Shared Structures in Campaign Donation Networks & Congressional Voting Patterns on Union-Related Bills

Nina-Marie Cielo* Cameron Eller† Jake Pellett‡ Tommy Schupp§

2024-01-12

Abstract

We explore the potential variations in congressional AFL-CIO scorescards among communities created by donor networks. The Null Hypothesis posits that these communities exhibit the same mean union score, while the Alternative Hypothesis suggests differences in mean union scores. To investigate these hypotheses, we conduct a comprehensive analysis of various donor network communities, employing statistical methods to assess and compare their union scores. We find that communities created by donor networks meaningfully represent different mean AFL-CIO scores.

1 Introduction

In the ever-evolving landscape of political campaigns, understanding the intricate web of connections between individual donors, political action committees (PACs), and candidates is paramount. This paper presents a comprehensive data analysis utilizing R Studio to unravel the dynamics of political contributions and community structures. The initial step involved coding to classify data into distinct categories, distinguishing between individuals and PACs based on a designated column.

The dataset was divided into two subsets: 'individuals_df' and 'pac_df.' For 'individuals_df,' a network analysis was conducted, establishing edges between candidates based on individual donations. Connections were weighted, reflecting the cumulative contributions from shared individual donors. A similar approach was employed for 'pac_df,' where connections between candidates were determined by PAC donations, and weights represented the financial support provided.

Following the network analyses, the research delved into exploring candidate communities within the network. The investigation extended to examining these communities' scorecard data, aiming to discern patterns and similarities among groups of candidates. This multifaceted approach offered a nuanced understanding of the relationships within and between different political entities.

The analytical phase included hypothesis testing, focusing on comparing the mean union scores of candidate communities. To evaluate whether communities exhibited similar or different means, ANOVA tests were employed, complemented by the use of standard error and other relevant statistical measures. Specifically, the paper delves into comparing means within communities with PACs, providing insights into potential disparities in political contributions.

*American University
†American University
‡American University
§American University

Furthermore, the analysis goes beyond statistical comparisons, delving into the nuanced characteristics of different candidate groups. Descriptive methods are utilized to elucidate the distinctive traits and behaviors within these political communities, shedding light on the intricate interplay of individual and PAC contributions.

This scientific data analysis paper, rooted in R Studio, provides a robust exploration of political contributions and community structures. By combining network analyses, hypothesis testing, and descriptive approaches, the study offers a comprehensive perspective on the complex relationships that define contemporary political landscapes.

2 Political Donations and Network Analysis

The last year has been marked by a growing number of coordinated labor movements and union action across several industries in the United States. Our research explores the nature of connections between labor union legislative scorecards from members of congress and the networks created by prolific wide-reaching large campaign contributors.

There are a number of political implications of the expectations placed upon a candidate by a donor, including but not limited to ideological transparency and the degree of a representative agenda. Conversely, the actions that a congressperson takes to attract donations to further their own interests is also a deep topic of study. La Raja, Raymond J., and Brian Schaffner (2015) theorize that candidates who receive very large funding from individuals are very likely to be ideologically centrist, while candidates who seek funding from large amounts of small donations are more likely to be ideologically extreme. Additionally, the theory extends to corporations and business groups to claim that they will tend to favor moderate and conservative officeholders who will bargain over specifics rather than overarching ideological change. In our research, we expect to find that large donations from corporations will skew toward conservative networks in congress, while large donations from individuals will flow to centrist networks.

The academic literary body surrounding lobbying and political donations is dense, but there is a distinct difference in the post Citizens United v Federal Election Commission (2010) world. Citizens United further emphasized an already disproportionate influence in massive donations and dark money in elections. Hansen and Rocca (2019) draw a clear line in analysis of individual campaign contributors in the pre and post Citizens landscape to argue that a flood of extremely large donations to super PACs have contributed to a reduction in academic research surrounding political donations. Our investigation seeks to construct a web of large donations from available sources to understand the financial connections between current members of congress and highly active large donors. The primary difference in a post Citizens analysis is that many of the largest donations, largely from CEOs, is that the data has become obfuscated by inflated dark money values.

Graph theory is a powerful tool for analyzing complex systems of interacting agents in the political context. The precedent set by the research of Porter et al. (2005) to project bipartite networks onto one-mode networks, where the strength of connections can be quantified using the normalized interlock, in measuring social networks in congressional committees allows for an effective method of congressional network analysis. Our chosen method of clustering is the walktrap algorithm, discussed in greater detail in the following sections.

3 Data and Methods

For this study, two primary datasets were utilized—Open Secrets donation data and Union Scores for the year 2022. The Open Secrets data underwent cleaning to isolate donors contributing over $1000 to House and Senate members in 2022. OpenSecrets' bulk data is freely available for educational purposes, but it is available "warts and all." The data is available only as .txt files that are delimited using a non-standardized coding. Campaign finance data from OpenSecrets is compiled, cleaned, and standardized from data from the Federal Election Commission (FEC). The data guidebook emphasizes that users should avoid double counting individual contributions to PACs, so we did not include individual donations to PACs in our analysis.

