

Reddit Text Analysis

Saeed Beztchi, Carissa Bleker, Ashley Cliff, Sean Whalen

Motivation

- Reddit is the third most popular social website in US, 6th most popular worldwide
- Provides numerous subreddits dedicated to many unique topics
- Each subreddit may have thousands of users posting submissions and commenting which differ in opinion

Motivation

- Impacts how information is shared, phrases are created, opinions on issues are formed and adopted
- Reddit provided a simple social media platform to build our tools - which can be applied to other platforms

Aim

1. Identify and extract themes within each subreddit
2. Link users to themes
3. Identify controversial themes occurring during a particular event

Data

Data Collection & Analysis

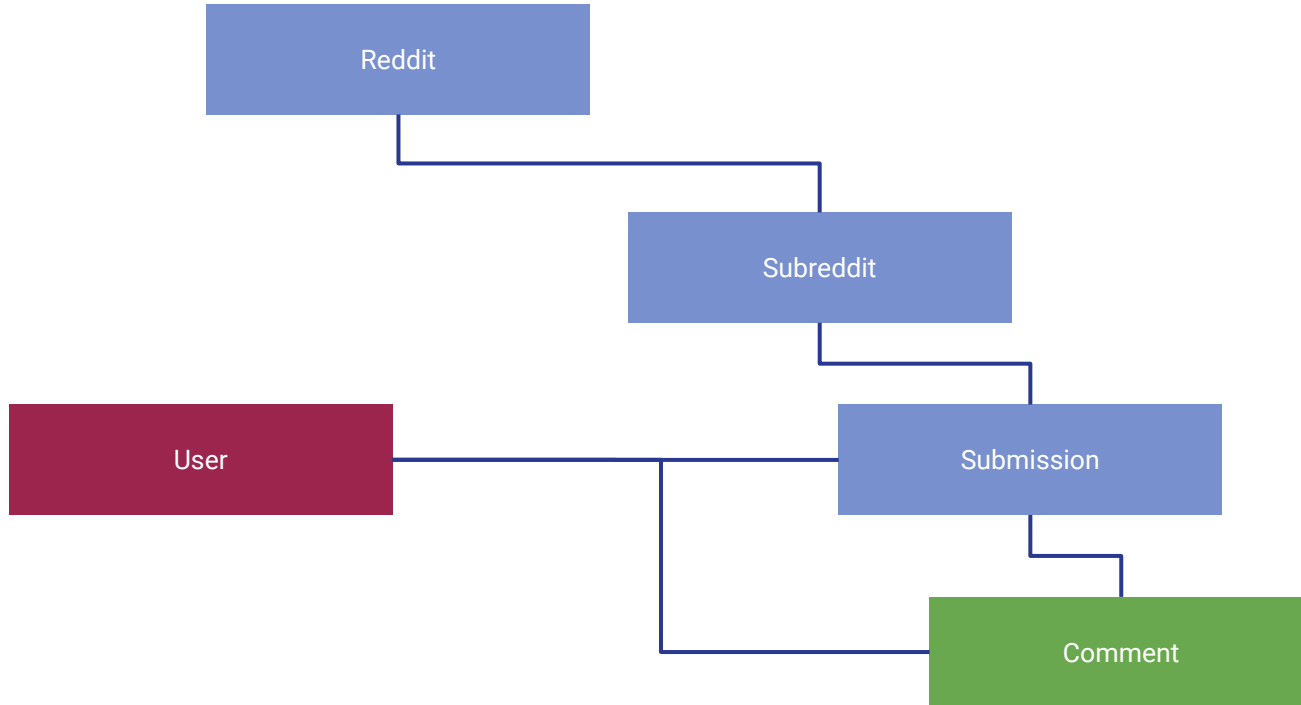
Tools:

- Python
- R
- PRAW
- Google BigQuery

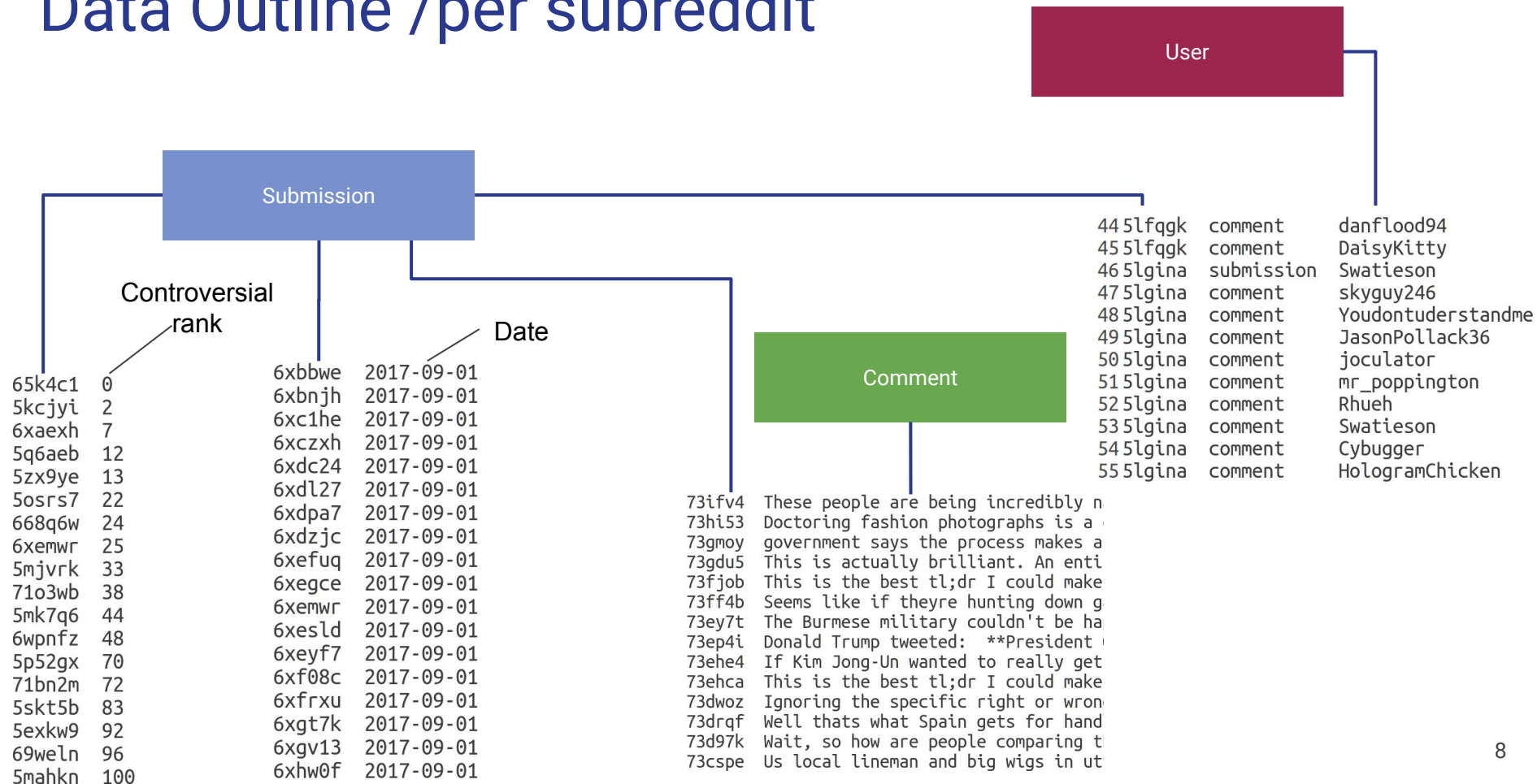
Subreddits:

