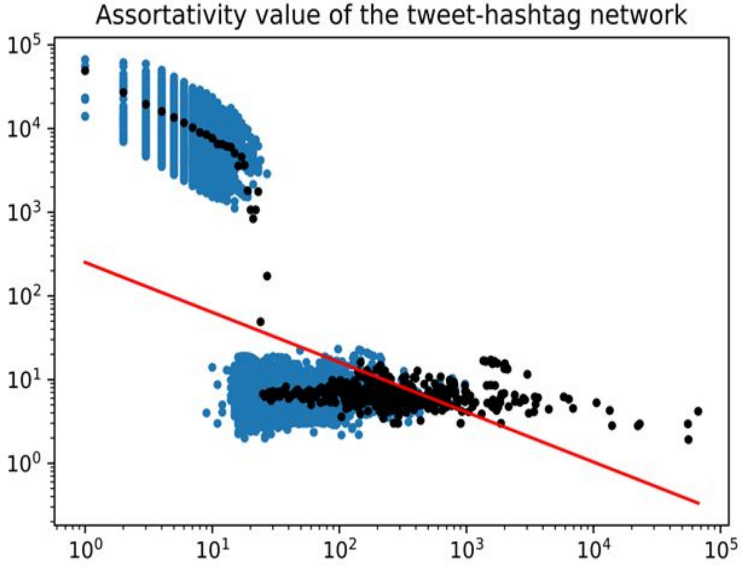
As a part of analysis there is comutation detection and ranking of hashtags that appear in prochoice prolife and mixed dataset.

Prochoice dataset

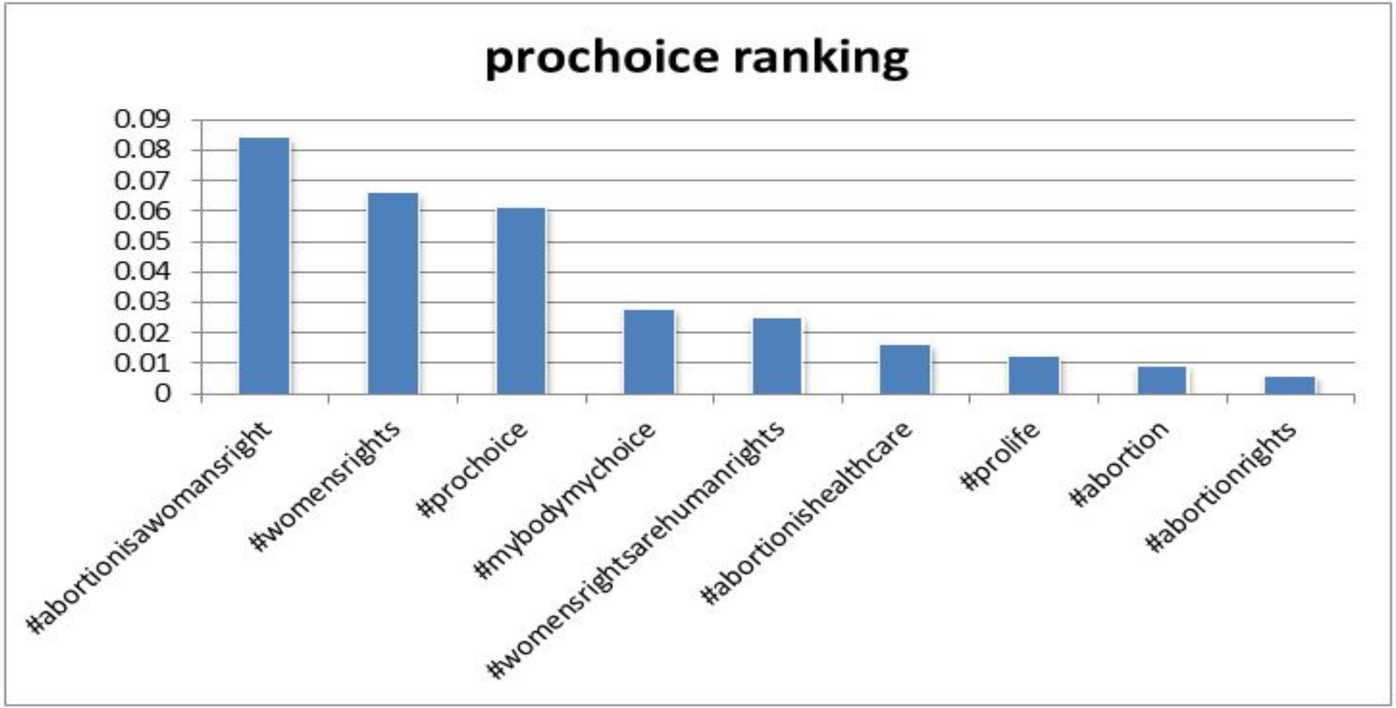
Degree distribution:



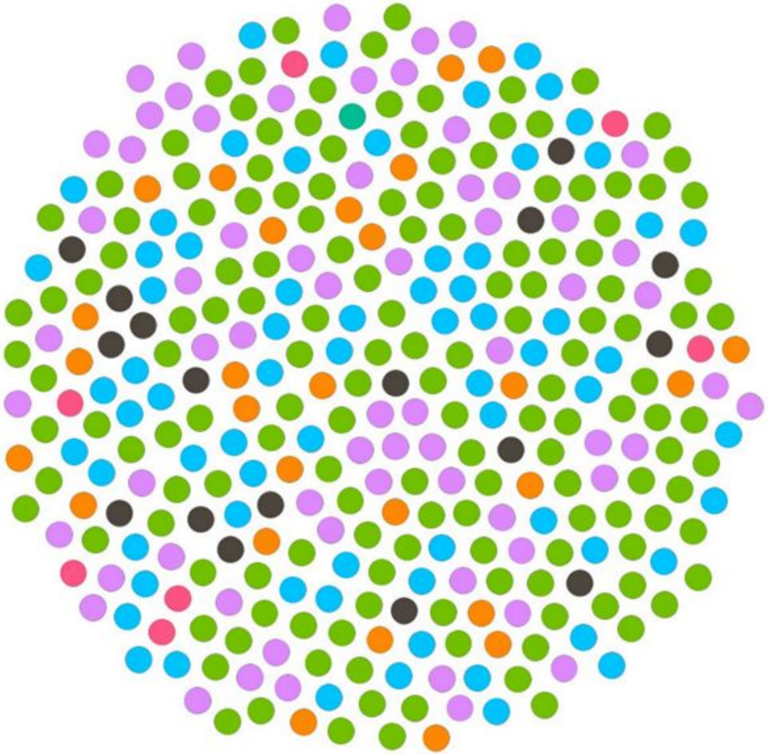
Assortativity:



Prochoice dataset ranking



Prochoice dataset comunity detection



In prochoice dataset can be detected 8 different communities, four of them are dominant with 34.76%, 29.92%, 23,14% and 10.26%.

Hashtags related to #boycottalabama #resist #abortionrights #abortionban (34.76%)

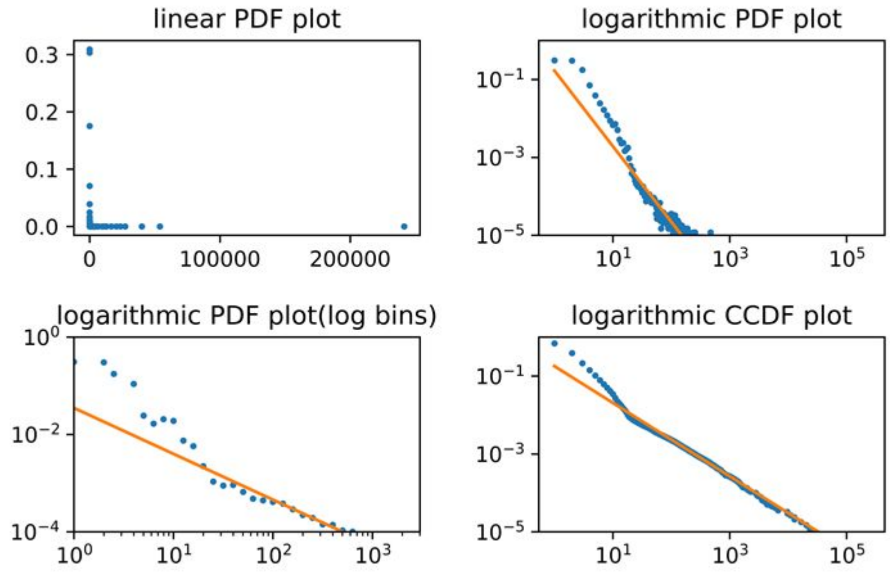
Hashtags related to #women #womenrights #humanrights (29.92%)

Hashtags related to #prochoice #proabortion #abortion (23.14%)

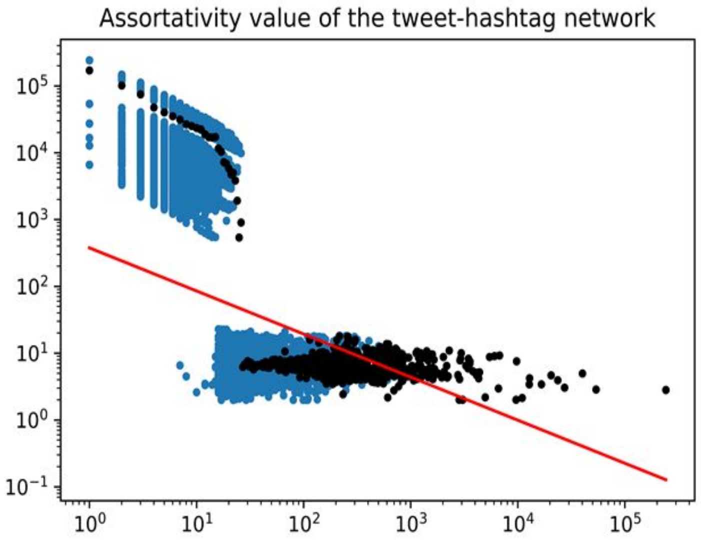
Hashtags related to #mybodymychoice #righttochoose #abortionishealthcare (10.26)

