What emotion makes people engage more with politicians' tweets?

2022 GWU Datathon Report

Xiang Li
Statistics department, GWU

Outline

1. Outcome and predictors
2. Results
3. Discussion and Next step

Code and reports are available at
https://github.com/xiangli2pro/GWU_Datathon_2022

Outcome

Binary outcome: *engage_active*

$$engage\_rate = \frac{1}{2}\left(\frac{favoriate\_count + retweet\_count}{followers\_count} + \frac{favoriate\_count + retweet\_count}{statuses\_count}\right)$$

$$engage\_active = ifelse(engage\_rate > engage\_thresh, 1,0).$$

engage_thresh: 75% percentile of engage rate of all tweets in year 2021

User-specific predictors

1. party (categorical): which party (Independent, Republican, Democratic) the politician belongs to. I web-scraped the social media accounts of the politicians from the public website then assign the party information to each user.

2. favor_perFriend (numerical): average number of favorites given by the user to the friends (people followed by the user).

3. listed_level (numerical): if the number of organizations the user belongs to is less than the 0.25-th percentile of the number among all users, the listed_level = 1. If the number is less than the 0.50-th percentile, the listed_level = 2. If the number is less than the 0.75-th percentile, the listed_level = 3. If the number is greater than the 0.75-th percentile, the listed_level = 4.

4. has_url (binary): whether or not url is included in the tweet.

5. has_media (binary): whether or not media (photo) is included in the tweet.

6. is_independent (binary): if the tweet is not quoted or retweeted or a reply, it's an independent tweet, otherwise it's not independent.

7. engageQuoted_active (binary): the engage activity of the quoted tweet. same definition as the engage_active.

8. engageRetweet_active (binary): the engage activity of the retweeted tweet. same definition as the engage_active.

9. hashtag_num (numeric): the number of hashtag in the tweet.

10. metion_num (numeric): the number of mentioned names in the tweet.

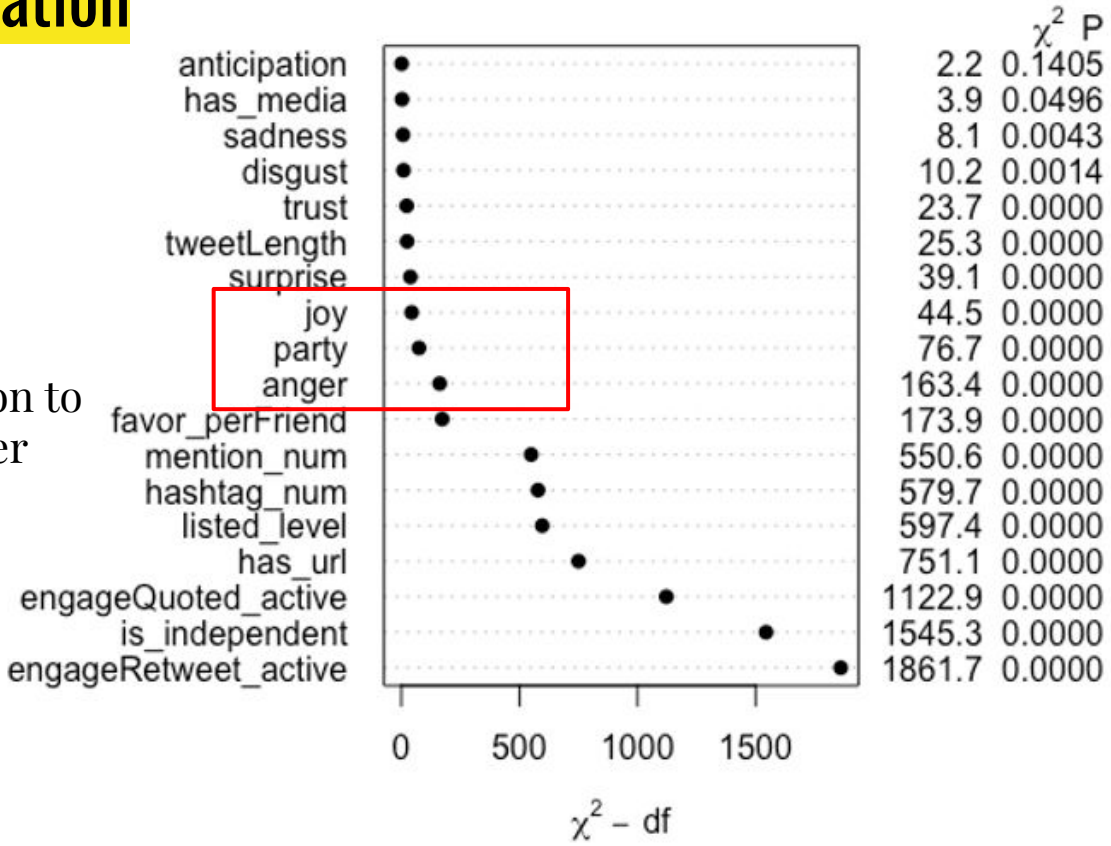
Emotion-specific predictors (NRC lexicon)