- World News
<https://www.reddit.com/r/worldnews/>
 - 1 year (Oct 2016 to Sept 2017)
- NBA <https://www.reddit.com/r/nba/>
 - 6 month (January-June 2017)

Data Outline /reddit



Data Outline /per subreddit



Analysis

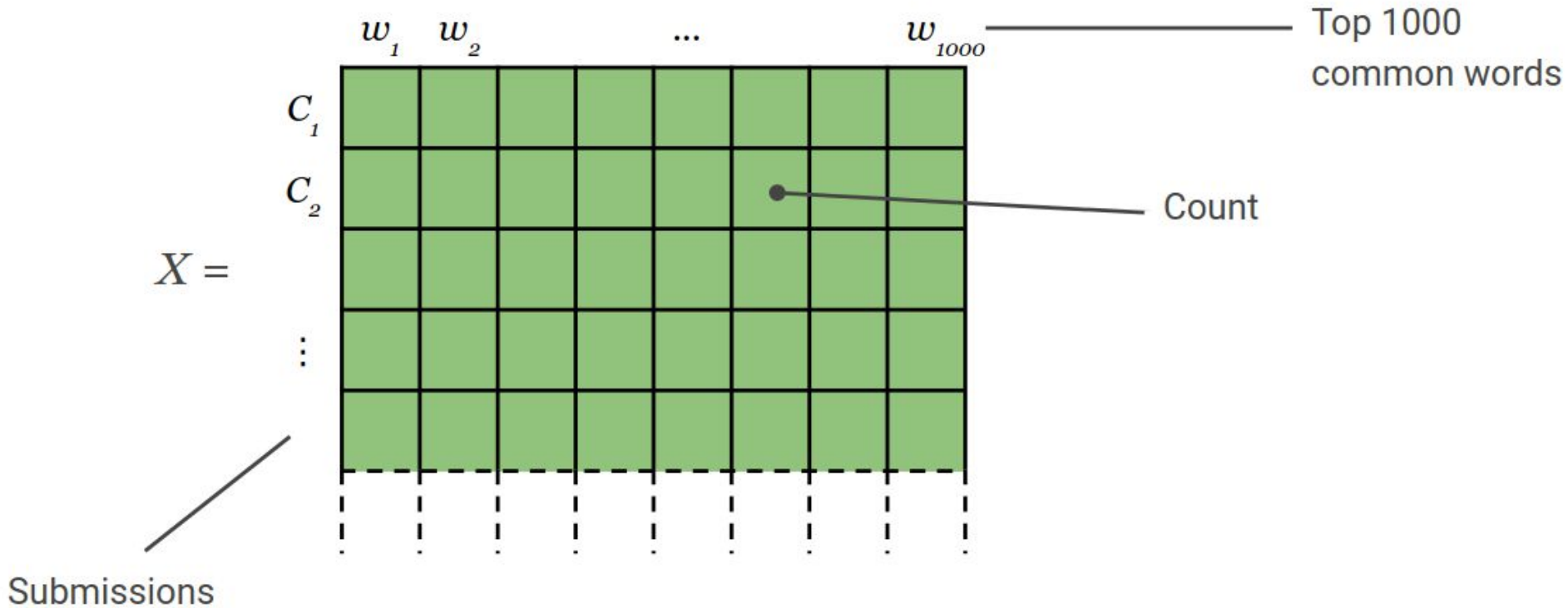
Non-Negative Matrix Factorization (NMF)

- Similar to dimensionality reduction methods such as PCA, SVD
- In simplest terms, for a given X find W and H to minimise:

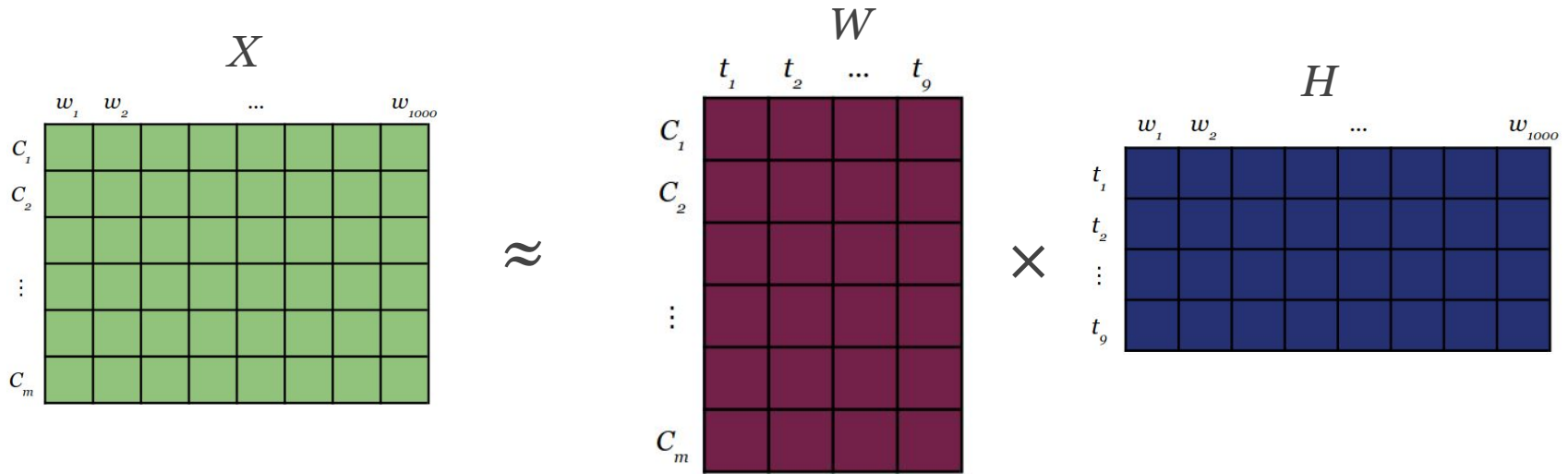
$$|| X - WH ||_F$$

- Then $X \approx WH$
- Matrices are all non-negative
 - Inherent to data
 - Easier to interpret

Non-Negative Matrix Factorization (NMF)



