

# alt-reddit

Nick Rubin





# Objective

**evaluate the content of individual subreddits  
independent of case-by-case moderation**

01

**reduce individual bias**

direct connection between  
subreddit and analytics team

02

**provide deeper insight**

far more information covered

# data/methodology

## dataset:

86.6 gb of comments from 2016-2019

pared down to ~19,000 comments (each) from r/politics and r/The\_Donald from 2016

## methods:

processing: lemmetization, n-grams, custom stop words

analysis: MALLET LDA, PCA, GENSIM

visualizations: pyLDAvis, T-SNE

## data manipulation:

>50 net upvotes (sorted descending)

> 5 words long

## tools:





# r/The\_Donald is The\_Quarantined

## **june 2019:**

reddit "quarantined" its subreddit "The\_Donald," the most popular section of its site for supporters of President Trump because of what it said were "threats of violence against police and public officials."

## **alt-right growth:**

since the 2016 election, Cornell University documented "the escalation and spread of anti-semitic memes and rhetoric across the Web.", finding large increase in anti-semitism (in some cases over double) from 2016-2018.



Are you sure you want to view this community?

This community is quarantined.

Significant issues with reporting and addressing rule-breaking content. Most recently the violations have included threats of violence against police and public officials.

By continuing to view this community, you can help moderators maintain the community and downvoting rule-breaking content.

Are you certain you want to continue?



**results**



# topics: r/politics

a selection of 3 of the 8 topics, selected for distinctness/clarity. displayed with the top 10 words and their respective weights

## Clinton emails

```
(4,
[('email', 0.03650766316514551),
 ('server', 0.021181332874117446),
 ('gt', 0.020061994144997417),
 ('fbi', 0.013905631134837265),
 ('report', 0.012226623041157224),
 ('secretary', 0.011473221973480283),
 ('state', 0.009729636645427931),
 ('state department', 0.009514379197520234),
 ('private', 0.009471327707938695),
 ('information', 0.008868606853797142)]),
```

## the economy

```
(0,
[('money', 0.017650909430962734),
 ('pay', 0.014822088184411091),
 ('american', 0.013816804600925795),
 ('job', 0.013255716089213073),
 ('work', 0.01038013746668537),
 ('year', 0.010029457146864918),
 ('make', 0.00848646373965493),
 ('country', 0.00820591948379857),
 ('high', 0.007621452284097816),
 ('give', 0.007270771964277365)]),
```

## the election

```
(6,
[('party', 0.03789316159375752),
 ('candidate', 0.03570187862430186),
 ('win', 0.030651238609337002),
 ('republican', 0.02364982229229577),
 ('democrat', 0.018786243018625904),
 ('supporter', 0.018625905240373053),
 ('lose', 0.01843884449907806),
 ('dnc', 0.017931108201277357),
 ('support', 0.017556986718687367),
 ('democratic', 0.017396648940434516)]),
```



# topics: r/The\_Donald

a selection of 3 of the 8 topics, selected for distinctness/clarity. displayed with the top 10 words and their respective weights

**fear**

```
(5,
[('country', 0.026102885055716684),
 ('muslim', 0.02572126392917112),
 ('gt', 0.01698214013127767),
 ('leave', 0.016142573652877423),
 ('world', 0.015570141963059074),
 ('islam', 0.012593497176003663),
 ('american', 0.011372309571057854),
 ('attack', 0.009807662952221035),
 ('rape', 0.00976950083956648),
 ('child', 0.008929934361166235)]),
```

**divisions**

```
(3,
[('people', 0.08011670543705397),
 ('woman', 0.029610061901194654),
 ('man', 0.027875251350392302),
 ('call', 0.02771754130031936),
 ('white', 0.02318337736072231),
 ('black', 0.02097543665970114),
 ('liberal', 0.02046287899696408),
 ('medium', 0.019437763671489965),
 ('racist', 0.018294365808461145),
 ('hear', 0.009856878129558806)]),
```