Data provided by OpenSecrets.org Center for Responsive Politics

The names were standardized, and the 'candidate_party' variable was simplified to 'D' for Democrats, 'R' for Republicans, and 'I' for Independents. A new variable, 'pac_or_individual,' was introduced to categorize PACs ('p') and individuals ('i'). Notably, we integrated two initially separate datasets for individuals and PACs.

We had many difficulties navigating this dataset. First of all, the files were extraordinarily large. Before cleaning, the original individual donations file was 15GB. We also had to use joins and merges to connect disparate datasets containing different information so that we could get the information we wanted. For example, the individual donors dataset did not have candidate names, only their ID numbers. So, we had to cross-reference with another source that corresponded ID numbers to candidate names.

Simultaneously, the Union Scores data for 2022 was processed, encompassing both yearly and lifetime scores. Ultimately, we settled on just selecting lifetime scores. Titles preceding names were removed, and percentage-based score values were converted to a standard format. The House and Senate data frames were merged, and encoding discrepancies between the Union Scores (latin1) and Open Secrets (utf-8) datasets were addressed.

To connect these datasets, we had to conduct a tremendous amount of data cleaning. In particular, we had to individually clean members of the AFL-CIO scorecard, since some members had suffixes, middle names, or nicknames associated to them. This undertaking was difficult and took a long time, but ultimately we wrangled the data the way we wanted it.

The AFL-CIO Union Score is a number created by the American Federation of Labor and Congress of Industrial Organizations (AFL-CIO), which represents a congressperson's favorability towards unions through their voting patterns. The AFL-CIO separates their score calculations based on the chamber the congressperson is in as the House and Senate hold different votes. Additionally, while the union score for House members is calculated using bills and resolutions, the union score for Senate members also takes into account nominations. The score itself is calculated by taking a percentage of how often a congressperson voted in favor of union interests out of how often that member actively voted. The score is also calculated for each year, and averaged into a lifetime score. For example, as Representative Mary Petola of Alaska did not take office until September 13 2022, her score calculation in the 2022 dataset for the House does not take into account the eight bills the score card tracked before this. The first bill tracked by the score card that she was eligible to vote on was the Presidential Election Reform Act (H.R. 8873), which held its roll call vote on

September 20, 2022. After she took office the score card tracked five bills, four of which she voted yes on and one that she abstained, making her score for 2022 100%.

For our paper, we utilized the lifetime scores of congresspeople as calculated in the 2022 data sets. In the House data set, the score card tracked twelve bills and one resolution. In the Senate data set, the score card tracked five nominations, six bills, and one resolution.

Notably, the union scores are not weighted by popularity of the legislation or nomination up for vote or any other factors. For example, a congressperson voting for a bill despite many of their peers not voting for it does not influence their score other than the fact that they voted yes. In future analysis weighting the union score by different legislation's popularity, as well as taking into account type (for example appropriations, resolutions, or even levels of nomination such as Supreme Court nominations), could help to identify the 'extremeness' and 'strength of convictions' of a candidate.

We leveraged the igraph package in R to construct network graphs, visualizing connections between Congress members based on donor contributions. Originally, we had a bipartite graph: directed connections from donors to Congresspeople. The method we used to project this is that two Congresspeople are connected if they have a mutual donor. The weight of the connection is the sum of the contributions of the mutual donor to both candidates. Nodes, representing Congress members, were sized proportionally to the total amount donated.

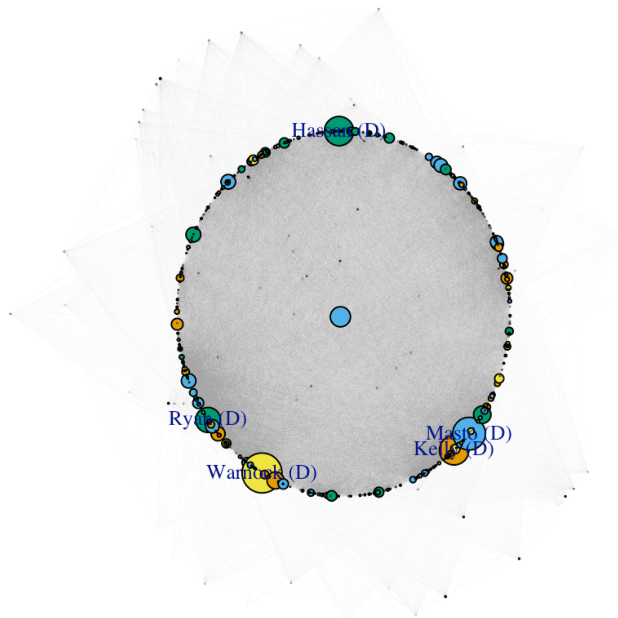
This took an extraordinarily long time to learn, but the results were fascinating. The graphs that we created were extremely complex and with more time, we would learn more methods of representing the information. This is our first time using graph theory, as we were deeply inspired by Dr. Noel, and we used the notes provided to us by him. Moreover, we were not aware of the LGL (large graph layout), which sped up our processing time by orders of magnitude. Originally, when we were running the code to draw these graphs, it took us 15 hours to draw a graph. With LGL, it sped it up into minutes. We also achieved performance speed ups by vectorizing the operations we used, which is demonstrated in the code we have in our GitHub.