Non-Negative Matrix Factorization (NMF)



/r/worldnews year

	Name	Words
0	Reddit	money work government life pay drug canada year drugs used food company problem day reddit live person fuck companies police
1	North Korea	nk korea north nuclear kim korean war nukes missile south china missiles japan military weapons strike regime test attack bomb
2	Russia	russia russian putin ukraine nato russians war sanctions propaganda military soviet europe syria government evidence nuclear countries media intelligence poland
3	Israeli–Palestinian conflict	israel palestinians israeli jews palestinian palestine jewish arab land peace state bank iran war west territory east civilians anti conflict
4	US politics	trump president obama election hillary clinton vote party republicans donald investigation republican campaign american media america democrats politics administration political
5	Climate change	climate energy solar coal change ice nuclear water earth global oil planet science power gas environment fuel humans space sea
6	EU/Brexit	eu brexit uk vote referendum germany europe voted party britain leave labour british france european countries trade government turkey immigration
7	Asia	china chinese india japan trade countries government war america military pakistan japanese indian economy american asia democracy south economic western
8	Syria Conflict	isis iran turkey saudi assad syria war iraq arabia erdogan syrian rebels military civilians al weapons turkish terrorists terrorist oil
9	Religion	muslims muslim islam religion women religious islamic church christians christian gay terrorism removed terrorist law jews men refugees police god

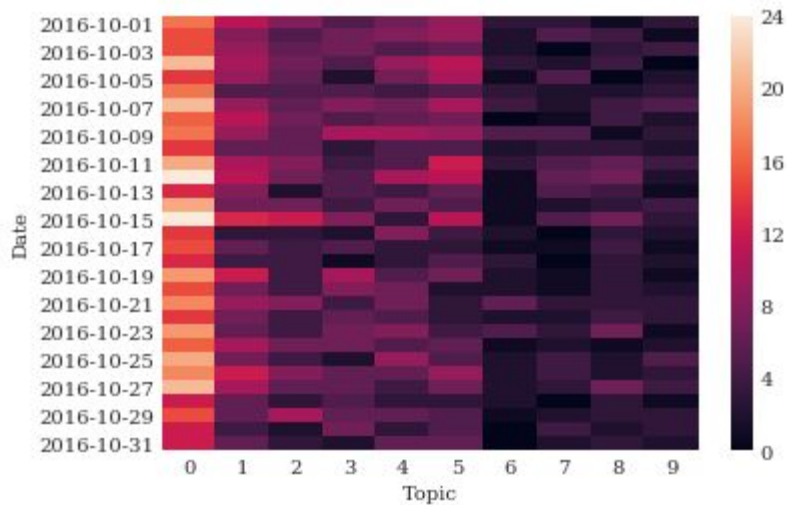
/r/worldnews October 2016

	Name	Words
0	Reddit	government money work reddit life year countries news new problem power used fuck saying live org pay doing human big
6	Philippines Drug War (?)	duterte philippines drug drugs china president police trump guy aid hitler military murder corruption obama corrupt killed elected killing voted
1	Russia	russia russian war nato putin ukraine russians syria nuclear military crimea assad nukes missile missiles soviet cold countries europe ussr
7	Israeli–Palestinian conflict	israel palestinians jews palestinian israeli jewish land arab peace war civilians state iran bank terrorist terrorism civilian british terrorists site
8	US politics	wikileaks assange leaks trump hillary clinton dnc emails election information russia russian internet media putin government evidence news campaign conspiracy
9	India-Pakistan Border (?)	pakistan india indian nuclear army removed terrorists nukes war military terrorist attack afghanistan china terrorism border muslims weapons energy christians
3	EU/Brexit	eu brexit uk referendum vote leave scotland voted trade britain british market europe immigration currency remain european leaving countries movement
2	Asia	china chinese japan korea japanese korean south military vietnam north america philippines war sea air india communist asia government space
5	Syria Conflict	isis syria assad iraq war turkey rebels saudi civilians syrian al iran terrorists erdogan arabia army turkish terrorist yemen military
4	Religion	muslims religion muslim islam religious women marriage christian christians church islamic refugees child ₁₄ rape married iran law gay sex children

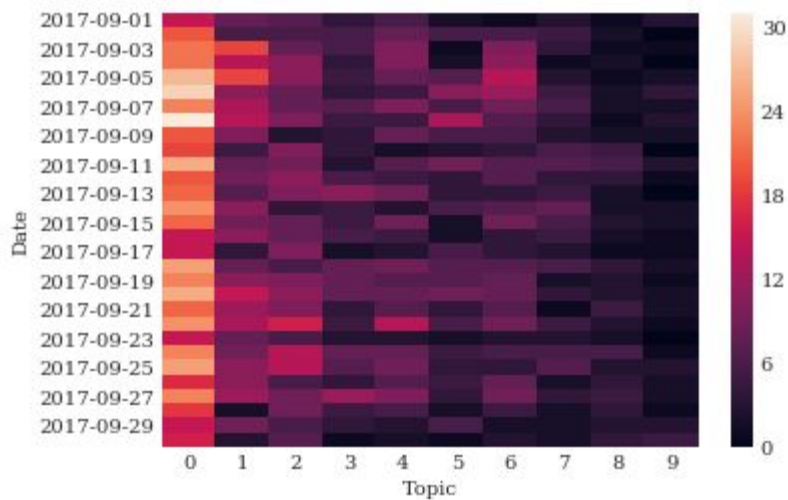
/r/worldnews September 2017

	Name	Words
0	Reddit	government money work country life canada new bad article used problem live look year day news reddit countries read person
1	North Korea	nk korea north nuclear kim war nukes korean china missile japan weapons sk south military missiles nuke sanctions russia seoul
4	Russia	russia russian putin ukraine russians nato facebook propaganda war ads election military media soviet news syria europe countries evidence influence
3	Iraqi Kurdistan-Israel	israel kurds turkey iran kurdish iraq isis saudi syria jews war independence arabia country state region military east land weapons
2	US politics	trump president election hillary obama clinton vote campaign donald news ads party evidence political twitter elected congress politics fake media
5	Climate change	climate hurricanes hurricane change warming energy water global solar science earth ocean sea planet humans data hit fuel natural increase
8	Catalonia independence	catalonia spain referendum catalan spanish independence vote eu constitution government democracy country brexit region independent uk majority law voting civil
6	Asia	china chinese nk trade government india cars country russia democracy economy bitcoin sanctions independence korea apple economic business companies countries
7	Rohingya Crisis	rohingya myanmar muslims muslim genocide ethnic india military islam religion refugees pakistan country islamic violence religious government indian peace isis
9	Puerto Rico	puerto rico pr hurricane island ships water act aid power help federal ship hurricanes infrastructure solar debt days military trump

