The Split Senate

Tara Chari[1] and Lior Pachter[1,2]

[1]Division of Biology and Biological Engineering,
California Institute of Technology, Pasadena, California

[2]Department of Computing and Mathematical Sciences,
California Institute of Technology, Pasadena, California

May 4, 2021

Abstract

The theory of split-systems provides a mathematical formalism for understanding and visualizing partitions of sets into two parts. In particular, split-networks, which generalize phylogenetic trees, are widely used in evolutionary biology. We show that these tools can be used to analyze and visualize voting patterns within a social structure. As an example, we consider United States Senate votes, and we show that the Neighbor-Net algorithm, coupled to SplitsTree visualization, provides an effective exploratory data analysis framework for elucidating voting patterns among senators. We also introduce a statistical approach to identify contributing votes underlying the patterns we identify, and explore shifts in inter- and intra- party structures over time. The analyses we describe should be broadly applicable to visualizing and studying coalitions in any voting organization.

Introduction

There are numerous techniques to assess political relationships within voting organizations based on the compilation of statistics from votes by members. Methods range from clustering techniques [11, 21, 47] and network analyses [1, 11, 22], to multi-dimensional scaling (MDS) [3, 41, 44] and dimensionality reduction [10, 34, 39, 42]. These approaches seek to highlight or define where individuals are most similar and divergent in their political behaviors. Applications include analysis and interpretation of historical voting data [1, 17, 25, 43], identification of voting coalitions among individuals [23, 28, 29, 37], and prediction of future voting patterns [46, 47]. However, within one visualization it can be difficult to represent the multi-level relationships that may exist between individuals and sub-groups of individuals, and interpretable quantitative information on the strength or proximity of these relationships is absent in the output from many methods [1, 7, 32, 37, 39, 40]. Additionally, for exploratory analysis it is challenging to obtain an unbiased visualization that provides internal structural information as well as a global overview of the political dataset of interest prior to more narrow, focused investigation [11, 33, 36, 37].

We propose the use of an algorithm popular in phylogenetics, but which has not been used in political science applications to date. The Neighbor-Net algorithm (NNet), is useful for visualizing

hierarchical structure based on splits in datasets, i.e. typically partitions into two parts of taxa in biology. The NNet algorithm creates an object called a circular split-system, along with a split network that realizes it in a 2D network representation [5, 18, 19]. These objects are also often referred to as phylogenetic networks [12]. The input to the algorithm is a matrix of distances among objects. In phylogenetics, distances are typically based on molecular distances from DNA alignments or other phylogenetically informative characters. NNet has also been used for studying relationships among languages, where distances are based on linguistic characters such as phonemes [4, 13]. The advance NNet represents over standard phylogenetic methods is that it can reveal signals in the data that are in conflict with a strictly hierarchical tree structure. This is important in phylogenetics where gene trees may conflict with species trees, and in linguistics where distinct characters may have been shared at different times, sometimes between spatially and temporally distant languages. Recently, NNet has also been used to analyze structure in single-cell gene expression data [49].

Though NNet has not previously been used for political data analysis, it is similar to commonly used MDS techniques in that its inputs are dissimilarity based measurements between the elements of the dataset [3, 5, 41, 44]. However, in contrast to MDS-based embedding methods which focus on recovering ideal points for individuals in a low dimensional space [38], NNet additionally defines structures between individuals and represents the relative strengths of these relationships within its network construction. Given a set of n elements in $M : \{m_1, ...m_n\}$, representing members of the Senate as m, and an element $\times$ feature matrix (features represented as votes here) R (Fig. 1a), a pairwise dissimilarity (distance) matrix δ is constructed (Fig. 1b). With this distance matrix NNet will generate a circular ordering of these elements $\pi = \{m_1, ...m_n\}$, where m_i and m_{i+1} are adjacent vertices on an n-cycle C_n comprised of the elements of M, and a split-system (Fig. 1c).

A split $A|B$ is a bi-partition of the set of elements in M, where $A \cup B = M, A, B \neq \emptyset$, and $A \cap B = \emptyset$, and a split-system is a collection of splits (Fig. 1c). NNet is an agglomerative algorithm which works to construct a circular ordering by iteratively joining the nodes of graph G whose vertices are composed of the elements of M, and defining splits of these nodes at each agglomeration (joining) step [5]. This produces a circular split-system Σ which can be graphically described as placing the elements "around a circle and consider[ing] the splits given by cutting the circle along a line" [5] (Fig. 1c).