For the graphs, some of the most influential individual large donors were the physician Karla Jurvetson, investor Seth Klarman, philanthropist Ari Nessel, philanthropist Cherna Moskowitz, and investor Steve Mandel. The most influential PAC donors were the Congressional Leadership Fund, the Senate Leadership Fund, the Senate Majority PAC, the National Republican Congressional Committee, and the Club for Growth Action.

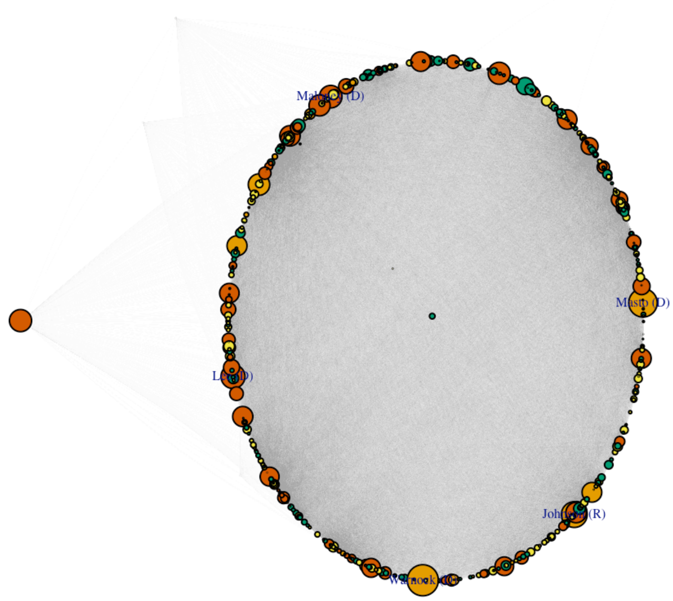
A walktrap function was employed to identify groups within the network. We chose a number of steps that provided us with a number of groups that aligned with our expectations (more than 2, but not too many). Many configurations of the steps parameter led to either far too many groups or simply two (Democrats and Republicans). The communities we landed on were then transformed into a data frame, enabling the association of union scores to specific Congress members.

This graphical representation facilitated a comprehensive exploration of relationships within and between these donor groups. The graphs could be a lot cleaner and more thoughtful.

The size of the vertex in the graph is a function of how well connected they are: it is a sum of their total edge weights to all other Congresspeople.

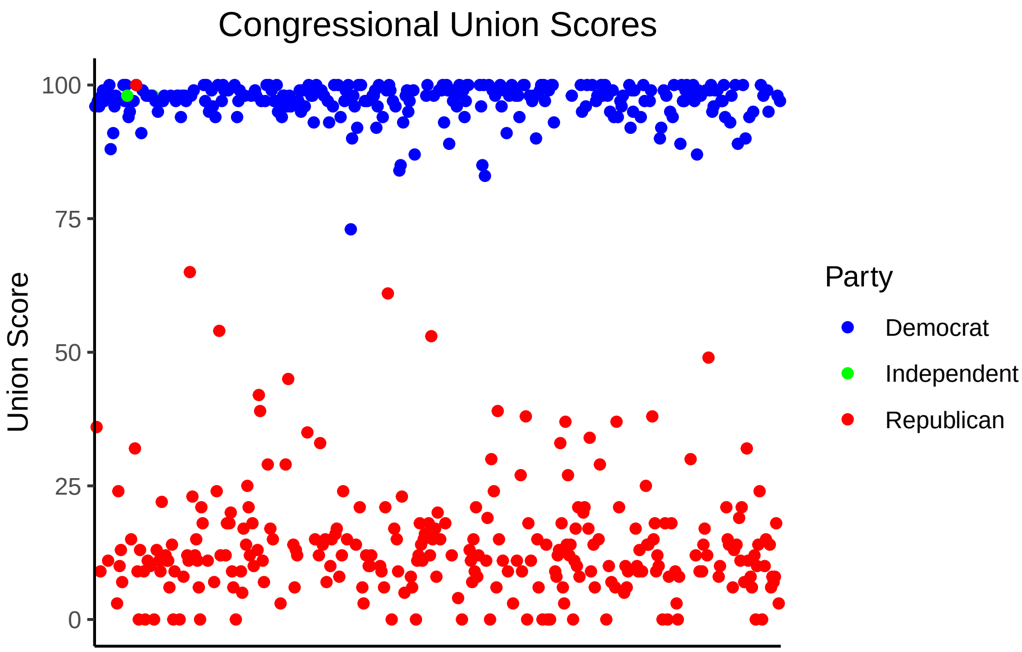


Above is the graph of the individual donation network. The candidates with the most prolific donor connections are highlighted by showing their names. Communities are distinguished by colors. The most connected Congresspeople were Raphael Warnock, Catherine Cortez Masto, Mark Kelly, Maggie Hassan, and Tim Ryan.



