Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Minh-Thy Tyler   April 10, 2024

The Effects of California's 2012 Election System Change on Polarization

by

Minh-Thy Tyler

Maggie Penn

Advisor

Quantitative Theory and Methods

Maggie Penn

Advisor

Jo Guldi

Committee Member

Sandeep Soni

Committee Member

2024

The Effects of California's 2012 Election System Change on Polarization

by

Minh-Thy Tyler

Maggie Penn

Advisor

An abstract of
A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Quantitative Theory and Methods

2024

Abstract

The discourse surrounding election systems has remained a point of discussion among political scholars and experts. Recent years have witnessed a notable trend where various states and cities across the United States have modified their election systems to find methods that mitigate political polarization, increase fairness in elections, and enact other reforms. One such significant transition occurred in California, where the electoral system transitioned from a closed primary system to a top-two primary system in 2012. This study explores the ramifications of California's change on the impact of political polarization among elected congressional representatives during federal elections in the 112th and 113th Congresses. Using Adam Bonica's DIME and DIME+ datasets, this research takes a multifaceted approach to offer insights into the dynamics of this change. First, it assesses the ideological shifts among representatives elected in 2010 who became incumbents in the 113th Congress following the 2012 election. Along with this, this research uses a difference-in-difference model to compare ideological shifts among representatives with those in similar states. The second approach conducts a network analysis based on voting behaviors across the two Congresses. Lastly, through sentiment analysis by topic, this study researches the potential changes in political rhetoric surrounding various issues. Despite these methodologies, the findings indicate that California's electoral reform did not have significant changes in political polarization among representatives in the two years after the 2012 election.

The Effects of California's 2012 Election System Change on Polarization

by

Minh-Thy Tyler

Maggie Penn

Advisor

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Quantitative Theory and Methods

2024

Table of Contents

Chapter 1: Introduction

The United States has a persistent electoral system problem. Advocates for election reform and political scientists have argued that partisan primaries used by the majority of federal elections across the nation are inherently flawed. Many scholars argue that this system exacerbates political polarization and gives rise to a variety of other problems. Advocates for nonpartisan primaries argue that it promotes moderation and decreases polarization and does not shut out independent voters like partisan primaries do (Lopez, 2023). As of 2024, only five states have successfully transitioned to federal or statewide nonpartisan primaries: Alaska, California, Nebraska, Washington, and Louisiana (Primary election types by state, 2023). In 2010, California passed Proposition 14, which mandated all running candidates in nonpresidential races to participate in the state's nonpartisan primary (California Proposition 14, Top-Two Primaries Amendment, 2010).

   The reform to implement the top-two nonpartisan primary system, also known as the jungle primary, was pushed by a group spearheaded by Governor Arnold Schwarzenegger. They hoped to broaden participation in primaries by allowing voters without party affiliation to vote, while also potentially fostering moderation among candidates competing among a broader electorate encompassing Republicans, Democrats, and independents (Nagourney, 2018). There are many journalists and scholars who contest the effectiveness of this reform. In a 2018 NPR interview, political scientist Christian Grose cited evidence suggesting a reduction in polarization within California's state legislature post-2012 (Grose, 2018). Conversely, journalist Harold Meyerson, in a 2014 Los Angeles Times Op-Ed, criticized the jungle primary for dividing votes and benefiting parties running fewer candidates, with minimal impact on removing extreme representatives in 2014 (Meyerson, 2014).

   This research will focus on the impact of California's transition to the top-two primary system on political polarization within its elected officials. It assesses political polarization through three methodologies: examining shifts in representative ideology among representatives, centrality measures via network analysis, and analyzing sentiment scores across various political topics. Through these methodologies, I define moderation in ideological change as a shift towards a more neutral position on a two-dimensional ideological scale. In network analysis, moderation is characterized as an increase in cross-party alignment on voting behavior among candidates. Moderation in sentiment is defined by the degree of similarity in tone between Democrats and Republicans.

Motivation

Exploring the impact of California's electoral reform on political polarization among elected congressional representatives is important to understand the changing political landscape shaped by electoral reforms. With only a couple of states using top-two nonpartisan primaries, California

can be a case study to understand the transition from partisan to nonpartisan primaries. By focusing on federal representatives elected in California in 2010 and 2012, this research seeks to uncover insights into the evolving nature of political discourse and ideology two years before and two years after the change in the election system.

Thesis Statement

This research will analyze political polarization among California representatives in the 112th and 113th Congress through their congressional speeches, voting history, and ideological score based on their donors to answer the following hypothesis: The implementation of a top-two primary system in replacement of a partisan primary establishes more moderation among candidates in California.

Chapter 2: Literature Review

California's Primary Election System History

Over the last three decades, California's electoral landscape for non-presidential elections has been heavily debated and undergone significant changes. Prior to 1996, the state operated under a closed primary system, which restricted participation in a political party's primary to only registered voters affiliated to that party. However, this changed in 1996 with the passage of Proposition 198, which transformed California's electoral system from a "closed" to an "open" primary system. This change allowed all registered voters, regardless of political affiliation, to cast their votes for any candidate in the primary. In 2000, the United States Supreme Court declared Proposition 198 unconstitutional, reinstating the closed primary system. The subsequent response in 2001 was the passage of Senate Bill 28. This legislation modified California's closed primary system to allow voters unaffiliated with a political party to partake in a primary election, provided they adhered to the party's rules and received notice from the Secretary of State. Under this revised closed primary, non-partisan voters who did not request a ballot form a participating party received a non-partisan ballot, only containing the names of candidates for nonpartisan offices and state and local measures to be voted on during the primary election (The County of El Dorado Elections Department, n.d.).

On June 8th, 2010 California voters approved Proposition 14, the "Top Two Candidates Open Primary Act," which replaced the state's modified closed primary with the top-two primary system (Alvarez & Sinclair, 2015). This system was similar to the one adopted in the late 1990s, except that this time, any two candidates regardless of party affiliation could advance to the general election. Reformers argue the theory that this change would create a more open primary and produce more centrist, moderate, and pragmatic candidates. The top-two primary was first implemented in California's 2012 election. Whether this transition has helped to decrease

polarization is contested, and there have been numerous studies about its effect on Congress. One study suggests that the top-two produced more competitive legislative elections in 2012, although the results were cautious about causality (Alvarez & Sinclair, 2015). A subsequent study, using a difference-in-difference design and investigating the average legislative polarization across California both pre- and post-primary reform, found that California's adoption of the top two primary system yielded mixed results in fostering moderate candidate selection. While the data suggests minimal impact on Republican moderation, there are indications of some effect within the Democratic realm, despite the fact that part of this influence may be attributed to concurrent redistricting efforts (McGhee & Shore, 2017). Given the relative newness of the top-two primary system in California, it is important to acknowledge the potential for evolving behavioral dynamics among candidates and voters as they become more accustomed to the system over time.

Network analysis for politician behavior

Politicians with similar voting patterns can be understood as having similar views and interests, thereby forming a political network. Alvarez & Sinclair (2012) explored this concept using legislative voting behavior from the California State Assembly. Their hypothesis centered around legislators elected during the years of the nonpartisan blanket primary in 1998 and 2000, positing that these individuals would be more centrally networked and more likely to collaborate with their peers. They connected each legislator if they voted similarly on the same bill, where the strength of their social network tie is determined by the total number of bills on which both legislators have agreed. From this, they were able to establish each legislator's centrality. They found that individuals who are members of the majority have increased legislative power centrality if they were first elected in a blanket primary year, but the primary had no effect on the legislative power centrality for members of the minority party. When it comes to agreeing with each other, the average rate of agreement during a blanket primary session was much higher than the average rate of agreement during a nonblanket primary session (Alvarez & Sinclair, 2012).

Similarly, Brito et al. (2020) explores a network-based framework to study the evolution of the Brazilian political system in terms of how politicians vote in proposals over time. They construct a network where deputies are represented as nodes, and edges are established between two deputies if their agreement percentage exceeds their disagreement percentage. This approach significantly influenced this research's methodology, particularly in the construction of edges and nodes in my network. However, the cut point percentage of agreement for the edges are customized to suit the research.

Natural Language Processing (NLP)

Over the years, congressional speeches have revealed significant differences in the usage of language. For example, Tucker et al. (2020) finds that the frequency of energy-related terms in congressional speeches has fluctuated, coinciding with significant events such as the

Organization of Arab Petroleum Exporting Countries (OAPEC) embargo in 1973 and the surge in gas prices in 2008. Interestingly, their research analysis shows differential usage patterns between Democratic and Republican speeches during these periods. Additionally, sentiment analysis, a NLP technique, evaluates the polarity and tonality of texts by discerning and evaluating expressions used to appraise them. In politics, sentiment analysis can gauge support for legislative initiatives or the polarization of debates (Haselmayer & Jenny, 2017). In their research of the 2013 Austrian national elections, Haselmayer and Jenny (2017) use sentiment analysis to evaluate the negative sentiment in political discourse. Natural Language Processing techniques are invaluable in understanding textual data.