October 2016



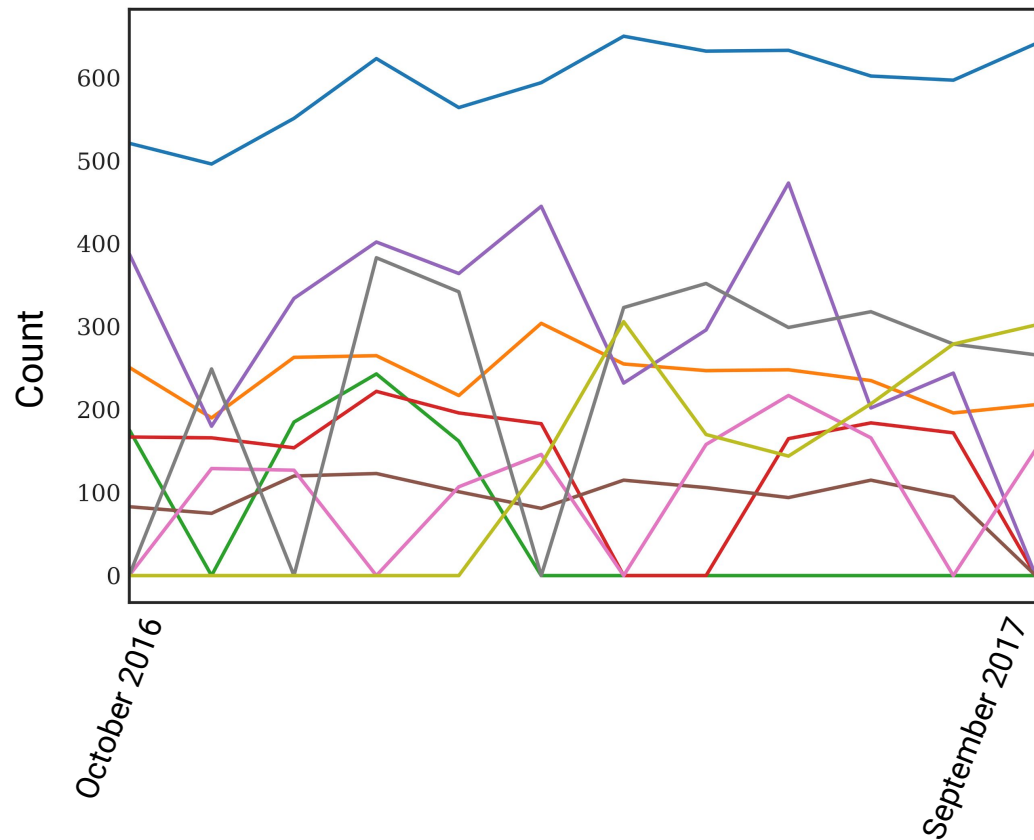
September 2017



Topics over year

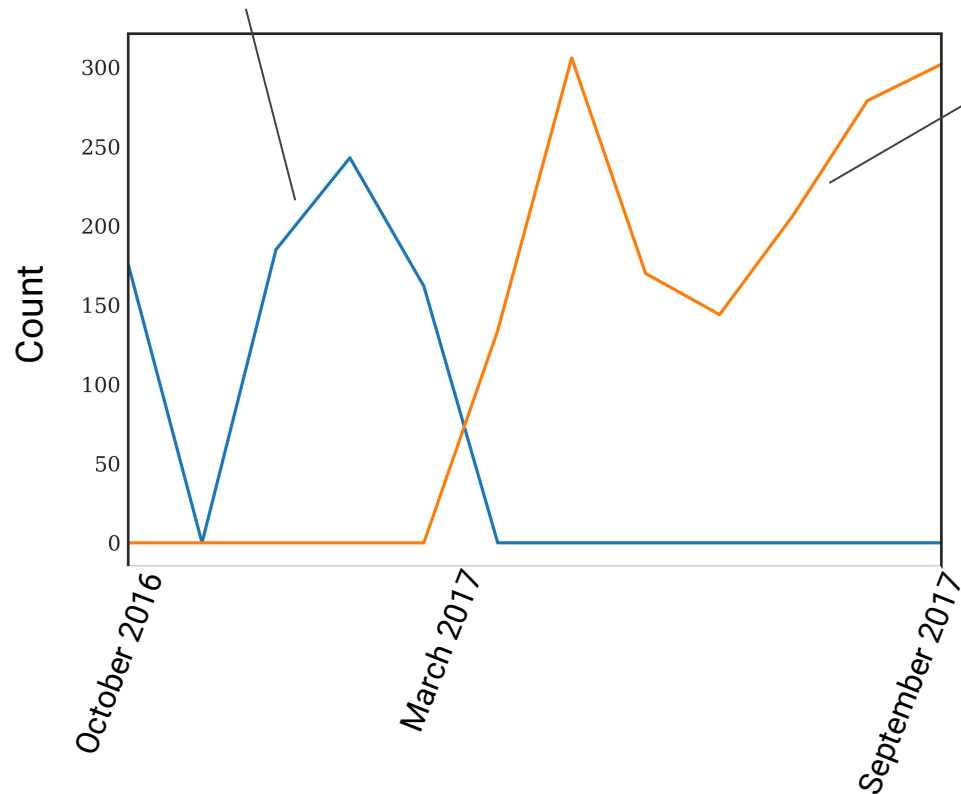
- Similar topics appear in different months
- Use word overlap between all topics in different months to determine which may be the same
- Aggregate across the year