Above is the graph of the PAC donation network. The candidates with the most prolific donor connections are highlighted by showing their names. Communities are distinguished by colors. The most connected Congresspeople were Raphael Warnock, Catherine Cortez Masto, Ron Johnson, Susie Lee, and Sean Patrick Maloney.

For statistical insights, we implemented code to perform ANOVA on the identified communities, assessing differences in mean Union Scores. We also conducted additional analyses, such as graphical representations of score and donation distributions.



For ease, we have renamed the communities we created based on the political characteristics of their members. We have not included the full communities, but those are available on our GitHub.

3.0.1 For Individual Contributions:

3.0.1.1 Community 1: Major Fundraisers

Major names: Bernie Sanders, Kyrsten Sinema, Alexandria Ocasio-Cortez, Joe Manchin, Chuck Schumer, Jon Ossoff, Tammy Duckworth

This group is comprised of both Senators and Representatives, primarily Democrats and some Independents, but mainly 'very nameable' players who are successful at fundraising!

3.0.1.2 Community 2: Extreme Ideological Statements

Major names: Elizabeth Warren, Ed Markey, Nancy Pelosi, Matt Gaetz, Lauren Boebert Marjorie Taylor Green

Individual donors generally donate to congresspeople with more extreme ideological positions, but they also cast their financial support to institutional players or congresspeople that they are making a clear statement to by donating. This is a more extreme group of donors, placing a focus on contemporary and new shifts in the poltiical landscape.

3.0.1.3 Community 3: The Fringe

Major names: Adam Schiff, Scott Perry, Connie Conway

A majority of the members in this group are congresspeople on the political fringe or are developing and shifting their niche. They are not engrossed in the current political machine, but they do have developments that are upcoming.

3.0.1.4 Community 4: The Battleground

Major names: Raphael Warnock, Hakeem Jeffries, Rand Paul, Rick Scott

A majority of these are either currently elected or are elected into competitive districts. The major connections in this community will be found in donors who are spending on current competitive elections.

3.0.2 For PAC Contributions:

3.0.2.1 Community 1: Battleground Senators

Major names: Bill Foster, Tim Ryan, Raphael Warnock, Jon Ossoff, Elizabeth Warren

A majority of the members in this grouping are Senators, and many are also either recently elected or have been elected in battleground states.

3.0.2.2 Community 2: New Era Conservatives

Major names: Ted Cruz, Lindsey Graham, Marjorie Taylor Greene, Lauren Boebert

This group was not used for analysis due to the relatively small group size, but it is comprised of the major players in the new era conservative movement.

3.0.2.3 Community 3: Range Senators

Major names: Tammy Baldwin, Rick Scott, Joe Manchin, Kyrsten Sinema.

This group holds both the most ideologically extreme senators and the most centrist senators who generally act as swing votes in key legislative moments, it is likely that ‘far’ left PACs and ‘far’ right PACs donate with the interest to sway them.

3.0.2.4 Community 4: Catch All

Major names: Nancy Pelosi, Jared Huffman, Sheila Lee, Al Green, Adam Schiff, Ayanna Pressley, Yvette Clarke, Ed Markey

The largest community features pillars of the Democratic party and the generally left leaning house of representative members. It seems that these PACs donate to candidates with the intention of having a majority of house members to pass democrat-agenda legislation.

3.0.2.5 Community 6: Centrist House Members

Major names: Chris Pappas, Tom Malinowksi, Susie Lee, Haley Stevens

This group is very similar to group 4, but featuring some pillars of the modern Republican strategy with a slight weight towards the ideological center. This group is also smaller than group 4, placing a bigger emphasis on members with generally strong positions on a single main issue.

4 Results

After organizing communities by donor networks for both large individual donors and large PAC donors, we tested the mean union scores of the communities using an Analysis of Variance (ANOVA) Test. Using an ANOVA depends on certain assumptions being fulfilled.

```
Warning: `cols` is now required when using `unnest()`.
i Please use `cols = c()`.
```