Chapter 3: Data Collection and Methodology

Data Collection

This research uses Adam Bonica's Database on Ideology, Money in Politics, and Elections (DIME), which contains over 500 million itemized political contributions made by individuals and organizations across local, state, and federal elections spanning from 1972 to 2022. It also includes Bonica's "DIME scores," a set of measures gauging ideology that has been extensively validated across various studies (Bonica, 2018). Along with this, Bonica's DIME PLUS is used, which extends on DIME by incorporating data on legislative voting, lawmaking, and political discourse. The DIME PLUS Congressional Text datasets contain all proceedings, floor debates, and extensions of remarks for Congress scraped from the GPO's Federal Digital System, while the Congressional Votes provides data on representatives voting patterns on legislative bills. These datasets span the period from 2003 to 2014.

I restructured and filtered Bonica's datasets to tailor the data to fit the specific needs of my research. In the Congressional Text data from DIME PLUS, a candidate name column was absent and only Bonica's unique candidate ID was in the dataset. I added a column correlating the unique candidates IDs with their respective names. Furthermore, to focus the analysis, I filtered to include only individuals from the House of Representatives representing California.

Ideological Change

The first aspect of my project revolves around evaluating the ideological change among candidates who were elected in the 2010 California election and subsequently ran as incumbents in the 2012 California election. The reason I chose to focus on incumbents is because the majority of representatives elected in both 2010 and 2012 were incumbents, making a paired comparison plausible. I measured their ideological position using Bonica's (2023) two-dimensional candidate recipient scores, which are period-specific estimations of candidate's ideology derived from donation-based metrics. I then use a paired sample t-test to discern any

statistically significant changes in their ideological score over the time period two years before the primary change and two years after the primary change.

Additionally, I use a difference-in-difference analysis to examine the influence of California's transition to a top-two primary system on the average ideology calculated from Bonica's (2023) candidate recipient scores of both Republicans and Democrats from 2006 to 2016. I included years beyond the scope of the 112th and 113th Congresses to provide the model with a broader dataset, enhancing its ability to make precise predictions regarding the average ideology of California's representatives post-2012. The selected control states–Nevada, New Jersey, New York, and Maryland–were chosen because of their closed or semi-closed primary systems, mirroring California's before the implementation of the top-two primary system. Further, these states had the same presidential election outcomes as California during these years. The treatment variable is California, while the post-treatment period spans from 2012 onward. The formula for the OLS model is as follows:

$$Y_{it} = \beta_0 + \beta_1 California_i + \beta_2 PostTreatment_t + \beta_3 (California_i * PostTreatment_t)$$

Where Y_{it} is the average ideology calculated from Bonica's candidate recipient scores. $California_i$ is the binary variable equal to 1 if the observation is from California, and 0 otherwise. $PostTreatment_t$ is the binary variable equal to 1 for observations 2012 onward, and 0 for observations before 2012. This difference-in-difference analysis controls for the following factors:

1. *California vs. Control States:* the treatment variable is whether the observation is from California or not from California. This binary variable allows for the distinction between the treatment group (California) and the control group (other states).
2. *Time Period:* This variable differentiate the post-treatment period (from 2012 onward) from the pre-treatment period (pre-2012) and assess how the treatment effect changes over time.
3. *Interactive term:* This variable is the combined effect of being in California during the post-treatment period. This is needed to estimate the differential impact of the treatment between California and the controlled states over time.

Network Analysis

The second aspect of my project involves constructing two groups of two networks with different cut points representing the voting behavior of the 112th and 113th Congresses. In these networks, nodes represent individual representatives, and edges reflect their voting alignments. The criteria for creating edges are as follows:

1. *Intra-party connections:* For the first group, an edge is formed between two representatives of the same party if they vote the same way on a bill more than 86% of the time. For the second group, the cut point for intra-party connection is 85%.

2. *Inter-party connections*: For the first group, an edge is established between two representatives of different parties if they vote the same way on a bill more than 17% of the time. For the second group, the cut point for inter-party connection is 18%.

These cut points are determined based on observations that if intra and inter-party connections were given equal importance, the network would predominantly form edges solely within their own parties. The cut points were adjusted to address this imbalance and detect subtle indications of polarization between parties. They reflect common probabilities observed within intra and inter-party contexts, and aim to capture the nuanced levels of polarization between parties.

If a congressperson abstains from voting on a bill, that bill will be excluded from the representatives' proportion. To make this clearer, consider a hypothetical scenario in the first group involving Representative A and B, both from the Pink Party, and bills 1, 2, and 3. If Representative A votes yes for bills 1 and 2 but abstains from voting on bill 3, and Representative B votes no on bill 1 and yes on bills 2 and 3, the proportion between Representative A and B would be calculated as 0.5. Because an intra-party edge is formed between two representatives if they have a proportion of more than 0.86, Representative A and B would not have an edge between them. The framework in the next approach to analyze the relationships among representatives can be summarized in the following:

1. *Community detection and analysis:* The Louvain method is used for community detection to assign representatives to distinct communities based on the network's structure. This is a heuristic method that optimizes modularity–a metric gauging the network's segmentation into cohesive groups (Blondel et al. 2008).
2. *Network characterization:* I extract metrics from the network to quantify political concepts like fragmentation and polarization. Fragmentation offers insights into the internal coherence of party affiliations and is measured by the average degree centrality of intra-party connections. Polarization is assessed by examining the average degree and betweenness centrality of inter-party connections. It is also measured among the detected communities by summing the inter-community edges.

Sentiment Analysis

The final component of my thesis involves evaluating sentiment scores on the speech and debate topics addressed by California Congress members. In the dataset provided by Bonica (2016), each textual document spoken by a member includes a weight assigned to each topic discussed. For my research, I filtered this data by assigning each speech or debate to the topic with the highest weight. Any speech where a topic other than the highest one comprises more than 20% of the weight is excluded from my analysis.

I preprocessed the text from the filtered dataset by eliminating all stopwords, such as "the," "a," "is," "are," etc., which are irrelevant in assessing the sentiment of the document. After this, I employed VADER Sentiment Analysis (vaderSentiment 3.3.2, 2018), a lexicon and

rule-based sentiment analysis tool, to evaluate the sentiment of each topic. I compounded the sentiment score of each document, offering a unified metric that captures the overall positivity or negativity of the text. A higher compound score suggests a more positive sentiment, while a lower score indicates a more negative sentiment.

I initially conducted a visual comparison of sentiment score changes for both Democrats and Republicans. I also visually examined the changes in Euclidean distance of sentiment scores by topic between the average scores for Republicans and Democrats. Given the limited representation of newly elected Congress representatives from California in the 112th Congress, only two individuals, I decided against comparing solely the newly elected members of the 112th and 113th Congress due to the scarcity of data points. Instead, I focused on comparing the sentiment scores of incumbents, categorized by party affiliation. This comparison was conducted across speeches addressing the four most prevalent topics discussed in Congress among California representatives: the economy, federal agencies and government regulation, Congress and procedural matters, and abortion and social conservatism. The remaining topics had much fewer speeches in comparison when filtering for those reelected in 2012, with the fifth most prevalent topic comprising only 81 documents.

Chapter 4: Findings and Analysis

Ideological Change

Figure 1 displays Bonica's period-specific candidate ideology score of elected representatives in 2010, two years prior to the top-two primary change, and for the same representatives who were reelected after the primary change in 2012. The light blue circles represent the representative's ideology in 2010, and dark blue represents the ideology in 2012. By convention, negative values indicate a more liberal stance, while positive values are a more conservative stance. Placements near zero suggest a more moderate position. For Democrats, an increase in ideology towards zero signifies greater moderation, while for Republicans, a decrease towards zero indicates a move towards moderation. Some candidates show signs of slight moderation. However, it appears that the overall impact is minimal. Table 1 presents the mean and median values of incumbent representative ideology before and after the primary change. The average ideology for incumbent Democrats shifted further from zero, both in terms of mean and median, following the election change. Conversely, incumbent Republicans experienced a shift towards zero in both average and median ideology points after the election change.