Topics over year



Topics over year

china chinese taiwan war trade korea nuclear

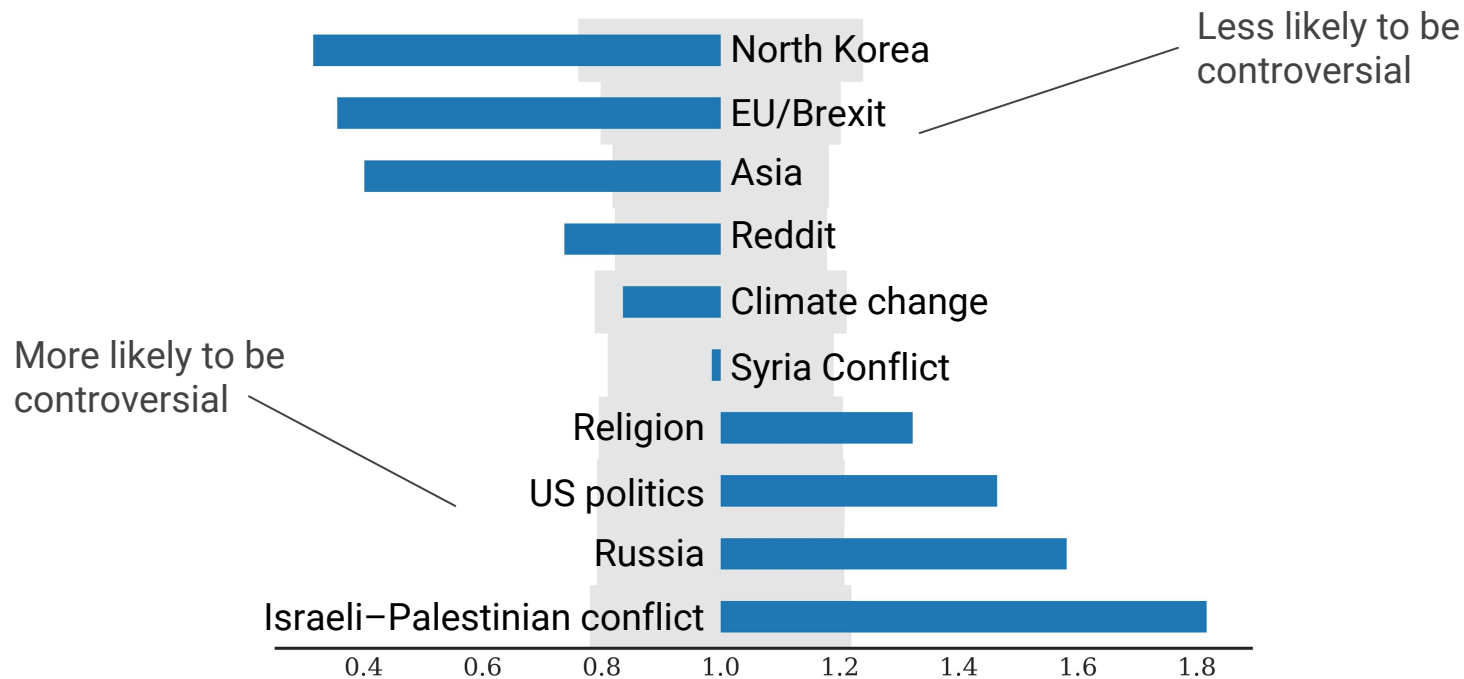


nk korea north korean kim nuclear

Controversial topics in /r/worldnews

- Reddit ranks submissions as ***controversial*** based on user interactions
- Pulled the highest ranked submissions for the year
- Compared the contributions of a topic in controversial to the background
- Background topic contributions found by random sampling all submissions for the year (10,000 iterations)

Controversial topics in /r/worldnews

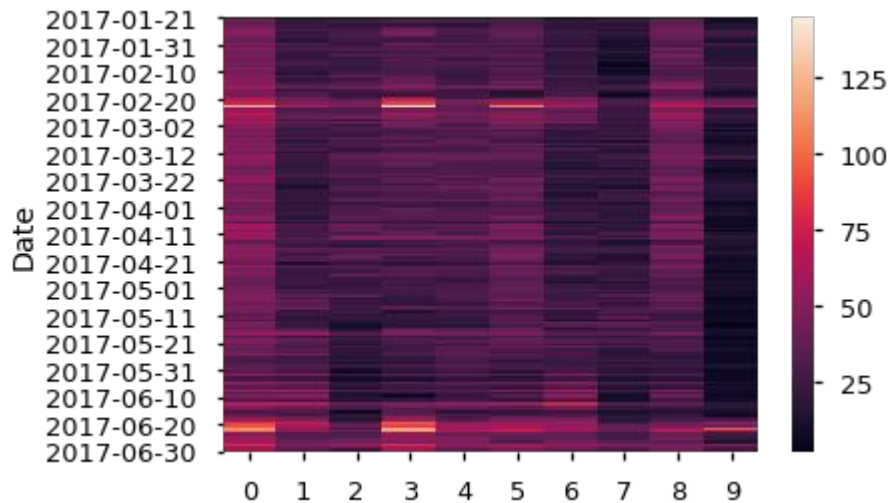


Topic best guess		NMF generated key words				
#0:	n/a	com	https	fuck	guy	http
#1:	cavs	lebron	cavs	kyrie	finals	cleveland
#2:	mvp race	harden	westbrook	russ	mvp	triple
#3:	trades	trade	pick	draft	picks	lakers
#4:	goats	kobe	shaq	jordan	player	lebron
#5:	playoff series	celtics	cavs	warriors	series	playoffs
#6:	warriors	kd	warriors	curry	steph	durant
#7:	spurs	kawhi	spurs	pop	lma	manu
#8:	refs suck	foul	refs	wall	ball	shot
#9:	knicks	melo	knicks	phil	kp	dolan

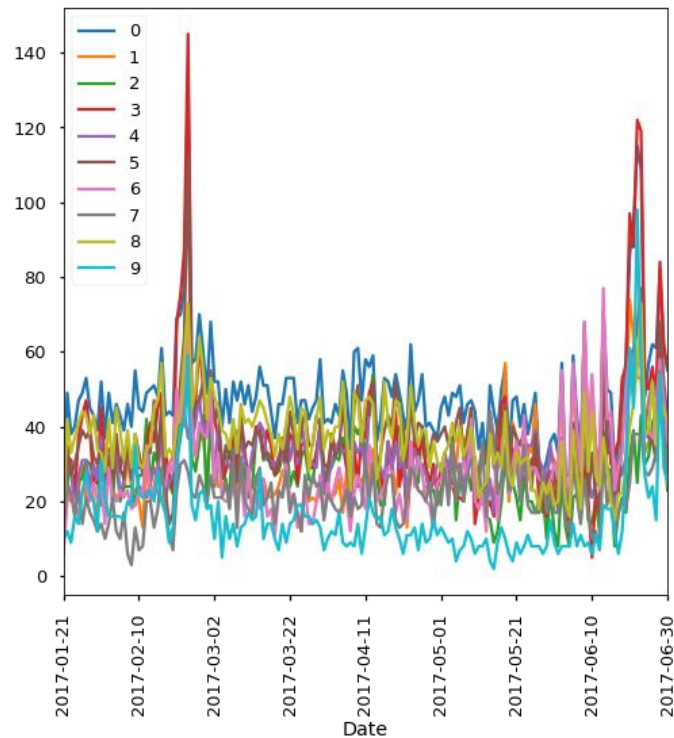
Second half of the 2016-17 season, from 01/2017-06/2017