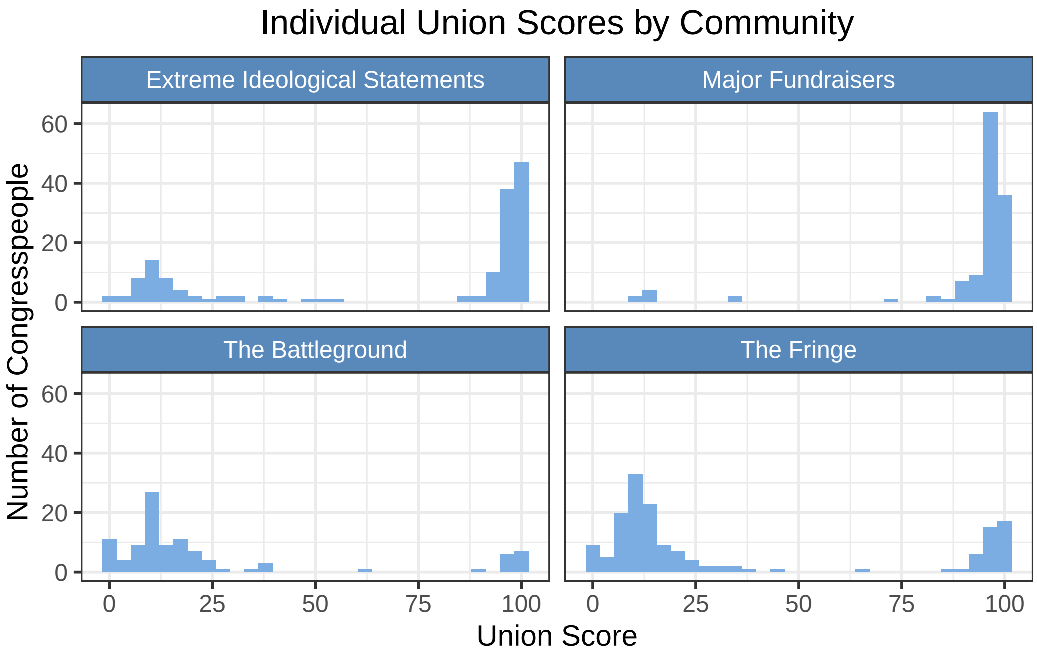
```
# A tibble: 539 x 6
   congresspeople         community          score1 lifetime_score1 name    party
   <chr>                  <chr>               <dbl>           <dbl> <chr>   <chr>
 1 Richard Blumenthal (D) Major Fundraisers     100              99 Richar~ D
 2 Mike Thompson (D)      Major Fundraisers     100              96 Mike T~ D
 3 Tom O'Halleran (D)     Major Fundraisers     100              90 Tom O'~ D
 4 Mark Kelly (D)         Major Fundraisers      92              97 Mark K~ D
 5 Scott Peters (D)       Major Fundraisers     100              89 Scott ~ D
 6 Henry Cuellar (D)      Major Fundraisers      92              73 Henry ~ D
 7 Jamaal Bowman (D)      Major Fundraisers     100              97 Jamaal~ D
 8 Susan Wild (D)         Major Fundraisers     100              99 Susan ~ D
 9 Ruben Gallego (D)      Major Fundraisers     100              98 Ruben ~ D
10 Jared Golden (D)       Major Fundraisers     100              92 Jared ~ D
# i 529 more rows
```

```
Warning: `cols` is now required when using `unnest()`.
i Please use `cols = c()`.
```

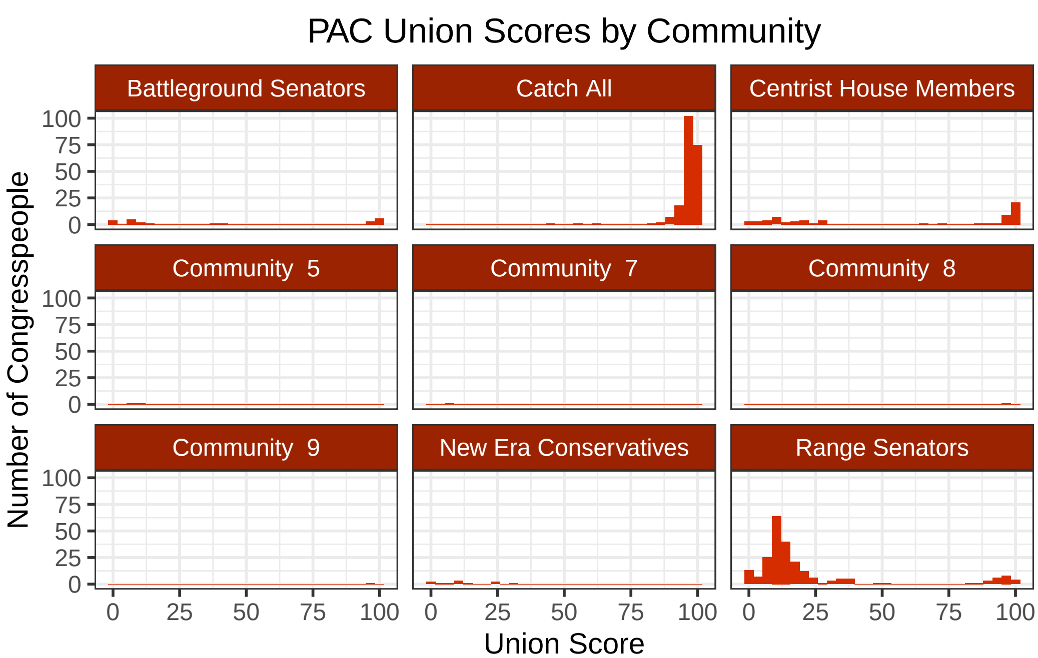
The ANOVA Assumptions are that 1. Each community has a normal population distribution of union score, 2. The union scores of congresspeople are independent, 3. The union scores between communities have an equal variance.