Figure 1: Average Incumbent Ideology Comparison
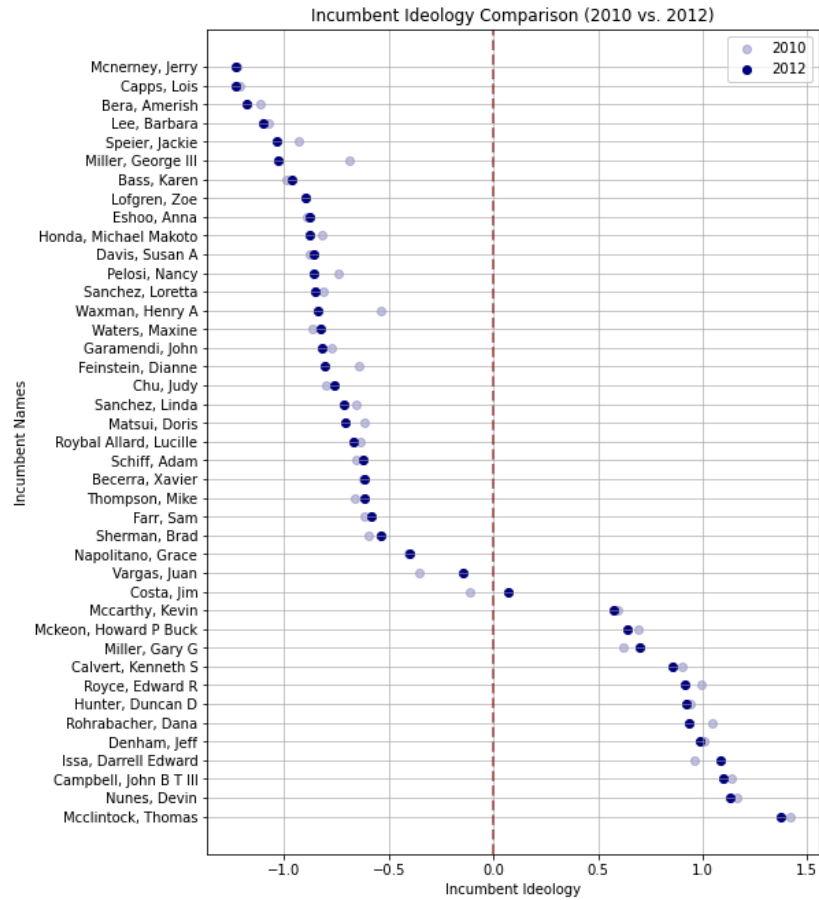
Table 1: Average and Median Ideal Points Before and After Reform

	Pre-reform Mean	Pre-reform Median	Post-reform Mean	Post-reform Median
Democrats	-0.751	-0.739	-0.778	-0.828
Republicans	0.956	0.975	0.935	0.929

To verify whether there was any statistically significant change in incumbent representative ideology, I conducted a paired t-test on the ideologies of Republicans and Democrats in the 112th Congress compared to their ideologies when they were elected in the 113th Congress. The t-statistic for Democrats is 1.292 with a corresponding p-value of 0.206. As the p-value exceeds the significance level of 0.05, we fail to reject the null hypothesis, suggesting no significant difference in ideologies for Democrats. Similarly, for Republicans, the t-statistic is 1.146 with a p-value of 0.272, leading to the same conclusion of failing to reject the null hypothesis and

indicating no significant difference in ideology for Republicans before and after the primary change.

        Tables 2 and 3 show the OLS regression results of the difference-in-difference analysis conducted on the average ideology of Democrats and Republicans. In Table 2, the coefficient associated with the difference-in-difference (DID) interaction term represents the difference in change in California's average ideology for Democrats after the implementation of the top-two primary system relative to the control group of Democrat representatives in similar states. With a p-value of 0.553, the indicator coefficient suggests that the effect of the top-two primary on the average ideology of California Democratic representatives is not statistically significant at the 0.05 significance level. Although the coefficient value of 0.0872 indicates a small positive effect of the top-two primary, it is not statistically distinguishable from zero.

        In Table 3, the indicator coefficient value is 0.0840, which indicates a small positive effect of the top-two primary on the average ideology of California Republican representatives. However, the p-value associated with this coefficient is 0.725, meaning that the effect is not statistically significant. These findings conclude that there is insufficient evidence to reject the null hypothesis, suggesting that the top-two primary system did not have a significant impact on the average ideology, calculated from Bonica's (2023) candidate scores, for both Democrats and Republicans in California.

Table 2: DID OLS Regression Results for Democrat Ideology Scores

	coef	std err	t	P > |t|	[0.025	0.975]
constant	-0.9071	0.046	-19.767	0.000	-1.001	-0.813
treatment	-2.009	0.103	-1.958	0.061	-0.412	0.010
post	0.0675	0.065	1.041	0.308	-0.066	0.201
DID	0.0872	0.145	0.601	0.553	-0.211	0.385

Table 3: DID OLS Regression Results for Republicans Ideology Scores

	coef	std err	t	P > |t|	[0.025	0.975]
constant	1.0598	0.075	14.170	0.000	0.906	1.214
treatment	-0.0025	0.167	-0.015	0.988	-0.346	0.341
post	-0.0750	0.106	-0.709	0.485	-0.292	0.142
DID	0.0840	0.237	0.355	0.725	-0.402	0.570

Network Analysis

Figure 2 and Figure 3 show the network of the 112th and 113th California representatives for the first group, where the intra-party weight is 86% and inter-party weight is 17%. Intra-party edges are navy and inter-party edges are green.