1. textLength (numeric): length of the cleaned tweet text.
2. anticipation (numeric): average number of words belong to anticipation in a tweet.
3. anger (numeric): average number of words belong to anger in a tweet.
4. fear (numeric): average number of words belong to fear in a tweet.
5. negative (numeric): average number of words belong to negative in a tweet.
6. joy (numeric): average number of words belong to joy in a tweet.
7. positive (numeric): average number of words belong to positive in a tweet.
8. trust (numeric): average number of words belong to trust in a tweet.
9. sadness (numeric): average number of words belong to sadness in a tweet.
10. surprise (numeric): average number of words belong to surprise in a tweet.
11. disgust (numeric): average number of words belong to disgust in a tweet.

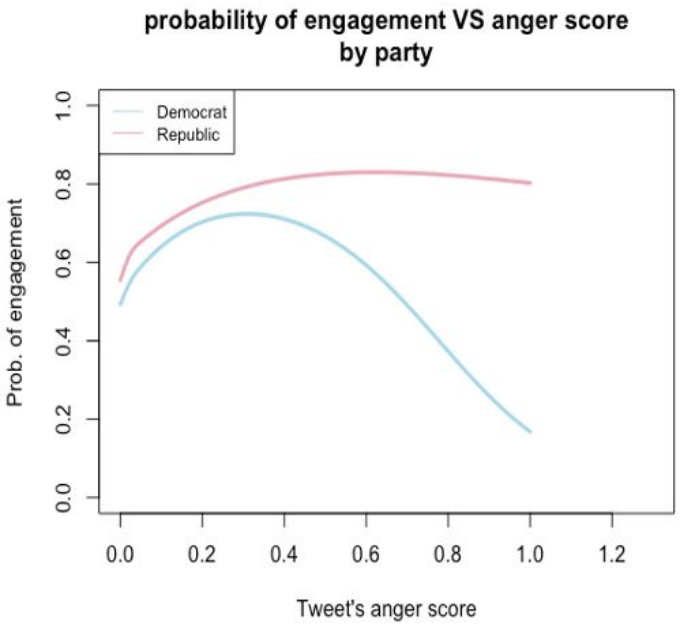
Model selection and validation

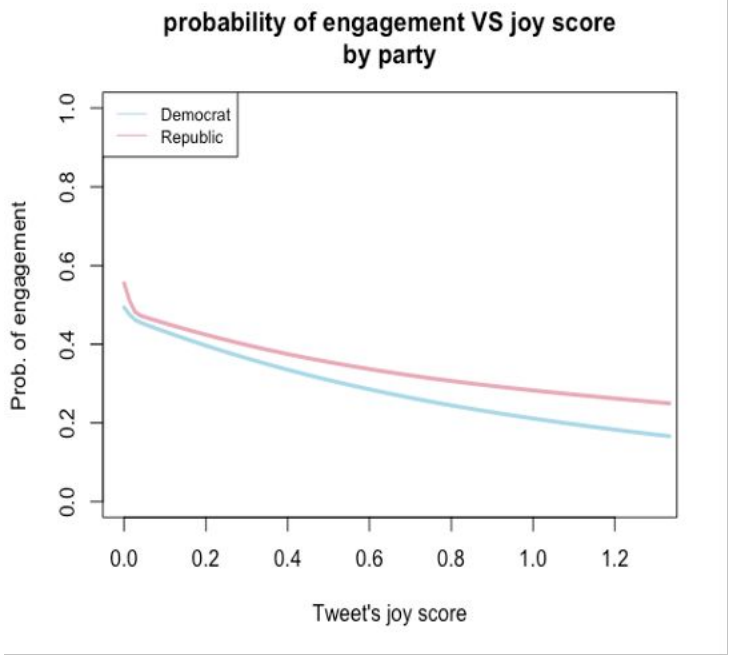
Logistic spline regression + non-linearity of emotions + party*emotions

Anger makes the most contribution to the outcome probability than other emotions.