The system can then be visualized as a planar splits graph [18] (Fig. 1d). The weights of the splits, λ, (represented by lengths in the planar graph)(Fig. 1d) are then obtained by performing non-negative least squares optimization for the constructed π and Σ so as to best match δ [5]. Together π, Σ, and λ constitute the 2D visualization of a circular split-network, as is shown in Fig. 1d.

A particularly useful interpretation of a split network is through its representation of 'feature diversity' [15]. For instance, features found for elements in A but not for B can be assigned to split $A|B$ and features exclusive to B not A can be assigned to a split $A|B \cup \rho$ where ρ represents an outgroup [31]. This implies that these distinguishing features can be inferred from the splits comprising the split network [31]. With these underlying facets of the circular split system created by NNet, we demonstrate its applicability to generating visualizations in the political arena, focusing on the multi-level, quantitative relationships and structures it reveals between the members of the

a Roll-call Vote (v) Matrix R

R	v_1	v_2	...	v_n
m_1	Nay	Yea		Yea
m_2	Abstain	Nay		Yea
m_3	Yea	Nay	...	Nay
m_4	Nay	Yea		Yea
m_5	Nay	Nay		Abstain
m_6	Abstain	Abstain		Nay

M = Set of senators m_i for $i = 1 \ldots 6$

b Distance Matrix δ

δ	m_1	m_2	m_3	m_4	m_5	m_6
m_1	0	21.5	15	5	22	11
m_2	21.5	0	15.5	23.5	4.5	12.5
m_3	15	15.5	0	12	13	16
m_4	5	23.5	12	0	23	14
m_5	22	4.5	13	23	0	15
m_6	11	12.5	16	14	15	0

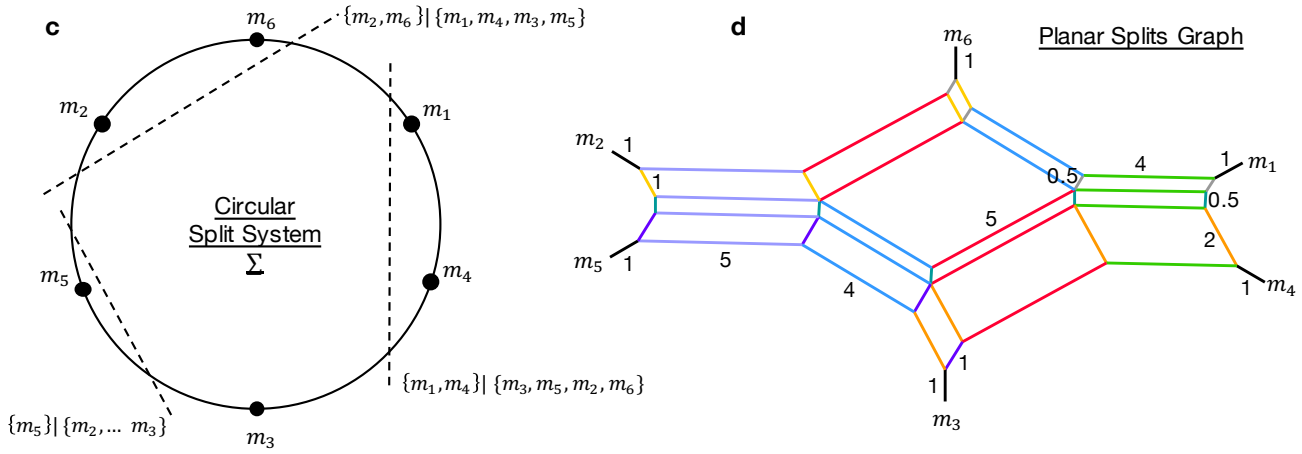

Figure 1: **a)** Matrix R of feature (vote) values for six elements (senators) m **b)** L_1 pairwise distance matrix between all six elements **c)** Circular split system with elements m shown in circular ordering, and splits defined as bi-sections of the circle **d)** Splits graph representation of circular system in **c** with split weights denoted along splits. Parallel splits (representing the same split) denoted by the same colors. Figure adapted from [5, 26]. [Code]

US Senate.