Active players/coaches/gms highlighted with team colors

r/nba

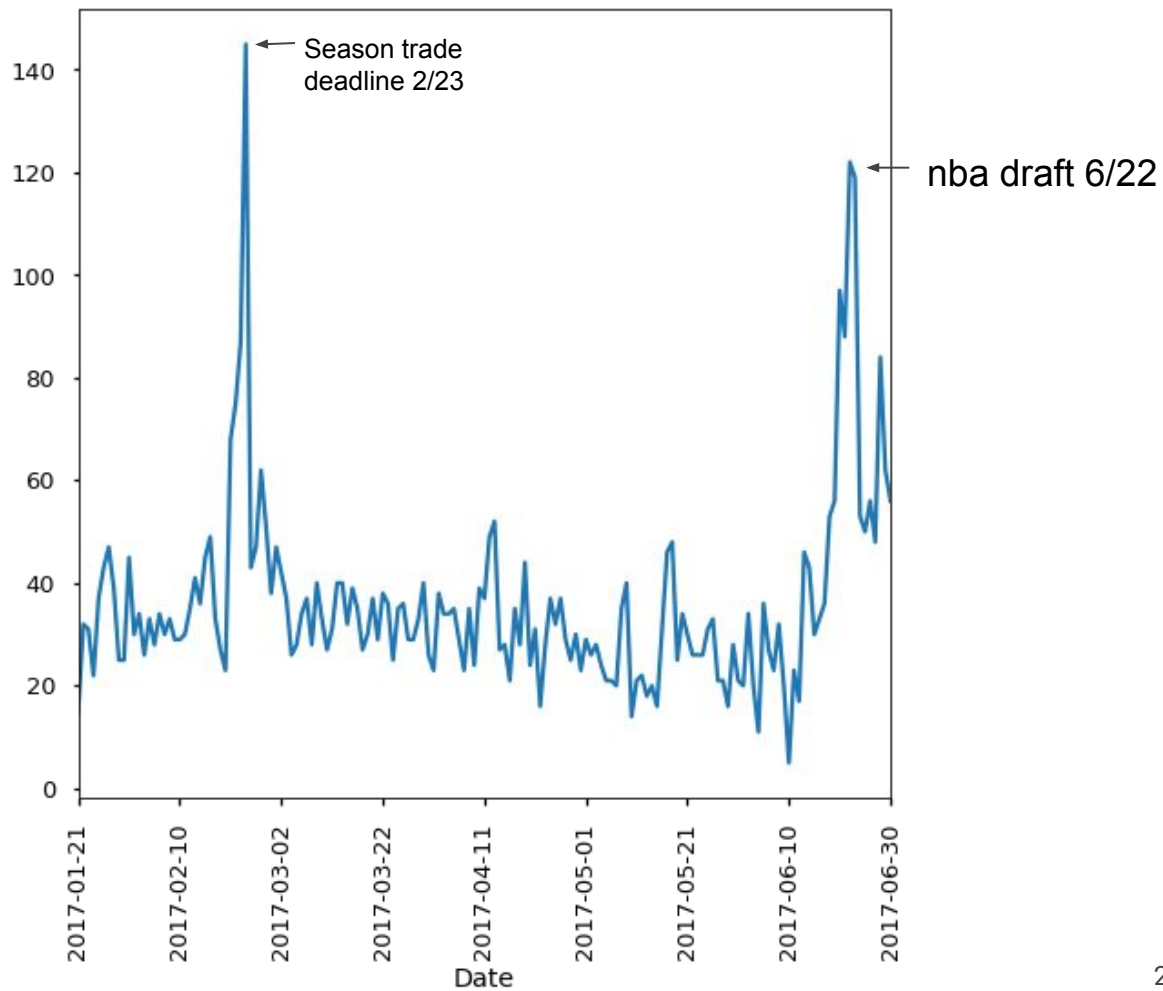


hot all the time: too general



r/nba

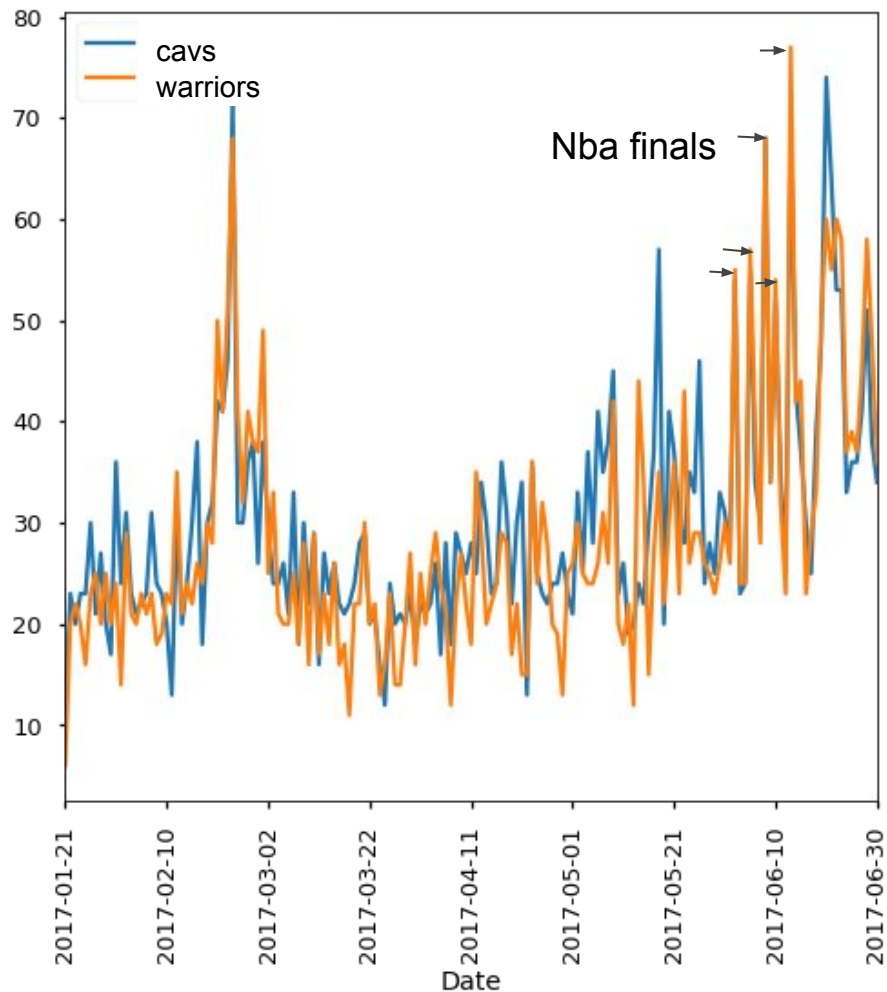
Trades theme vs. time



r/nba

Warriors and Cavaliers
themes vs. time

(the two best teams of the season)














r/nba

What posts had the greatest contribution to reddit's discussion on who the 16-17' MVP should be?