Figure 2: Group 1 Network of California Representatives in 112th Congress
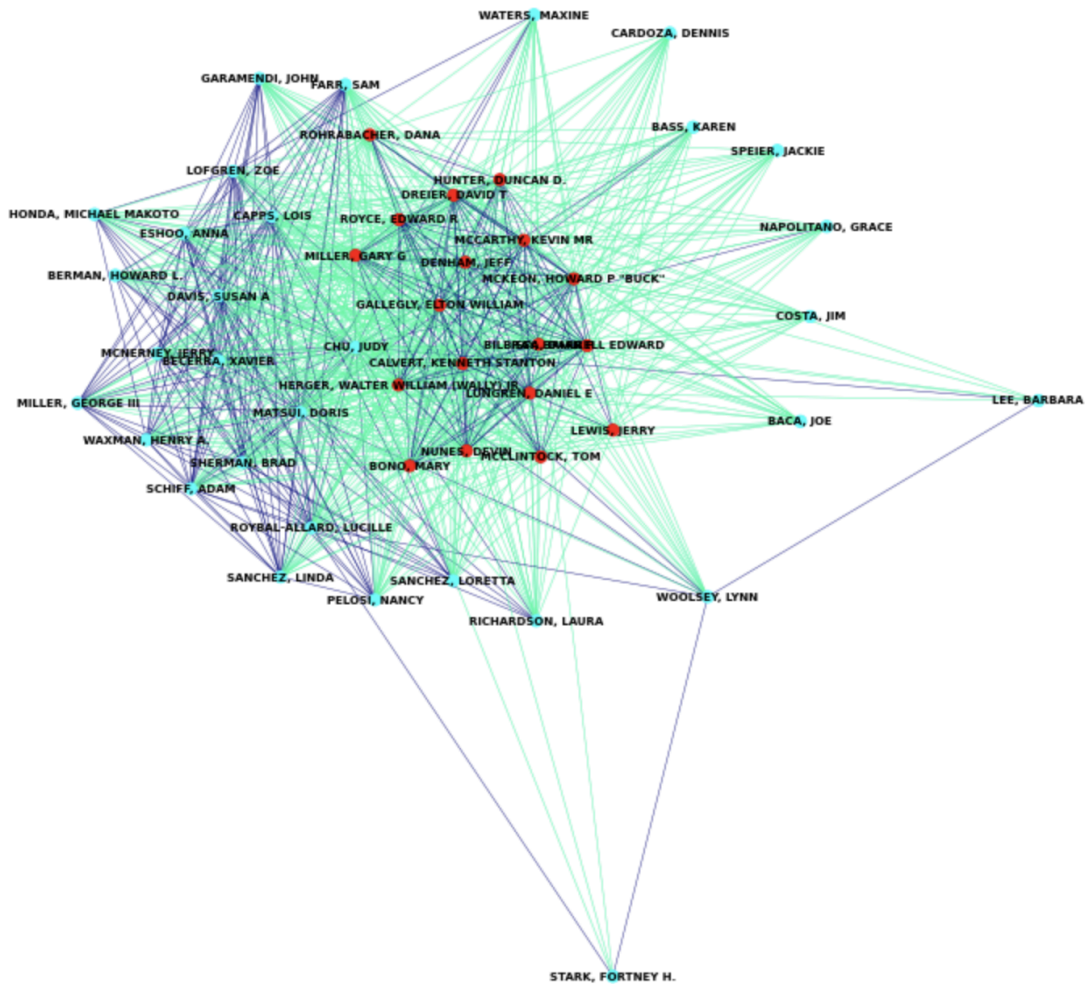
Figure 3: Group 1 Network of California Representatives in 113th Congress
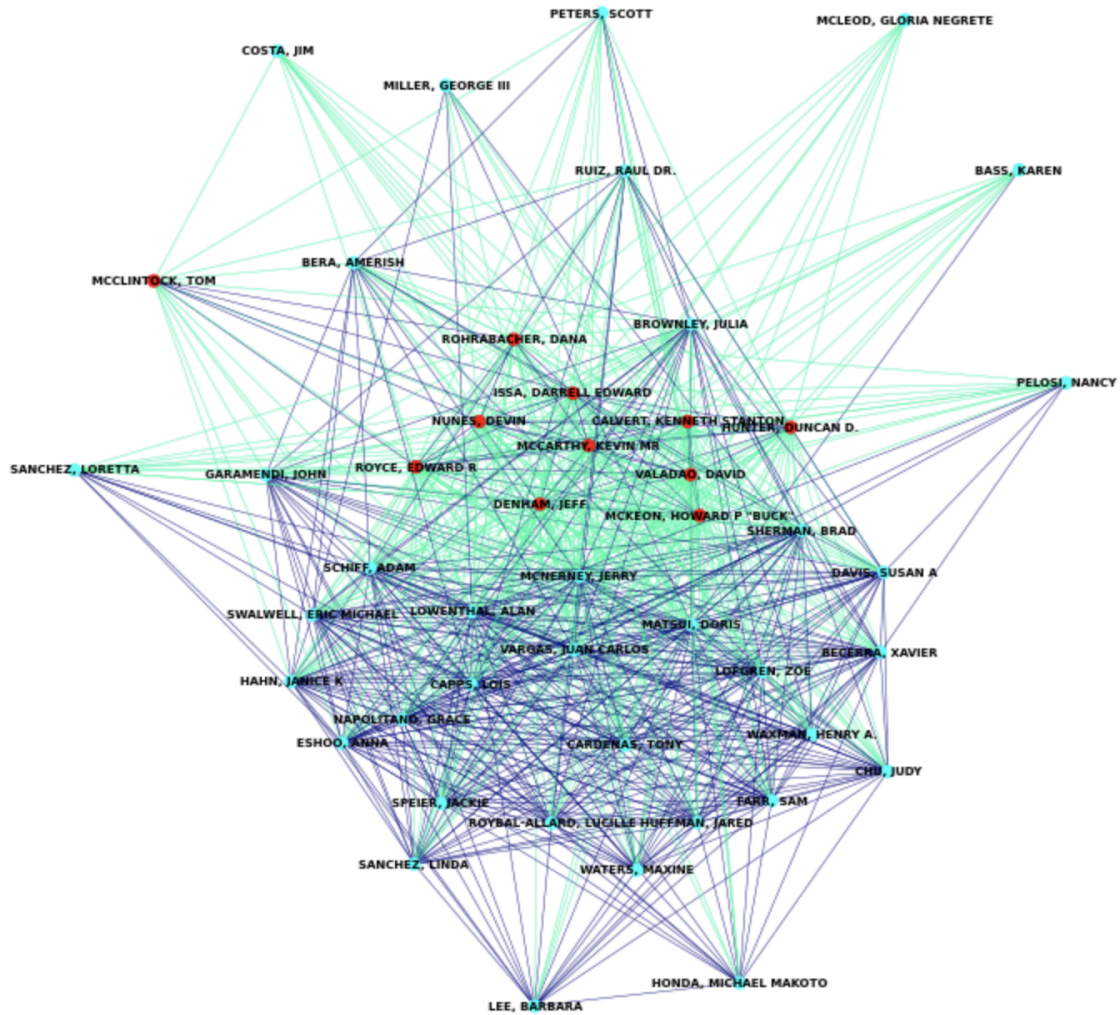
In the community detection algorithm, the community is where representatives in Congress are grouped together based on their connections in the network. The algorithm identifies cohesive communities within the network where representatives are densely connected to each other but sparsely connected to representatives clustered in other groups. The community detection algorithm's output shown in Table 4 and Table 5 aims to uncover patterns of relationships within the network of representatives across the 112th and 113th Congresses. While these tables indicate minimal variation in community assignments for California representatives across the different Congresses, in the 112th Congress, three Republicans were grouped in Community 1 alongside other Democrats. In contrast, in the 113th Congress, all Republicans were clustered into a single community, suggestive of higher polarization than in the 112th.

Table 4: Group 1 Network Community of 112th California Representatives

Party	Community 0	Community 1
D	'RICHARDSON, LAURA', 'SPEIER, JACKIE', 'BASS, KAREN', 'WATERS, MAXINE', 'WOOLSEY, LYNN', 'SANCHEZ, LORETTA', 'LEE, BARBARA', 'NAPOLITANO, GRACE', 'BACA, JOE', 'STARK, FORTNEY H.', 'BERMAN, HOWARD L.', 'HONDA, MICHAEL MAKOTO' ,'CARDOZA, DENNIS', 'COSTA, JIM'	'CHU, JUDY', 'GARAMENDI, JOHN', 'MATSUI, DORIS', 'ESHOO, ANNA', 'FARR, SAM', 'BECERRA, XAVIER', 'ROYBAL-ALLARD, LUCILLE', 'LOFGREN, ZOE', 'SHERMAN, BRAD', 'CAPPS, LOIS', 'MILLER, GEORGE III', 'WAXMAN, HENRY A.', 'PELOSI, NANCY', 'ROHRABACHER, DANA', 'SCHIFF, ADAM', 'DAVIS, SUSAN A', 'SANCHEZ, LINDA', 'MCNERNEY, JERRY'
R	'HUNTER, DUNCAN D.', 'DENHAM, JEFF', 'MCKEON, HOWARD P "BUCK"', 'ROYCE, EDWARD R', 'CALVERT, KENNETH STANTON', 'BILBRAY, BRIAN P', 'BONO, MARY','MILLER, GARY G','LUNGREN, DANIEL E', 'DREIER, DAVID T', 'GALLEGLY, ELTON WILLIAM', 'HERGER, WALTER WILLIAM (WALLY) JR', 'ISSA, DARRELL EDWARD', 'NUNES, DEVIN', 'MCCARTHY, KEVIN MR'	'MCCLINTOCK, TOM', 'LEWIS, JERRY', 'ROHRABACHER, DANA'

Table 5: Group 1 Network Community of 113th California Representatives

Party	Community 0	Community 1
D	'SPEIER, JACKIE', 'CHU, JUDY', 'HAHN, JANICE K', 'WATERS, MAXINE', 'MATSUI, DORIS', 'ESHOO, ANNA', 'FARR, SAM', 'BECERRA, XAVIER', 'ROYBAL-ALLARD, LUCILLE', 'LOFGREN, ZOE', 'SHERMAN, BRAD', 'CAPPS, LOIS', 'LEE, BARBARA', 'NAPOLITANO, GRACE', 'WAXMAN, HENRY A.', 'LOWENTHAL, ALAN', 'HUFFMAN, JARED', 'SWALWELL, ERIC MICHAEL', 'CARDENAS, TONY', 'VARGAS, JUAN CARLOS', 'HONDA, MICHAEL MAKOTO', 'SCHIFF, ADAM', 'DAVIS, SUSAN A', 'SANCHEZ, LINDA', 'MCNERNEY, JERRY'	'GARAMENDI, JOHN', 'BASS, KAREN', 'SANCHEZ, LORETTA', 'MILLER, GEORGE III', 'BERA, AMERISH', 'BROWNLEY, JULIA', 'MCLEOD, GLORIA NEGRETE', 'RUIZ, RAUL DR.', 'PETERS, SCOTT', 'PELOSI, NANCY', 'COSTA, JIM',
R		'HUNTER, DUNCAN D.', 'DENHAM, JEFF', 'MCKEON, HOWARD P "BUCK"', 'ROYCE, EDWARD R', 'CALVERT, KENNETH STANTON', 'MCCLINTOCK, TOM', 'VALADAO, DAVID', 'ROHRABACHER, DANA', 'ISSA, DARRELL EDWARD', 'NUNES, DEVIN', 'MCCARTHY, KEVIN MR'

In the 112th Congress community, the total count of inter-community edges amounted to 339, whereas in the 113th Congress, the sum of edges was 311. There were 28 fewer inter-community edges in the 113th Congress, suggesting a heightened level of polarization among the communities compared to the 112th Congress.

Analyzing centrality measures, where I took the subgraph of solely Republicans and solely Democrats, the average inter-party betweenness centrality for Republicans and Democrats can be shown in Table 6.