Model selection and validation



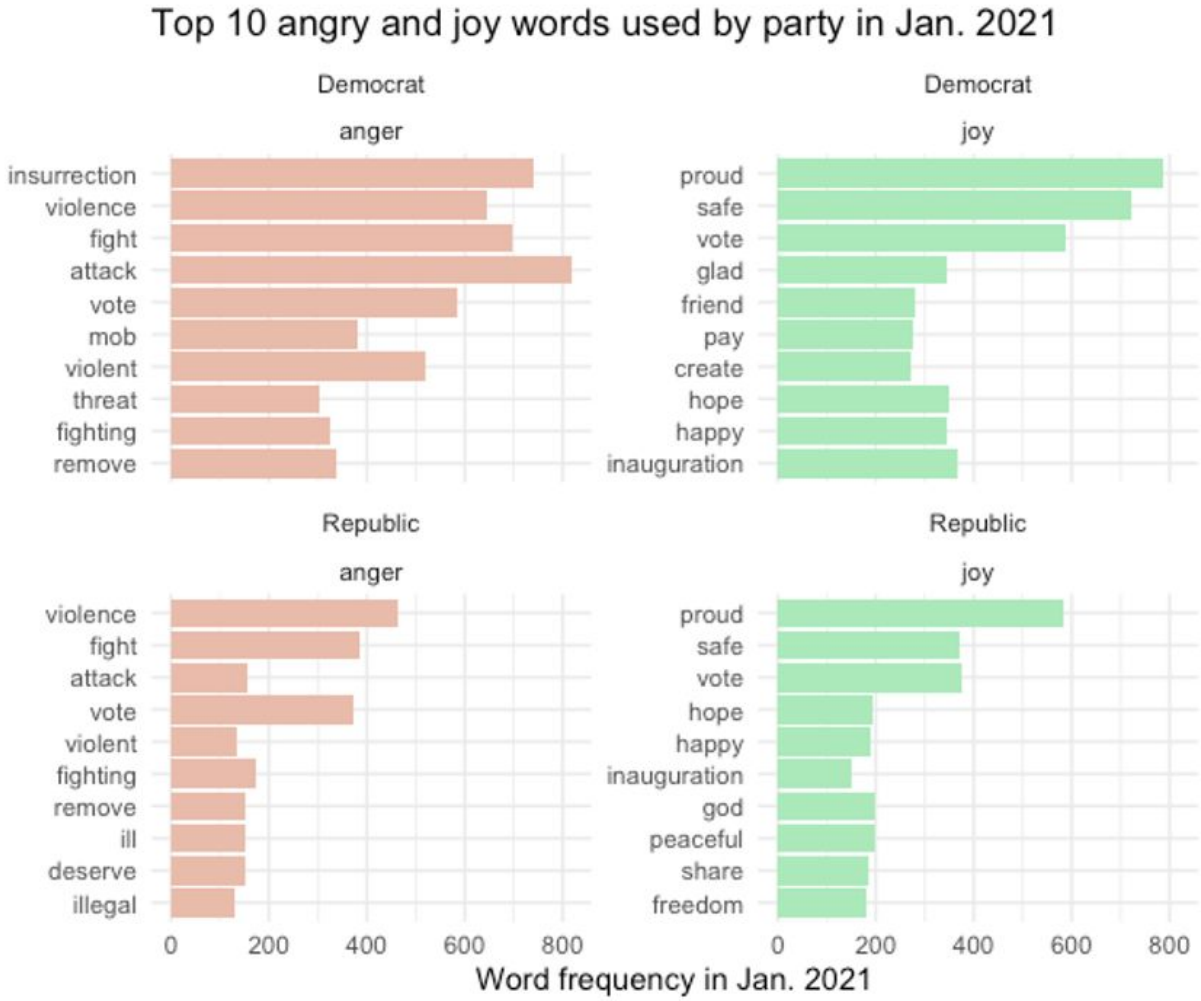


Logistic spline model assuming non-linearity on anger and joy.

