Battleground State:
Characterizing and Understanding Political Bias in News Discourse on Reddit
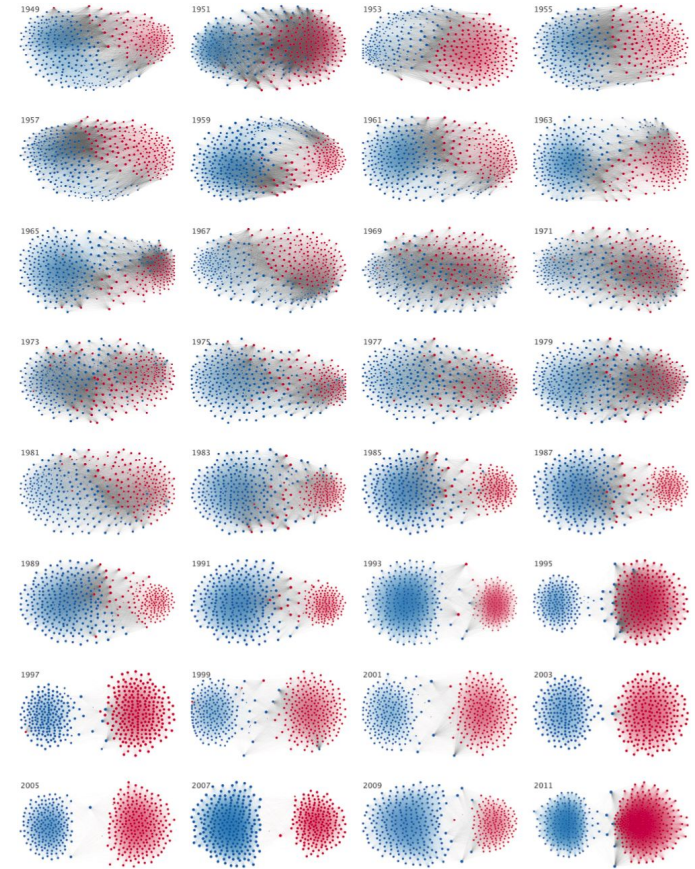
Mansi Ganatra & Samar Haider

12/2/2019

Background

The political landscape in the United States has grown increasingly polarized over the last half century

This trend of hyper-partisan politics has led to less cross-aisle cooperation than ever before

But political strategy in Congress is driven by the pressure on each member from their constituents

Andirs et. al., The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives, PLOS One 2015
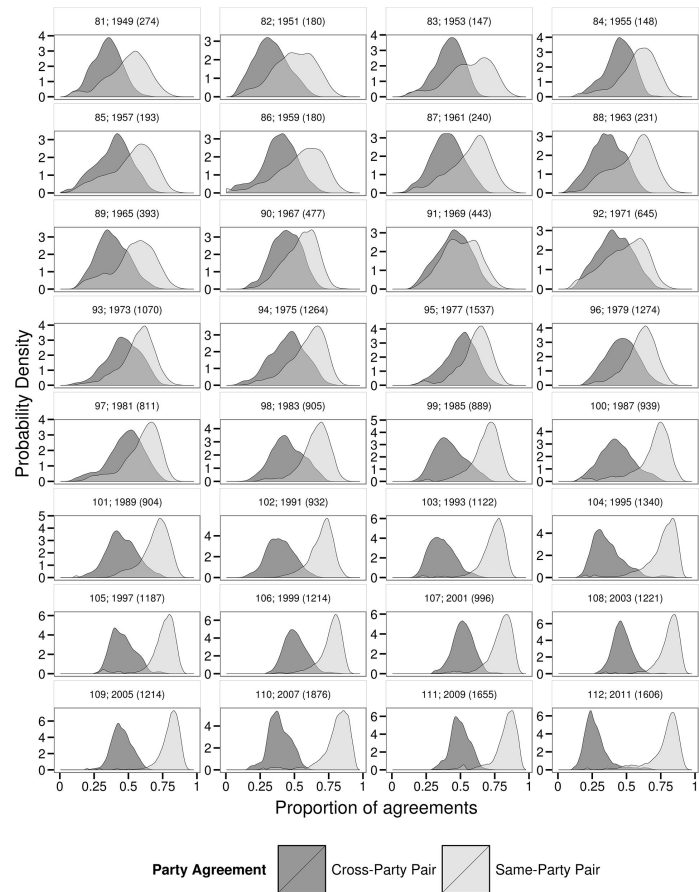
Background

This pressure in turn is driven by the general population's reaction to news and political commentary

The rise of digital media for news dissemination has seen an increase in the formation of echo chambers

Such an imbalanced news diet is not healthy for the governance of a republic

Andirs et. al., The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives, PLOS One 2015

Who will win the presidency?

Chance of winning

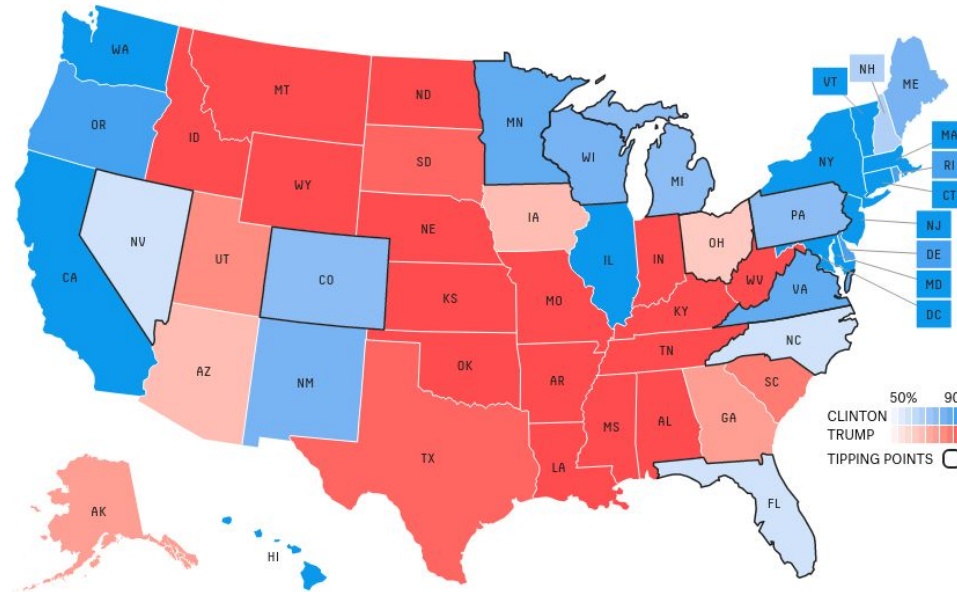**Hillary Clinton**
71.4%

Donald Trump
28.6%

CLINTON 50% 90
TRUMP
TIPPING POINTS ◯

Our focus

In this project we investigate the nature of news readership and political discourse on the Internet

We use activity on Reddit as a lens through which to understand the conversation surrounding the 2016 U.S. Presidential Election, which saw a 'surprise' victory for Donald Trump

We analyze the conversation on both neutral and partisan subreddits during the last three months of 2016 to uncover any latent biases in it

Prior work

Most previous work centered around the U.S. elections has focused on Twitter as a testbed for understanding the online conversation

While more popular, Twitter does not lend itself to the long-form discussions that take place on Reddit, which can often take the form of argumentation and play a role in swaying people's opinions

Work that uses Reddit data focuses on right- and left-wing communities' subreddits, and does not investigate their impact and engagement on neutral ground

reddit.com 6

Research Questions

- Is there a media bias in content shared on news and political discourse subreddits?
- Does the identity of a media outlet reporting a news story impact a post's popularity and engagement levels?
- Do people from either side of the aisle differ in their level of engagement on neutral news and political discourse subreddits?
- Is there a difference in the sentiment of comments mentioning entities from both sides?
- Is there a difference in the semantic representation of political entities in comments from local and global subreddits?