Results

Generation and Validation of the Circular Split System

For the current 116th Senate, the split network output of NNet, using the distance matrix δ generated from Senate votes, is shown in Fig. 2a. To better understand voting patterns by party we also generated the split network from Democrat (including Independent) and Republican senators separately (Fig. 2b,c). Note that the split network produced by running NNet run on a subset of a matrix will be the same as the restriction of the split network produced by running NNet on the full matrix.

We first verified that the generated split weights λ represented the same magnitudes of dissimilarity between pairs of senators as encapsulated in the input matrix δ, constructed directly from the vote matrix. By Pearson correlation analysis of the pairwise distances calculated from λ, using (2), and δ (Fig. 3) we found a correlation of 0.994. This demonstrates that the split network representation of the votes is highly concordant with the raw voting matrix, confirming that the circular split system is a good model for the voting structure. We did find several 'outlier' pairwise distances, outlined in black, which lie on the rim of the convex hull encompassing the distances (Fig. 3). These represent pairs of senators where their pairwise distances are more discordant with the split network. Individuals with repeated representation in these outlier points included Senators Warren, Booker, and Gillibrand. That the circular split system does not reflect their voting patterns perfectly is also evident in the longer lengths of their respective splits in Fig. 2a relative to the other senators.

Having inferred the circular split-system representation and split weights for the 116th Senate, we next examined individual relationships and neighbors across all members (Fig. 2a). As expected, there is a split dividing members of the two major parties. The split network also reveals member's nearest neighbors based on their voting behaviors, and noted 'mavericks' or 'centrists', such as Sen. Collins (Rep.) and Sen. Manchin (Dem.), stand out in their distant, centered placement relative to the rest of the Senate [42] (Fig. 2a).

While many individual's nearest neighbors are maintained regardless of whether the split-system is generated with only votes from within-party members or all senators' votes, there are a few neighbor differences dependent on the inclusion or exclusion of the other party's voting data. For example within the Republican party, while Senators Perdue and Sullivan remain neighbors in both split-networks (Fig 2a,b, marked as 3), Sen. Tillis joins their ranks only in the Senate-wide split-system. These changes suggest a delineation between those who vote similarly with respect to senators within their party, versus with respect to how they vote against the other party, and can be helpful for determining whether inclusion/exclusion of the other party's voting data is useful for a particular investigation or question.