Prolife dataset

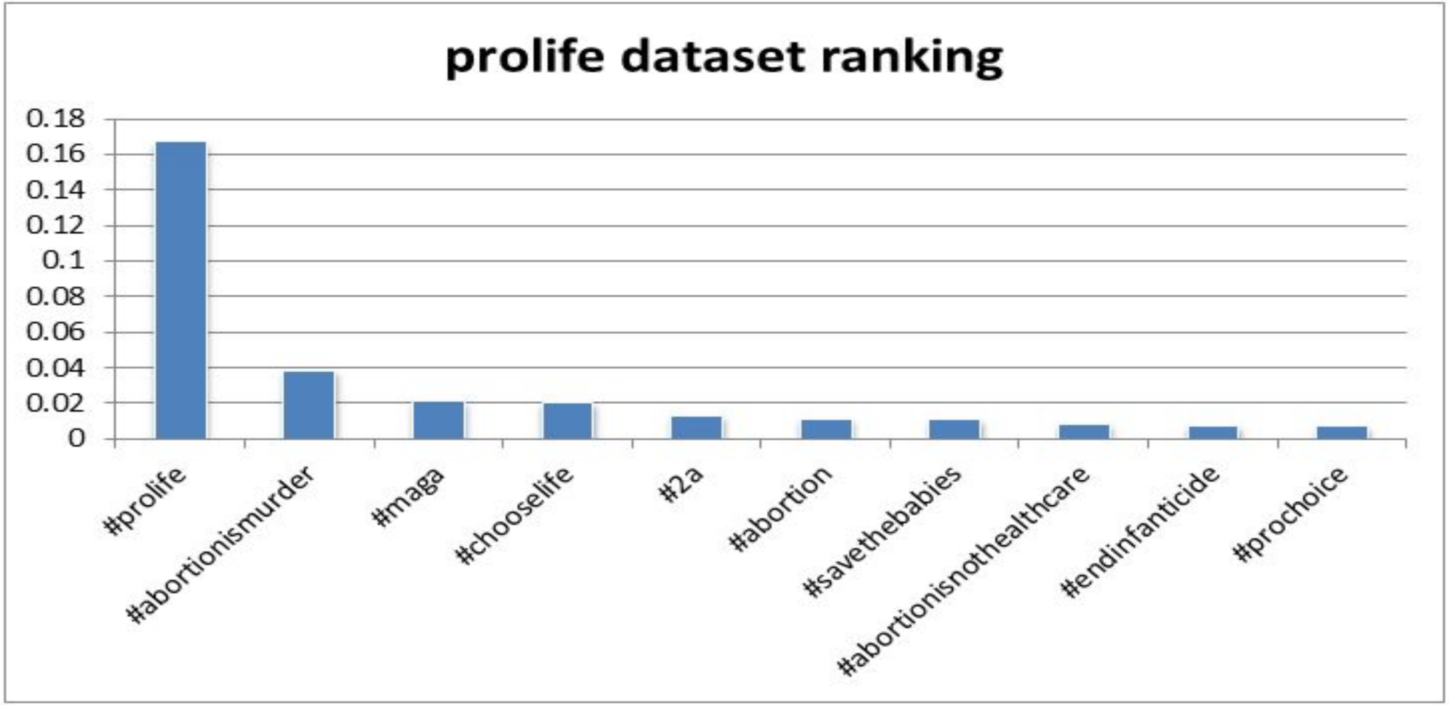
Degree distribution:



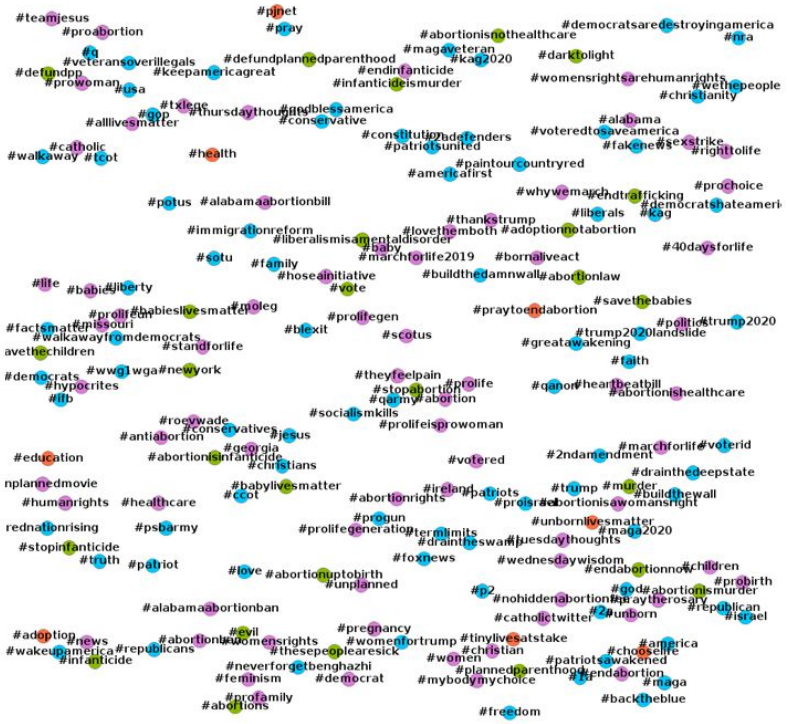
Assortativity:



Prolife dataset ranking



Prolife dataset comunity detection



After graph analysis there can be detected four diferent communities:

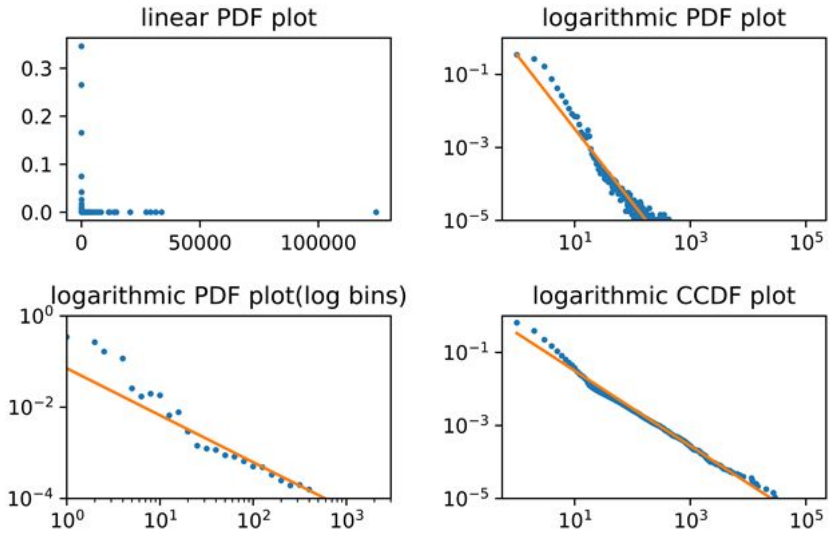
Religion oriented prolife hashtags #life #catholic #prayer

Empathy oriented #adoption #chooselife #chooselove #health

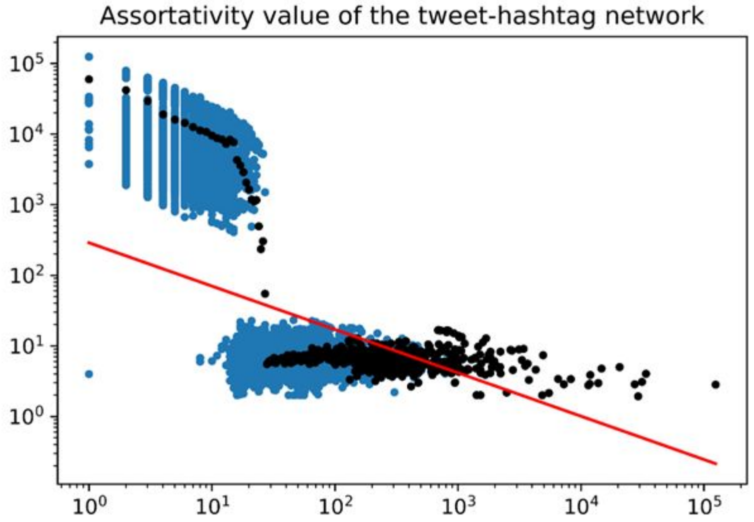
Criticizing prolife hashtags #murder #babykiller #infanticide

Hashtags related to usa such as #maga #trump2020 #walkawayfromdemocrats

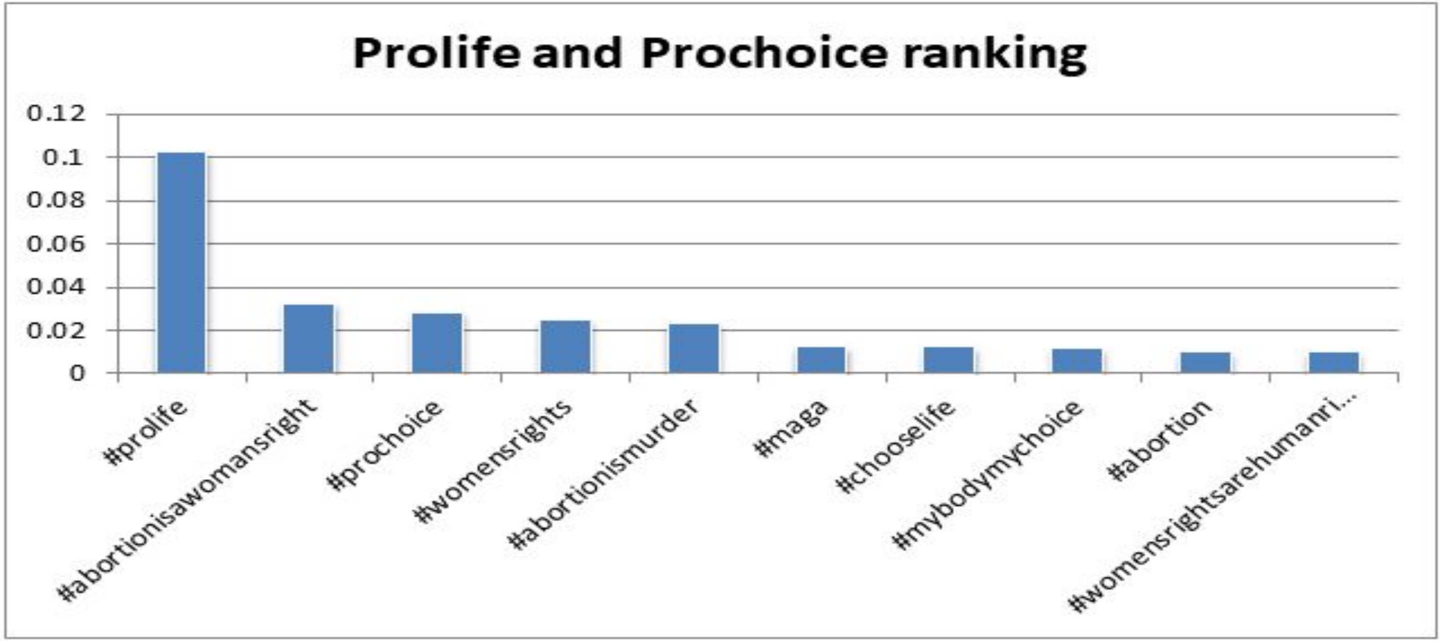
Degree distribution:



Assortativity:

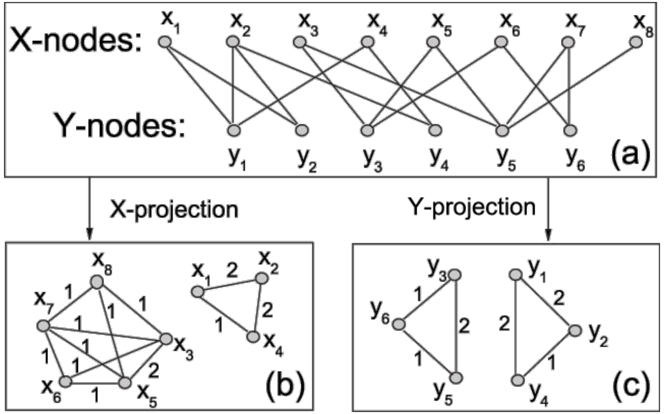


Mixed dataset ranking



Bipartite: Words-Hashtags

- The goal of this view on the data was to create bipartite network of words and hashtags and observe how they are used together in each of the data sets (prolife, prochoice and mixed).