Data

Timeframe:

October, November, December 2016

Subreddits:

news, worldnews, politics

hillaryclinton, The_Donald

democrats, Republican

Liberal, Conservative

Data

Source:

pushshift.io

Size:

Dumps of all submissions and comments made on Reddit for all three
months ~ **160 GB**

Filtered:

1,145,997 submissions

16,234,098 comments

# RQ 1: Media bias in submissions on neutral, news subreddits

We use media bias ratings obtained from AllSides to classify submissions as quoting news from either left-, right-, or center-leaning media outlets

We then analyze the distribution of submissions from each political leaning as well as the level of engagement with it, across the r/news, r/worldnews, and r/politics subreddits


AllSides™
Top Online News Media Bias Ratings
AllSides media bias ratings are based on multi-partisan, scientific analysis.
All ratings are based on online content only — not TV, print, or radio content.
Ratings do not reflect accuracy or credibility; they speak to perspective only.
Visit AllSides.com to view over 600 media bias ratings.

L	L	C	R	R
ALTERNET	abc	AP		THE AMERICAN SPECTATOR
CNN opinion	The Atlantic	BBC		BREITBART B
NOW!	BuzzFeedNEWS	Bloomberg	FOX NEWS online news only	THE BLAZE
THE DAILY BEAST	CBS	The CHRISTIAN SCIENCE MONITOR		CBN
THE HUFFINGTON POST	CNN online news only	npr online news only	reason	THE DAILY CALLER
The Intercept_	The Economist	REUTERS	THE WALL STREET JOURNAL opinion	Daily Mail
JACOBIN	theguardian		Washington Examiner	DAILY WIRE
MotherJones	NBC	THE HILL	The Washington Times	FOX NEWS opinion
msnbc	The New York Times online news only	USA TODAY		the FEDERALIST
THE NEW YORKER	npr opinion	THE WALL STREET JOURNAL online news only		NATIONAL REVIEW
The New York Times opinion	POLITICO			NEW YORK POST
The Nation	TIME			NEWSMAX
SLATE Vox	The Washington Post			

L LEFT L LEAN LEFT C CENTER R LEAN RIGHT R RIGHT

Version 1.1 | AllSides 2019

Media bias: number of submissions by leaning

Left-leaning websites are used as sources more often than either center- or right-leaning ones in U.S.-centric subreddits. However, r/worldnews is more centrist.
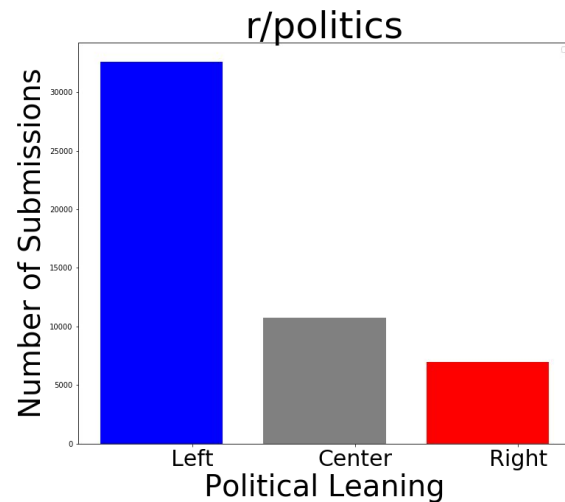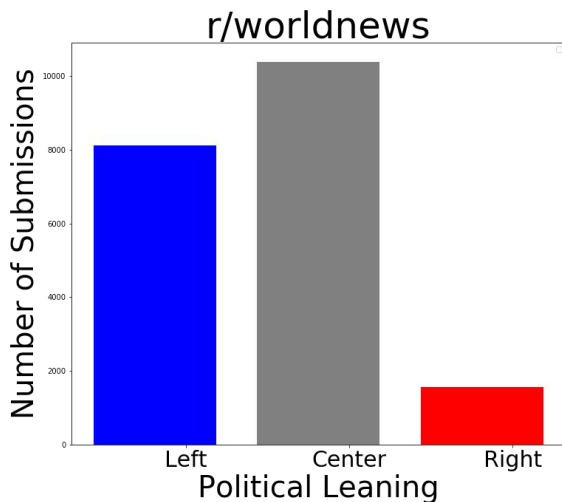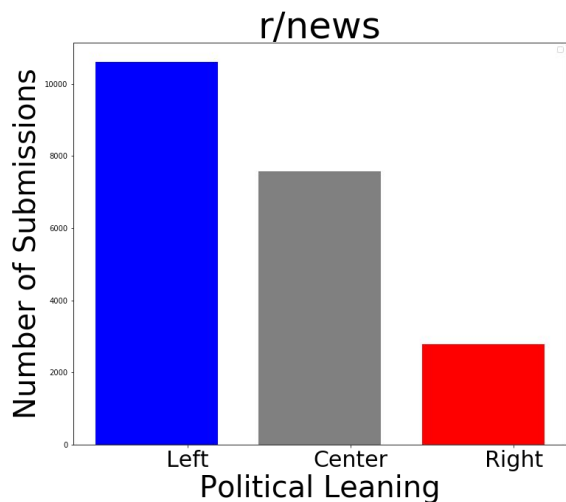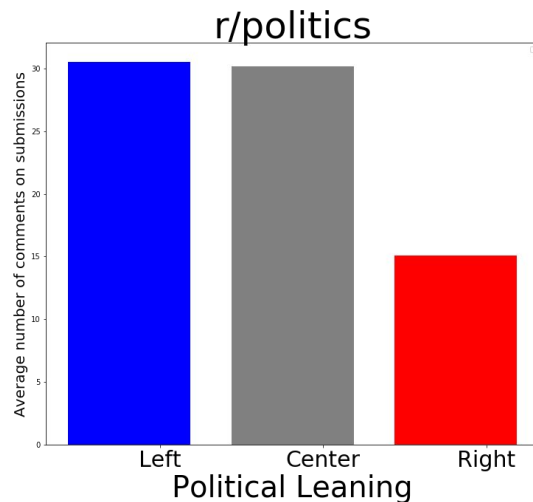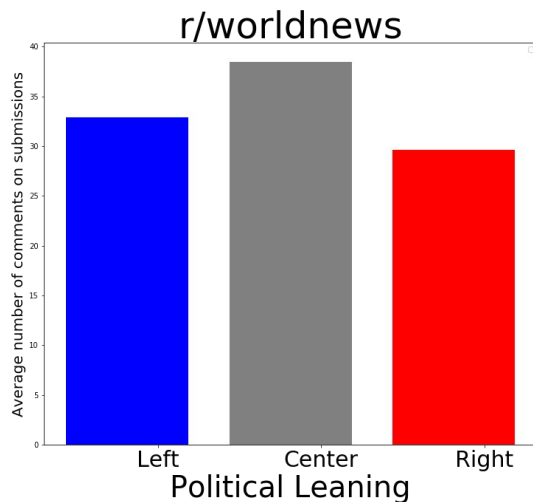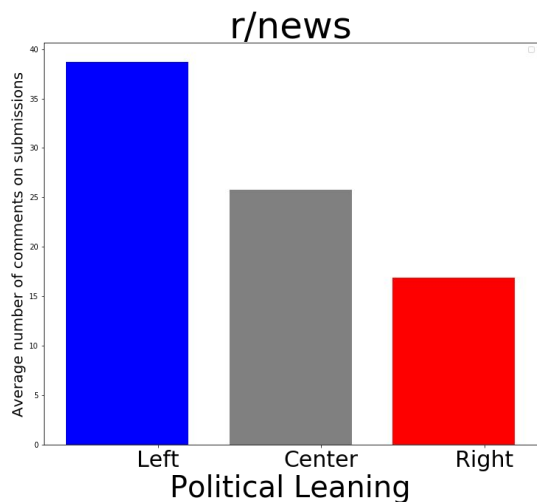
Media bias: popularity of submissions by leaning