Results

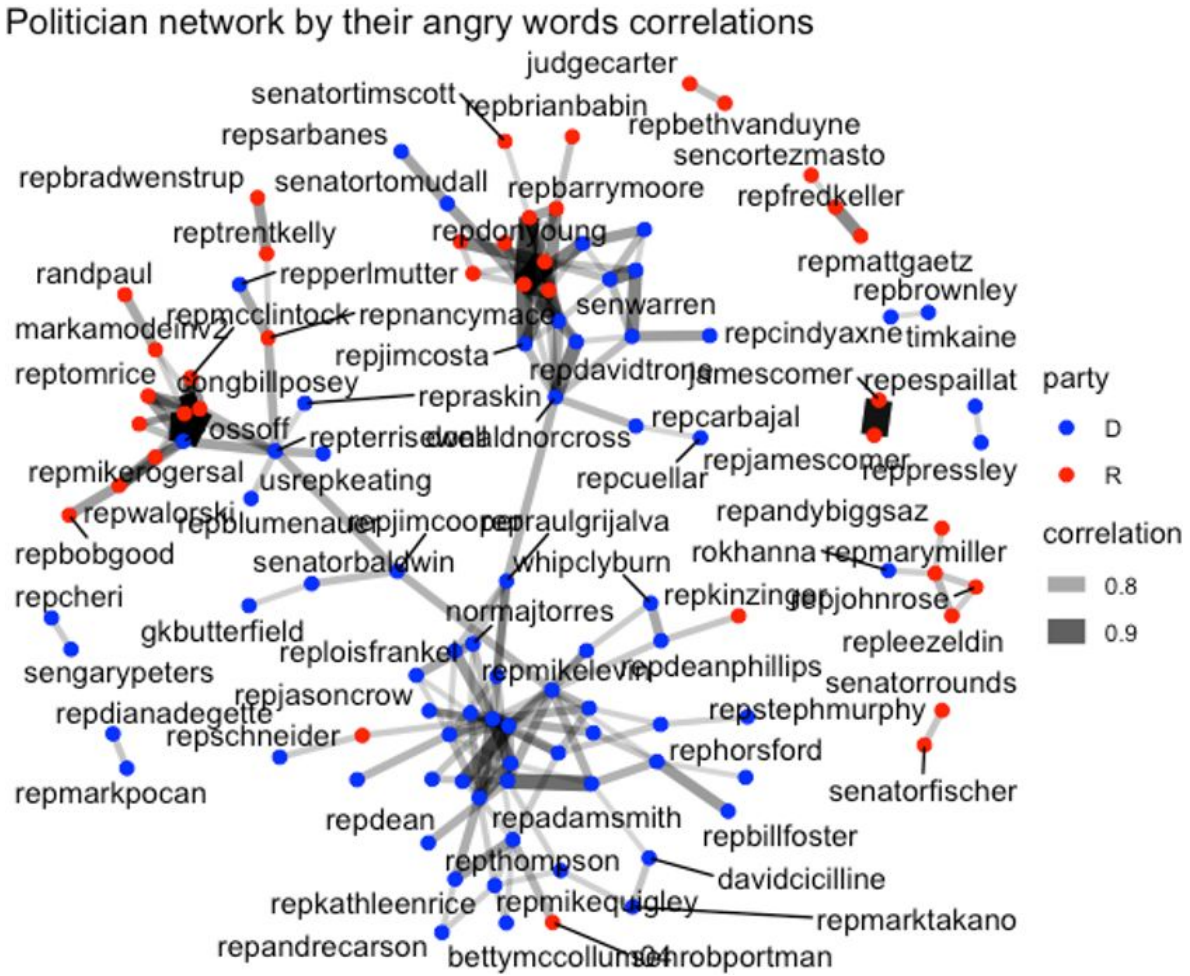
Democrats and Republicans are both angry to the capitol attack but use different narratives.

Democrats and Republicans are both happy to the inauguration.