Table 6: Inter-Party Betweenness Centrality

	Mean (Republicans)	Mean (Democrats)
112th	0.025069	0.005634
113th	0.059113	0.002630

The average inter-party betweenness centrality for Republicans increased 0.034 from the 112th to the 113th Congress, suggesting a trend towards greater moderation as Republicans aligned more frequently with Democrats in their voting patterns. Conversely, the average inter-party betweenness centrality for Democrats witnessed a slight decrease of 0.003 during the same period, indicating a divergence from shared voting patterns with Republicans. This suggests that Republicans had increased mediation among Democrats and more bipartisanship, while Democrats had the opposite over the same period.

Table 7: Inter-party and Intra-Party Degree Centrality

	Mean (Republicans)	Mean (Democrats)
112th (Inter)	0.594907	0.345430
113th (Inter)	0.697628	0.213164
112th (Intra)	0.297980	0.239003
113th (Intra)	0.181818	0.449495

In terms of average inter-party degree centrality, Republicans exhibit a trend similar to their betweenness measures. The average degree centrality increased by 0.103 from the 112th to the 113th Congress, indicating heightened direct interactions with representatives from different parties and better connectivity. In contrast, Democrats experienced a decrease in their average inter-party degree centrality by 0.132 during the same period, suggesting reduced connectivity with Republicans.

Interestingly, Republicans witnessed a decrease in their intra-party average degree centrality by 0.115, indicative of increased fragmentation and weakened intra-party communication from the 112th to the 113th Congress. In contrast, Democrats saw an increase in their average intra-party degree centrality by 0.210, signifying decreased fragmentation and greater cohesion within the party over the same period.

For Group 2, intra-party edges are formed when the proportion of shared votes among representatives are greater than 0.85, whereas inter-party edges are defined when this proportion exceeds 0.18. Notably, Group 2's network exhibited similar trends in all the centrality averages among parties as Group 1 as evidenced by Appendix 3 and 4. However, what distinguishes Group 2 is the outcome of the community detection algorithm. In the 112th Congress, the community detection algorithm categorized the representatives into three communities for this group, suggesting the emergence of an additional independent community beyond the typical two communities it usually identifies.

Table 8: Group 2 Network Community of 112th California Representatives

Party	Community 0	Community 1	Community 2
D	'CHU, JUDY', 'WATERS, MAXINE', 'MATSUI, DORIS', 'WOOLSEY, LYNN', 'BECERRA, XAVIER', 'ROYBAL-ALLARD, LUCILLE', 'LEE, BARBARA', 'STARK, FORTNEY H.', 'MILLER, GEORGE III', 'HONDA, MICHAEL MAKOTO', 'SANCHEZ, LINDA'	'GARAMENDI, JOHN', "ESHOO, ANNA', 'FARR, SAM', 'LOFGREN, ZOE', 'SHERMAN, BRAD', 'SANCHEZ, LORETTA', 'CAPPS, LOIS', 'WAXMAN, HENRY A.', 'PELOSI, NANCY', 'SCHIFF, ADAM', 'DAVIS, SUSAN A', 'MCNERNEY, JERRY'	'RICHARDSON, LAURA', 'SPEIER, JACKIE', 'BASS, KAREN', 'NAPOLITANO, GRACE', 'BACA, JOE', DAVID T', 'BERMAN, HOWARD L.', 'CARDOZA, DENNIS', 'COSTA, JIM'
R	'BILBRAY, BRIAN P', 'BONO, MARY'	HUNTER, DUNCAN D.', 'ROYCE, EDWARD R', 'MCCLINTOCK, TOM', ROHRABACHER, DANA'	DENHAM, JEFF', 'MCKEON, HOWARD P "BUCK"', 'CALVERT, KENNETH STANTON', 'MILLER, GARY G', 'LEWIS, JERRY', 'LUNGREN, DANIEL E', 'DREIER, 'GALLEGLY, ELTON WILLIAM', 'HERGER, WALTER WILLIAM (WALLY) JR', 'ISSA, DARRELL EDWARD', 'NUNES, DEVIN', 'MCCARTHY, KEVIN MR'

Table 9: Group 2 Network Community of 113th California Representatives

Party	Community 0	Community 1
D	'SPEIER, JACKIE', 'CHU, JUDY', 'HAHN, JANICE K', 'WATERS, MAXINE', 'MATSUI, DORIS', 'ESHOO, ANNA', 'FARR, SAM', 'BECERRA, XAVIER', 'ROYBAL-ALLARD, LUCILLE', 'LOFGREN, ZOE', 'SHERMAN, BRAD', 'SANCHEZ, LORETTA', 'CAPPS, LOIS', 'LEE, BARBARA', 'NAPOLITANO, GRACE', 'MILLER, GEORGE III', 'WAXMAN, HENRY A.', 'LOWENTHAL, ALAN', 'HUFFMAN, JARED', 'SWALWELL, ERIC MICHAEL', 'CARDENAS, TONY', 'VARGAS, JUAN CARLOS', 'HONDA, MICHAEL MAKOTO', 'SCHIFF, ADAM', 'DAVIS, SUSAN A', 'SANCHEZ, LINDA'	'GARAMENDI, JOHN', 'BASS, KAREN', 'BERA, AMERISH', 'BROWNLEY, JULIA', 'MCLEOD, GLORIA NEGRETE', 'RUIZ, RAUL DR.', 'PETERS, SCOTT', 'PELOSI, NANCY','COSTA, JIM', 'MCNERNEY, JERRY'
R		'HUNTER, DUNCAN D.', 'DENHAM, JEFF', 'MCKEON, HOWARD P "BUCK"', 'ROYCE, EDWARD R', 'CALVERT, KENNETH STANTON', 'MCCLINTOCK, TOM', 'VALADAO, DAVID', 'ROHRABACHER, DANA', 'ISSA, DARRELL EDWARD', 'NUNES, DEVIN', 'MCCARTHY, KEVIN MR'

The total count of inter-community edges for the 112th Congress is 486, 159 edges higher than the 327 inter-community edges observed in the 113th Congress. There were more inter-community edges formed in the 112th Congress, suggesting that there was less polarization among communities compared to the 113th Congress.

Sentiment Analysis

Table 10 displays the top ten words for positive, negative, and neutral sentiment in both the 112th and 113th Congresses. The positive words remain consistent between the two Congresses, as do many of the negative words. For neutral sentiment, both Congresses prominently feature military-related terms such as "gulf," "war," and "strike" in the 112th Congress and the top neutral words in the 113th Congress including "veteran" and "military."

Table 10: Positive, Negative and Neutral Words for 112th and 113th Congress

112th, Positive	113th, Positive	112th, Negative	113th, Negative	112th, Neutral	113th, Neutral
year	year	year	year	line	veteran
state	state	state	state	strike	claim
job	community	job	government	program	military
california	california	war	child	year	day
community	service	woman	congress	war	national
work	work	republican	woman	gulf	park
program	family	million	family	gulf war	national park
health	program	cut	million	section	medical
service	country	health	country	2	condition
support	make	program	republican	illness	medical condition

Examples of speeches showing the sentiment score from Congressperson Thomas McClintock (R) and Congressperson Barbara Lee (D) from the 112th and 113th Congresses can be found in Appendix 7.

Figure 4 illustrates the average sentiment scores among Democrats by topic across the 112th and 113th Congresses, while Figure 5 depicts the corresponding scores among Republicans. Notably, the average sentiment score for women's issues experienced the most significant change for Democrats, becoming notably more positive between the 112th and 113th Congresses. Among Republicans, the average sentiment score for guns had the most substantial shift, displaying a more positive trend in the 113th Congress.