Left-leaning websites receive more upvotes than others in r/news but fare identically to center-leaning ones in r/politics. r/worldnews remains centrist.
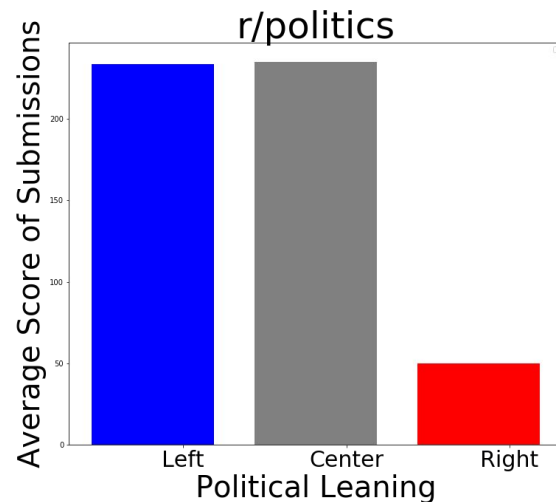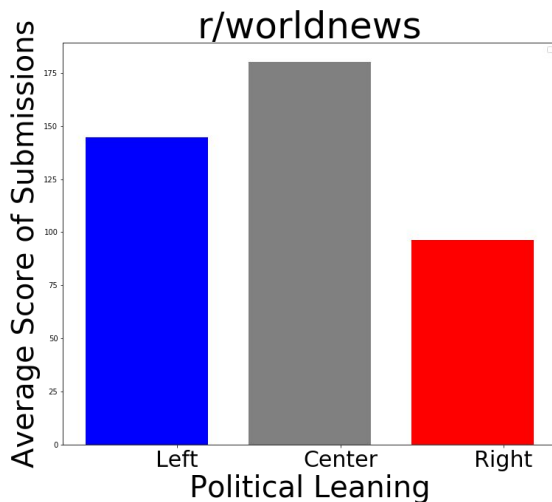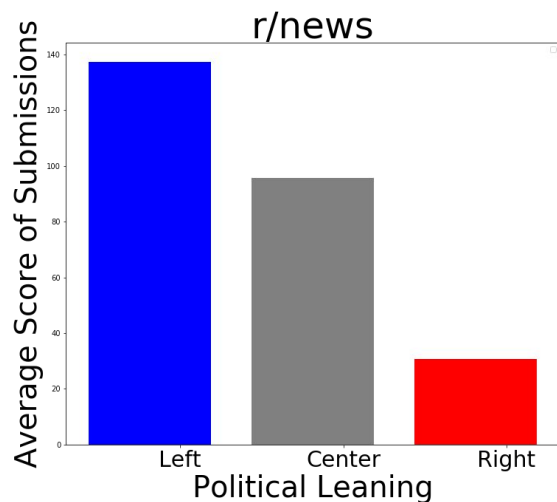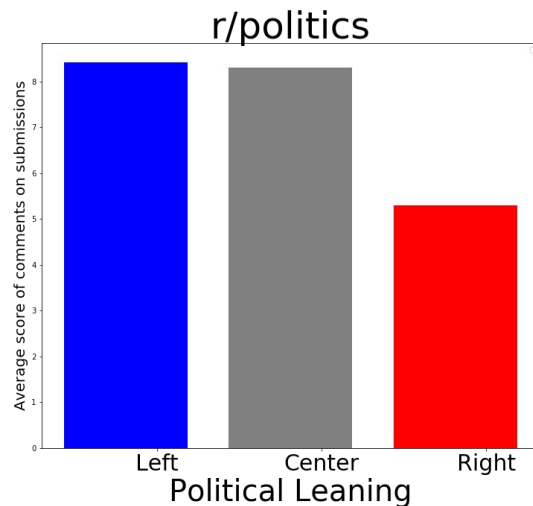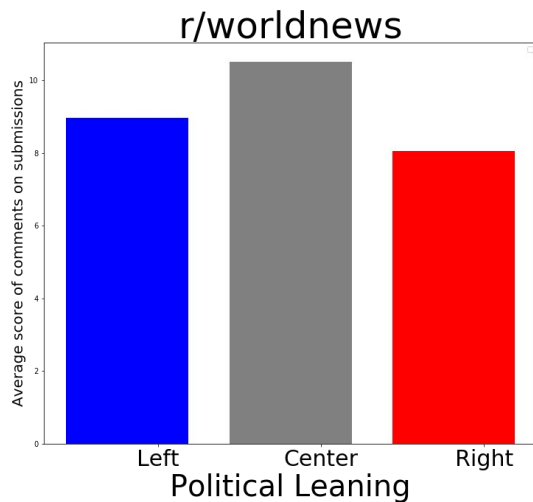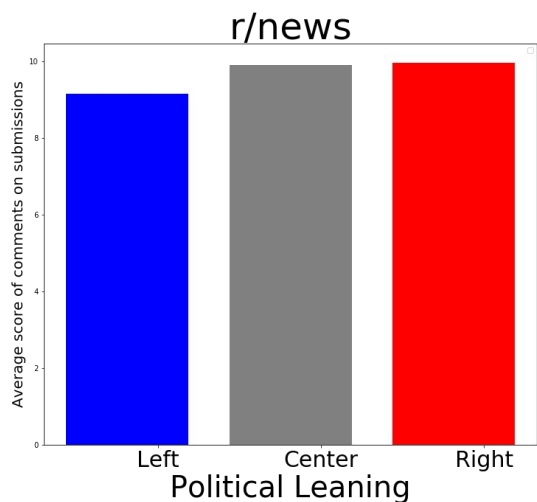
Media bias: engagement on submissions by leaning

Average number of comments on submissions.