We begin by addressing these assumptions.

First, we examine the union score histograms of the large individual donor communities. Community 1 and Community 4 both seem to be unimodal and normally distributed, however, both communities have some skew. Community 2 and Community 3 are both bimodal with peaks on both ends of the union score spectrum. They do not seem to fit the normality assumption very well.



Now, we examine the union score histograms of the large PAC donor communities. First, note that Community 2, Community 5, Community 7, Community 8, and Community 9 are all extremely small, between 1 and 10 members each. For that reason, we drop those communities from the analysis, as they are too small to gain meaningful insight. Community 3 and Community 4 are both unimodal and have some light skew. Meanwhile, Community 6 and Community 1 are more evenly distributed. Like the individual section, these do not seem to fit the normality assumptions very well.

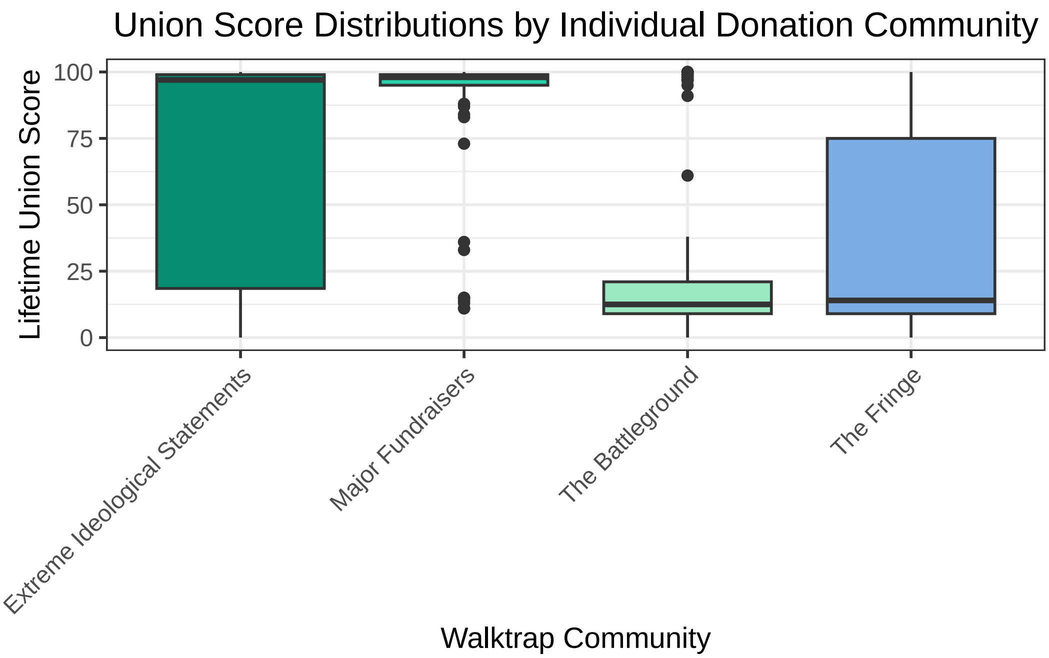


Even though the communities have skewed distributions, we treat them as normally distributed.

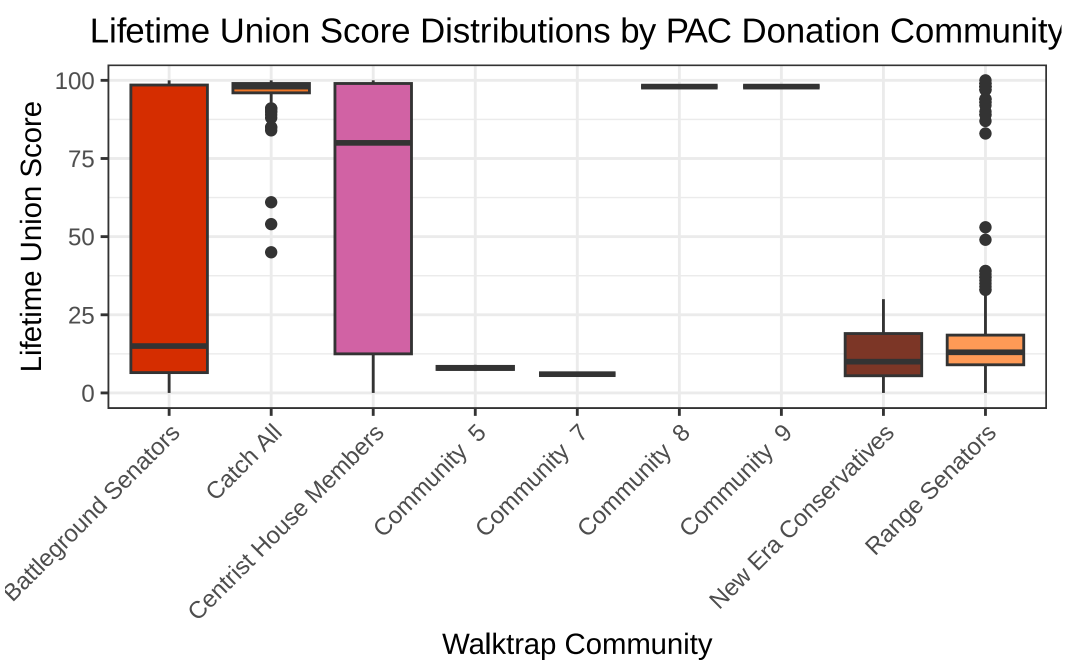
In regards to the independence of union scores, it is very unlikely that they are independent.