Sorted by correlation with MVP theme, pulled post titles



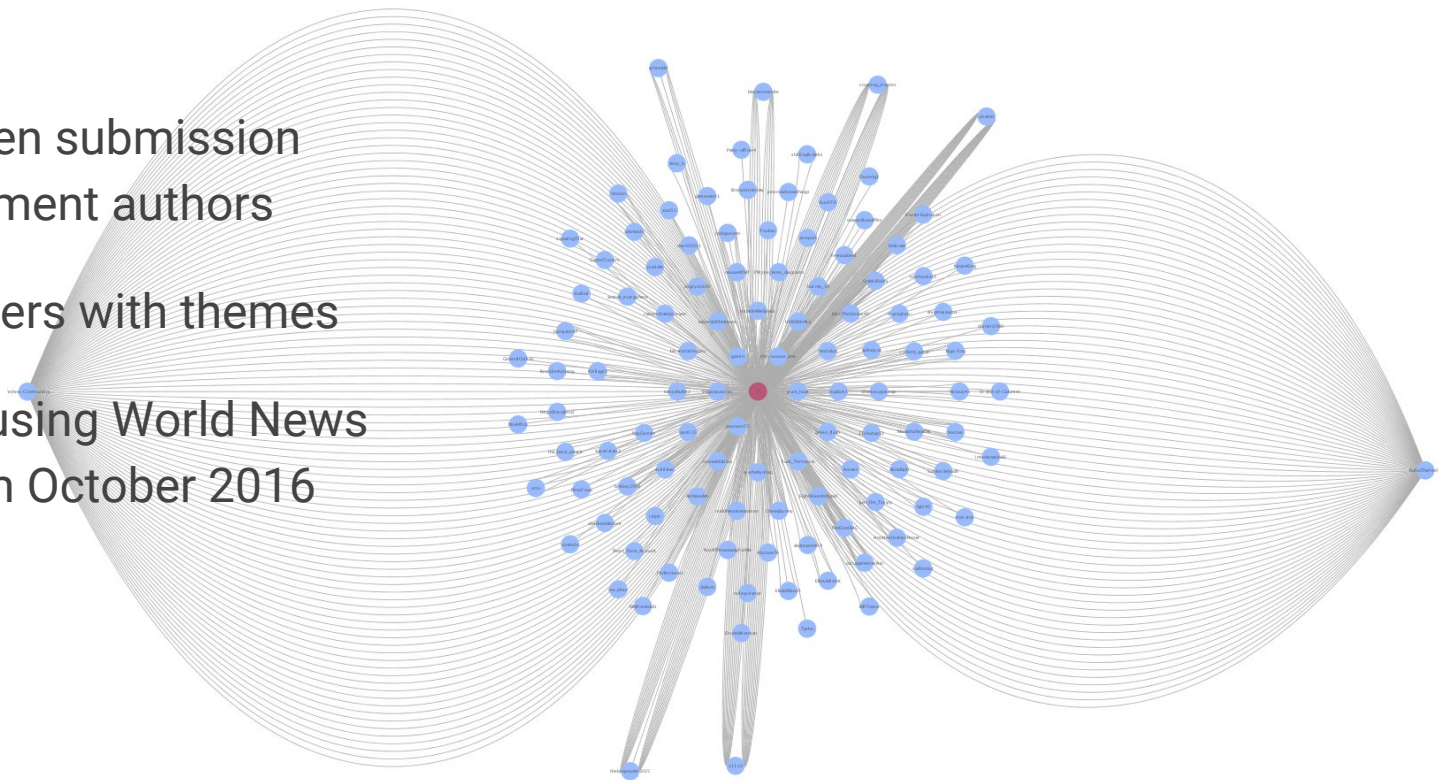
1.  [Historic Westbrook Triple Double game facts](#) (self.nba)
622 submitted 8 months ago * by  Grolgar
2.  [MVP Race: R. Westbrook vs. J. Harden](#) (i.imgur.com)
5988  submitted 7 months ago by  AsPeHeat
2129 comments share save hide givegold report crosspost
3.  [\[Feigen\] Harden scored or had assists on 72 points today. He is averaging 59 points](#)
2892 submitted 7 months ago by  rednation97
596 comments share save hide givegold report crosspost
4.  [\[FiveThirtyEight\] The Case For James Harden, MVP](#) (fivethirtyeight.c
739 submitted 7 months ago by  el_monstruo
618 comments share save hide givegold report crosspost
5.  [\[Kristian Winfield\] James Harden hasn't played like an MVP in crunch time: 36% FG, 27% 3pt](#)
439 (sbnation.com)
submitted 8 months ago by  BearsNecessity
306 comments share save hide givegold report crosspost

User Analysis

Difference between submission authors and comment authors

Association of users with themes

Example graphs using World News submissions from October 2016



Submission Authors and Topics

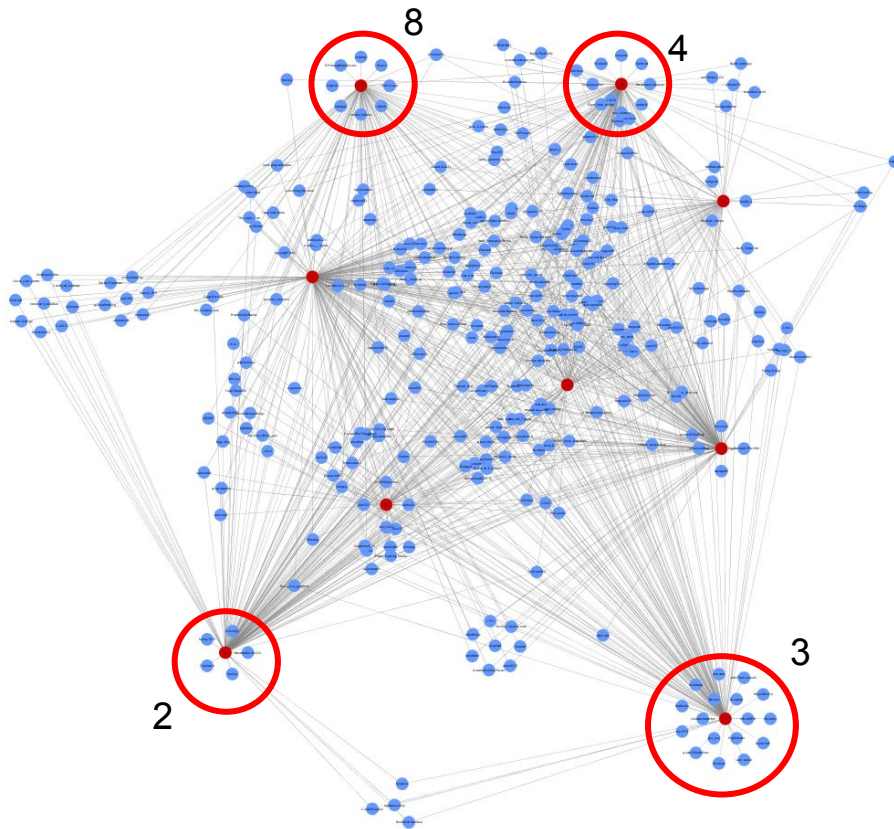
Hypothesis: Users tend to gravitate towards a single topic when posting submissions