Figure 4: Average Sentiment Scores for Democrats by Topic
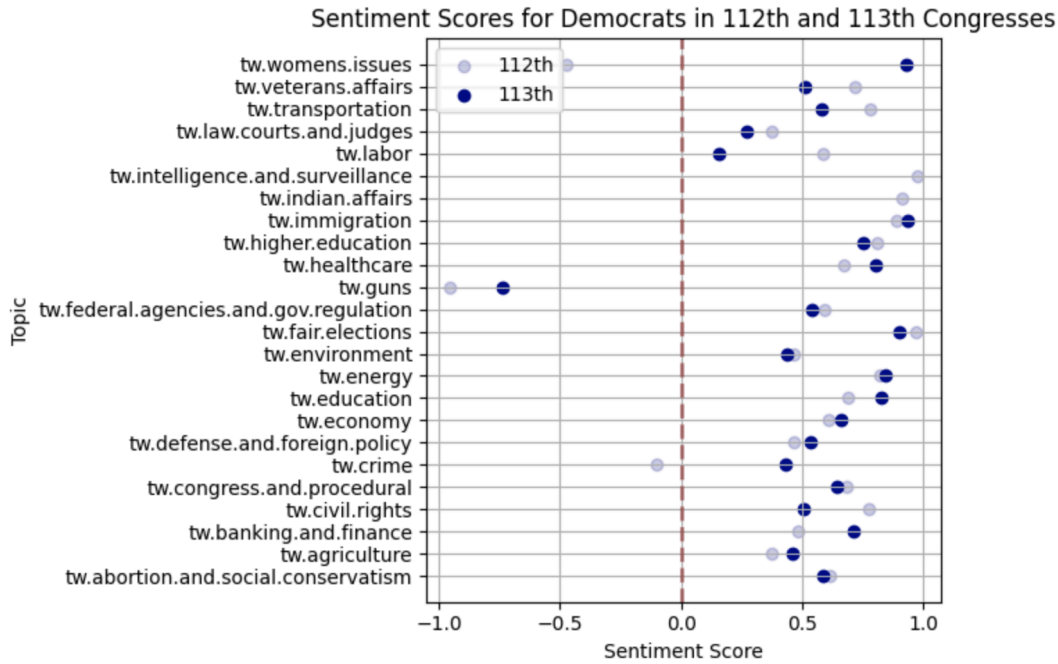
Figure 5: Average Sentiment Scores for Republicans by Topic
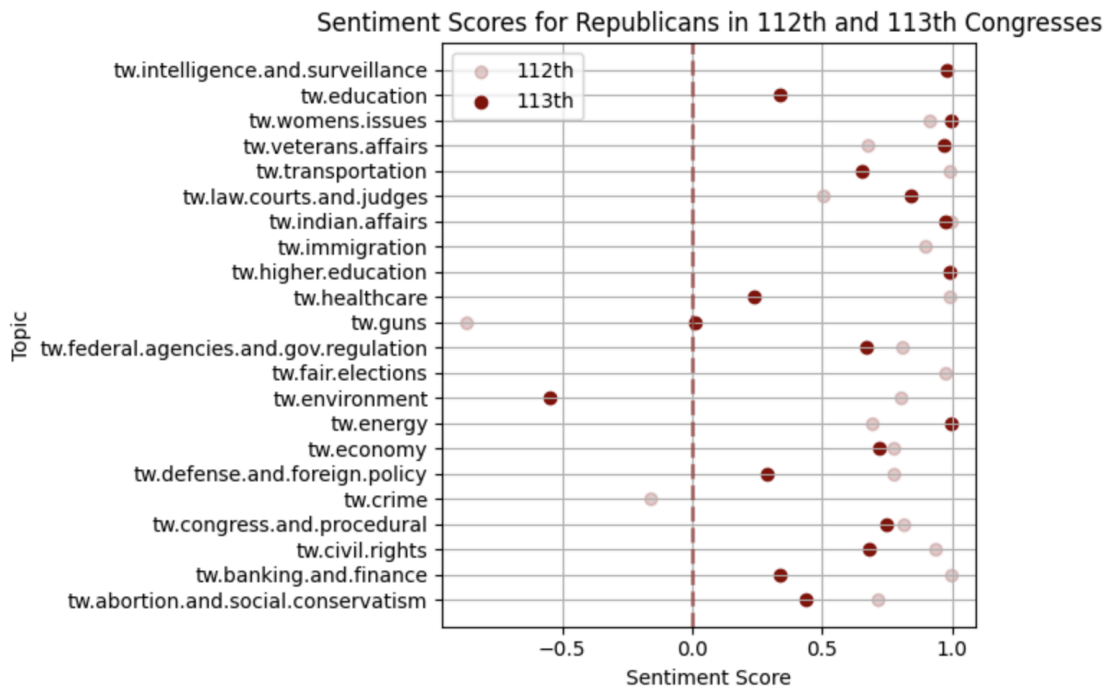
Figure 6 compares the Euclidean distance of the average sentiment score by topic between Republicans and Democrats, showing about half of the topics increasing in distance from the 112th to the 113th Congress, hinting at a discernible polarization in sentiment. However, notably, the average sentiment change in women's issues witnessed a drastic decrease from the 112th Congress to the 113th Congress. Nonetheless, it is important to note the limited relevance for comparison due to the limited 22 documents classified under women's issues among California representatives in both the 112th and 113th Congresses combined.

Figure 6: Euclidean Distance of Average Sentiment Scores by Topic between Republicans and Democrats
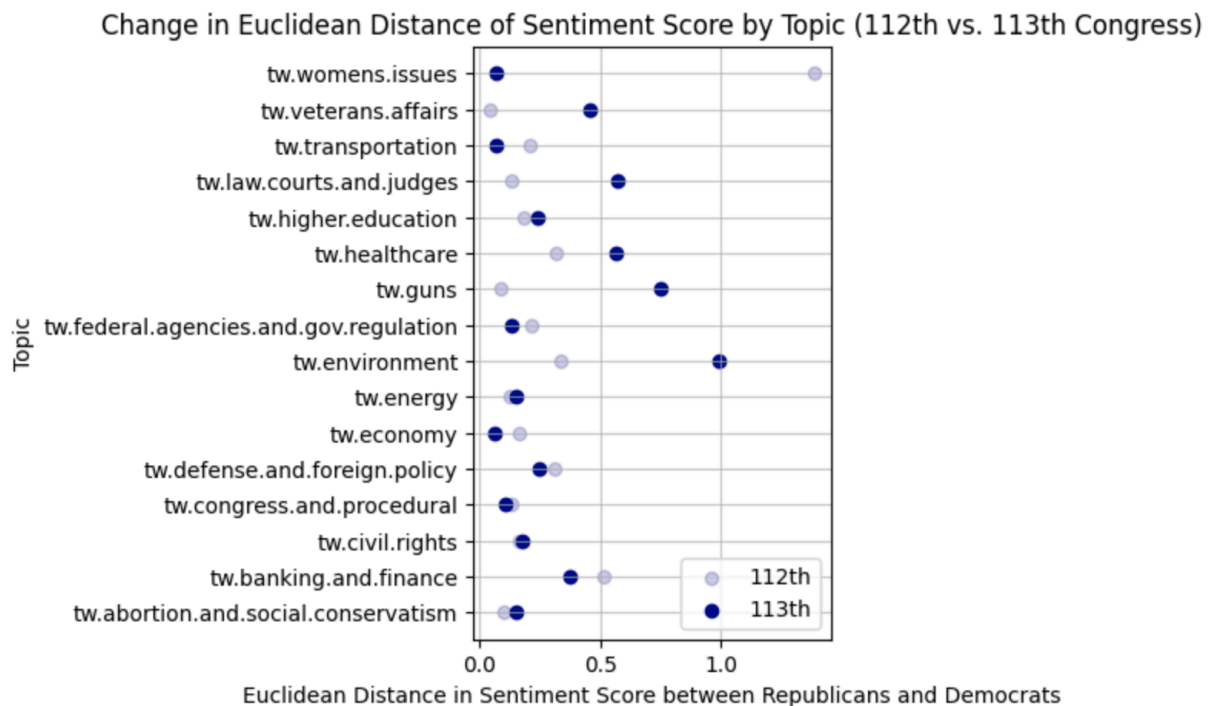
Table 9 shows the outcomes of the Wilcoxon Test conducted among Democrats and Republicans for the topics economy, federal agencies and government regulation, Congress and procedural matters, as well as abortion and social conservatism to determine whether any noteworthy changes occurred in sentiment analysis. According to these findings, no significant alterations were observed for these topics.

Table 11: Wilcoxon Test Values for Sentiment by Topic

Topic	Democrat p-value	Republican p-value
Economy	0.771	0.734
Federal agencies and government regulation	0.287	0.910
Abortion and social conservatism	0.119	0.910
Congress and procedural	0.325	0.734

Chapter 5: Discussion

Discussion

The result of this research suggests that California's adoption of the top-two primary system had minimal impact on polarization during the two years following its implementation in 2012. This conclusion comes from an analysis of data pertaining to California representatives in the 112th and 113th Congresses, encompassing their congressional speeches, voting records, and ideological leanings gauged by donor affiliations. This research does not provide evidence for the hypothesis that the introduction of a top-two open primary system would foster greater moderation among candidates. However, our network analysis reveals a trend in the voting behavior of Republican representatives from California. Specifically, there appears to be an increase in average inter-party betweenness and degree centrality in the 113th Congress.

There are several reasons that could contribute to the lack of strong observable changes in polarization during the two-year period following California's implementation of the top-two open primary system. First, it's important to acknowledge that significant shifts in political dynamics often require time to show, and the scope of this research is limited to examining the outcomes of elected representatives from 2012 to 2014. Additionally, it's possible that the top-two primary's impact of reducing polarization was stronger at the state legislature level. Further, the effects of the top-two primary in mitigating polarization may be strengthened when complemented by other reform measures. For instance, Alaska's adoption of ranked-choice voting alongside the top-four primary shows a dual approach that could potentially yield more substantial reductions in polarization.