Media bias: engagement on submissions by leaning
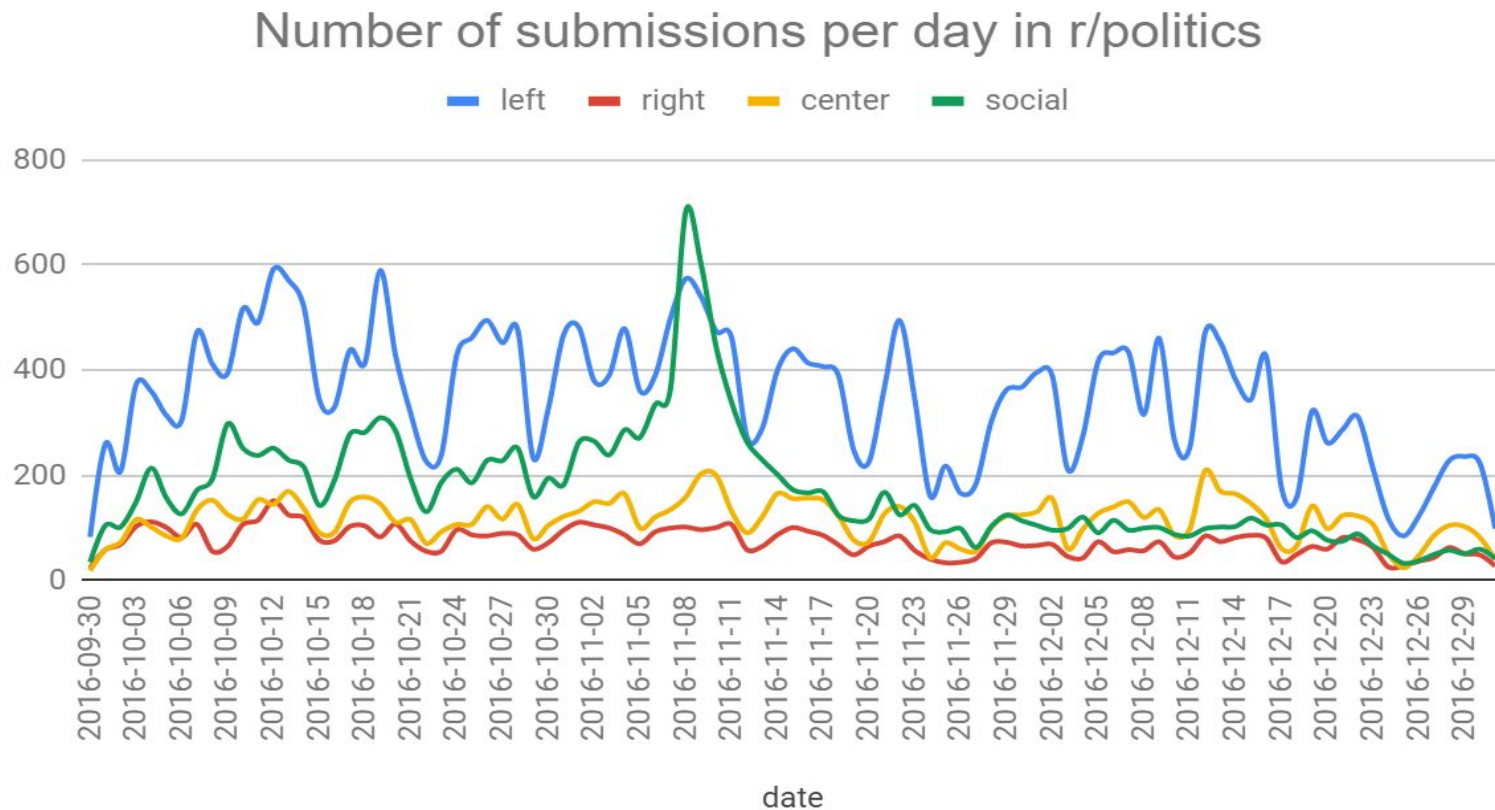
Average upvotes per comment on submissions.

Following the election: tracking activity over time
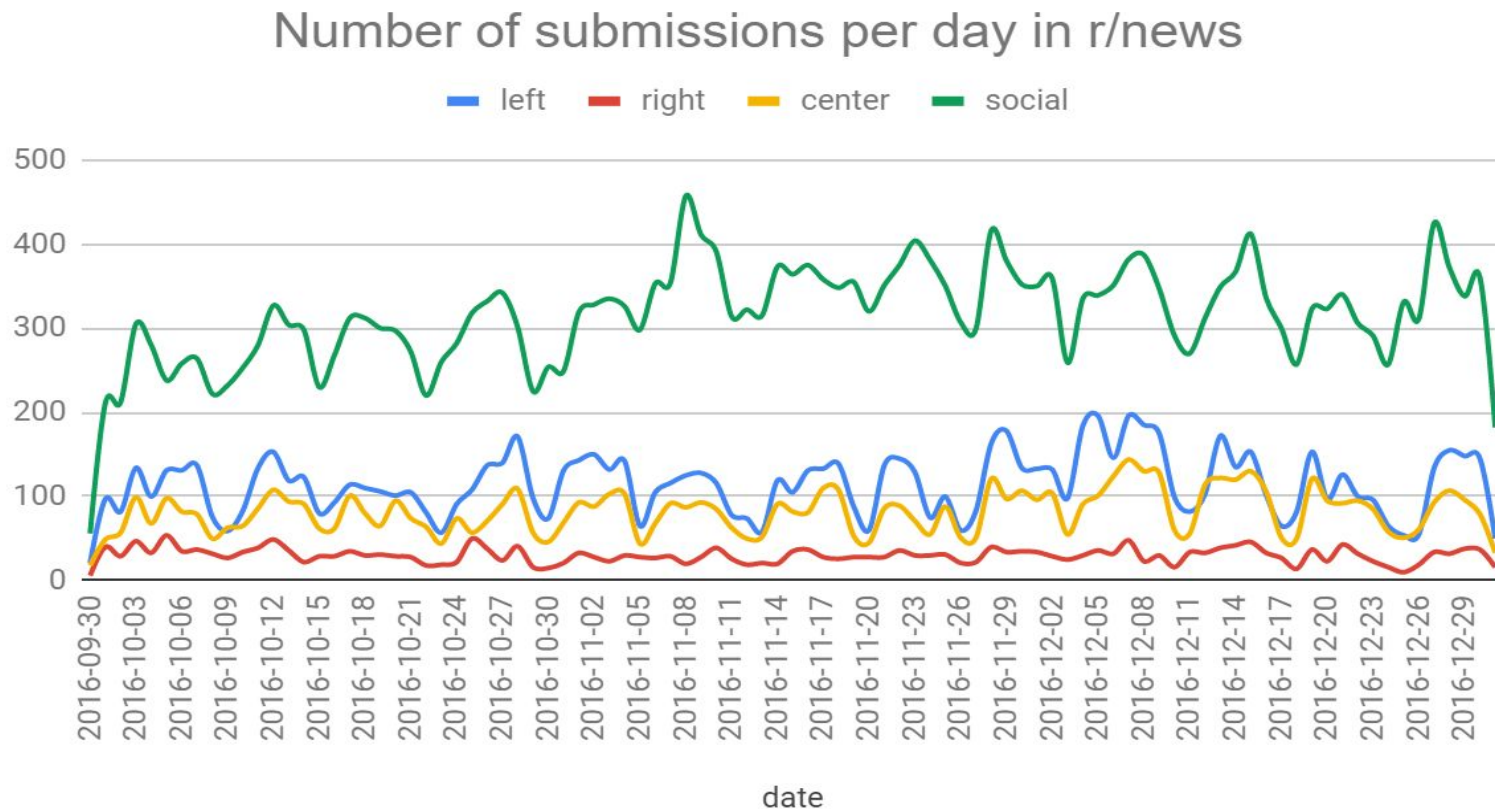
Number of submissions per day in r/news

