Bipartisanship in Congressional Voting Networks

Evan Jiang (evanjiang@college), Hannah Park (hannahpark@college)

December 10, 2024

1 Introduction

1.1 Background

The U.S. House of Representatives is the primary legislative body in American democracy. Over recent decades, it has become increasingly partisan, meaning that members of opposing political parties rarely collaborate or vote together. This trend contributes to legislative gridlock, where important bills fail to pass due to ideological divides. Bipartisanship—where members from both parties work together—is crucial for addressing national challenges effectively. However, identifying bipartisan representatives is not straightforward, as traditional methods often oversimplify the complex dynamics of voting behavior.

The motivation for this work lies in the need for tools to ”cut through the noise” of congressional voting data. Instead of relying on broad perceptions or simplified metrics, we aim to use rigorous statistical methods to identify representatives who consistently demonstrate bipartisanship. This analysis has practical implications, such as identifying legislators who might mediate between divided factions or highlighting patterns of collaboration that could inform future policymaking.

Our primary question is: *How can statistical network modeling techniques be used to identify bipartisan representatives in the U.S. House of Representatives?* By answering this, we aim to provide a systematic and replicable framework for analyzing bipartisanship.

1.2 Related Work

The study of congressional voting behavior has long been of interest to political scientists and statisticians alike. From a political science perspective, tools like DW-NOMINATE [3] have been widely used to place representatives on a left-right ideological spectrum. These methods focus on ideology and polarization but often fail to capture nuanced behaviors, such as the occasional bipartisan actions of representatives. Other studies rely on qualitative analyses or survey-based approaches to understand bipartisanship, but these are limited by their subjective nature and lack of scalability.

From a statistical modeling perspective, networks provide a framework for analyzing relationships between representatives. The Stochastic Block Model (SBM) is one of the earliest approaches to community detection in networks [1]. It groups representatives into distinct communities based on their voting behavior, assuming that representatives within a community are equally likely to vote similarly. However, the SBM struggles with two key limitations: it cannot handle degree heterogeneity, where representatives have vastly different numbers of connections (votes), and it assumes representatives belong to only one community. These assumptions are particularly problematic in congressional voting networks, where influential representatives vote more frequently, and bipartisan representatives align with multiple communities on different issues.

Graphon models, often used for very large networks, provide another approach to modeling connections between nodes. These models assume exchangeability of nodes, meaning any two representatives are equally likely to connect. However, this assumption fails to capture the structured nature of congressional voting, which is influenced by party affiliations and issue-specific alignments.

The Degree-Corrected Stochastic Block Model (DCSBM) improves on the SBM by addressing degree heterogeneity [2]. It allows for representatives to have varying levels of influence, which makes it more realistic for congressional data. Despite this improvement, the DCSBM still assumes that representatives belong exclusively to one community, making it unable to model the mixed memberships seen in bipartisan behavior.

The Degree-Corrected Mixed Membership Model (DCMM) builds on these earlier methods and offers the most flexibility for our analysis. It accommodates both degree heterogeneity and mixed memberships, allowing representatives to belong to multiple communities to varying degrees. Mathematically, the DCMM models the probability of an edge (shared vote) between two nodes i and j as:

$$P(A_{ij} = 1) = \theta_i \theta_j \times \pi_i' P \pi_j,$$

where π_i and π_j are mixed membership vectors that describe how much each representative belongs to different communities, θ_i and θ_j represent each representative's influence, and P is the community interaction matrix that quantifies the strength of connections between communities. This flexibility makes the DCMM particularly suited to congressional voting networks, where representatives may align with their party on some issues while crossing party lines on others.

2 Methodology

To implement the degree-corrected mixed membership model (DCMM) on congressional voting data, we follow a systematic process that includes data collection, preprocessing, exploratory analysis, graph construction, and model application. All of our code is uploaded and documented **here**.