Limitations and Future Research

It is important to acknowledge the constraints inherent to this study. One limitation is that the bulk of this analysis focuses exclusively on incumbents from the 112th and 113th Congresses, examining whether the adoption of the top-two primary influenced their polarization. This will introduce bias because there is a higher chance of reelection among incumbents, which would minimize the effect of the change in voting rule. Because of California's redistricting in 2010, this research did not directly compare candidates elected from individual districts before and after this transition. For future investigations, using a genetic matching algorithm to match California districts to similar districts where the top two primary was not in place could offer more robust insights for comparing California representatives who participated in the system.

Chapter 6: Conclusion

In this paper, I explored the repercussions of California's top-two primary system on polarization among federal representatives from the state. I approached this through multiple lenses. First, I examined ideological change using Bonica's candidate recipient scores, gauging how donors view candidates. The second way is through network analysis, showing how candidates are interacting with each other through voting patterns. The last way is through sentiment analysis, which shows how candidates are projecting themselves. The network analysis revealed a pattern of Republicans aligning more closely with Democrats in the 113th Congress compared to the 112th Congress. In contrast, Democrats aligned less with Republicans in the 113th Congress. This observation suggests that there may be varying degrees of modernization and extremism within each party, particularly in California, a predominantly Democratic state where voters tend to lean blue. However, despite these insights, the findings yielded inconclusive evidence regarding the system's effectiveness in promoting moderation among both Republican and Democratic members of the House of Representatives in California.

Bibliography

Alvarez, R., & Sinclair, B. (2012). Electoral Institutions and Legislative Behavior: The Effects of Primary Processes. *Sage Publications*, 544-557.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 1-11.

Bonica, A. (2016). DIME Plus. [Computer file]. Stanford, CA: Stanford University Libraries. ‹https://data.stanford.edu/dime-plus›.

Bonica, A. (2018). Compendium of Validation Results for DIME scores.

Bonica, A. (2023). Database on Ideology, Money in Politics, and Elections: Public version 3.1 [Computer file]. Stanford, CA: Stanford University Libraries. https://data.stanford.edu/dime.

Brito, A. C., Silva, F. N., & Amancio, D. R. (2020). A complex network approach to political analysis: Application to the Brazilian Chamber of Deputies. *PLoS One*, 15(3).

California Proposition 14, Top-Two Primaries Amendment (2010, June) Retrieved from Ballotpedia website: https://ballotpedia.org/California_Proposition_14,_Top-Two_Primaries_Amendment_(June_2010)

C. Tucker, E., Capps, C. J., & Shamir, L. (2020). A data science approach to 138 years of congressional speeches. *Heliyon*, 1-8.

Grose, C. (2018, June 5). How California's "Jungle Primary" System Works (A. Shapiro, Interviewer). Retrieved from https://www.npr.org/2018/06/05/617250124/how-californias-jungle-primary-system-works

Haselmayer, M., & Jenny, M. (2017). Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *Qual Quant*, 2623-2646.

Lopez, A. (2023, September 18). The U.S. has a "primary problem," say advocates who call for new election systems. Retrieved from NPR website: https://www.npr.org/2023/09/18/1199318220/nonpartisan-open-primaries-explainer

McGhee, E., & Shor, B. (2017). Has the Top Two Primary Elected More Moderates? *Cambridge University Press*, 1053-1066.

Meyerson, H. (2014, June 21). OP-Ed: California's jungle primary: Tried it. Dump it. Retrieved from LA Times website: https://www.latimes.com/opinion/op-ed/la-oe-meyerson-california-jungle-primary-20140622-story.html

Nagourney, A. (2018, May 24). Here's How California's 'Jungle Primary' System Works. Retrieved from New York Times website: https://www.nytimes.com/2018/05/24/us/california-primary-election-rules-system.html

OpenAI. (2024). ChatGPT [Computer software]. Retrieved from https://openai.com/chatgpt

Primary election types by state. (2023, August). Retrieved from Ballotpedia website:

https://ballotpedia.org/Primary_election_types_by_state

Alvarez, R. M. & Sinclair, J. A. (2015). *Nonpartisan Primary Election Reform: Mitigating Mischief*. Cambridge University Press.

The County of El Dorado Elections Department. *History Behind California's Primary Election System*.

vaderSentiment 3.3.2. (2018, April 23). Retrieved from PyPI website: https://pypi.org/project/vaderSentiment/

Appendix

Appendix 1: DID OLS Regression Results for Democrat Ideology Scores

```
                          OLS Regression Results
==============================================================================
Dep. Variable:        average_ideology   R-squared:                       0.217
Model:                             OLS   Adj. R-squared:                  0.127
Method:                  Least Squares   F-statistic:                     2.401
Date:                 Mon, 04 Mar 2024   Prob (F-statistic):             0.0906
Time:                         19:19:59   Log-Likelihood:                 14.751
No. Observations:                   30   AIC:                            -21.50
Df Residuals:                       26   BIC:                            -15.90
Df Model:                            3
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.9071      0.046    -19.767      0.000      -1.001      -0.813
treatment     -0.2009      0.103     -1.958      0.061      -0.412       0.010
post           0.0675      0.065      1.041      0.308      -0.066       0.201
did            0.0872      0.145      0.601      0.553      -0.211       0.385
==============================================================================
Omnibus:                        3.390   Durbin-Watson:                   1.516
Prob(Omnibus):                  0.184   Jarque-Bera (JB):                1.931
Skew:                           0.462   Prob(JB):                        0.381
Kurtosis:                       3.832   Cond. No.                         6.85
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Appendix 2: DID OLS Regression Results for Republican Ideology Scores