Following the election: tracking activity over time

Number of submissions per day in r/worldnews

left   right   center   social

Following the election: tracking activity over time

Number of submissions per day in r/politics
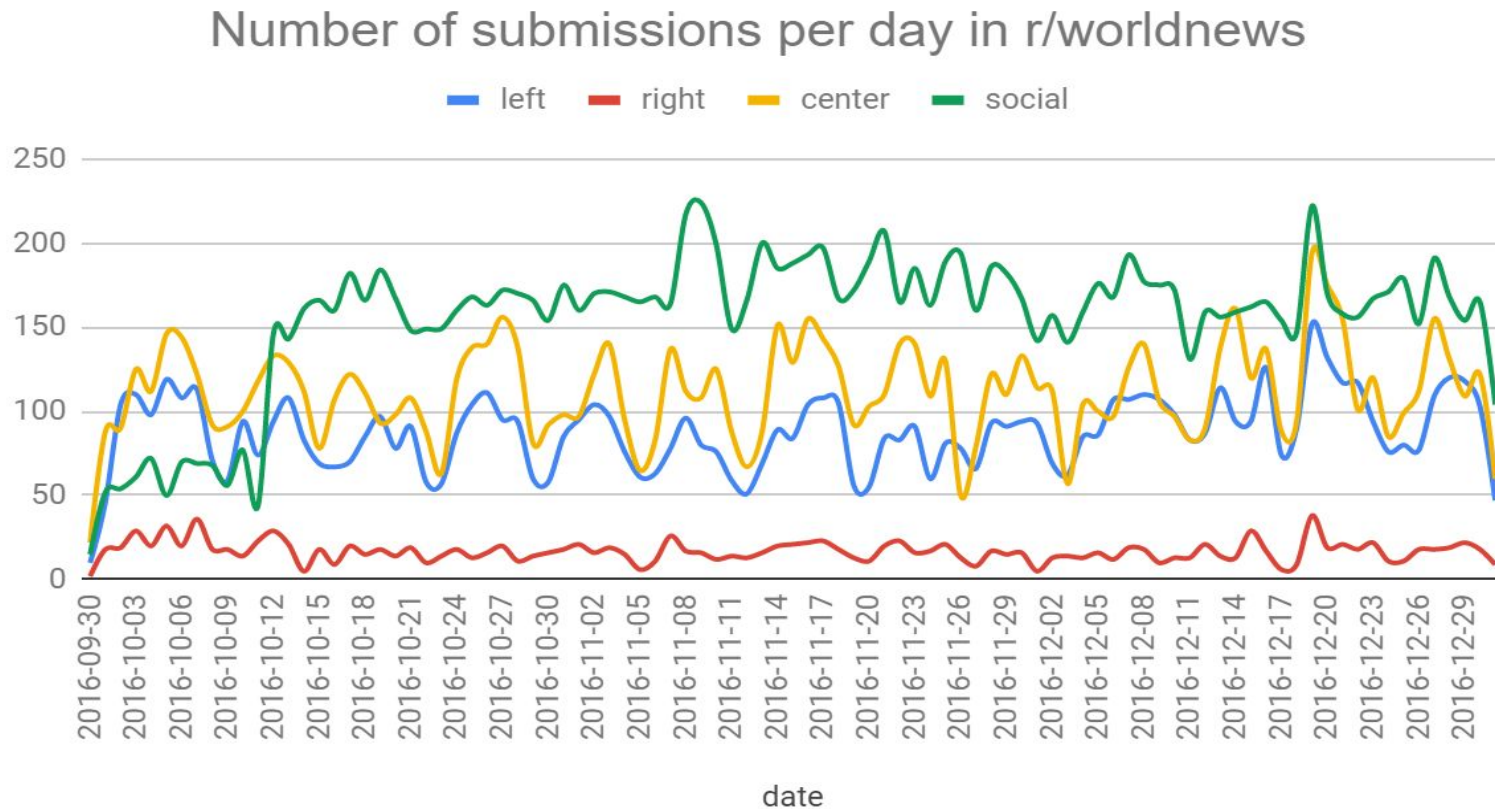
RQ 2: Does the identity of a media outlet predict a submission's popularity and engagement?

We use media bias ratings obtained from AllSides to classify submissions as quoting news from either left-, right-, or center-leaning media outlets

We then create a dataset with features that include a semantic representation of submissions' title, the political leaning of their source, its historic popularity, and the time of day the submission is made
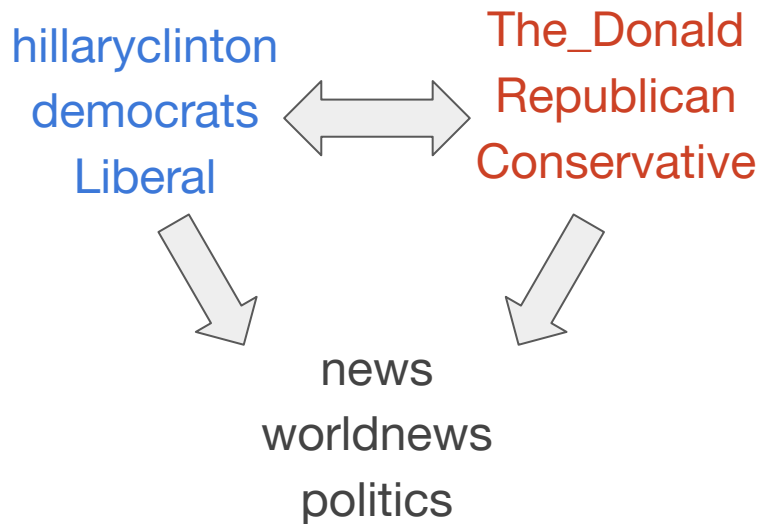

https://www.allsides.com/media-bias/media-bias-chart
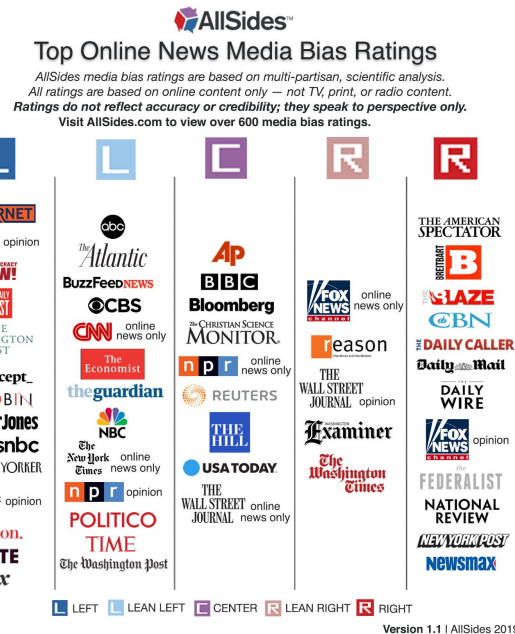
An unfair correlation? Political leaning of news source predicts expected popularity and engagement on r/news