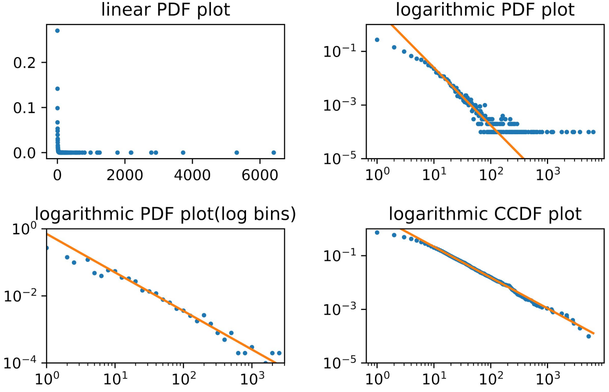


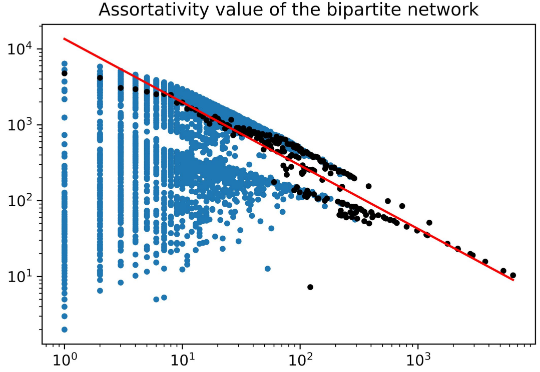
Bipartite: Words-Hashtags

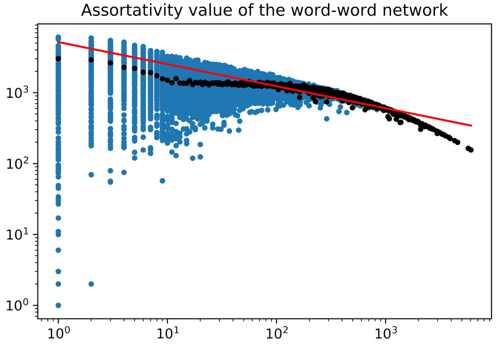
- Building the networks
  - Words were lemmatized and stopwords were removed.
  - Two sets *unique words* and *unique hashtags* were also built, recording the number of appearances in tweets every unique word and hashtag made. These sets were reduced using an arbitrary limit of 10.
- Results
  - Prochoice data set:
    - Unique hashtags, reduced: 3669
    - Unique words, reduced: 9740
  - Prolife data set:
    - Unique hashtags, reduced: 3324
    - Unique words, reduced: 12069
  - Mixed data set:
    - Unique hashtags, reduced 3165
    - Unique words, reduced 9518

Network analysis

- For every analyzed network, the degree distribution showed a power law with gamma ranging from 2,04 to 2,30. For the word projection dataset a higher value of minimum degree had to be used (200-250 as opposed to 50 used for others).
- All networks showed disasortativity, with word projection for prolife and mixed dataset leaning more to the neutral.

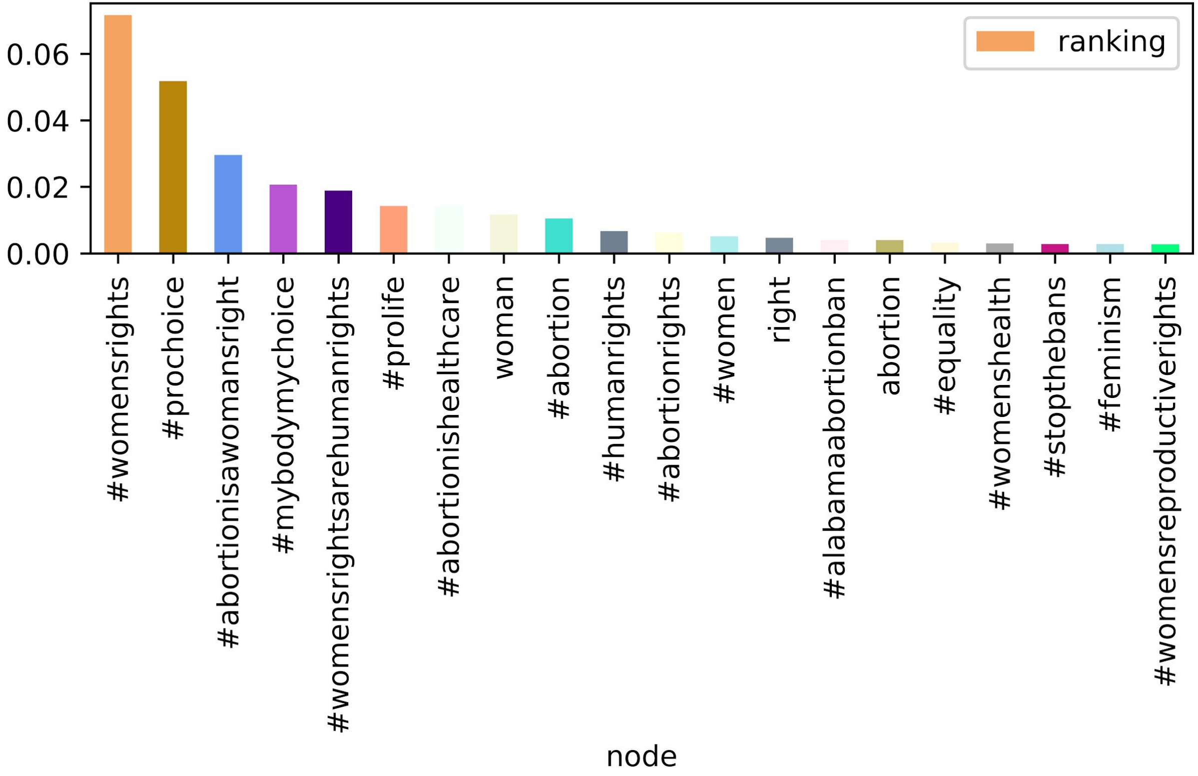






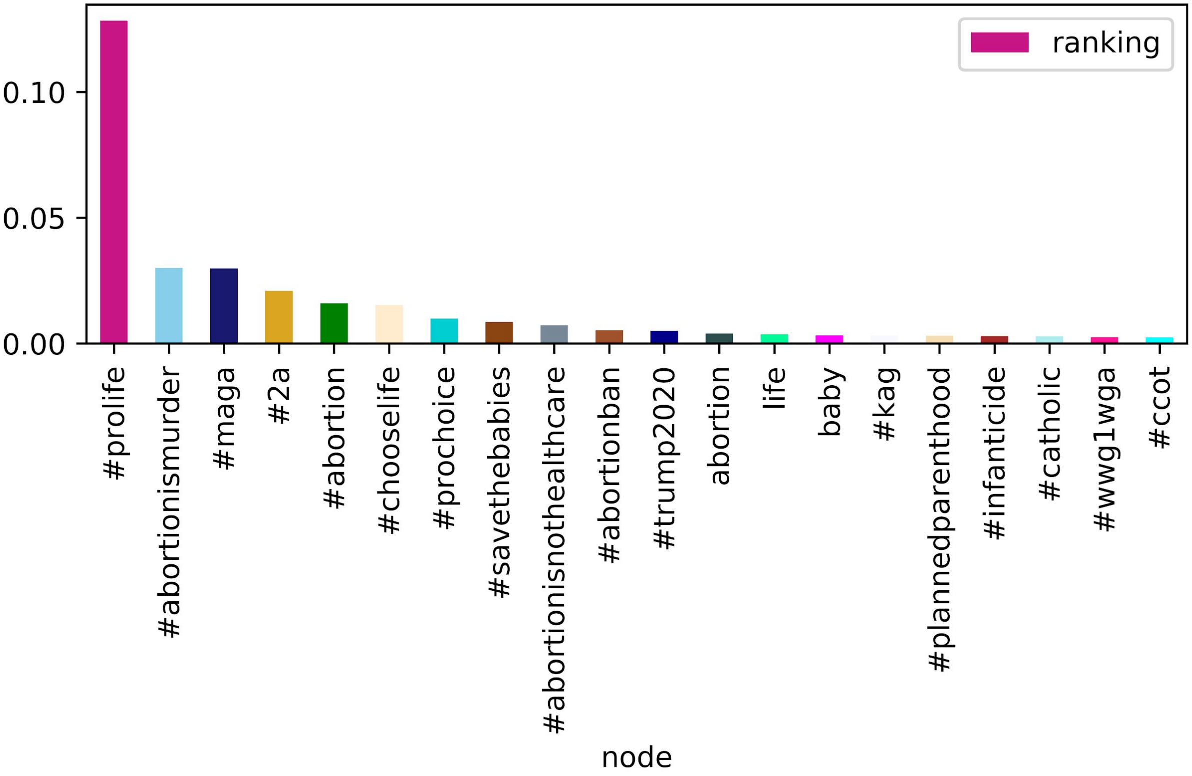
Bipartite - PageRank

Bipartite, prochoice:



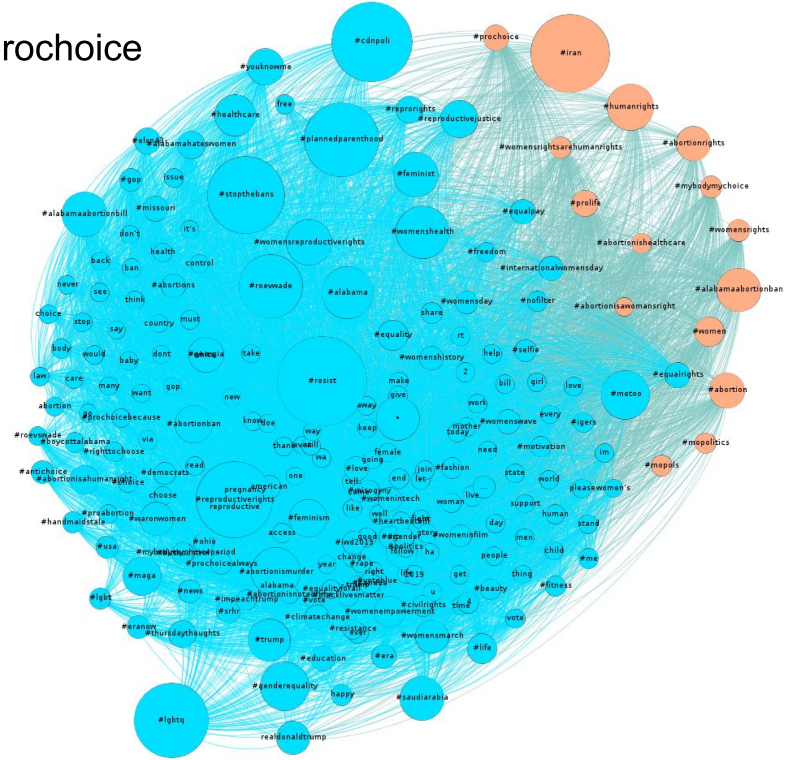
Bipartite - PageRank

Bipartite, prolife:

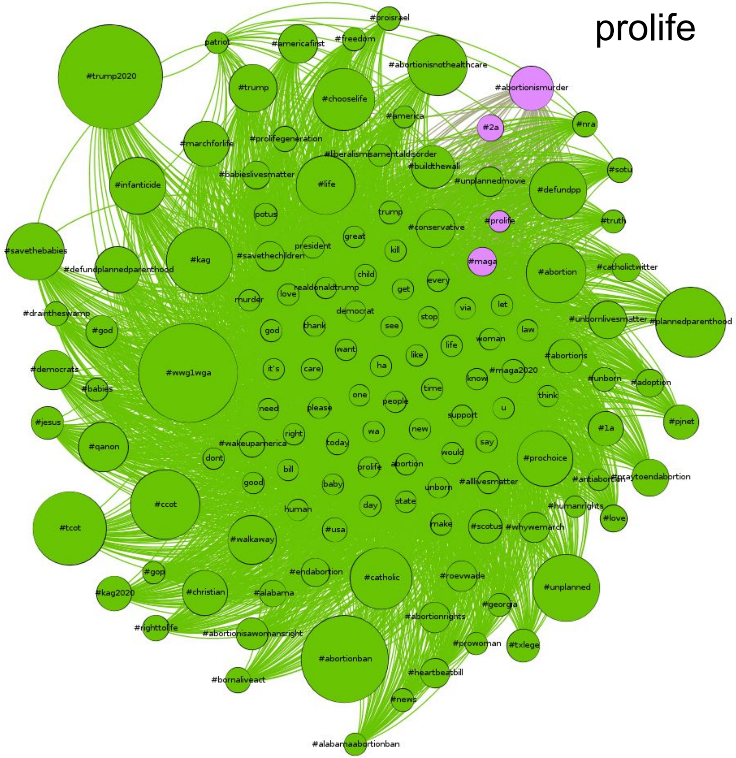


Bipartite - Community detection

prochoice

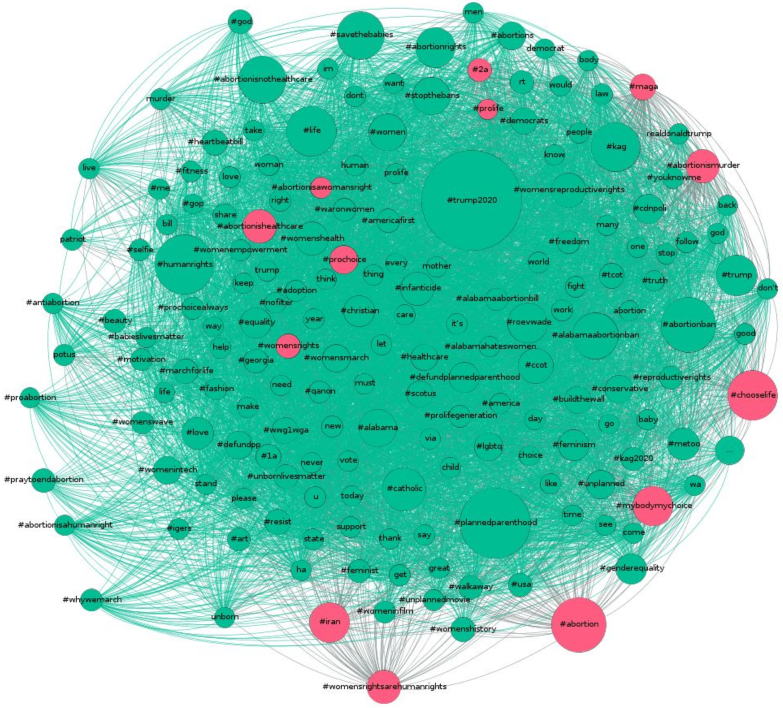


prolife



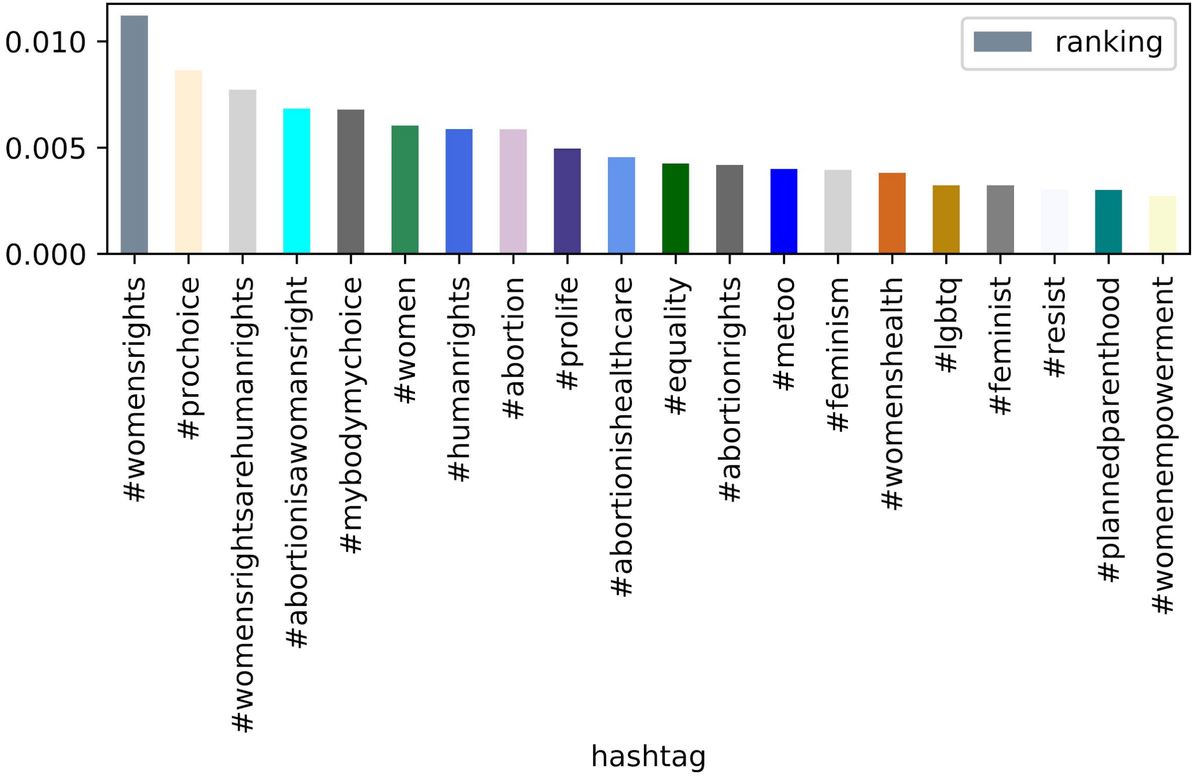
Bipartite - Community detection

Mixed



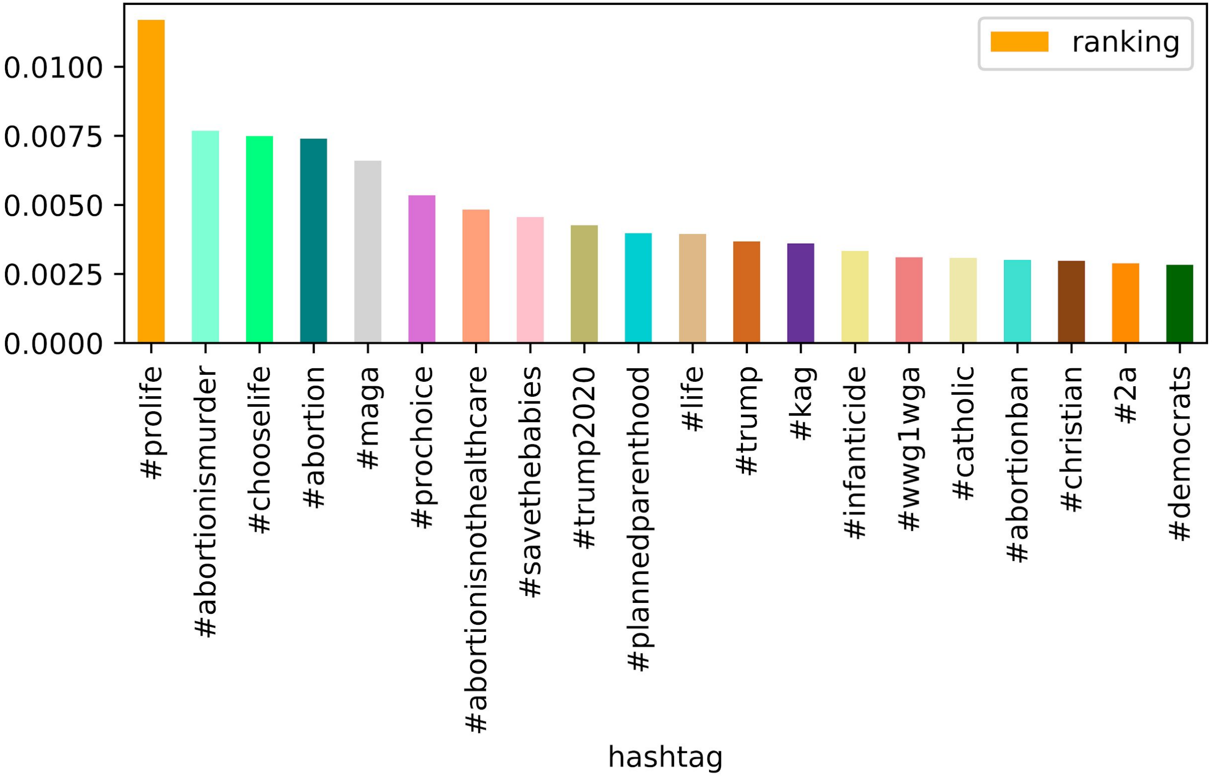
Hashtag projection - PageRank

Hashtag projection,
prochoice:



Hashtag projection - PageRank

Hashtag projection, prolife:

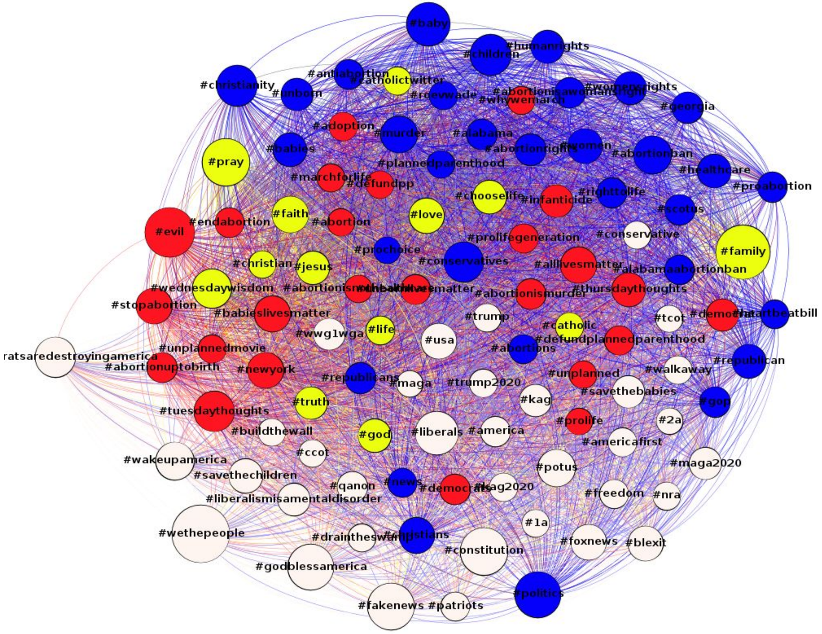


Hashtag projection - community detection

Prochoice

- The biggest two communities consisted of prochoice and prolife hashtags (55,64%) - teal community, and a set of what could be called feminism movement/gender equality hashtags (42,37%) - pink community.