**values**

```
(0,
[('government', 0.010790284983833371),
 ('gt', 0.009699977441912927),
 ('family', 0.008985638017896083),
 ('support', 0.008684863523573201),
 ('read', 0.00823370178208888),
 ('law', 0.00819610497029852),
 ('idea', 0.008045717723137079),
 ('kind', 0.007820136852394917),
 ('person', 0.0076321527934431165),
 ('find', 0.006955410181216633)]),
```

# pyLDAvis

r/politics



# pyLDAvis

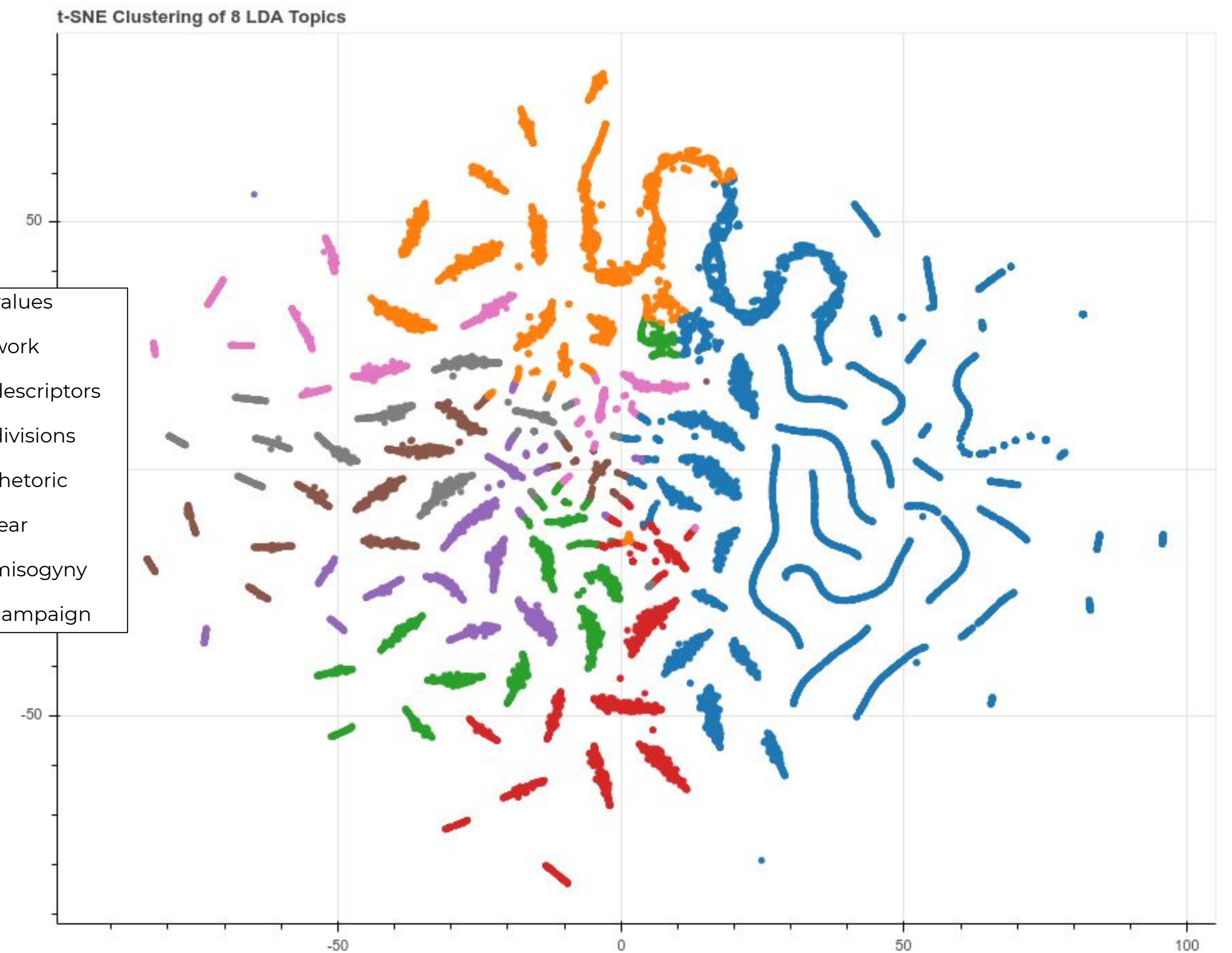
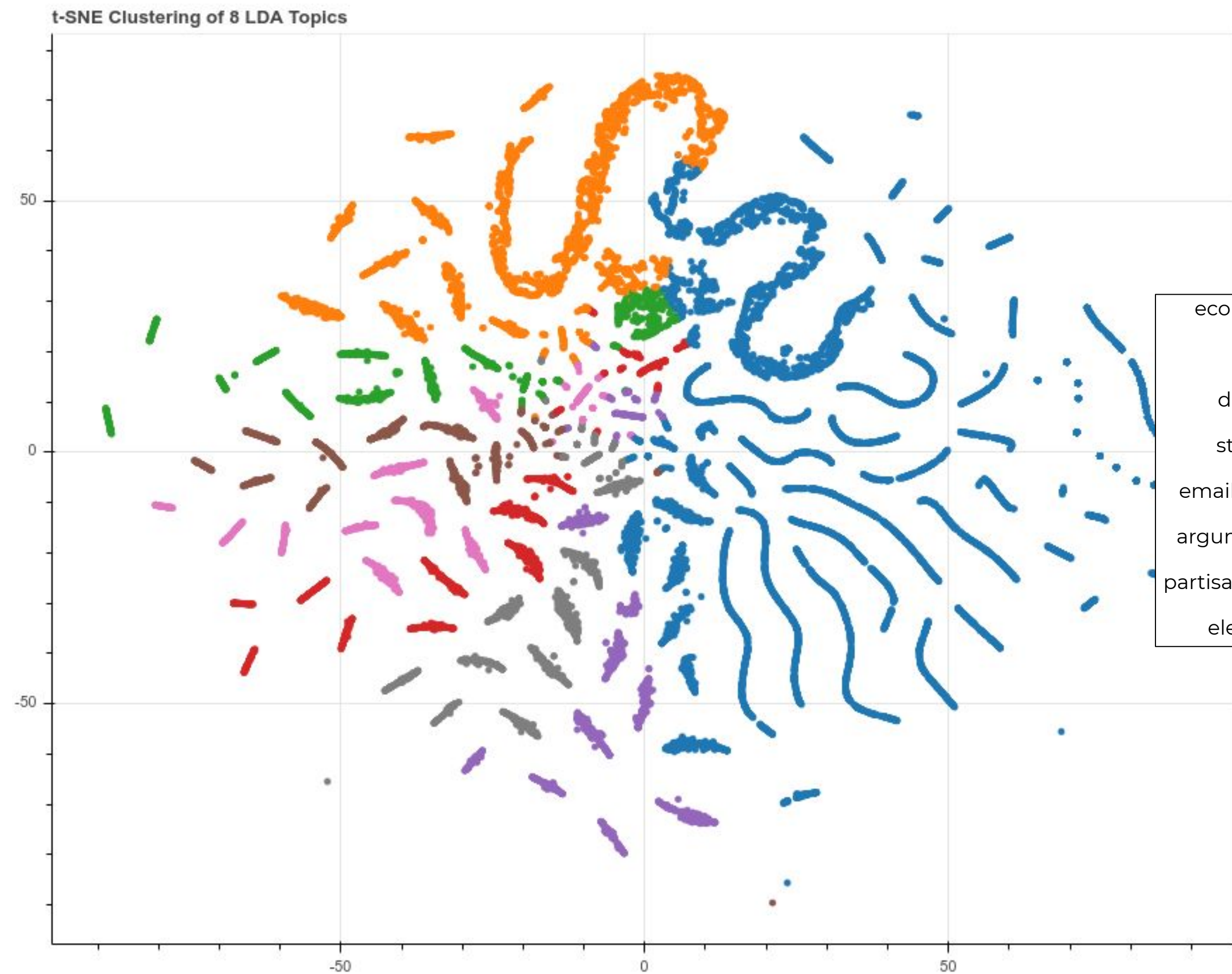
r/The\_Donald