feature	# comments	# upvotes
center	0.01960	0.01528
left	**0.04158**	**0.03076**
other	-0.01770	-0.01026
right	**0.00541**	**-0.00072**
evening	-0.01392	-0.01322
morning	0.01254	0.01148
night	-0.00727	-0.00657
noon	0.01138	0.01086

source leaning

submission time

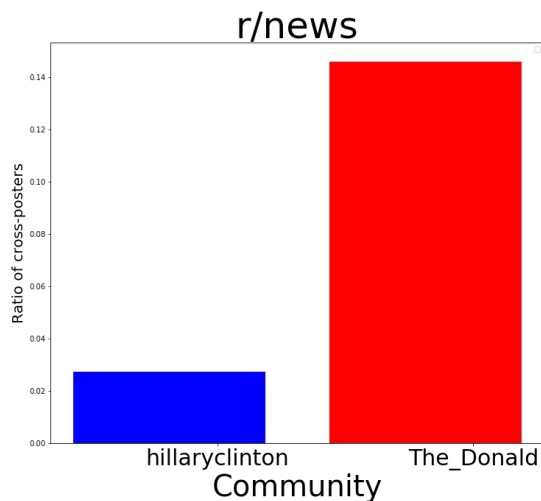# RQ 2: Community engagement on neutral ground

We extract all cross-posting users, i.e., those who post in both partisan subreddits as well as in neutral subreddits

We compare the representation of users from both sides of the aisle in the neutral subreddits

hillaryclinton
democrats
Liberal

The_Donald
Republican
Conservative

news
worldnews
politics

Community engagement: representation on neutral ground

Users from **The_Donald** post more frequently than those from **hillaryclinton**.
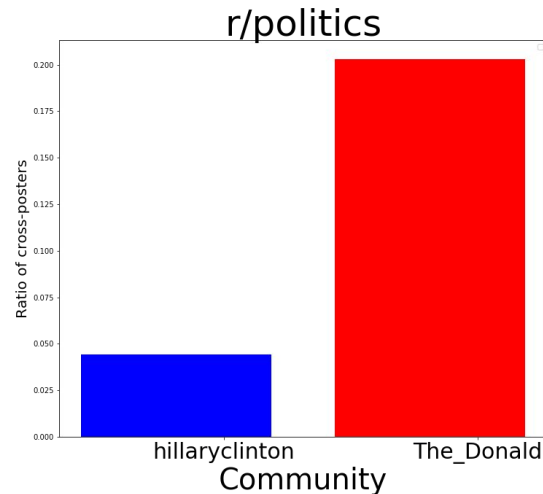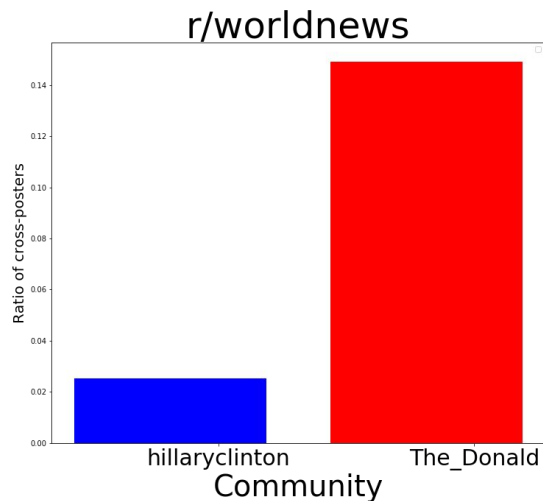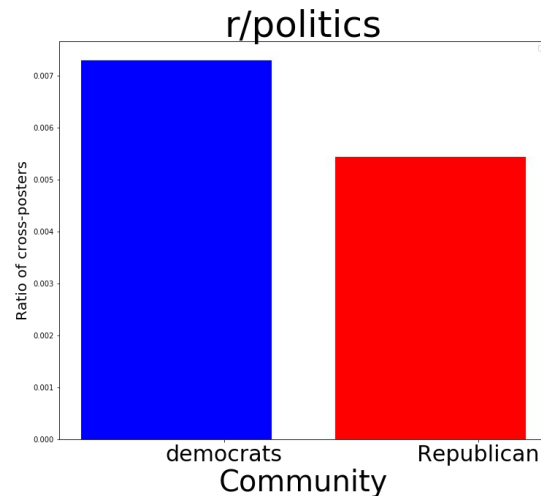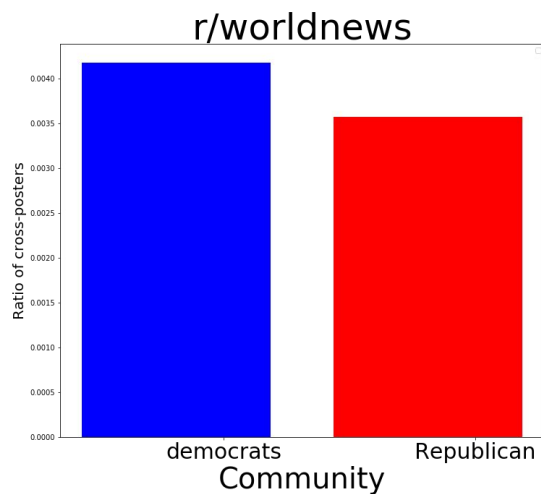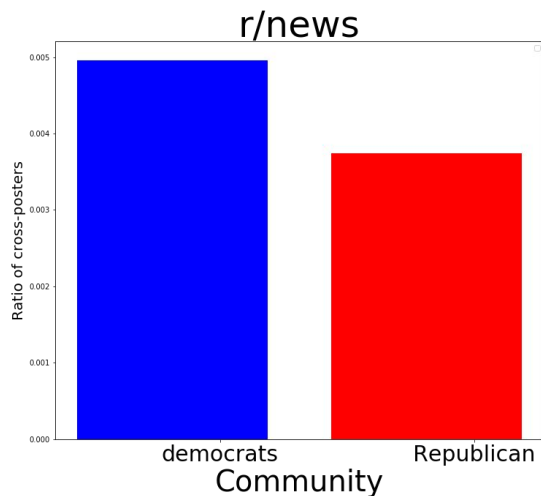
Community engagement: representation on neutral ground