Because of the nature of Congress, voting tends to occur in coalitions and by party line. However, to perform this analysis, we can assume that the observations are independent.

Now, we examine the variances of the communities using boxplots.

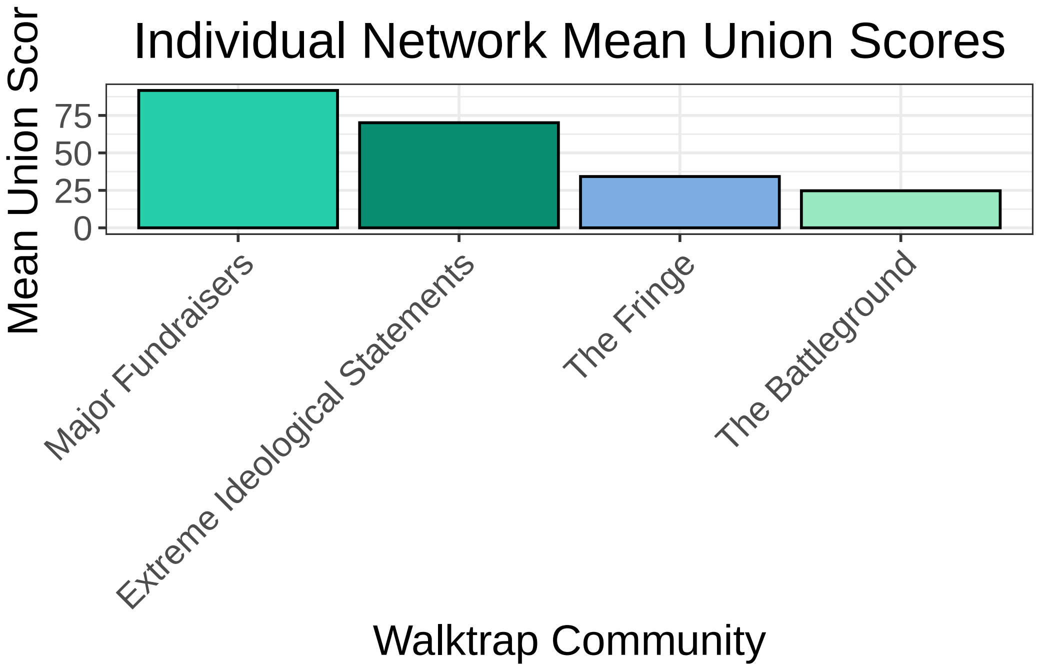


In the boxplots of large individual donation communities, we can see that the variances of Community 1 and Community 4 seem to be approximately equal. The variances of Community 2 and Community 3 also seem to be approximately equal. However, the variances of Community 1 and Community 4 are not equal to Community 2 and Community 3.



In the boxplots of large PAC donations communities, we find that Community 2, Community 3, and Community 4 have similar variances. We also find that Community 1 and Community 6 have similar variances. Also, Community 5, Community 7, Community 8, and Community 9 have similar variances. Again, the variances between these groupings of communities are quite different.

Though the ANOVA assumptions are violated, we still think that the test is appropriate to test the means of these communities. Future research could apply more complex non-parametric tests.

