The Language of Justice: Decoding the Supreme Court

Kekai Apana & Adrian Pan

Abstract

Our project sought to investigate whether linguistic patterns and voting history can predict justices' votes in the Roberts Court. Using machine learning, we analyzed and predicted justice voting patterns. We concluded that oral arguments strongly influence voting behavior, offering insights into the predictability of judicial decisions based on linguistic patterns and voting histories.

Research Questions

The Supreme Court of the United States, currently led by Chief Justice John Roberts, **plays a pivotal role in American society**. The Roberts court has issued many landmark decisions with **profound consequences**. Each justice comes to their own conclusions based on various parts of the cases, most importantly **the oral arguments**, in which the justices converse with each other and the arguing parties. Based on this information, we wondered:

1. Can we predict how the Justices in the Roberts Court will vote on a case based on their linguistic patterns in oral arguments and voting history?
2. Do certain justice's have similar arguments and does that relate to how they vote?

Data Collection

We used the **Cornell Supreme Court Oral Arguments Corpus**, a collection of SCOTUS oral arguments from 1955 to 2019. It includes over 1.7 million utterances, or comments, from 7,700 cases. In total, we analyzed 125,917 utterances from 14 justices and 1,166 cases. We extracted the following info using the ConvoKit package on the corpus' utterance class, joined with data from the cases JSON file

- Utterance text
- Utterance ID
- Response ID
- Speaker
- Case ID
- Speaker Type
- Year
- Case Title
- Court
- Votes
- Votes Side

Any case that did not have utterance data was excluded.

Data Exploration

Figure A is a heat map, showing the voting similarity between the justices in the Roberts court. **Figure B** is a network graph, showing the language similarities between the different justices in the Roberts Court. **Figure C** shows the graphs for the top 10 words that each Justice used in their oral arguments during their time in the Roberts Court.

A.

Voting Similarity Among Justices in the Roberts Court

B.

Network Graph: Justices and Shared Words

C.
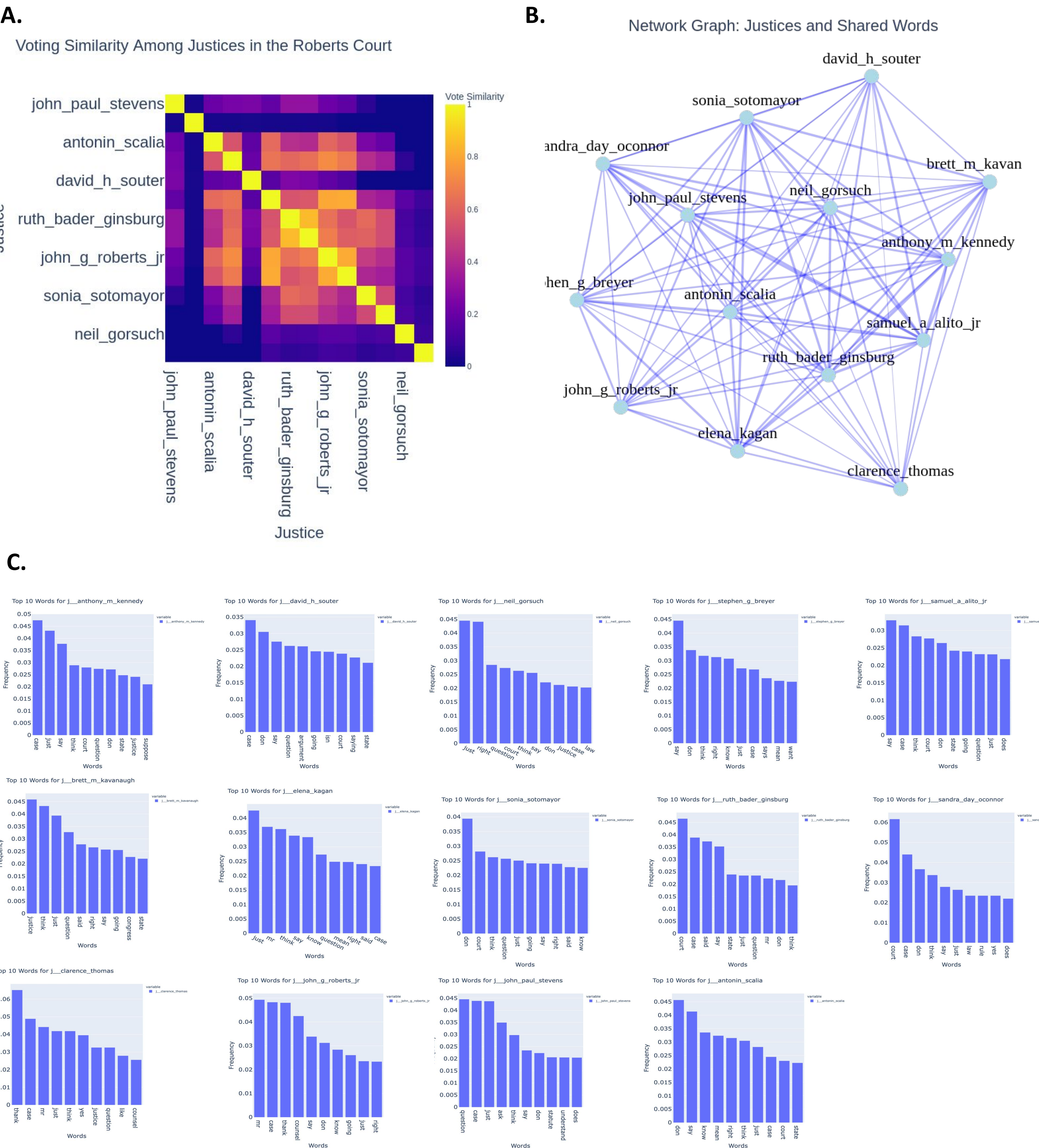
Data Analysis

We used a **Linear Support Vector Classifier** (LinearSVC) to predict Roberts Court justices' votes based on their oral argument language and voting history. Text was preprocessed, and features were transformed using OneHotEncoder (voting data) and TfidfVectorizer (text). Results showed varying prediction accuracy across justices, indicating that linguistic patterns correlate with voting behavior to differing degrees. Higher accuracy suggests stronger linguistic influence, while lower accuracy implies that other factors may impact voting.

	justice	scores
0	j__john_g_roberts_jr	0.880681
1	j__david_h_souter	0.694998
2	j__anthony_m_kennedy	0.901835
3	j__ruth_bader_ginsburg	0.773389
4	j__john_paul_stevens	0.752118
5	j__samuel_a_alito_jr	0.762248
6	j__antonin_scalia	0.785604
7	j__stephen_g_breyer	0.775854
8	j__sandra_day_oconnor	0.713968
9	j__clarence_thomas	0.890909
10	j__sonia_sotomayor	0.775000
11	j__elena_kagan	0.806749
12	j__neil_gorsuch	0.786325
13	j__brett_m_kavanaugh	0.918345

Conclusion

Our findings could imply that the justices with higher prediction accuracies rely heavily on the arguments presented during oral discussions when making their decisions. The varying accuracies across justices also highlight the complexity of judicial decision-making, where factors beyond linguistic patterns, such as legal principles, personal beliefs, and external influences, may play significant roles. Overall, our analysis provides valuable insights into the relationship between linguistic patterns and voting behavior in the Roberts Court, offering a data-driven perspective on judicial decision-making.