Users from **Conservative** post more frequently than those from **Liberal**.
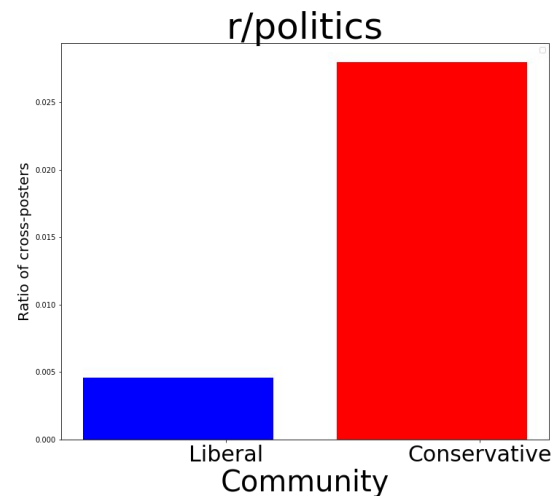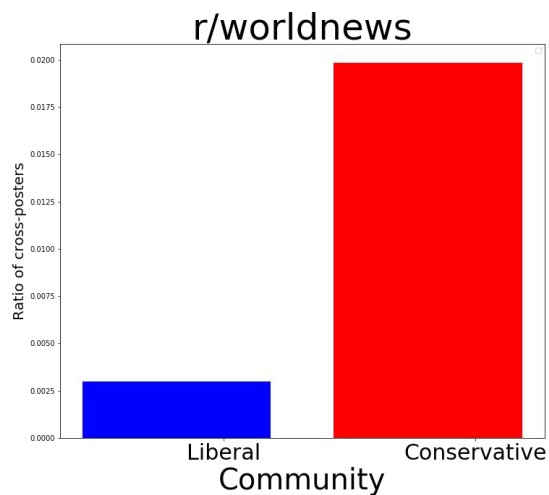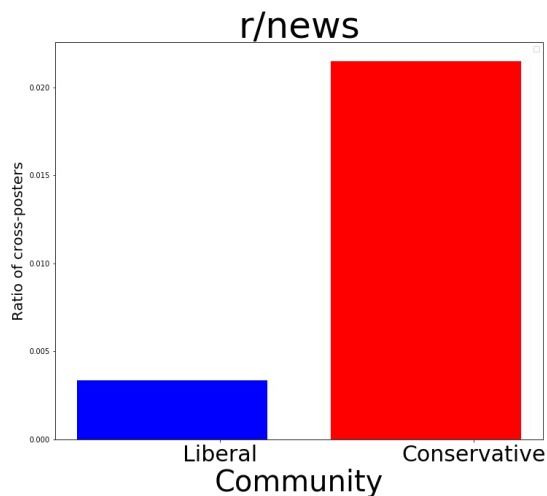
Community engagement: representation on neutral ground

Users from **democrats** post more frequently than those from **Republican**, although their numbers are comparable.

# RQ 4: Mining sentiment surrounding political actors

We extract all comments that contain
mentions of political actors and entities
from both parties

We compare average sentiment expressed
by the community for these actors and entities
in locally- and globally-focused subreddits

Hutto and Gilbert, VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, ICWSM 2014
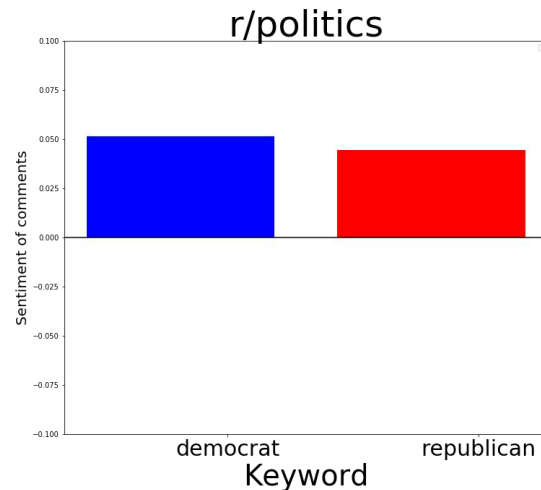
Mining sentiment: comparing local and global attitudes towards political actors and entities

Hillary is disliked more than Trump locally, but even more so in r/worldnews (!)
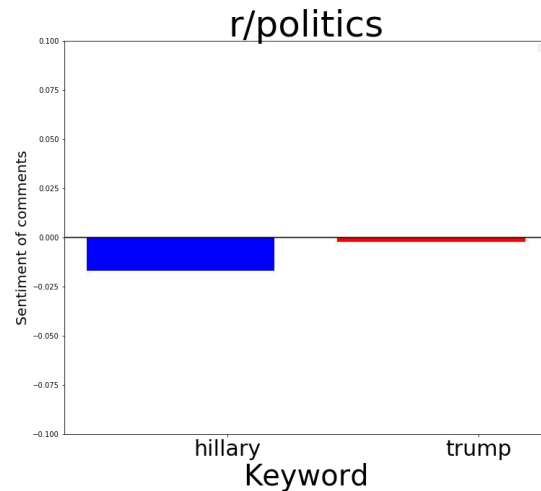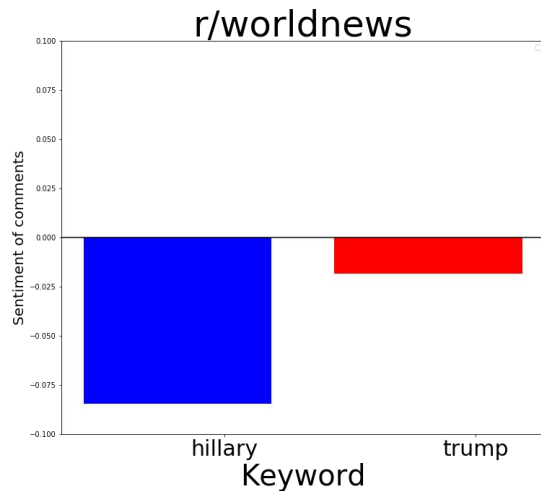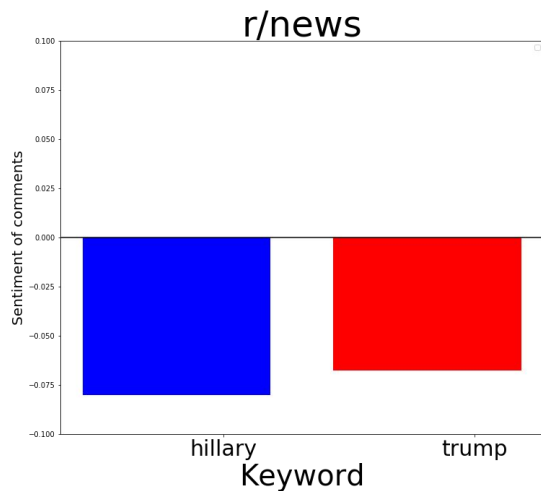
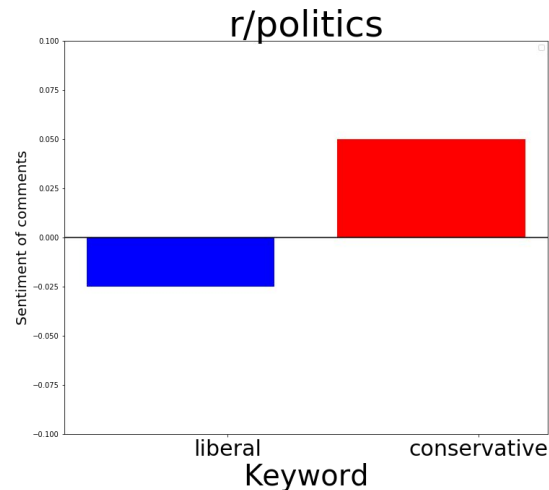# **Mining sentiment:** comparing local and global attitudes towards political actors and entities

