Methods Supplement for “Twitter found conservative politicians receive disproportionate amplification via its algorithmic timeline. Here’s one possible explanation why"

Megan A. Brown

October 27, 2021

1 Data Collection

Every day, we collect a 10% sample of all tweets on Twitter, including retweets. For this research, we are specifically interested in tweets by Members of Congress. For each day of tweets from January 1, 2021 through October 23, 2021, we filter for retweets of tweets originally authored by the politicians of interest. Because we only collect a 10% sample of tweets, we expect that we only get a 10% sample of retweets for any given tweet. However, due to the high follower count and engagement that the majority of these accounts receive, we expect that we collect the majority of tweets by the accounts in the list that remained active during the period. For each retweet of a tweet by the accounts in question, we retain the timestamp of the retweet, the engagement metrics (number of retweets, quote tweets, and replies) for the original tweet at the time of retweet, and the original tweet ID. For each member of Congress, we use their nominate scores[1] to estimate their ideology.

2 Data Definitions

For each original tweet by a member of Congress, we retain the latest retweet (and therefore, the most up-to-date engagement metrics we have for that retweet). For each tweet, we calculate the “ratio” as follows:

$$ratio = \frac{Quotes + Replies - Retweets}{Quote + Replies + Retweets}$$

For each member of congress, we calculate the ratio for each of their tweets and plot the distribution. For Figure 1, we select the 20 members of congress with the highest median ratio per tweet. For each member, we plot the interquartile range with a box and whiskers at 1.5x the interquartile range. Outliers greater than 1.5x the interquartile range are plotted independently with points.

We then calculate the quartiles of the ratio for all tweets in the corpus and bin tweets according to these quartiles. We pool each tweet by its ratio and then plot the distribution of congress

[1] https://voteview.com/data

member ideology for the bin (higher quartiles are equal to higher ratios). In Figure 2, we show the distribution of member ideology of the tweets in that quartile. As we move from the least ratioed tweets on the left to the most ratio tweets on the right, the median ideology of the member of congress tweeting becomes more conservative.

Finally, we use linear regression to estimate the relationship between member ideology and the ratios of their tweets. For each member, we calculate the median ratio of their tweets. We estimate a linear regression model, predicting the median as a function of member ideology. We find that the more conservative a member is, the higher their average ratio for tweets. This holds regardless of party. In Figure 3, we show the median ratio for each member based on their DW-Nominate ideology and the best fit line for each party.