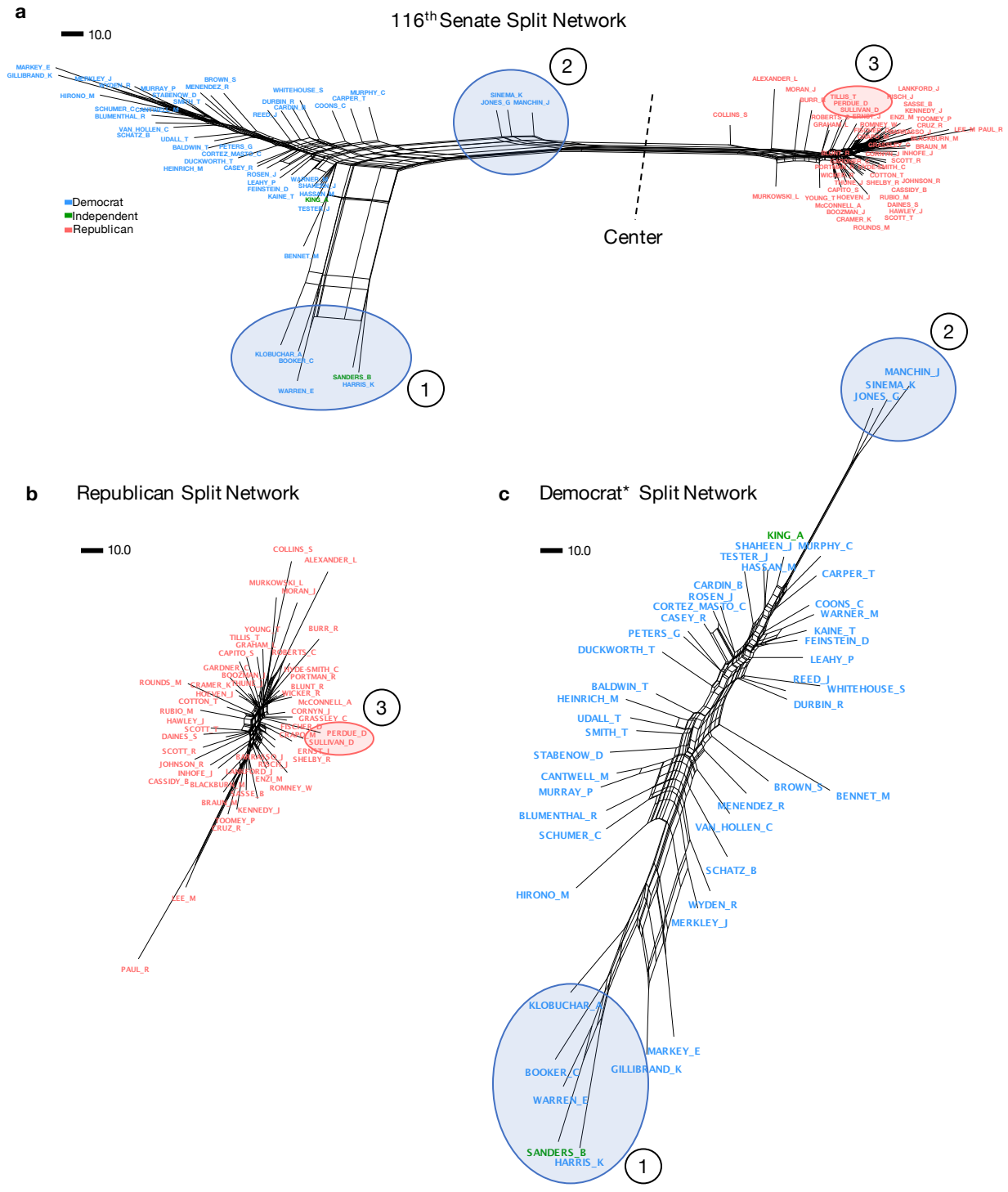

Figure 2: **a)** Splits graph representation of Senate-wide split network, for the 116th Congress. Republican, Democrat, and Independent members shown with colors. Nearest-neighbors and apparent 'coalitions' of members shown in circles. **b)** Splits graph for split network of only Republican votes **c)** Splits graph for split network of only Democrat (inclusive of Independent) votes. [Code]

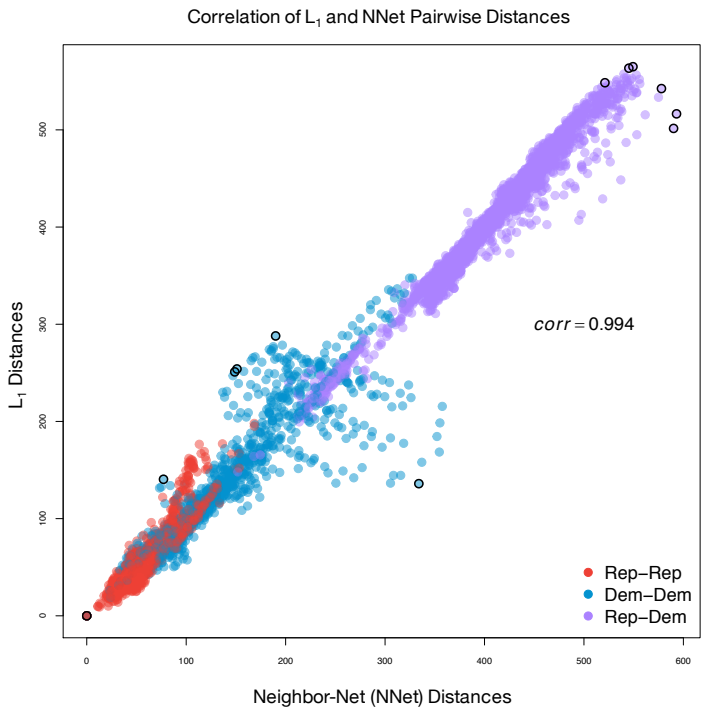

Figure 3: Pearson correlation between pairwise L_1 distances and pairwise split weight distances (Methods). Each dot represents distance between a pair of senators. Pairs forming the edge of the correlation plot, including points with the most discordant distances, are circled in black. [Code]

Analysis of Coalitions within the Senate

Beyond pairs of individuals, the Senate-wide diagram highlights apparent coalitions within the greater Senate structure, visible by clustering of individuals in the circular order, and in larger relative magnitudes of split weights (lengths) separating groups of individuals from the rest of the system (Fig. 2a, denoted 1 and 2). An interesting and notable example is the split of Democratic primary candidates from the rest of the Senate (Fig. 2a, denoted 1). Of the seven main incumbent senators to run in the 2020 Democratic Party presidential primary [6], five consistently cluster together in both the Senate-wide and intra-party circular split-systems (Fig. 2a,c). It should be noted that Senators Bennett and Gillibrand do not consistently cluster together with the rest of the candidates in both diagrams, again suggesting separation by voting behavior when votes of the opposing party are under consideration.