Democrats disliked and Republicans liked in r/worldnews
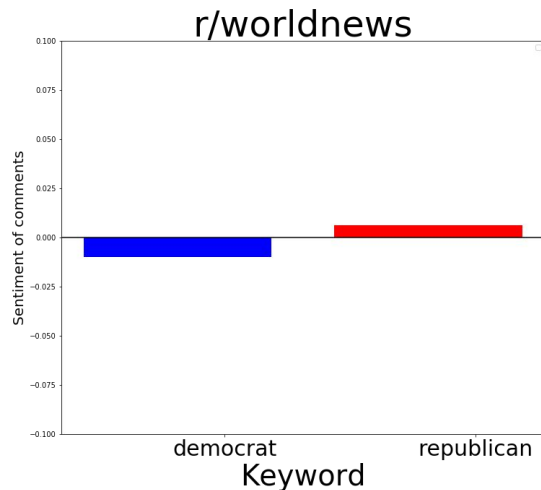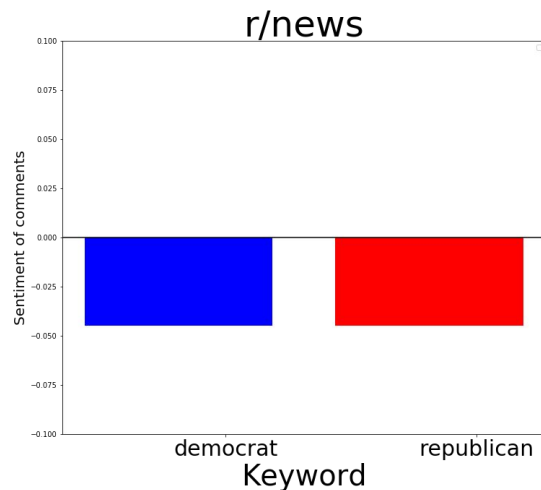
Mining sentiment: comparing local and global attitudes towards political actors and entities

Liberals disliked and conservatives liked in r/worldnews
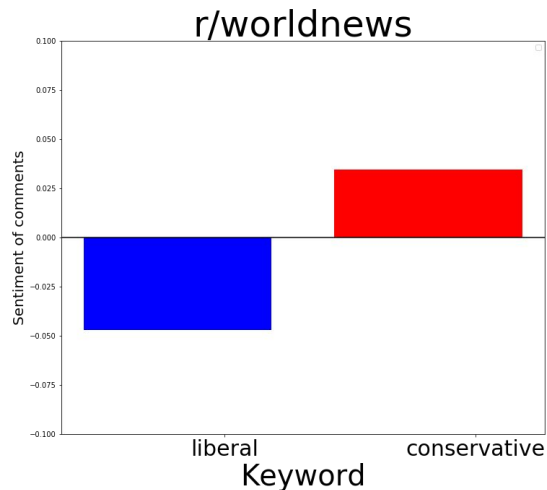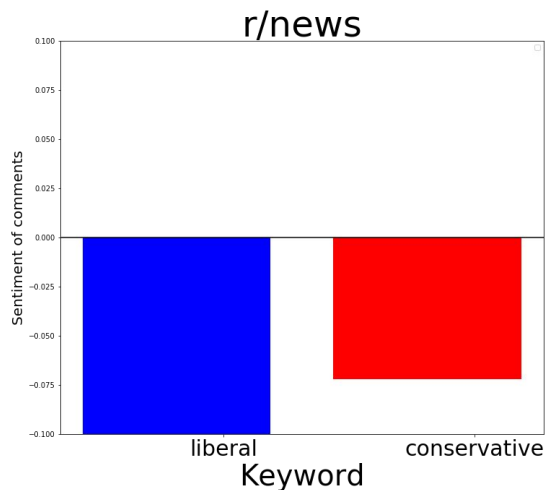
RQ 5: Comparing semantic representations of political entities in local and global discourse

We train word embeddings on comment
corpora collected from r/news and r/worldnews

We compare the semantic representations of entities in
local and global discourse, as well as in right-wing and left-wing
Conversations, by inspecting their nearest neighbors in the space

Mikolov et. al., Distributed representations of words and phrases and their compositionality, NeurIPS 2013

Comparing semantic representations of political entities: r/news vs /worldnews

r/news

hillary: clinton, hilary, hrc, bernie, trump, sanders, shillary, stein, obama, dnc, killary, clintons, hillarys, romney, djt, jill, mccain, superdelegates, kasich, hitlery

r/worldnews

hillary: clinton, hilary, hrc, bernie, sanders, trump, obama, clintons, dnc, gop, killary, romney, dws, shillary, she, donna, rnc, democrats, candidate, jeb

Comparing semantic representations of political entities: r/news vs /worldnews

r/news

trump: hillary, hrc, hilary, drumpf, obama, trumpler, clinton, bernie, djt, tump, sanders, ukip, trumps, killary, shillary, trumo, putin, gop, hilliary, hitlery

r/worldnews

trump: hillary, obama, putin, hilary, clinton, drumpf, duterte, hrc, he, bernie, netanyahu, sarkozy, bibi, du30, sanders, him, trumpler, trumps, gop, duerte

Comparing semantic representations of political entities: right- vs left-leaning subreddits

r/hillaryclinton, r/democrats, r/Liberal

hillary: hrc, trump, hilary, clinton, her, obama, she, bernie, sanders, bill, kaine, him, tpp, candidate, stein, he, dnc, chelsea, hers, ellison

r/The_Donald, r/Republican, r/Conservative

hillary: hilary, hrc, killary, shillary, hiliary, hitlery, hilliary, trump, illary, crooked, clinton, she, bernie, hilldog, hilllary, obama, hillery, hitlary, sanders, killery

Comparing semantic representations of political entities: right- vs left-leaning subreddits

r/hillaryclinton, r/democrats r/Liberal

trump: hillary, hrc, clinton, him, drumpf, pence, he, trumps, dt, obama, romney, rubio, bernie, putin, assad, hilary, his, sander, sanders, himself

r/The_Donald, r/Republican, r/Conservative

trump: djt, hillary, shillary, hilary, killary, bernie, dt, him, he, hrc, pence, tump, hitlery, romney, obama, sanders, clinton, gop, hilldog, hiliary

Conclusion & future work

- Left-leaning media outlets are quoted more often in local U.S. news and political subreddits, and they receive more engagement as well
- The global news subreddits are more centrist than U.S. ones, which means we may have a lesson to learn from the rest of the world
- Social media rises as a popular source for news during important, time-sensitive events, e.g. the result of the 2016 election
- The right-wing community is overrepresented in comparison to the left-wing community in the news and politics subreddits
- We find clear differences in the sentiment associated with and semantic representations of political actors and entities in local and global subreddits

Conclusion & future work

- Thus far we have been dealing with a late pivot in the project and wrangling with a large amount of data; in the future we plan to:
- Expand the list of media outlets classified by political leaning to improve coverage in submissions
- Evaluate *controversiality* of submissions with respect to their news source and political subject
- Create communities of users by connecting them based on activity in common subreddits across the entire website, and use it to discover other subreddits that are correlated with people's political leaning
- Use the lessons learned here to inform the development of a framework to encourage a well-balanced diet in news recommendation

Thank You