Hashtag projection - community detection

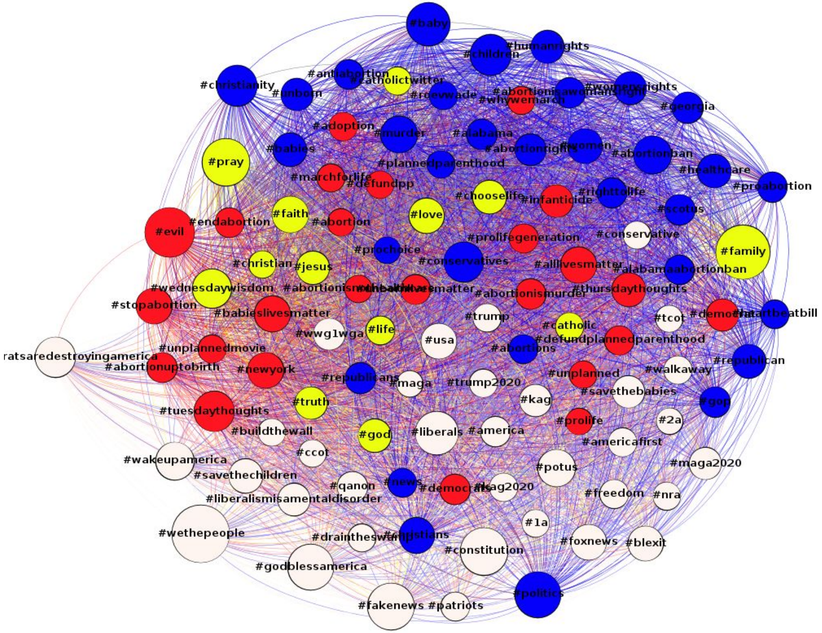


Prolife

There were four key communities in the prolife hashtag projection network:

- prolife (red, 29,6%) - all highly ranked prolife hashtags are in this community, except for #savethebabies
- USA politics(white, 26.38%) - all highly ranked prolife hashtags that have to do with USA politics (pro Trump) are in this group, along with other politics related hashtags and #savethebabies
- prochoice (blue, 23.59%) - all highly ranked prochoice hashtags are in this community, but it does include hashtags such as #christianity, #republican, #conservative and #gop.

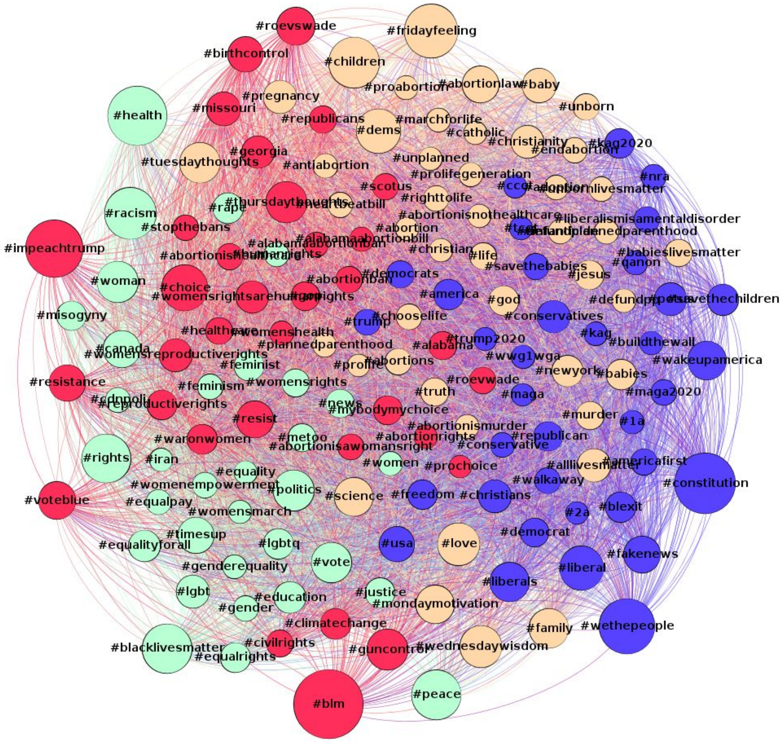
Hashtag projection - community detection



Prolife

- religion (yellow, 19.65%) - it includes a lot of religion related hashtags and has a prolife sentiment
- the Nancy Pelosi community (dark green, 0.75%, not pictured) - Nancy Pelosi is a hub in this community, which contains a lot of news related hashtags

Hashtag projection - community detection

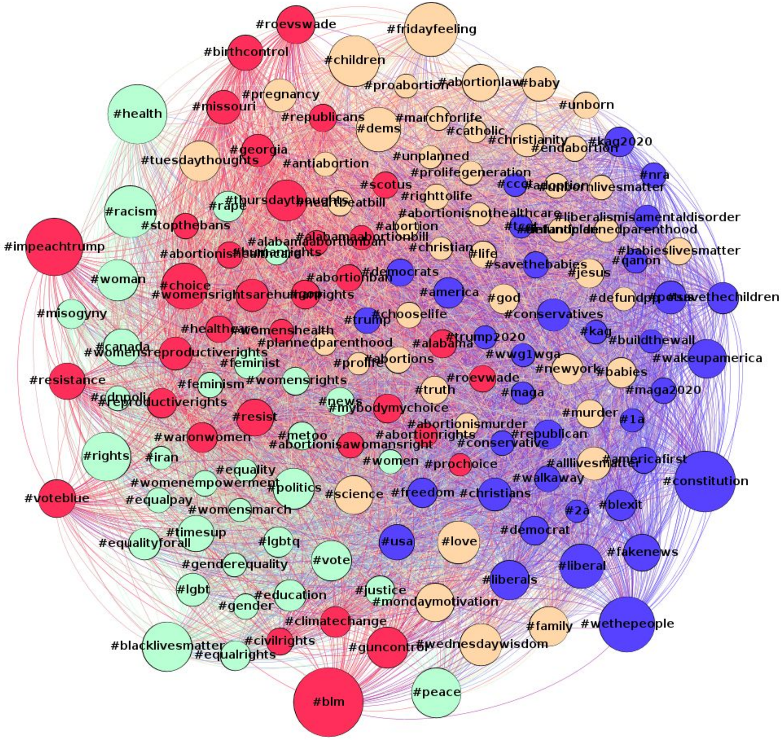


Mixed

There were four key communities in the mixed hashtag projection network:

- Gender equality/human rights/feminism group (green, 29,59%) - it contained a lot of hashtags that are in correlation with human right movements, such as #lgbtq, #blacklivesmatter, #racism, as well as feminism hashtags
- Prochoice group (red, 26,14%) - this group contained the prochoice hashtags from the hashtags used for data collection, as well as liberal/democrat and anti Trump hashtags (#impeachtrump, #civilrights, #guncontrol, #alabamaabortionrights, #voteblue)

Hashtag projection - community detection



Mixed

- Prolife group (orange, 23.64%) - this group contained prolife hashtags from the hashtags used for data collection, as well as religious/profamily hashtags
- Pro-Trump USA hashtags (blue, 17,69%) - this group contained republican/conservative and pro Trump hashtags such as #MAGA, #kag, #buildthewall etc

Word Projection - Community Detection

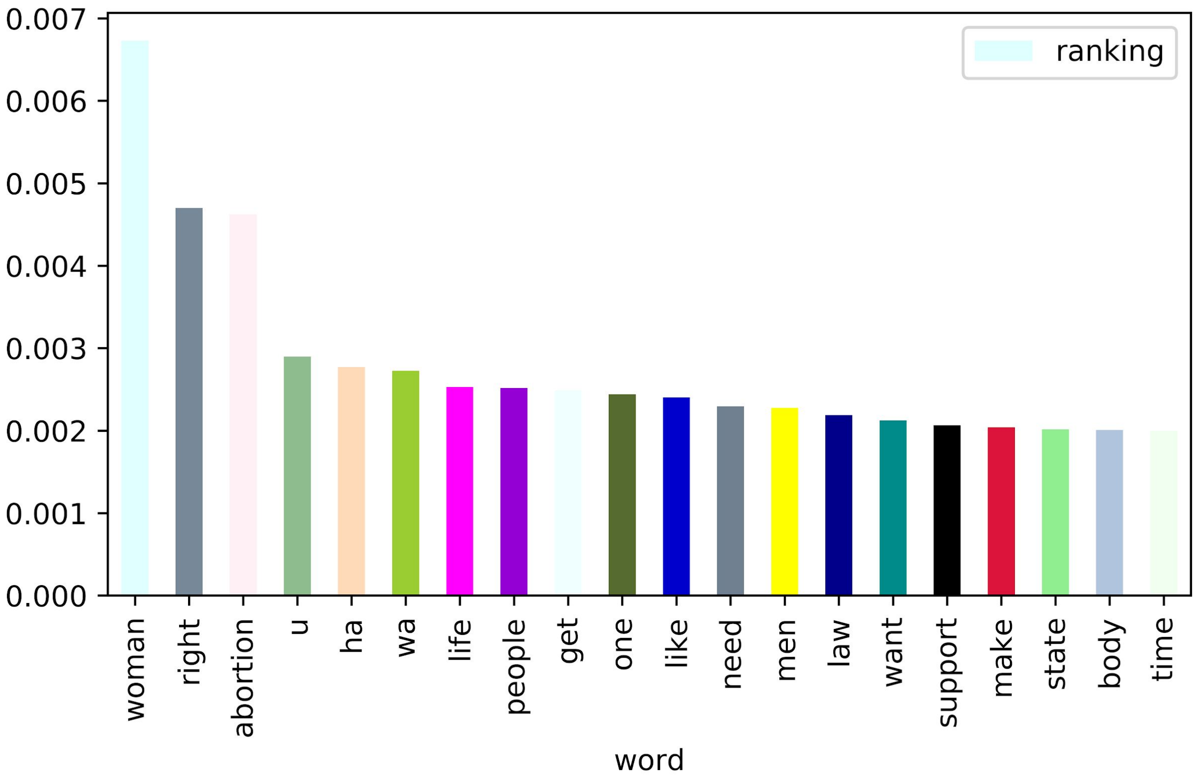
- There is no strong community detection. Since word projection is built from the bipartite network of hashtags and words we can deduce that this shows both sides use the same hashtags but can also be due to network structure (large but less hubs).

- This network is the largest in size, because the set of all words is greater than the set of all hashtags.