To verify whether this sequential ordering of these candidates was significant we used the Wald-Wolfowitz runs test [48] to determine the likelihood that this particular ordering was random (the null hypothesis). For this test, the circular ordering of senators can be represented as a linear ordering with senators that were Democratic Primary candidates represented as 0's and the other senators as 1's. To test for significant difference from the null hypothesis of a random ordering we found the probability of observing less than seven runs (at least five candidates clustered together) occurring in any ordering of the binarized senator representations. A run denotes a contiguous stretch of the ordering with senators from the same category (0 or 1). In both the Senate-wide and Democrat-only circular orderings, p-values were <0.001, revealing a statistically significant departure from randomness in the non-random ordering of these senate members

The inherent feature-representation of split-systems described previously also facilitates mapping of the splits of interest back to the features (votes) that underlie that split. For instance, given the split of five Democratic Primary candidates, we traced back the split to the votes contributing to their unique voting pattern by first extracting votes where all candidates voted the same. Of these votes, we found a particular set in which a majority of the rest of the party did not vote in accordance with these senators (Fig. 4a), temporally clustered in the latter half of 2019 (Fig 4a). These votes with the largest discrepancy were all abstentions by these senators, behavior which aligns with the previously noted trend of presidential candidates abstaining during campaign periods [2] (Fig. 4a).

For these (or any) splits of interest we can assign a statistical interpretation to how the votes contribute to the splits of interest by ranking them by p-value as described in the Methods. For this particular split of the five Primary candidates we see that the ranking results (Fig. 4b) are concordant with the votes of low intra-party agreement (Fig. 4a). The clustered abstentions have the highest likelihoods of contributing to splits, among other Yea or Nay votes also contributing to this split.

The p-value assignments also allow for investigation of distinct behaviors in the votes contributing to a split of interest. With the ranked votes we fit a LOESS (Local Regression) curve to the p-values (Fig. 4b, dashed line). This demonstrates the apparent temporal progression the contributing votes follow, with an upward trend in p-values leading to the abstention period, and a decrease in rankings following that time period (Fig. 4b).