2: Asia

3: EU Brexit

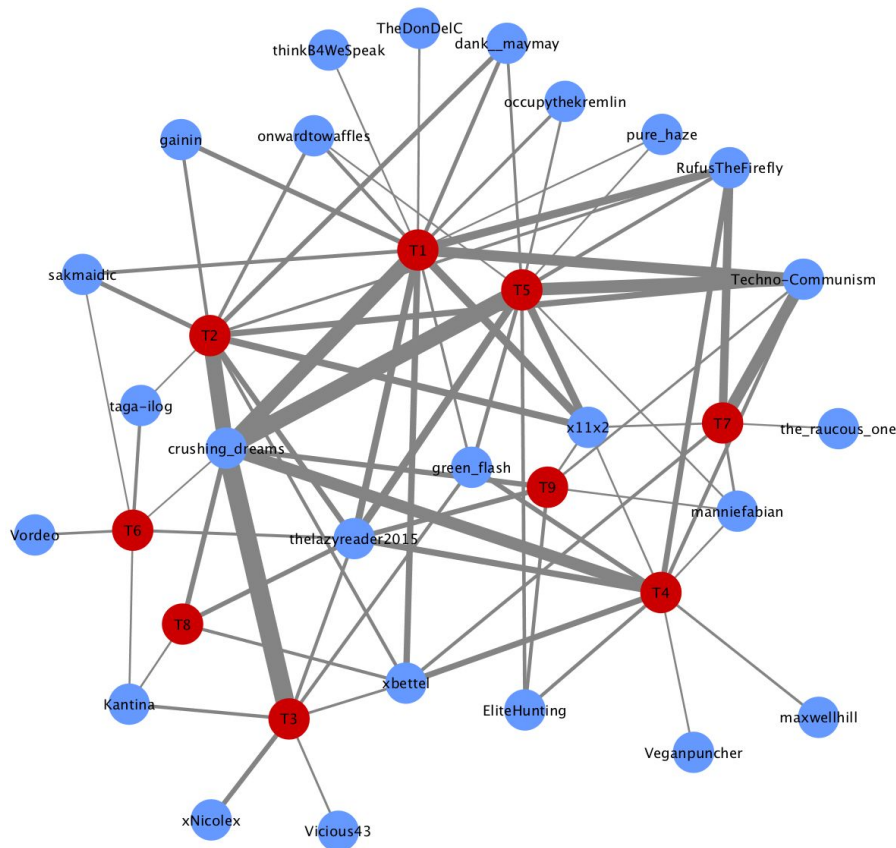
4: Religion

8: US Politics



Submission and Comment Authors and Topics

Edge thickness indicates total submission and comment count for each user to each topic



Conclusion & Future Work

Conclusion & Future Work

Conclusion

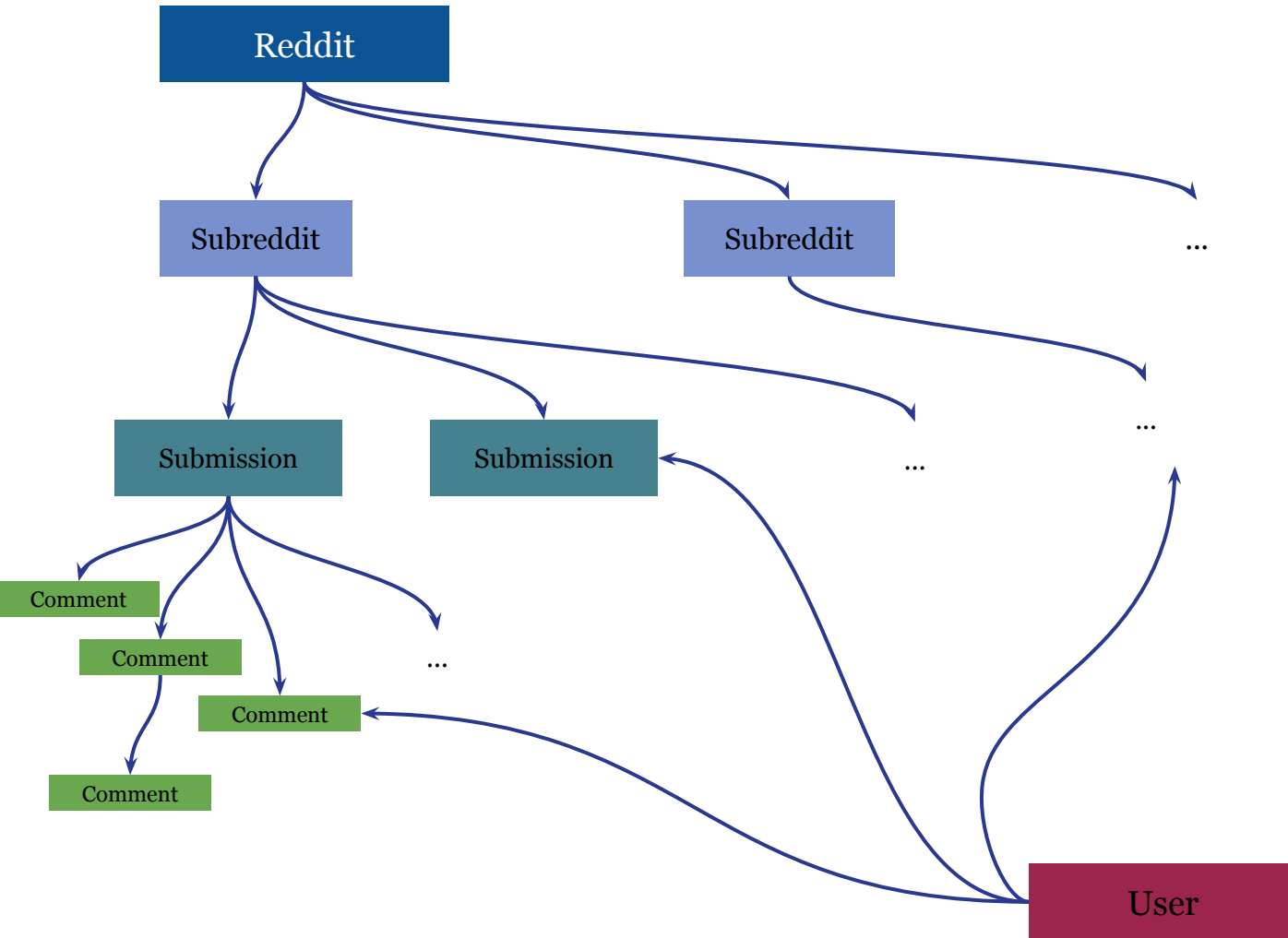
- We can computationally track the rise and fall of **meaningful** themes on reddit
 - Meaning: if you want to maximize post karma, post about themes on the rise!
- Users do not keep to a single topic when commenting and posting
- Controversial topics show which topics divide people

Future Work

- Use authors-to-themes analysis to identify authors with an agenda to push or suppress themes
 - Flag users who spam only one theme as a probable bot
- After finding themes, attempt to identify sides to the issue

Questions?

Backup slides



Non-Negative Matrix Factorization (NMF)