# T-SNE

Clustering

r/politics

r/The\_Donald





# recommendations

## resolutions

- Can effectively utilize topic modeling and Mallet LDA to assist moderators in quickly evaluating subreddit content
- 

## next steps

- create dashboard of individual subreddit topics modeling over time to examine proliferation of hate speech proportionally between subreddits
- build app to allow quick evaluation
- use username data to map individual participation across multiple subreddits

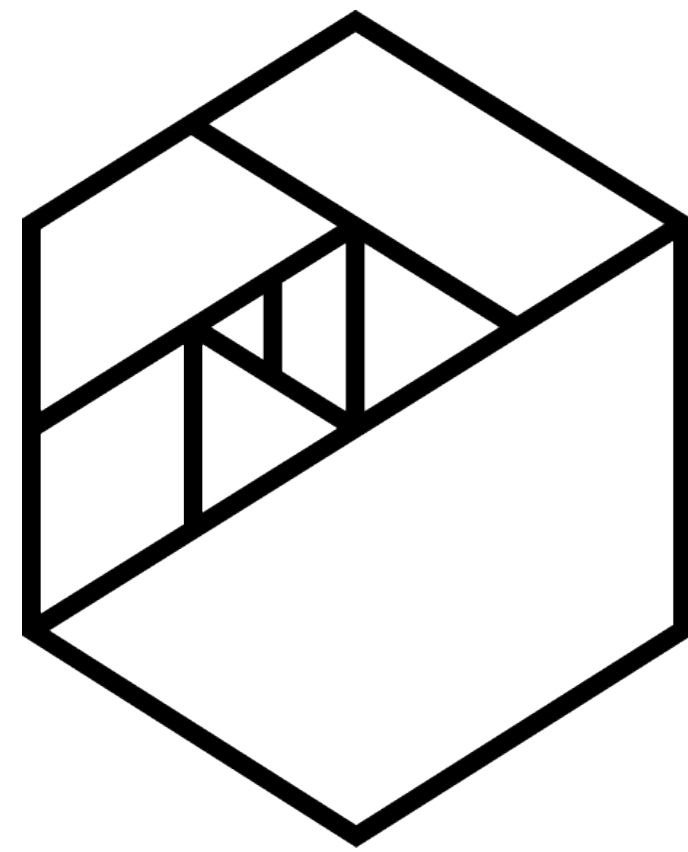
R  
E  
D  
D  
I  
T

R E D D I T

# Thank you!

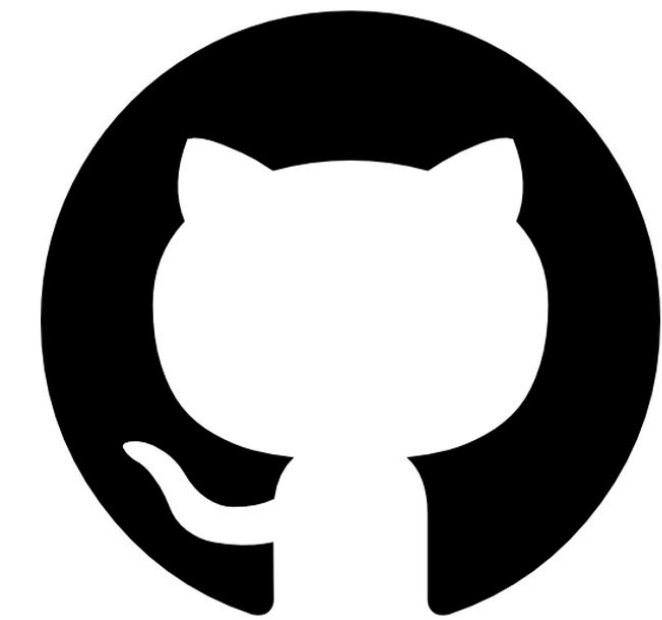


nick-rubin



data scientist

**METIS**



nick-rubin

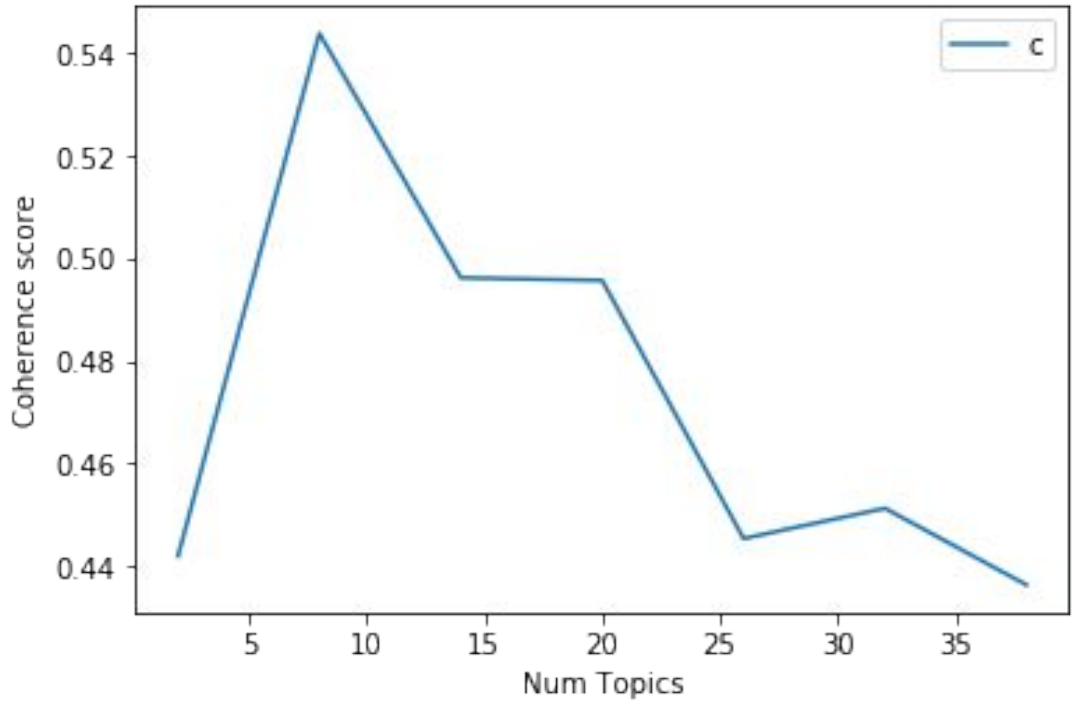
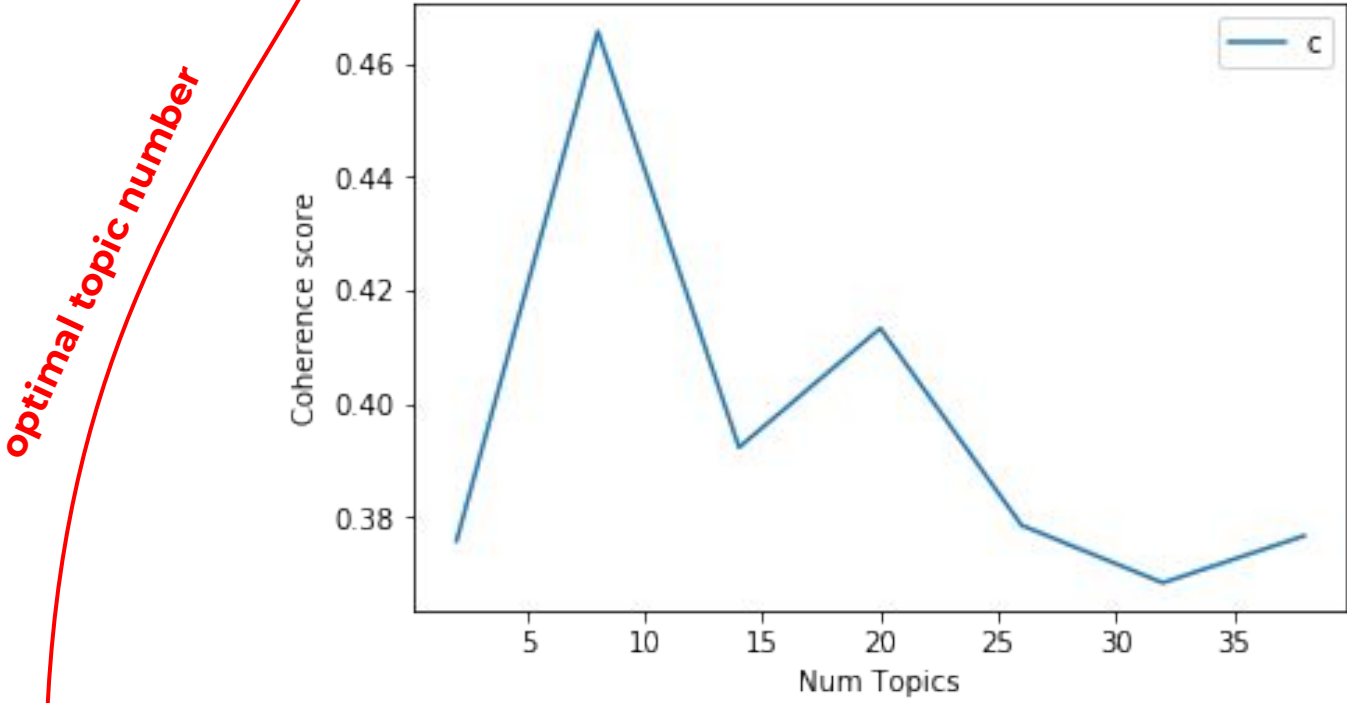