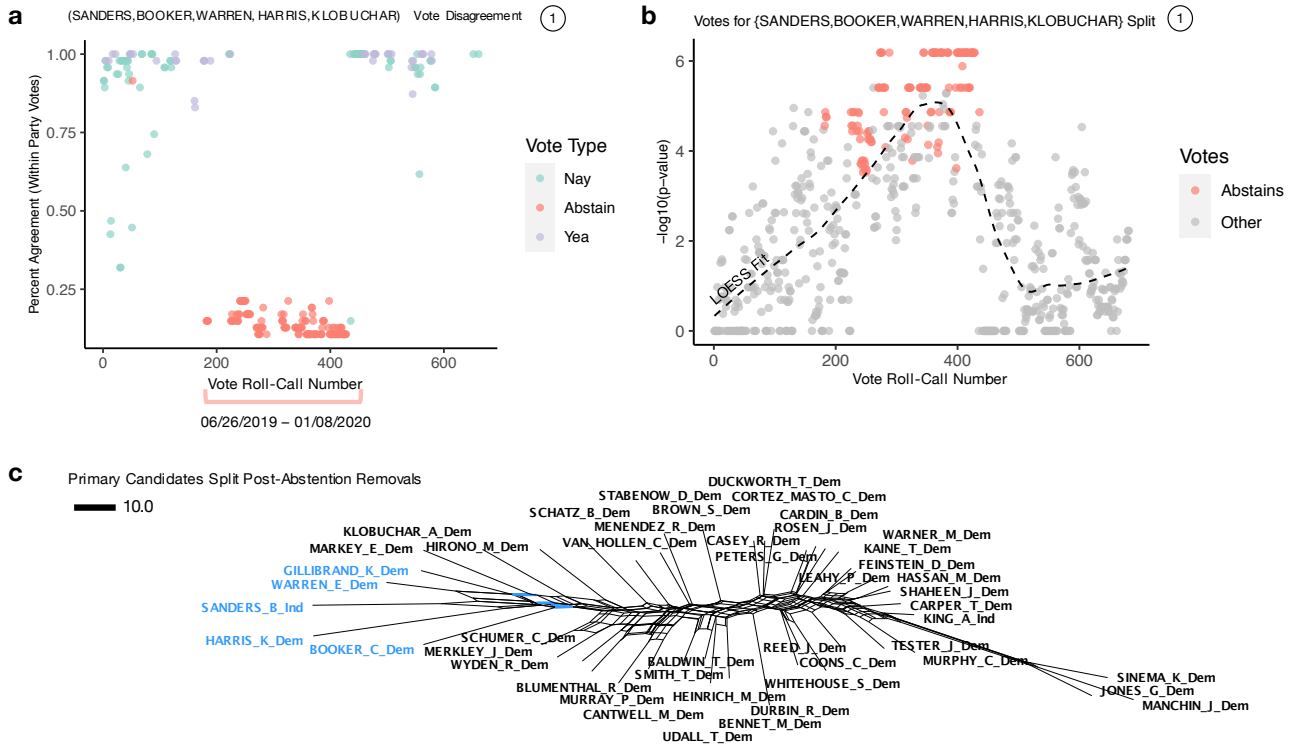

Figure 4: **a)** For roll-call votes where these five Primary candidates voted the same, percent disagreement within the party is shown (fraction of remaining members of the party who voted differently). Votes colored by the vote cast by the candidates. **b)** All roll-call votes ranked by p-value for the given split of the five Primary candidates. Abstentions with low agreement from **a** colored in ranking. Raw p-values reported here. LOESS (Local Regression) fit for p-value rankings over time (roll-call votes) shown by dashed line **c)** Splits graph for splits network of only Democrat (inclusive of Independent) votes after two iterative removals of low-agreement abstentions for clustered Primary candidates. [Code]

We then removed these clustered abstentions to discern who these five Senate members vote similarly to outside of this abstention time period. After a second removal of low agreement abstentions, for the split of Senators Booker, Sanders, Warren and Harris, who remained clustered despite the initial removal, we see that this group remains split from the rest of the party by voting behavior, with Sen. Gillibrand (Fig. 4c).

This analysis is also not limited to any particular split. Thus we next applied these techniques to another apparent 'coalition' within the party structure (Fig. 5).