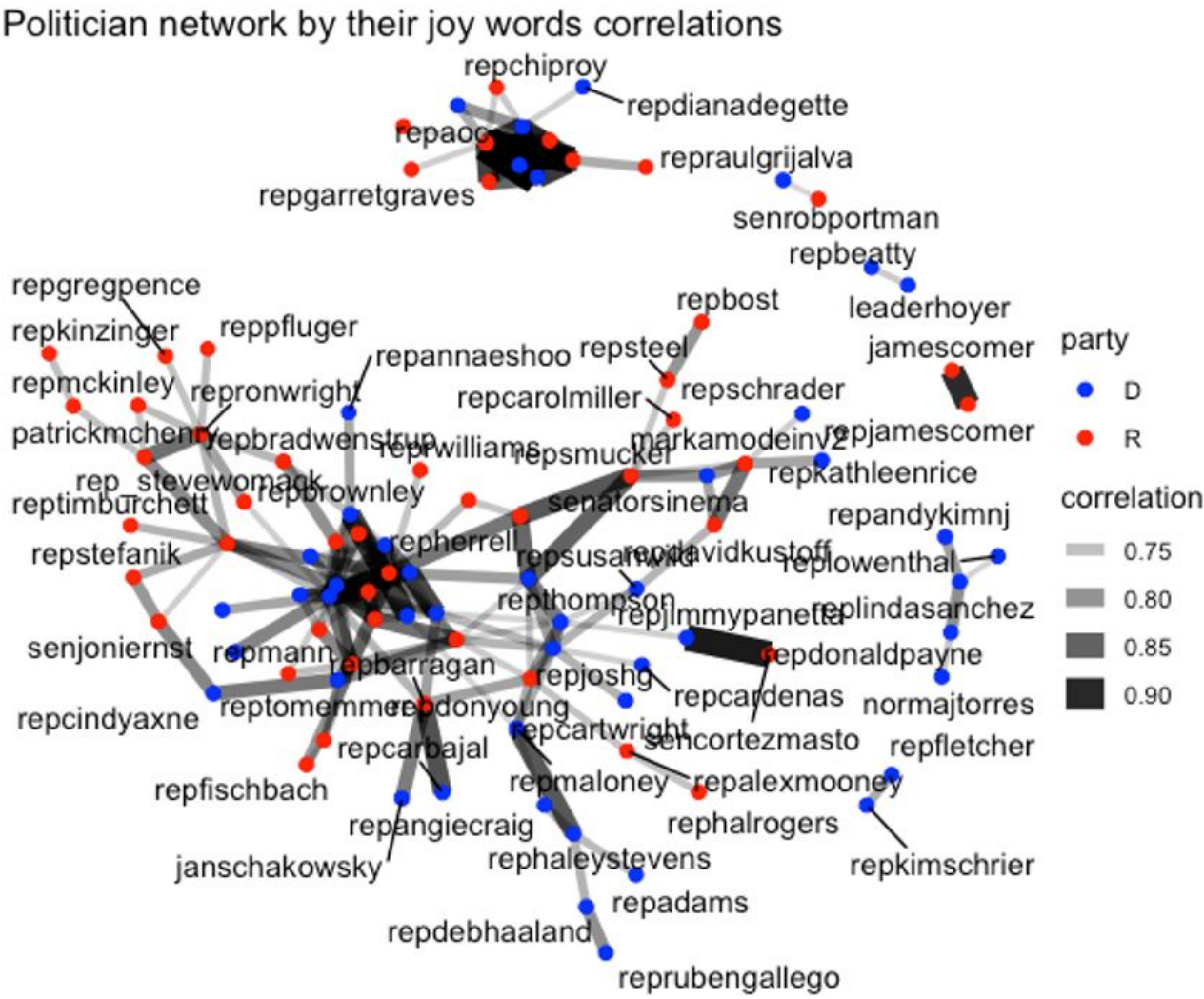
Results

Democrats and Republicans are divisive on the tweets content that invokes anger.



Results

Democrats and Republicans are less divisive on the tweets content that invokes joy.



Discussion and next steps

Question: should we encourage politicians to use angry tones in their tweets to get more engagements?

Next steps:

1. Verify the conclusion using tweets from other time periods.
2. Focus on the predictive accuracy of the model.