```
                          OLS Regression Results
==============================================================================
Dep. Variable:        average_ideology   R-squared:                       0.023
Model:                             OLS   Adj. R-squared:                 -0.090
Method:                  Least Squares   F-statistic:                    0.2053
Date:                 Mon, 04 Mar 2024   Prob (F-statistic):              0.892
Time:                         19:20:15   Log-Likelihood:               0.096340
No. Observations:                   30   AIC:                             7.807
Df Residuals:                       26   BIC:                             13.41
Df Model:                            3
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.0598      0.075     14.170      0.000       0.906       1.214
treatment     -0.0025      0.167     -0.015      0.988      -0.346       0.341
post          -0.0750      0.106     -0.709      0.485      -0.292       0.142
did            0.0840      0.237      0.355      0.725      -0.402       0.570
==============================================================================
Omnibus:                        0.917   Durbin-Watson:                   0.444
Prob(Omnibus):                  0.632   Jarque-Bera (JB):                0.935
Skew:                           0.352   Prob(JB):                        0.627
Kurtosis:                       2.497   Cond. No.                         6.85
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```
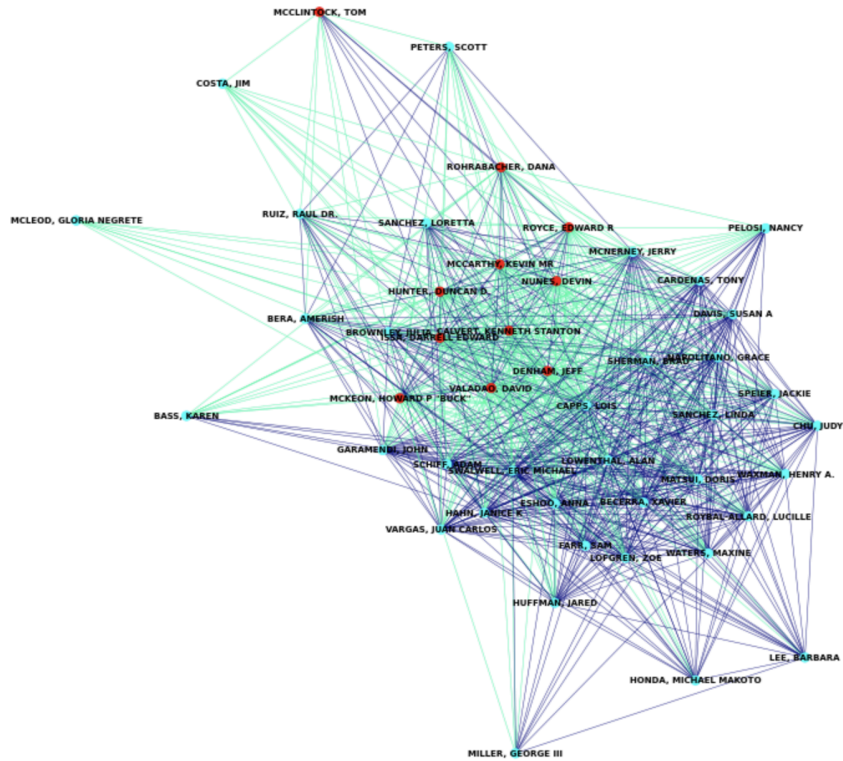
Appendix 3: Group 2 Network of California Representatives in 112th Congress
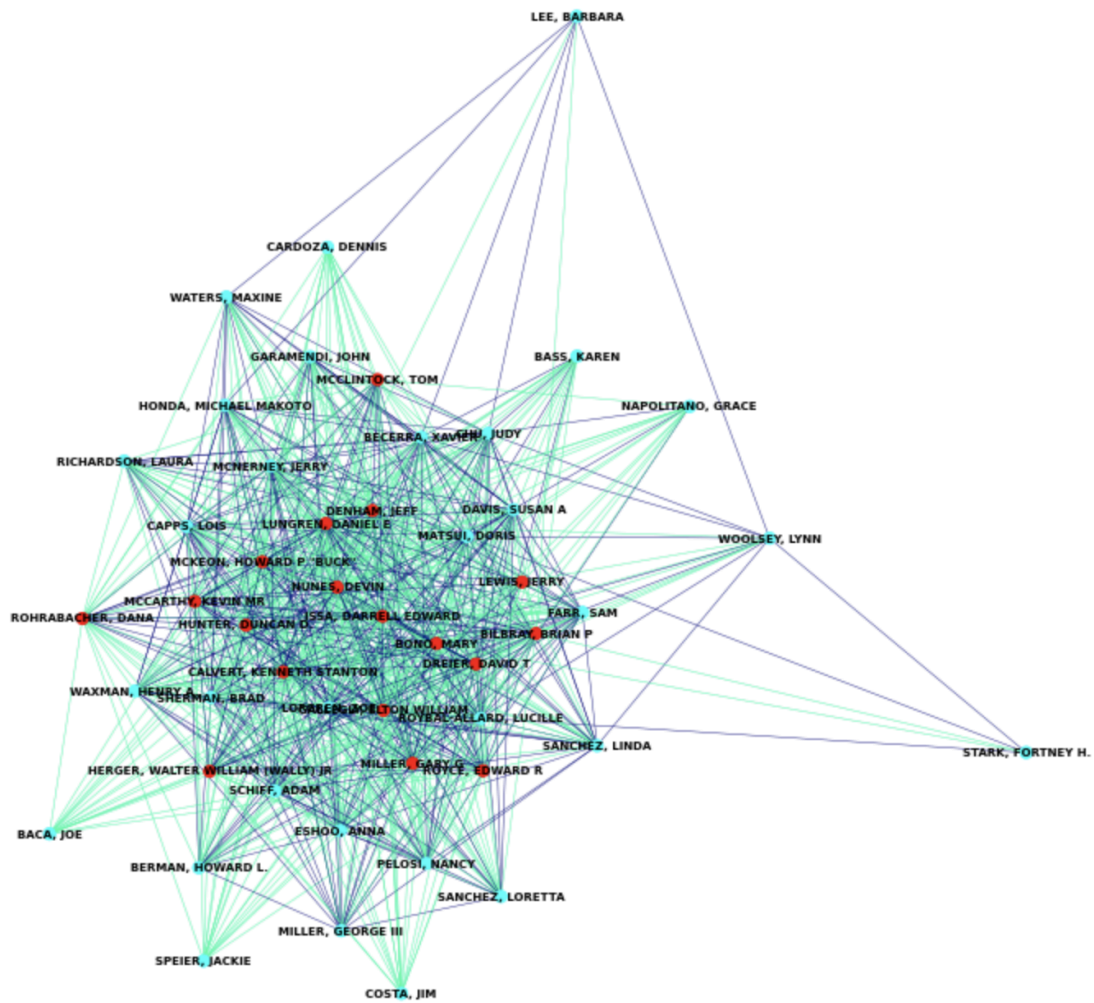
Appendix 4: Group 2 Network of California Representatives in 113th Congress

Appendix 5: Group 2 Inter-party Betweenness Centrality

	Mean (Republicans)	Mean (Democrats)
112th	0.026793	0.006635
113th	0.061045	0.003221

Appendix 6: Group 2 Inter-party and Intra-Party Degree Centrality

	Mean (Republicans)	Mean (Democrats)
112th (Inter)	0.554398	0.321909
113th (Inter)	0.654130	0.199879
112th (Intra)	0.330808	0.299120
113th (Intra)	0.194215	0.513889

Appendix 7: Speeches from Thomas McClintock and Barbara Lee in the 112th and 113th Congresses

112th Congress

McClintock, Thomas (R), Sentiment score: -0.7650
I thank the gentleman for yielding. Mr. Speaker, this vote stands as a defining moment in this crisis. Every rating agency has warned that an increase in the debt limit without a credible plan to balance the budget will do great damage to our nation's credit. And worse, fiscal experts warn that without such a plan, we risk a sovereign debt crisis within the next 2 years. This measure gives the President everything he has asked for—the $2.4 trillion debt increase to pay for the bills that he and the Congress have recklessly racked up. But it also calls for a constitutionally enforceable workout plan to place our nation back on the course to fiscal solvency, the centerpiece of which is a balanced budget amendment that has been proposed in one form or another since the birth of our Constitution and that 49 states have adopted. Now, the gentleman from Maryland reminds us that only a few of those 49 states have both a balanced budget requirement and a two-thirds vote for tax increases. My home state of California happens to be one of them. California's deficits, as bad as they are, have been proportionally roughly half the size of those that the federal government has run up in the same period. These budget protections work—maybe not perfectly, but they do work. And I might add that when California also had a real spending limit, as this measure calls for, California enjoyed an era of balanced budgets, prudent reserves, no tax increases, and steady economic growth.

Lee, Barbara (D), Sentiment score: -0.9776
I want to thank the gentlelady for yielding time and for her leadership on an issue as critical as extending a safety net to those desperately looking for jobs and needing this bridge over troubled waters at this point. Mr. Speaker, the Republican bill would gut unemployment benefits for the millions of Americans who are looking for work, at a time when there are roughly four people for every one job. It would reduce unemployment benefits from 99 weeks to 59 weeks during a serious crisis among our long-term unemployed. This makes no economic sense and, quite frankly, it is heartless. The Lee-Scott amendment would have replaced these Republican Christmas-time cuts with real extensions of unemployment benefits, adding an additional 14 weeks of unemployment insurance for the millions of Americans who have already exhausted their benefits. However, the Republicans did not allow any amendments—no fixes were allowed to the heartless and senseless cuts in this bill. This bill is truly a sham. It's a shame, and it's a disgrace. It will cost our nation jobs and is a slap in the face to job seekers. We should be focusing on reigniting the American dream, not making it more of a nightmare for people, as this bill would do.

113th Congress

McClintock, Thomas (R), Sentiment Score: -0.3818
I thank the gentleman for yielding. Madam Chairwoman, a family that earns $27,000 but spends $36,000 and has accrued a credit card debt of $165,000 is undoubtedly teetering on the brink of financial ruin. Proportionally, that is precisely where our federal government stands today. If that family were to seek advice from a credit counselor, the first piece of advice they would receive is to sit down and create a budget. They would need to make some incredibly tough decisions, and it might take several years to regain solvency. However, the United States Senate has not passed a budget in nearly four years, and our President has only presented budgets that lack seriousness, continue reckless spending, and never achieve balance. This bill simply requires that if the President cannot balance the budget this year, he informs us of how long it will take and what actions need to be taken to achieve balance. We would expect this level of responsibility from any family, and we should demand the same from our government.

Lee, Barbara (D), Sentiment Score: 0.9879
Madam Speaker, the Affordable Care Act is already making significant strides in improving access to quality, affordable healthcare for millions of Americans, including my constituents. In my home state, Covered California stands as a shining example of the ACA's potential success when fully implemented. The state's exchange is enrolling thousands of Californians into healthcare coverage daily. Residents in my district, spanning from young adults to seniors, as well as children with preexisting conditions and those unable to afford health insurance, are benefitting from critical protections and savings thanks to the ACA. For instance, one constituent recently reached out after reviewing plans offered through covered California. They found coverage that was not only more affordable but also better suited to their needs. This constituent expressed their reluctance to revert to a system where insurance companies have more control, leading to higher premiums for reduced coverage. Each time Republicans vote to repeal or dismantle the Affordable Care Act, they send a clear message that it's acceptable to charge women more simply for being women, to deny coverage to victims of domestic violence, and to allow insurance companies to prioritize profits by increasing premiums.