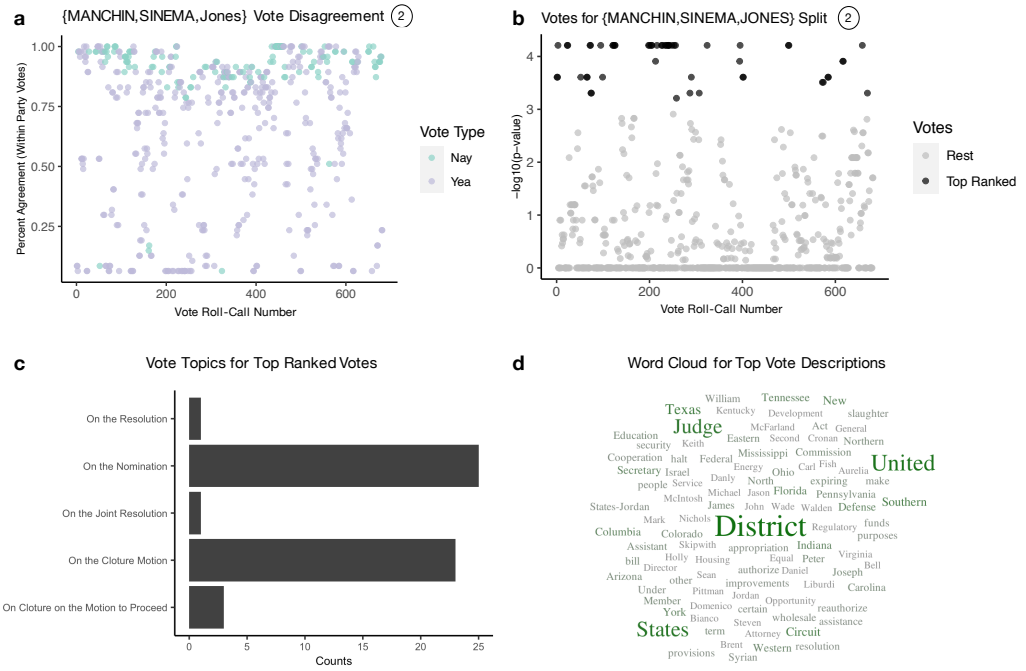

Figure 5: **a)** For roll-call votes where Senators Manchin, Sinema, and Jones voted the same, percent disagreement within the party is shown (fraction of remaining members of the party who voted differently). Votes colored by the vote cast by the candidates. **b)** All roll-call votes ranked by p-value for the given split of the Senate members. 'Top Ranked' votes (-log10(p-values) above 3) highlighted **c)** Counts of vote topics for top votes colored in **b** shown **d)** Word cloud for top vote descriptions shown. Generated from worditout.com. [Code]

We focused on the split of Senators Manchin, Sinema, and Jones, who cluster on the opposite end of the Democrat split network from the candidate senators, (Fig. 2c, denoted as 2) and are situated between both major parties in the Senate-wide split network (Fig. 2a, denoted as 2). Votes separating these senators from the rest of their party, obtained by the p-value ranking described above, are scattered across time (Fig. 5a,b), in contrast to the clustered abstentions of the previous split of senators analyzed. By tracing back the contributing votes for this split, we can additionally quantify the prevalence of particular topics in the highest-ranked votes contributing to the split (Fig. 5c). With detailed descriptions of the top votes, we can visualize the representative content of these votes as well (Fig. 5d). Although p-values are used here for ordinal purposes, they can be corrected for multiple testing to determine which votes are significantly associated with a split.

Temporal Variation in Distributions of Party-Specific Voting Agreement

From the split networks in Figures 2 and 4c,d, we note a variety of structures within the Senate and the individual parties, with particularly dense areas and sparse regions of individuals denoting areas of high or low voting agreement. To assess and visualize this agreement across Senate members we denoted the 'center' (Fig. 2a) split as described in the Methods to make relative quantifications of the spread of member's voting behaviors. This also provides a comparative metric for how 'left' or 'right' of center members are [20]. This assignment of distances from the center is not limited to the 116th Congress, and thus we explored the dynamics of this metric over time for all Senates over the last 30 years (Fig. 6a).

By aggregating distances for each of the main parties, we visualized if or how the spread and magnitude of voting agreement within and between parties has changed over time as a product of their constituents. What we observed fits with previously reported trends of increasing partisanship in the Senate [1, 24, 32], at least within the last six years. This is demonstrated by upward shifts in the median party distances, i.e. increasing distances of each party's members from the center. The larger spread of center distances observed in the Democratic party in recent Senates versus a tightening of the Republican distances also suggests differing levels of voting unification within each party [30]. This is also in contrast to earlier senates, where greater 'unification' (tighter distance distributions) in the Democratic party is demonstrated (Fig. 6a 101st, 102nd). Shifts in party-specific voting unification were further investigated by examining the distribution of center distances ranges (the difference between the highest and lowest distance) for each Senate session with respect to the party in the Senate majority (Fig. 6b). This revealed a significant difference in the range or spread of voting behaviors within each of the two main parties (Independents included with Democrat senators) when the opposing party was in the majority versus minority (Fig. 6b). Though there are many factors which can contribute to greater or lesser party unity [45], this suggests a relationship between voting behavior and the party's standing in the Senate, possibly related to recent observations on the ability of the majority party to influence the legislative agenda of the chamber floor particularly when the party is ideologically cohesive [9].

We additionally investigated these agreement distributions at the level of their constituent members, as visualized for the 116th Senate (Fig. 6c). At the level of individual senators we can note the differences in magnitude of the center distances among non-Republican senators versus Republican senators and place each senator within this greater distribution. These individual distances were then compared to the coordinates of the ideal points assigned to each senator by DW-NOMINATE, demonstrating a high correlation of ~ 0.8 to this benchmark methodology within each party (Fig. 6d) [8]. Here we utilized the first dimension of the DW-NOMINATE coordinates, as it tends to be the most interpretable and commonly utilized part of the embedding space [14]. This highlights the ability of NNet to not only replicate the spectrum of 'left' and 'right' within the Senate, as the DW-NOMINATE coordinates reveal [8, 14], but also provide the structure of coalitions within which these preferences reside.