# appendix



```
[0,
[('money', 0.017650909430962734),
('pay', 0.014822088184411091),
('american', 0.013816804600925795),
('job', 0.013255716089213073),
('work', 0.01038013746668537),
('year', 0.010029457146864918),
('make', 0.00848646373965493),
('country', 0.00820591948379857),
('high', 0.007621452284097816),
('give', 0.007270771964277365)]],
(1,
[('time', 0.028463838741831875),
('politic', 0.023784472256063796),
('article', 0.02217853582899546),
('campaign', 0.01849595747037324),
('news', 0.01827444899767416),
('post', 0.016834643925130136),
('woman', 0.015976298593421197),
('man', 0.013844279543692547),
('watch', 0.012653671502934987),
('make', 0.012598294384760218)]],
(2,
[('debate', 0.030536281147287694),
('make', 0.02435598505492036),
('give', 0.017023906508975475),
('start', 0.016602522684495885),
('lie', 0.015647386015675478),
('good', 0.015619293760710173),
('agree', 0.012051577380116302),
('put', 0.01163019355563671),
('literally', 0.011433547770879569),
('question', 0.01106834845633059)]],
(3,
[('medium', 0.019873334789255296),
('time', 0.01733457086700153),
('president', 0.016570211836645556),
('bill', 0.012393535706486132),
('year', 0.012175147412098712),
('long', 0.011547281065734877),
('attack', 0.011083205940161607),
('find', 0.01094671325616947),
('show', 0.009499890805852807),
('big', 0.009090412753876393)]],
(4,
[('server', 0.021101552074117407),
('gt', 0.020061994144997417),
('fbi', 0.013905631134837265),
('report', 0.012226623041157224),
('secretary', 0.011473221973480283),
('state', 0.009729636645427931),
('state department', 0.009514379197520234),
('private', 0.009471327707938695),
('information', 0.008868606853797142)]],
(5,
[('people', 0.07521246843520052),
('thing', 0.03432077982025035),
('fuck', 0.02473594178500638),
('shit', 0.018300795568709442),
('lot', 0.01642727198674957),
('good', 0.01509679871840126),
('guy', 0.015015341171359527),
('bad', 0.01287029243259388),
('issue', 0.012544462244426947),
('talk', 0.010480871052703032)]],
(6,
[('party', 0.03789316159375752),
('candidate', 0.03570187862430186),
('win', 0.030651238609337002),
('republican', 0.02364982229229577),
('democrat', 0.018786243018625904),
('supporter', 0.018625905240373053),
('lose', 0.01843884449907806),
('dnc', 0.017931108201277357),
('support', 0.017556986718687367),
('democratic', 0.017396648940434516)]],
(7,
[('vote', 0.07182847854691814),
('state', 0.023607420212079774),
('supporter', 0.022498000464408266),
('delegate', 0.017079904022291596),
('convention', 0.014654660852963183),
('change', 0.013725844320028897),
('election', 0.013055032379576356),
('people', 0.01246162181686834),
('voter', 0.012255218142882943),
('rule', 0.011945612631904847)]])
Coherence Score: 0.5327257673598975
```