- The strongest result we can see in analysis of word projection network is in its ranking. Mainly, we can see that the prochoice dataset values *woman, abortion* and *right* the most, while the prolife dataset values *abortion, baby* and *life*.

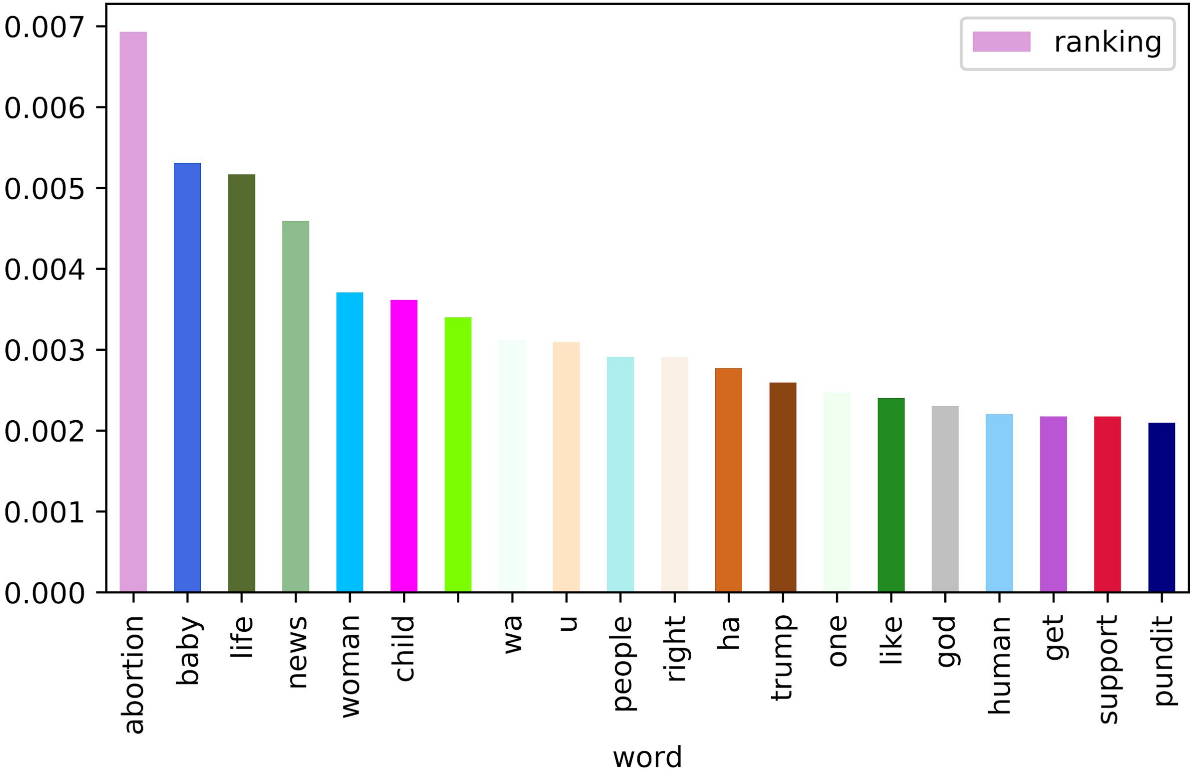
Word Projection - PageRank

Word projection, prochoice:



Word Projection - PageRank

Word projection, prolife:



Final conclusion

- comparing the different networks
- meaning of the results obtained
- communities identified

Conclusion

- The strongest results were found when observing the hashtag projection of the bipartite network
- The bipartite network analysis managed to observe communities without strong prolife (prochoice) in prochoice (prolife)
- The communities that are identified as discussing abortion on twitter are on one hand people who are religious, conservative, support Trump and the Republican party - the prolifers, and on the other hand people who want equal rights, female empowerment but are also discussing other human rights issues such as racism and LGBTQ rights - the prochoicers.
- The weakest results were obtained when observing the word projection of the bipartite network
- Except for a small filipino community and mentions of countries throughout the dataset, this analysis shows that the USA related topics dominate the discussion on abortion on Twitter, in the English speaking word.

Bipartite Network Tweets -Words

The aim of this analysis was to create the bipartite network of tweets and words on the three datasets, prochoice , prolife and on the merged dataset and to compare between their differences .

Bipartite Network Tweets-Words

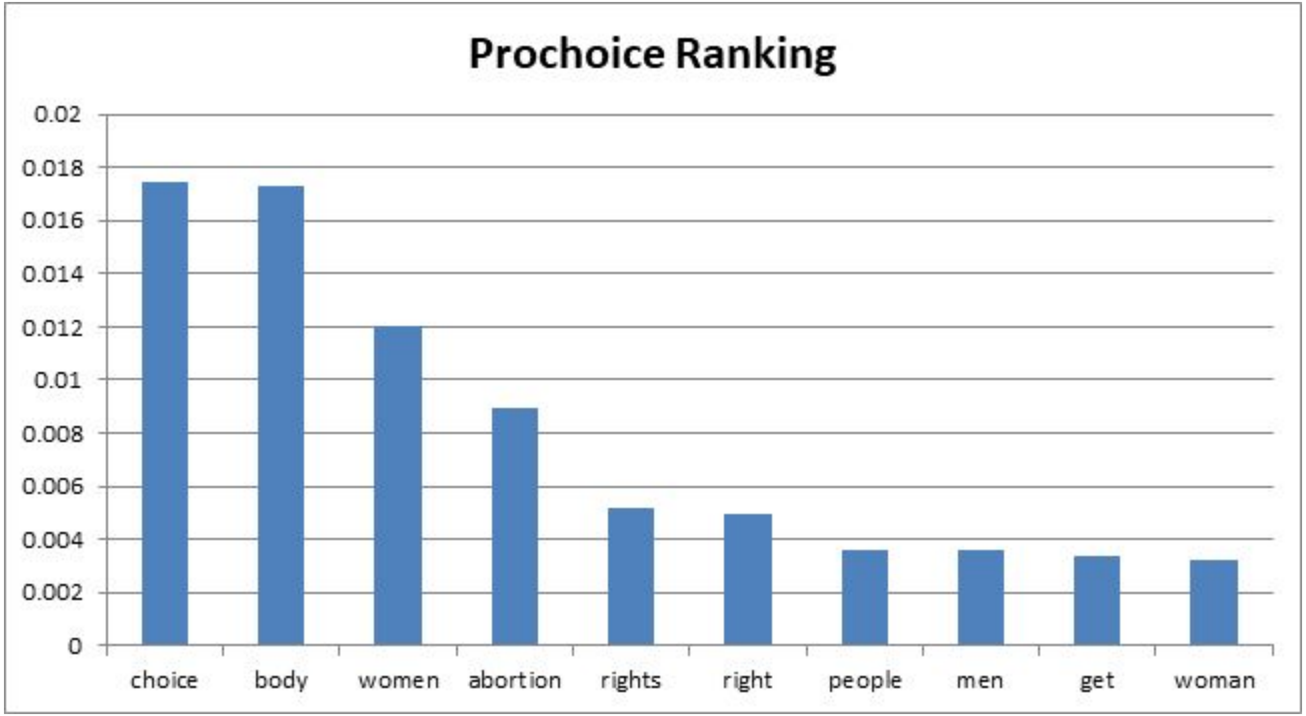
As in the first part of the pre-processing of the data, functions to remove the punctuation but not ‘#’, to remove numbers , urls and mentions were used. Needs to be mentioned that all the hashtags were disgarded.

Bipartite Network Tweets-Words

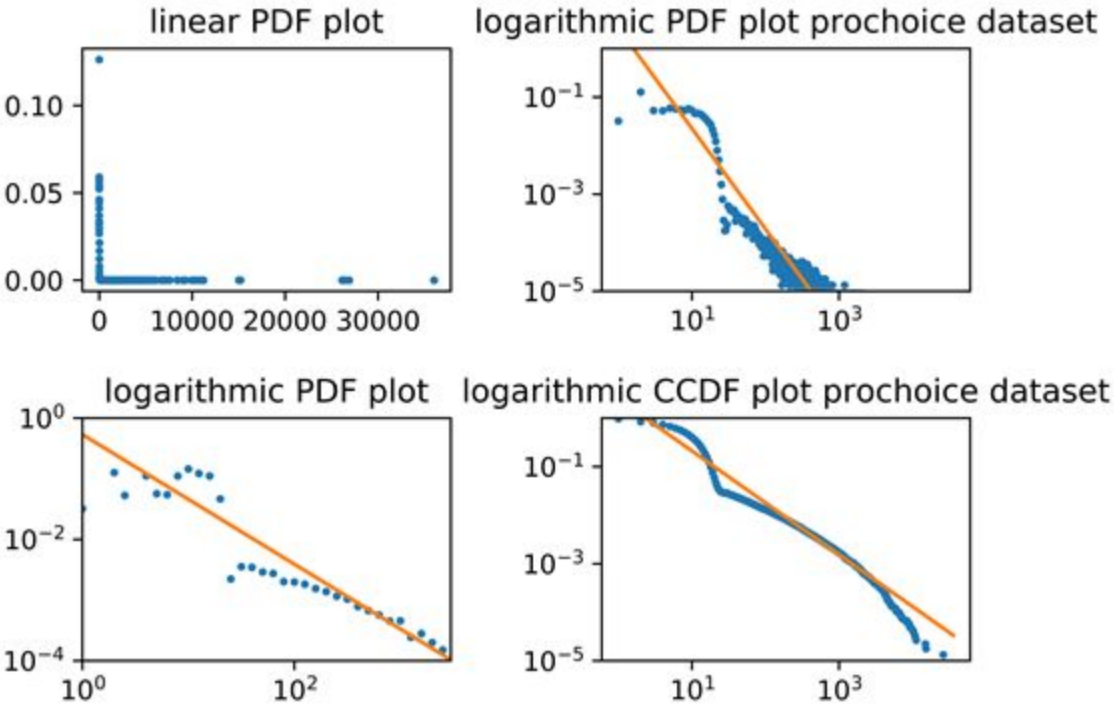
- Network analysis

My aim was to create the bipartite network of words and tweets , so the first step would be to extract the words from the tweets by preprocessing them, using the libraries and functions mentioned above. Then to identify the most relevant words used in tweets, taking into consideration the fact that we have divided the topic of abortion in two different approaches. Also, I have tried to identify communities into the tweet-words bipartite network in three different datasets, prochoice , prolife and in the merged dataset .