2.1 Data Sourcing

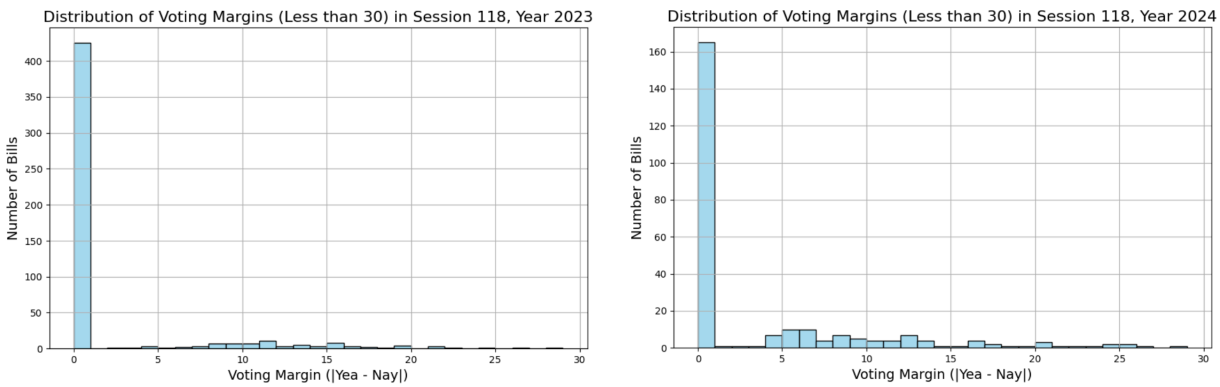
We use roll call voting data from the U.S. House of Representatives, focusing on legislative sessions from recent years. Each vote in the dataset records whether a representative voted "yea," "nay," or abstained. The data is sourced from GovTrack (`https://www.govtrack.us/`), which provides structured voting records. To ensure meaningful analysis, we filter the dataset to include only substantive bills, such as policy-related legislation, and exclude ceremonial resolutions or procedural votes. Filtering is done with GPT-3.5, which identifies and excludes bills based on their descriptions.

2.2 Data Processing

Before conducting the analysis, we preprocess the data to handle inconsistencies and missing values. This includes excluding representatives with insufficient voting records (e.g., due to absences), filtering for only "yea" and "nay" votes to calculate similarity between representatives, and encoding votes numerically for calculations.

2.3 Exploratory Data Analysis

To validate our research question, we first plot the distribution of voting margins (percentage difference between "yea" and "nay" votes) for all bills in Figure 1:



(a) Original images.

(b) Reconstructed images.

Figure 1: Voting margin distributions for bills in 2023 and 2024.

The tight voting margins indicate contentious issues, showing the importance of individual representatives in these scenarios. Then, using the cleaned data, we compute pairwise voting similarity for representatives. The similarity score between representatives i and j is defined as:

$$S_{ij} = \frac{\text{Number of identical votes between } i \text{ and } j}{\text{Total votes both } i \text{ and } j \text{ participated in}}.$$

Representatives are ranked based on their average similarity scores, providing a baseline for identifying potential bipartisan actors (see Figure 2).

2.4 Graph Construction

To analyze the voting network, we construct a binary similarity matrix. The matrix is defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } S_{ij} \geq \text{threshold}, \\ 0 & \text{otherwise.} \end{cases}$$

We choose a similarity threshold of 0.5 based on observations from the EDA. This threshold ensures that only meaningful connections are included in the graph while avoiding overfitting. Each

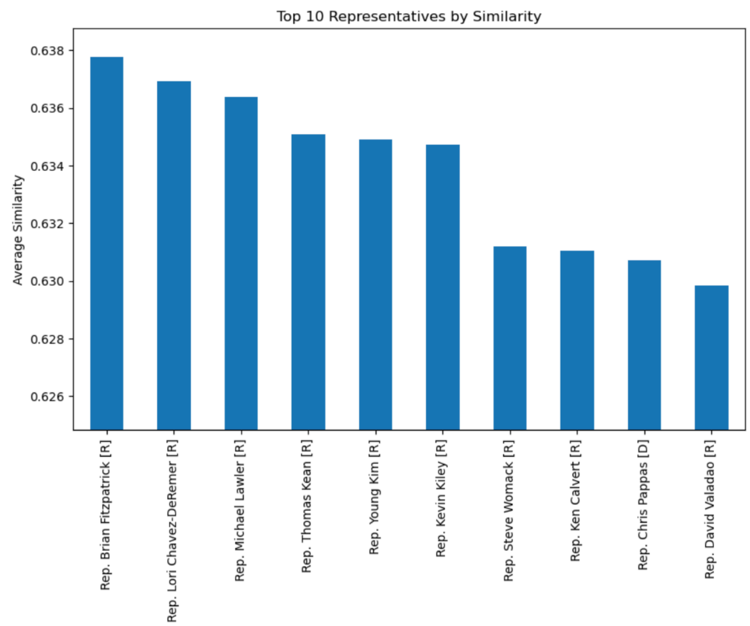


Figure 2: Representatives by highest average voting similarity score.

entry A_{ij} represents whether representatives i and j share a significant voting similarity. The resulting matrix is converted into a graph structure using the `igraph` package in R. In this graph, nodes are representatives, and undirected edges are a significant similarity in voting behavior, which we visualize in Figure 4.