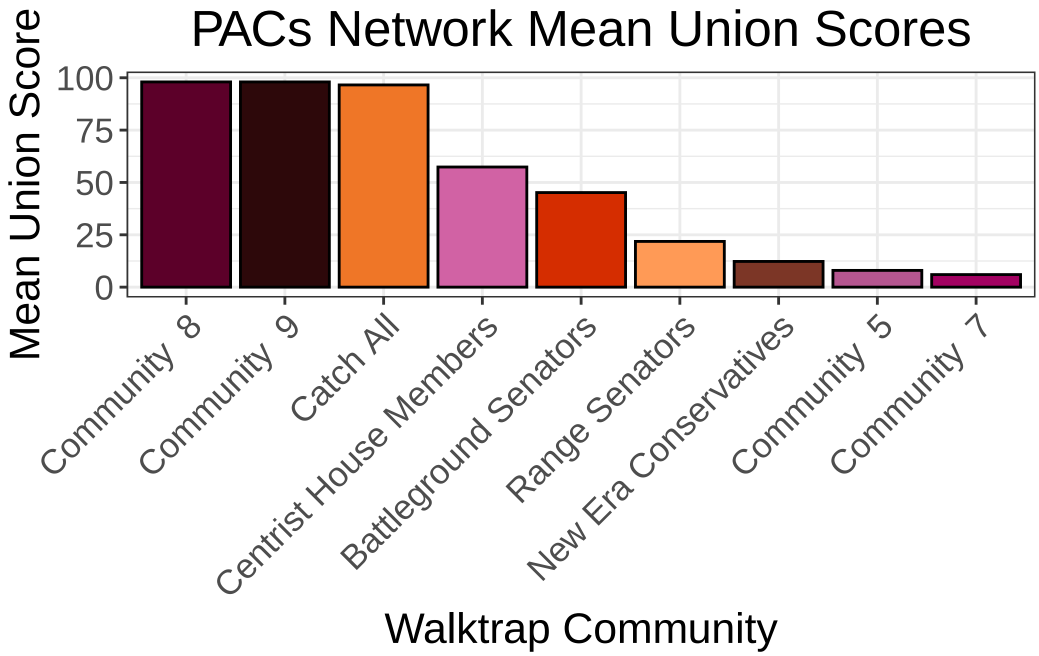


Examining the means of the large individual donor communities, we see that the mean union score of Community 1 and Community 2 is favorable towards unions. Conversely, Community 3 and Community 4 are less favorable towards unions.

```
individual_anova = aov(lifetime_score ~ community, data = individuals)
summary(individual_anova)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
community     3 367608  122536   110.3 <2e-16 ***
Residuals   535 594267    1111
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find that, with F-Value 110.3 and level of significance p<0.001, we reject that the means of these communities are the same. This means that for the distinct communities of congresspeople generated by mutual large individual donors, voting patterns on bills related to unions are distinct. In particular, this shows that structure in relational campaign donation behavior encodes information about the voting patterns of congresspeople towards unions.




```r
pacs_anova = aov(lifetime_score ~ community, data = pacs)
summary(pacs_anova)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
community     8 641578   80197   132.2 <2e-16 ***
Residuals   531 322078     607
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find that, with F-Value 329 and level of significance $p<0.001$, we reject that means of the communities created by PAC donation networks are the same. This means that for the distinct communities of congresspeople generated by mutual large PAC donors, voting patterns on bills related to unions are distinct. In particular, this shows that structure in relational campaign donation behavior encodes information about the voting patterns of congresspeople towards unions.

Between the two groups of donors, these results show that union scores are legitimate ways of distinguishing communities created by large PAC donations and large individual donations networks. This is a meaningful result because it shows that donation patterns are important in finding out information about congressional voting behavior.

5 Discussion

In this study, we delved into the intricate voting behavior of Congresspeople, particularly concerning unions. The mystery surrounding how members act on legislation, especially in relation to unions, has long intrigued political researchers. Our findings underscore the significance of understanding donor networks to decipher this complex behavior. Essentially, our research illuminates that comprehending donor networks is a more accessible route to understanding voting behavior on unions.

This crucial insight serves as a cornerstone for understanding the complexities of congressional decision-making.

However, our study is not without limitations. Time constraints posed challenges, restricting the depth of our analysis. Additionally, computing power limitations influenced the complexity of our methods. A notable limitation stems from missing data due to the Citizens United decision and changes in campaign donations research. Despite these constraints, we propose that graph theory could offer valuable insights into the connections among donors, mitigating the impact of missing data.

Looking ahead, our research lays a foundation for future investigations. Incorporating small donations into the analysis could further enrich our understanding of donor networks. Time series analysis, exploring the directionality of voting behavior, is a logical next step for a more dynamic perspective. Investigating dark money and tracking donations through Super PACs would provide a more comprehensive view of the financial landscape influencing congressional decisions. Furthermore, engaging with other scorecards or metrics could offer additional dimensions for analysis.

6 References

Hansen, W. L., & Rocca, M. S. (2019). The Impact of "Citizens United" on Large Corporations and Their Employees. Political Research Quarterly, 72(2), 403–419. https://doi.org/10.1177/1065912918793230

La Raja, R. J., & Schaffner, B. F. (2015). The Hydraulics of Campaign Money. In Campaign Finance and Political Polarization: When Purists Prevail (pp. 108–133). University of Michigan Press. http://www.jstor.org/stable/j.ctvdtpj2w.9

Porter, M. A., Mucha, P. J., Newman, M. E. J., Warmbrand, C. M., & Levin, S. A. (2005). A Network Analysis of Committees in the U.S. House of Representatives. Proceedings of the National Academy of Sciences - PNAS, 102(20), 7057–7062. https://doi.org/10.1073/pnas.0500191102

OpenSecrets (2022). Campaign Finance Data. Center for Responsible Politics. https://www.opensecrets.org/bulk-data/downloads

American Federation of Labor and Congress of Industrial Organizations (2023). Legislative Voting Records. https://aflcio.org/scorecard/legislators