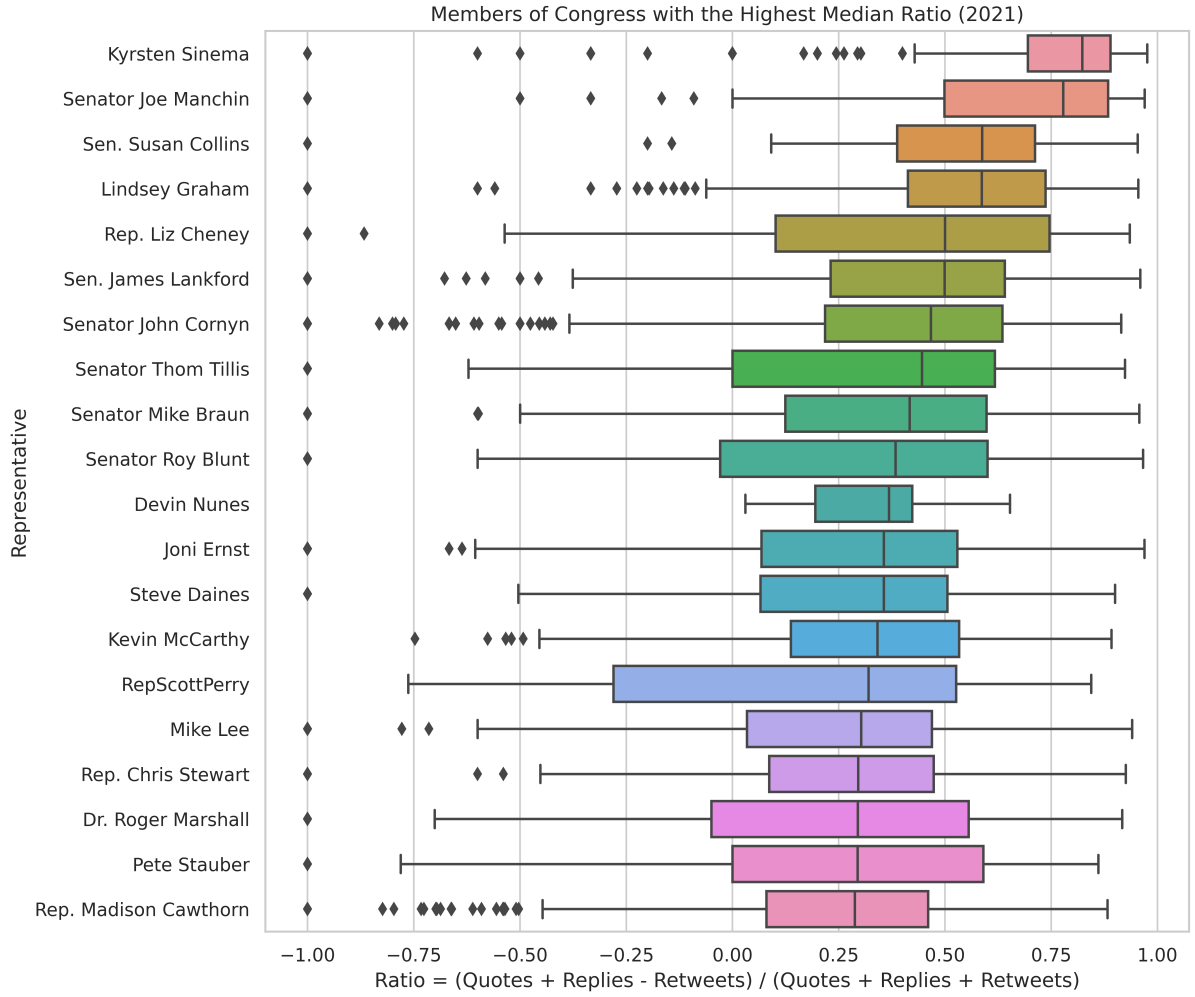

Figure 1: Distribution of ratios for tweets by the top 20 “ratioed” members of congress. Data: NYU Center for Social Media and Politics Figure: Megan A. Brown, NYU Center for Social Media and Politics