We also construct a separate similarity matrix with continuous, non-binary entries to observe how voting similarities vary in a more granular way. Unlike the binary matrix, which only captures whether representatives meet the 0.5 similarity threshold, the continuous matrix retains the exact similarity values between representatives, allowing for finer distinctions in their voting behavior.

2.5 Model Application

DCMM is implemented on the constructed graphs to uncover latent communities and quantify representatives' influence and memberships in each community. The DCMM assigns each representative an influence θ_i and mixed membership vector π_i, where $\pi_i = (\pi_{i1}, \pi_{i2})$ and π_{ik} represents the degree to which representative i belongs to community k. The model also estimates a community interaction matrix P, where P_{kl} quantifies the connection strength between communities k and l.

3 Discussion

3.1 Binary DCMM

Based on the results, we see that the DCMM identified two primary communities, as visualized by the clustering on the left and right of Figure 3. These clusters strongly align with the true

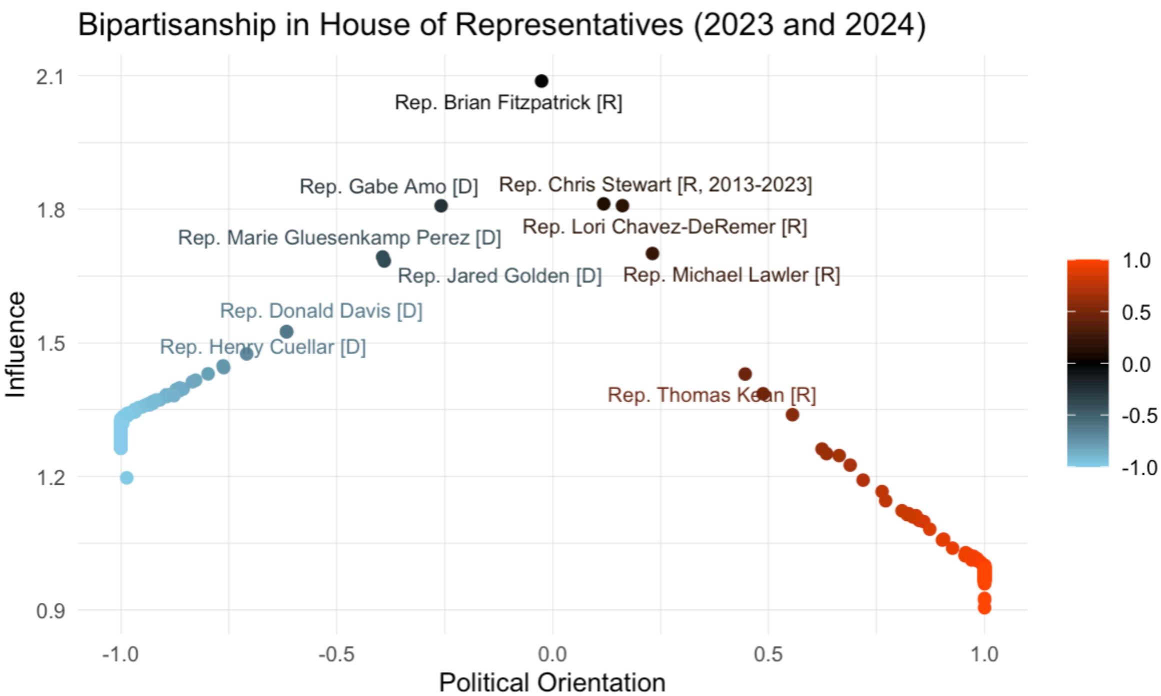


Figure 3: Results from DCMM showing representatives' influence and membership.

party affiliations of the representatives with an accuracy of 98.9%, where Representatives with higher `Partisan Tendency` scores near +1 are predominantly Republicans, while those near -1 are primarily Democrats.

However, a few representatives with `Partisan Tendency` values closer to 0 demonstrate mixed membership in both communities. These individuals often engage in voting patterns that bridge the partisan divide, indicating their potential role as bipartisan actors. For example, representatives like Rep. Brian Fitzpatrick [R] and Rep. Henry Cuellar [D] are positioned closer to the center, reflecting their moderate or bipartisan tendencies. These representatives do not strictly adhere to their party's voting bloc, indicating their unique position within the legislative process.