Insights for the Prochoice Dataset



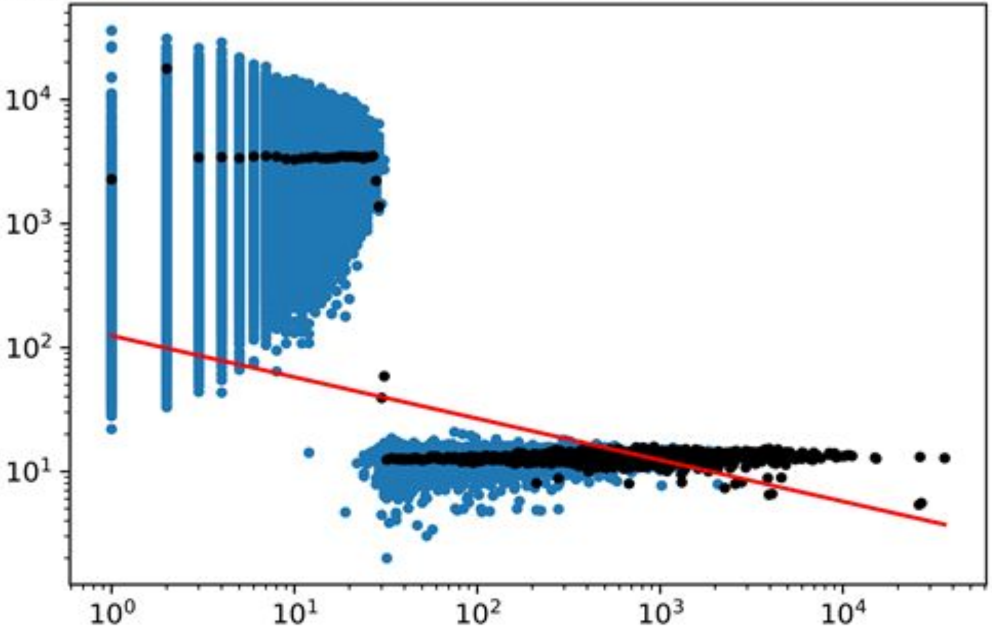
Insights for Prochoice Dataset



The value of gamma in this case is 2.068. Since gamma is between the value 2 and 3 , this network is a scale free one.

Insights for Prochoise Datasets

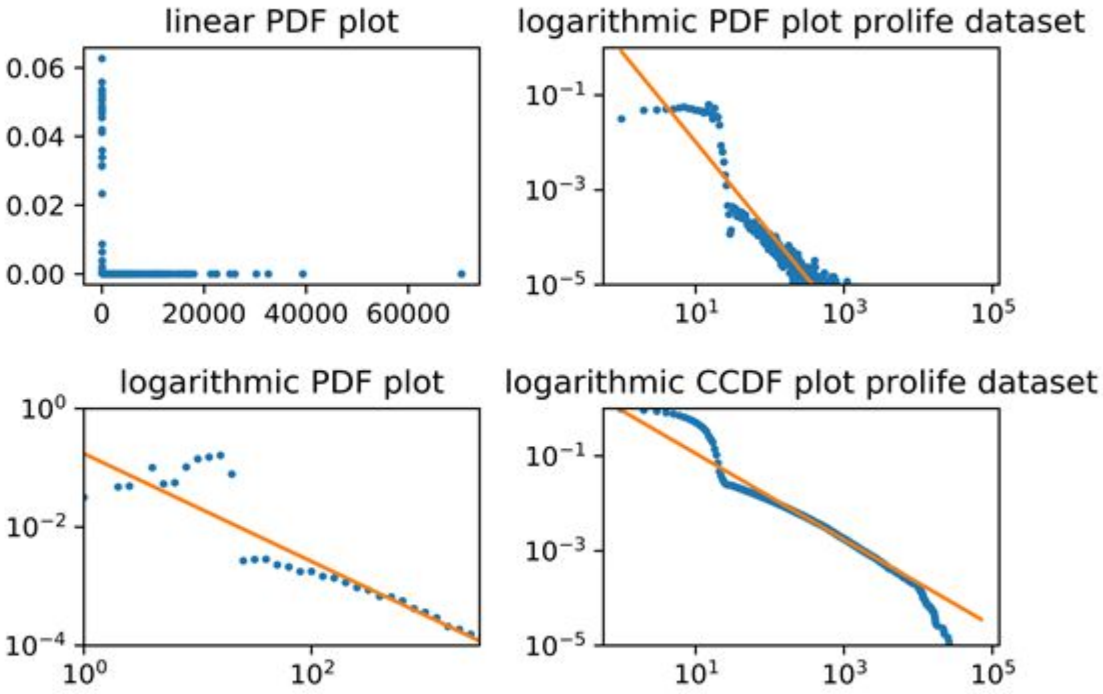
Assortativity value of the tweets-word network on prochoise dataset



Assortativity explains to what extends the nodes in the network are associated with other words in the network. I can point out that this network is disassortative because high degree nodes tend to connect to nodes with lower degrees.

Insights for the Prolife Dataset

Insights for the Prolife Dataset

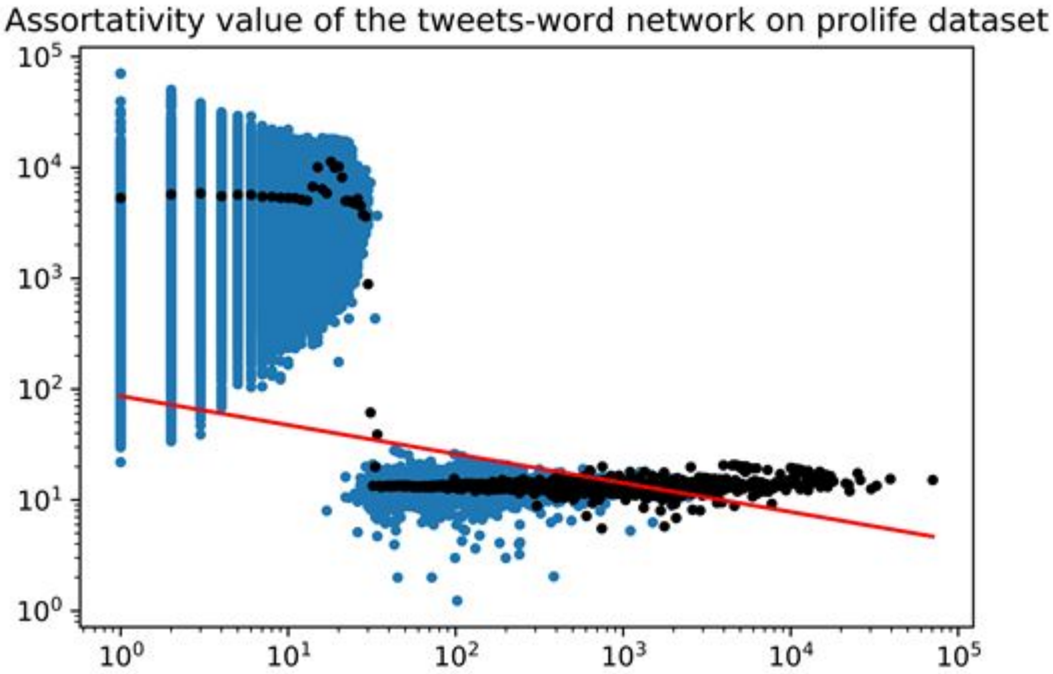


The value of gamma in the case of the prochoice dataset is 1.909.In this case , the gamma is lower than 2 , which means that the network is not totally scale free.

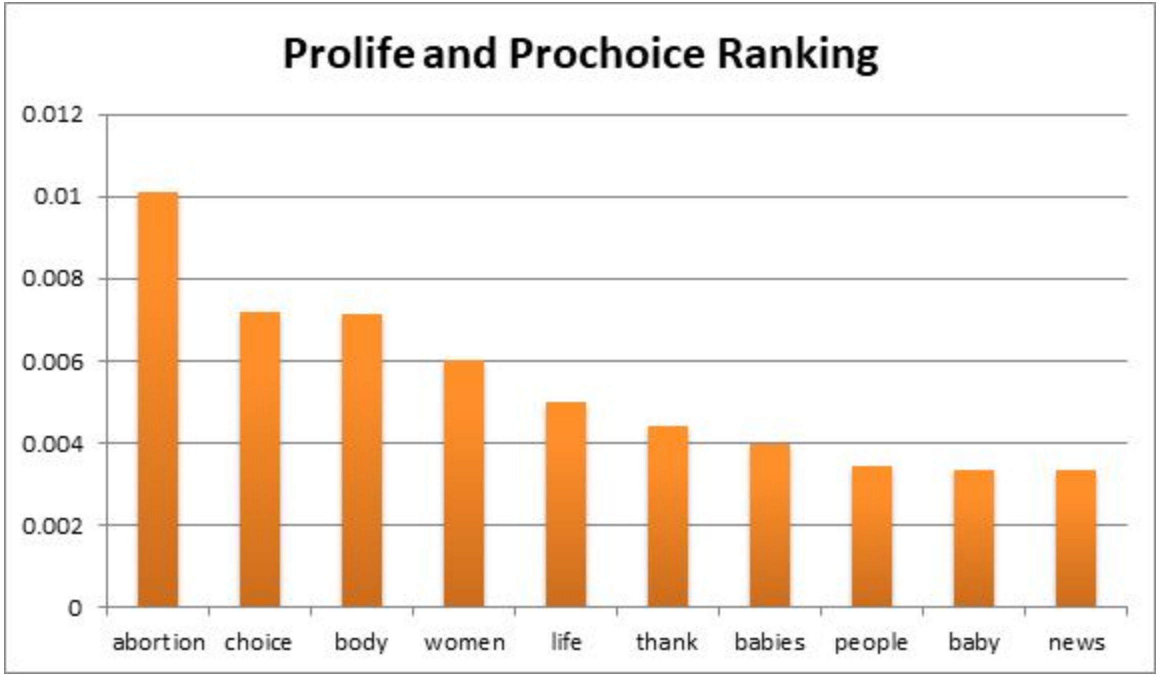
In order for the network to be scale free , the value of gamma should be around 2 and 3.

The number of links grows faster than the number of nodes, this leads to possessing the small world property.