	r/politics	r/The_Donald
Coherence	0.53	0.56
Perplexity	-8.34	-8.58



```
[0,
[('government', 0.010790284983833371),
('gt', 0.009699977441912927),
('family', 0.008985638017896083),
('support', 0.008684863523573201),
('read', 0.00823370178208888),
('law', 0.00819610497029852),
('idea', 0.008045717723137079),
('kind', 0.007820136852394917),
('person', 0.0076321527934431165),
('find', 0.006955410181216633)]],
(1,
[('pay', 0.01668305969584626),
('money', 0.0160021184837709),
('job', 0.014110615116894909),
('work', 0.013278353635469472),
('year', 0.011954301278656276),
('state', 0.010592418854505561),
('part', 0.010024967844442763),
('people', 0.009987137777105243),
('change', 0.009608837103730044),
('news', 0.009268366497692367)]],
(2,
[('shit', 0.03478014414122601),
('thing', 0.021418738359381326),
('good', 0.01890841363673172),
('day', 0.018179609684994737),
('show', 0.017653251275406914),
('real', 0.01587173050449429),
('life', 0.015061948335897643),
('watch', 0.01449510081787999),
('talk', 0.0141306988420115),
('people', 0.08011670543705397),
('woman', 0.029610061901194654),
('man', 0.027875251350392302),
('call', 0.02771754130031936),
('white', 0.02318337736072231),
('black', 0.02097543665970114),
('liberal', 0.02046287899696408),
('medium', 0.019437763671489965),
('racist', 0.018294365808461145),
('hear', 0.009856878129558806)]],
(4,
[('make', 0.07110703567855951),
('guy', 0.0333027675891964),
('time', 0.027342447482494165),
('great', 0.026258752917639214),
('feel', 0.015796932310770256),
('thing', 0.013504501500500166),
('wall', 0.012837612537512505),
('america', 0.011670556852284094),
('give', 0.011212070690230077),
('bad', 0.01108702900966989)]],
(5,
[('country', 0.026102885055716684),
('muslim', 0.02572126392917112),
('gt', 0.01698214013127767),
('leave', 0.016142573652877423),
('world', 0.015570141963059074),
('islam', 0.012593497176003663),
('american', 0.011372309571057854),
('attack', 0.009807662952221035),
('rape', 0.00976950083956648),
('child', 0.008929934361166235)]],
(6,
[('fuck', 0.06823266219239374),
('love', 0.029165630955340128),
('good', 0.025271356367553236),
('post', 0.021667080951197282),
('cuck', 0.01918137376750352),
('edit', 0.015577098351147569),
('literally', 0.014789957742977877),
('sjw', 0.013422818791946308),
('sli mg', 0.012345679012345678),
('mod', 0.011807109122545364)]],
(7,
[('vote', 0.029609475032010245),
('win', 0.022647247119078104),
('president', 0.01828585147247119),
('supporter', 0.017765685019206147),
('run', 0.015564980793854034),
('campaign', 0.013644366197183098),
('lose', 0.013644366197183098),
('debate', 0.013204225352112676),
('candidate', 0.01088348271446863),
('hope', 0.01052336747759283)]])
```



# references

<https://qz.com/1083444/analysis-of-500-million-reddit-comments-shows-how-the-alt-right-made-the-alt-left-a-thing/>

<https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>

<https://www.cnn.com/2019/06/26/tech/the-donald-subreddit-quarantined/index.html>

<https://qz.com/1056319/what-is-the-alt-right-a-linguistic-data-analysis-of-3-billion-reddit-comments-shows-a-disparate-group-that-is-quickly-uniting/>

<https://arxiv.org/abs/1809.01644>