An interesting observation from above is the association between higher bipartisanship (`Partisan Tendency` closer to 0) and greater influence (`Influence` values). Representatives with moderate scores, indicating bipartisan behavior, tend to have higher influence within the voting network. This could suggest that bipartisan representatives often play key roles in facilitating legislative negotiations or bridging party divides, which increases their centrality in the network.

The results show a moderate correlation with our baseline from EDA and the Georgetown Bipartisan Index, a measure that evaluates representatives' voting behavior and cross-party collaboration. Many of the representatives identified as bipartisan by the DCMM model also rank highly on the Georgetown Bipartisan Index. For instance, the representatives highlighted near the center of the plot, such as Rep. Donald Davis [D] and Rep. Michael Lawler [R], are consistently recognized for their bipartisan efforts.

3.2 Continuous DCMM

The continuous DCMM provides a more nuanced view of congressional voting patterns by using a similarity matrix with continuous values rather than binary entries. This approach sharpens partisan clustering, with fewer representatives near the center of the `Partisan Tendency` axis (see Figure 5). The increased granularity reduces the likelihood of misrepresenting representatives as bipartisan based on occasional cross-party votes, presenting a more partisan perspective.

Interestingly, the correlation between influence and bipartisanship is weaker in this model compared to the binary DCMM. Highly influential representatives, such as Rep. Steve Womack [R] and Rep. Michael “Mike” Simpson [R], are firmly partisan, diverging from the Georgetown Bipartisan Index’s focus on cross-party collaboration. The model reveals a predominance of highly influential Republicans and highlights sharper polarization, reflecting broader trends in Congress with the Republicans controlling the House in the 118th Session.

4 Future Work

This paper demonstrates the utility of statistical network modeling in understanding congressional voting behavior. Next steps include expanding the analysis across multiple congressional sessions, which would reveal trends in bipartisanship and polarization over time. Temporal network models could track changes in community structures, identifying important events such as leadership changes or landmark legislation that influence voting behavior.

Moreover, allowing for more than two communities ($K \geq 3$) could uncover ideological nuances, such as progressive or moderate factions within parties, or latent third-party influences. This would provide a more detailed map of voting behavior, capturing cross-party coalitions and internal divisions.

References

[1] Paul W. Holland, Kathryn B. Laskey, and Samuel Leinhardt. "Stochastic Blockmodels: First Steps". In: *Social Networks* (1983). URL: https://www.stat.cmu.edu/~brian/780/bibliography/04%20Blockmodels/Holland%20-%201983%20-%20Stochastic%20blockmodels,%20first%20steps.pdf.

[2] Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. "Mixed Membership Estimation for Social Networks". In: *arXiv preprint arXiv:1708.07852* (2017). URL: https://arxiv.org/pdf/1708.07852.

[3] Keith T Poole and Howard Rosenthal. "A Spatial Model for Legislative Roll Call Analysis". In: *American Journal of Political Science* (1985). URL: http://www.jstor.org/stable/2111172.

5 Appendix

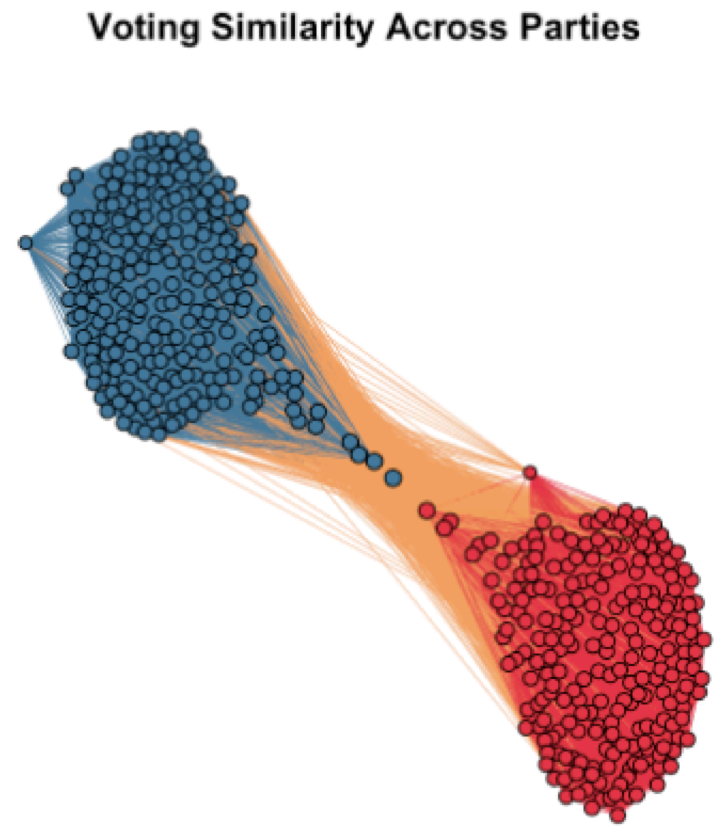


Figure 4: Network visualization from binary similarity matrix.