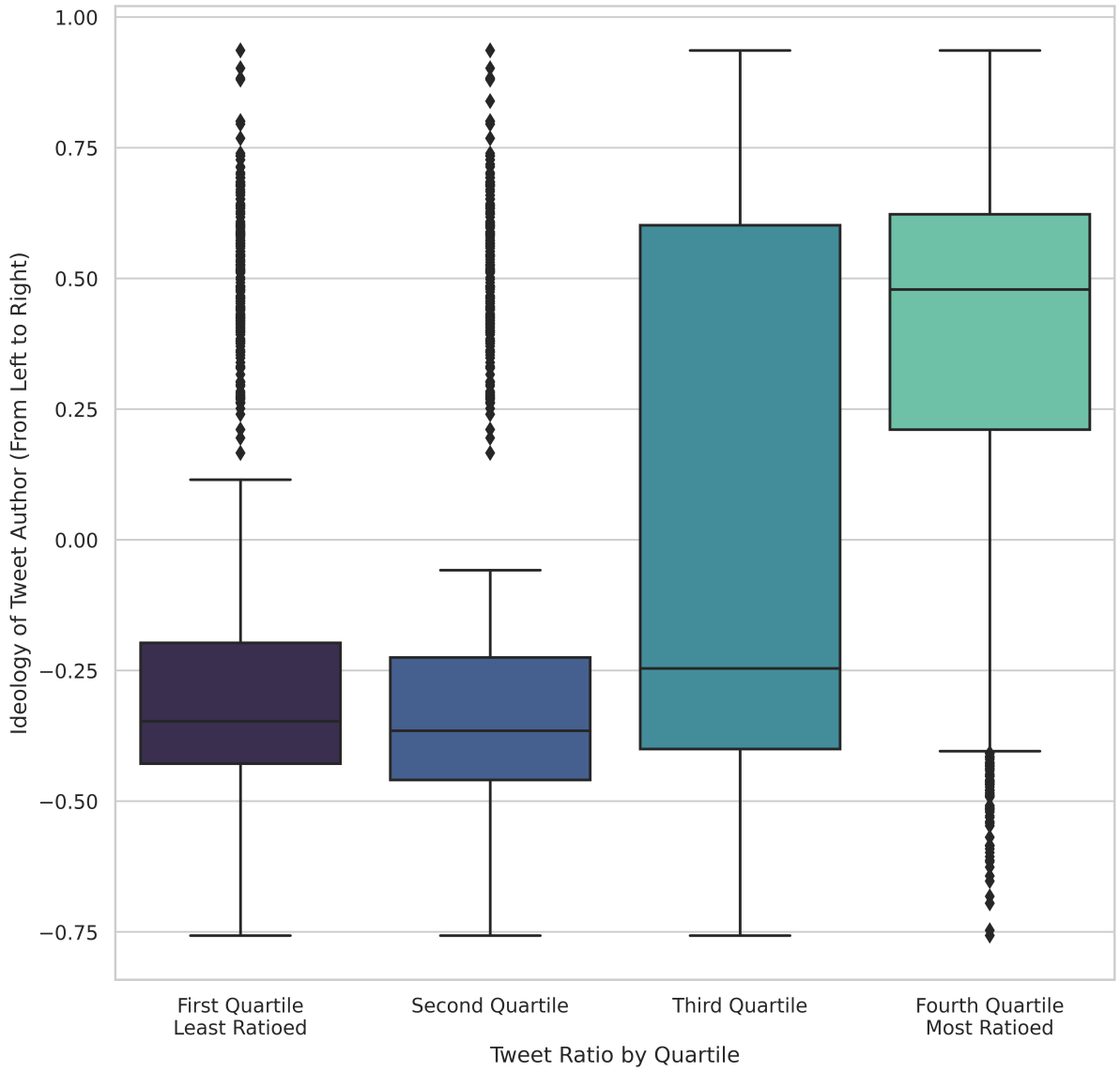

Figure 2: Distribution of congress member ideology by ratio quartile. Data: Twitter - NYU Center for Social Media and Politics, Ideology - VoteView, Figure: Megan A. Brown, NYU Center for Social Media and Politics

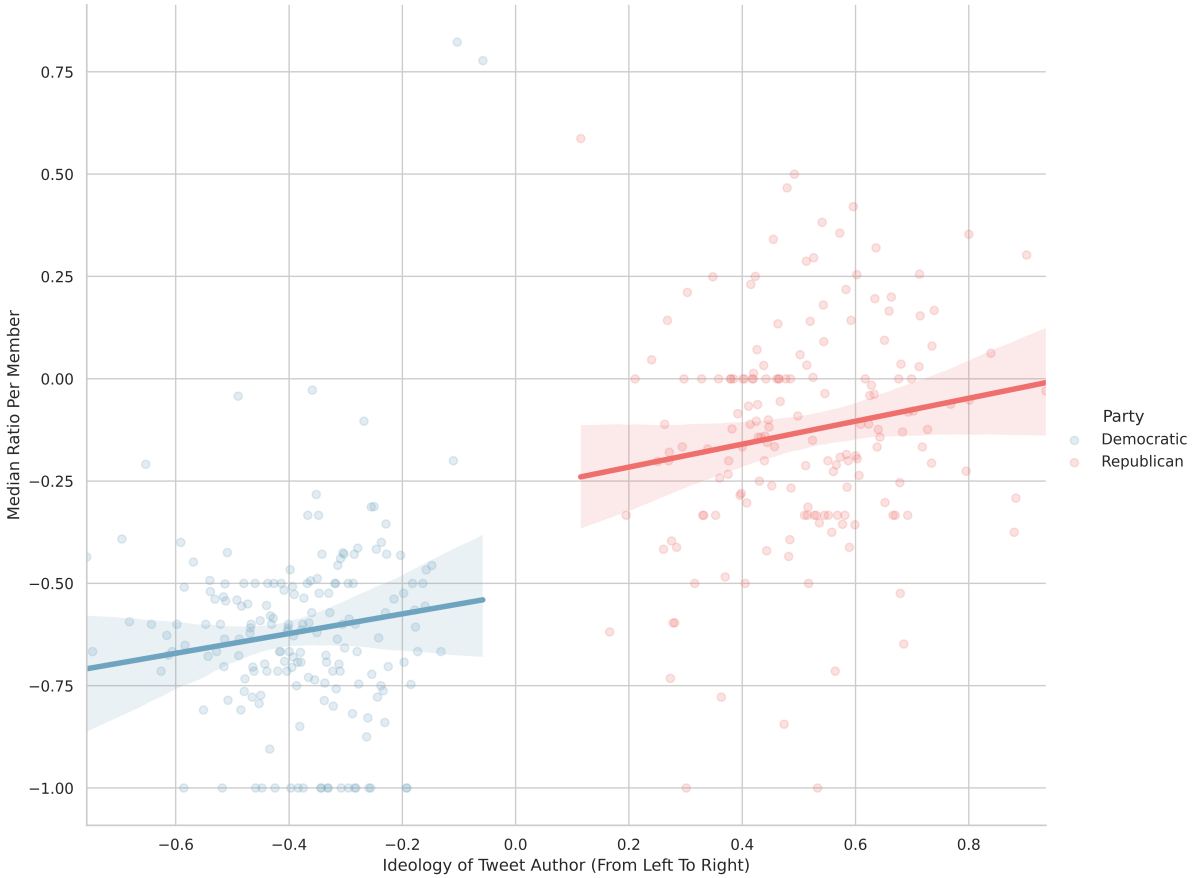

Figure 3: Median ratio as a function of ideology. We plot for each member their DW-NOMINATE ideology score and the median ratio for their tweets (points). We plot the estimated regression split by party (Democrats are blue and Republicans are red). Data: Twitter - NYU Center for Social Media and Politics, Ideology - VoteView, Figure: Megan A. Brown, NYU Center for Social Media and Politics