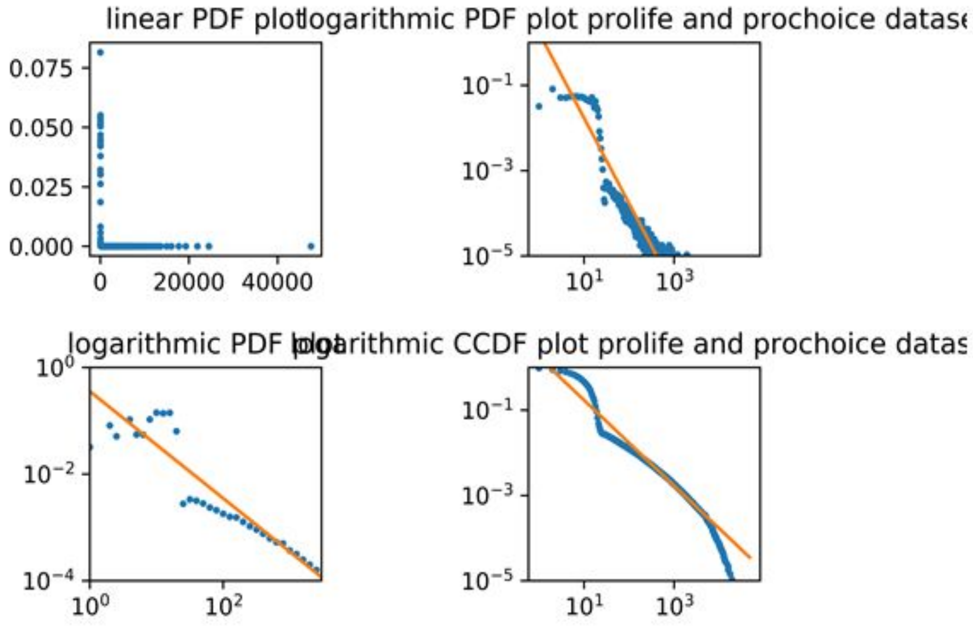
Insights for the Prolife Dataset

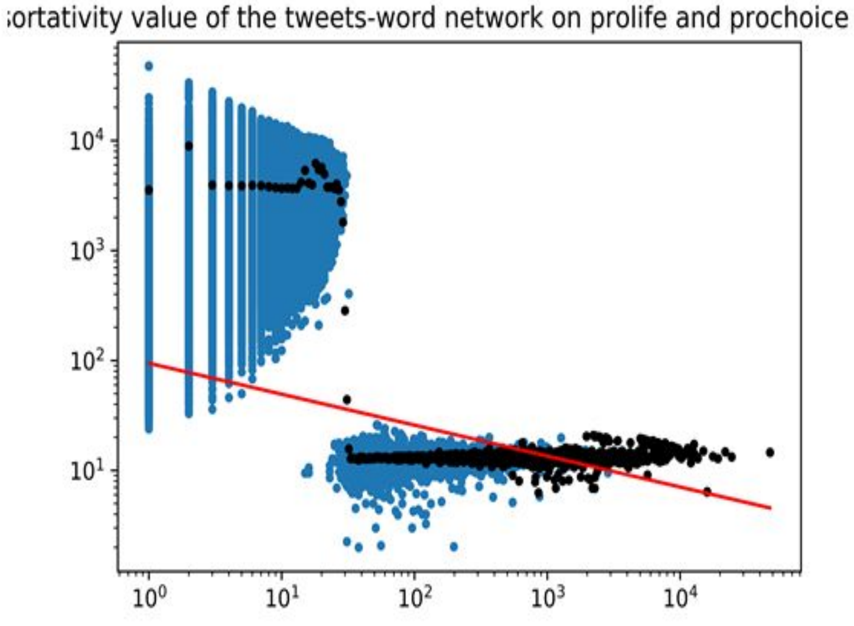


Insights for the merged datasets

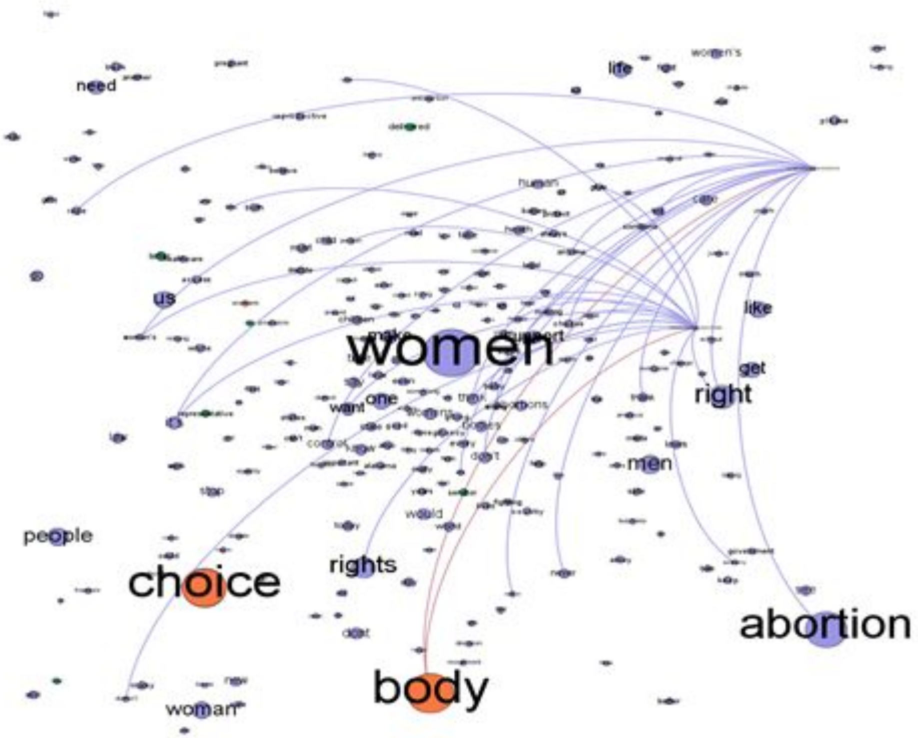


Insights for the merged datasets



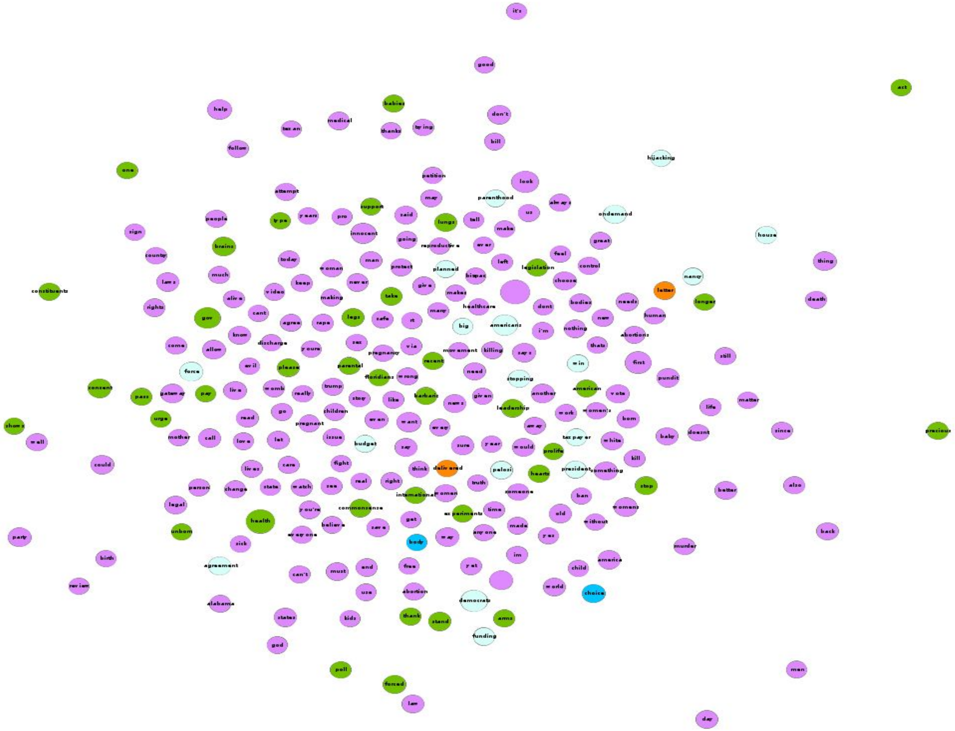


Bipartite Community Detection



For the prochoice dataset , applying a modularity value of 1.7 two communities were detected.

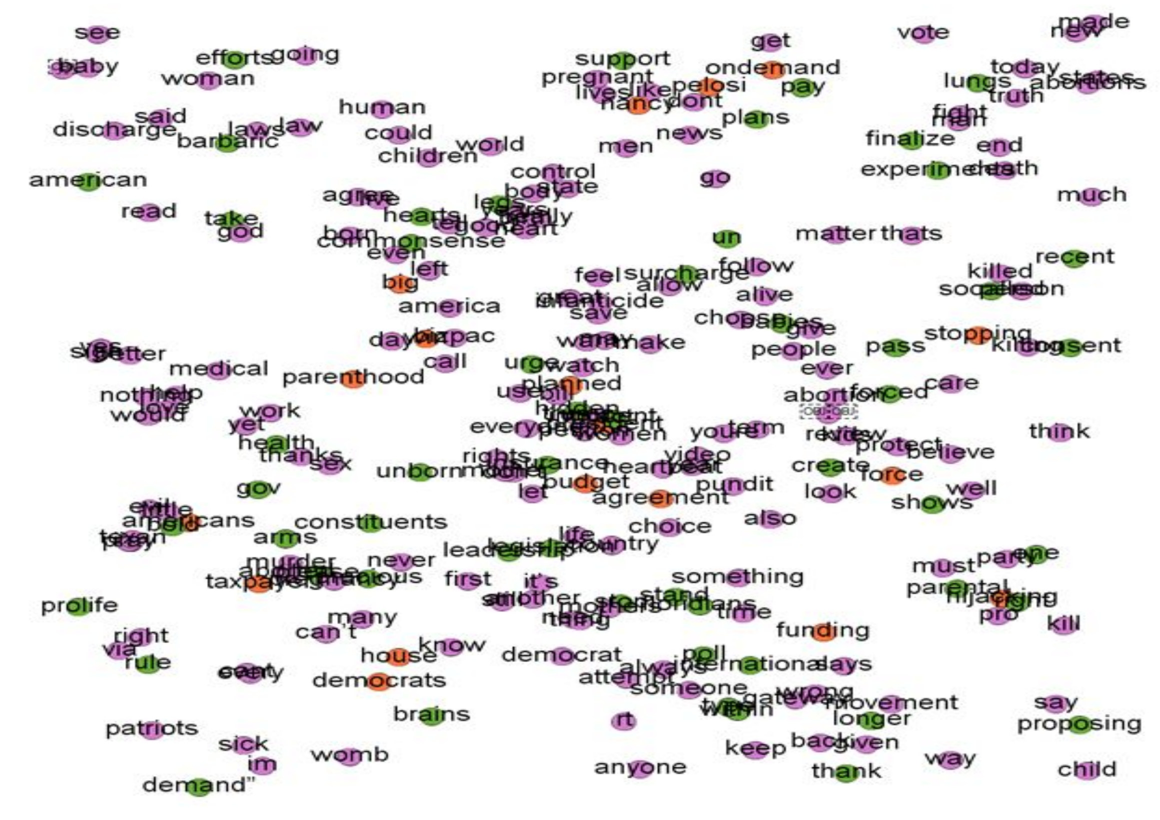
Bipartite Community Detection



For technical issues , this hereby shoud be the bipartite community detector for the prolife dataset . The only thing we can detect from the graph, are six communities detected, eventhough only four are apparent after applying degree filter.

But, i can point out some of the most high centrality nodes were : choice, child,saves

Bipartite Community Detection



In the merged dataset eight communities were detected.

Conclusion and Interpretation:

- The strongest results were found while observing hashtag projections

  → makes sense because hashtags on Twitter are used to underline statements

- Interests of Pro-Life community and Pro-Choice community highly different

  → Pro-Life: political orientations and movements (supporting Trump and his party), religion (being christian)

  → Pro-Choice: equal rights, female empowerment, other human rights topics like racism and LGBTQ

Conclusion and Interpretation:

- In Pro-Life data set only strongly polarized Pro-Life hashtags appeared; on Pro-Choice data set strongly polarized hashtags of both movements appeared

  → Pro-Choice community really wants to enter into a dialogue while Pro-Life mostly uses topic for political reasons

Conclusion, Interpretation and Outlook:

- Pro-Life data set likely to be biased because of upcoming presidential elections
- Collected data based on states gave an idea that in more gender equal states the post are more Pro-Choice than in less gender equal states

Outlook:

Future researches using data not only from 2019 and not only in English language to create less distorted idea of abortion sentiment