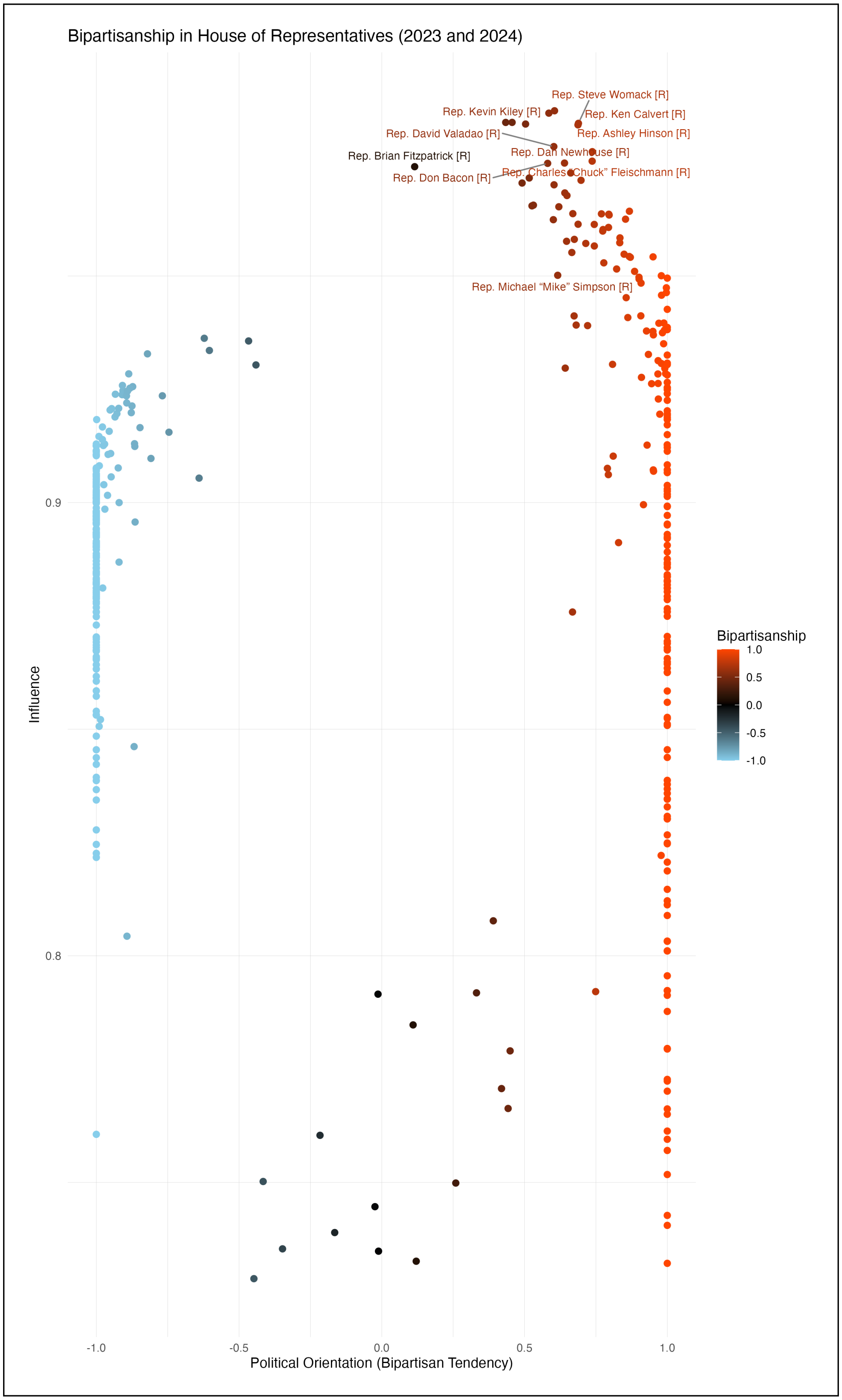


Figure 5: Results from DCMM showing representatives' influence and membership.