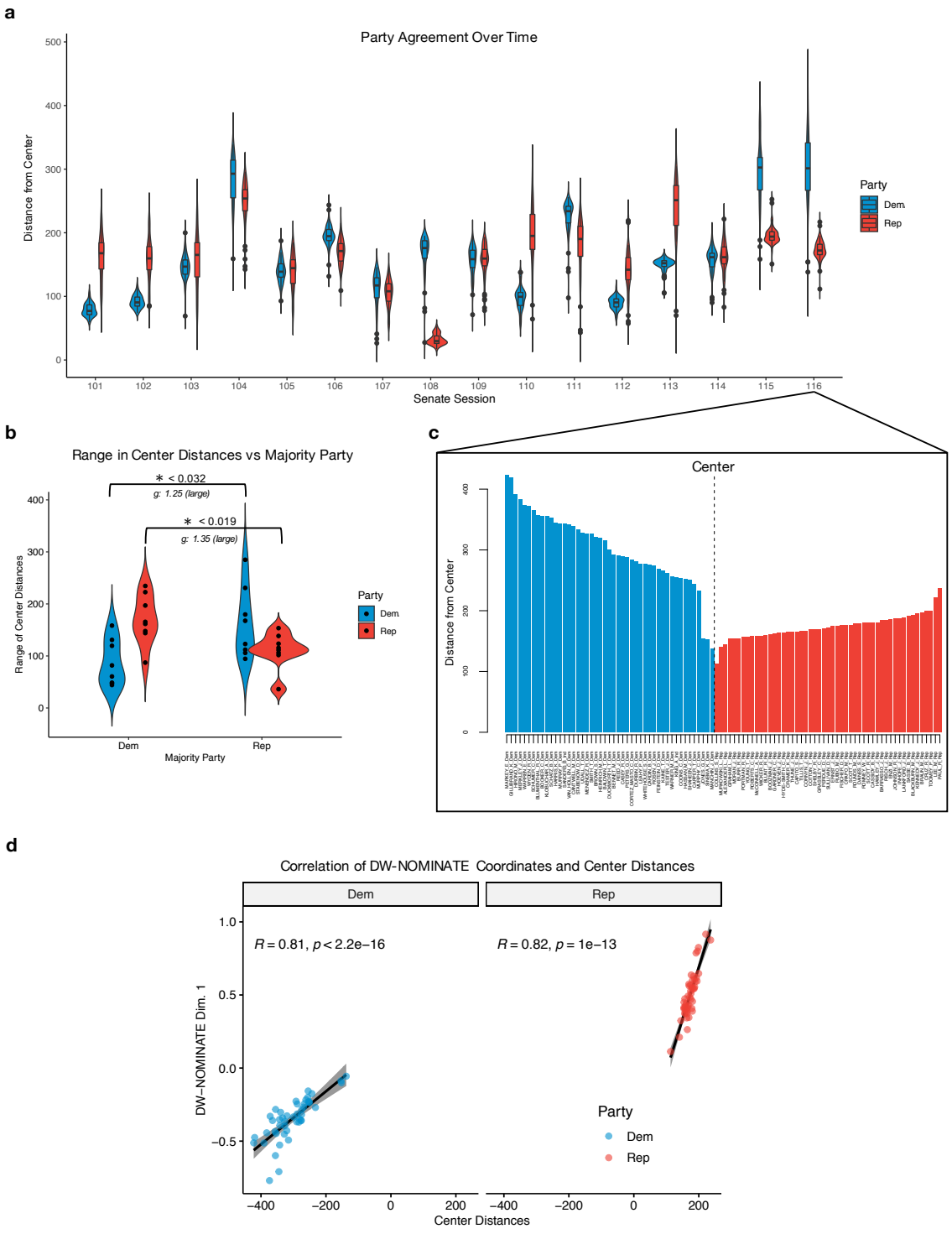

Figure 6: **a)** Distribution of members distances from center shown for the 101st to 116th US Senates, within each of the main parties. **b)** Distribution of center distance ranges, calculated for each party per Senate session. Mann-Whitney U-test used to determine if party-specific ranges differ in sessions with either party in the majority. $n = 16$. * denotes p-value <0.05. g denotes Hedges' g measure of effect size. **c)** Center distances for the 116th Senate shown for each senator. **d)** Spearman correlation between DW-NOMINATE coordinates (first dimension) and center distances, by party, for senators in the 116th senate. Center distances on opposing sides of the center split are negated. [Code]

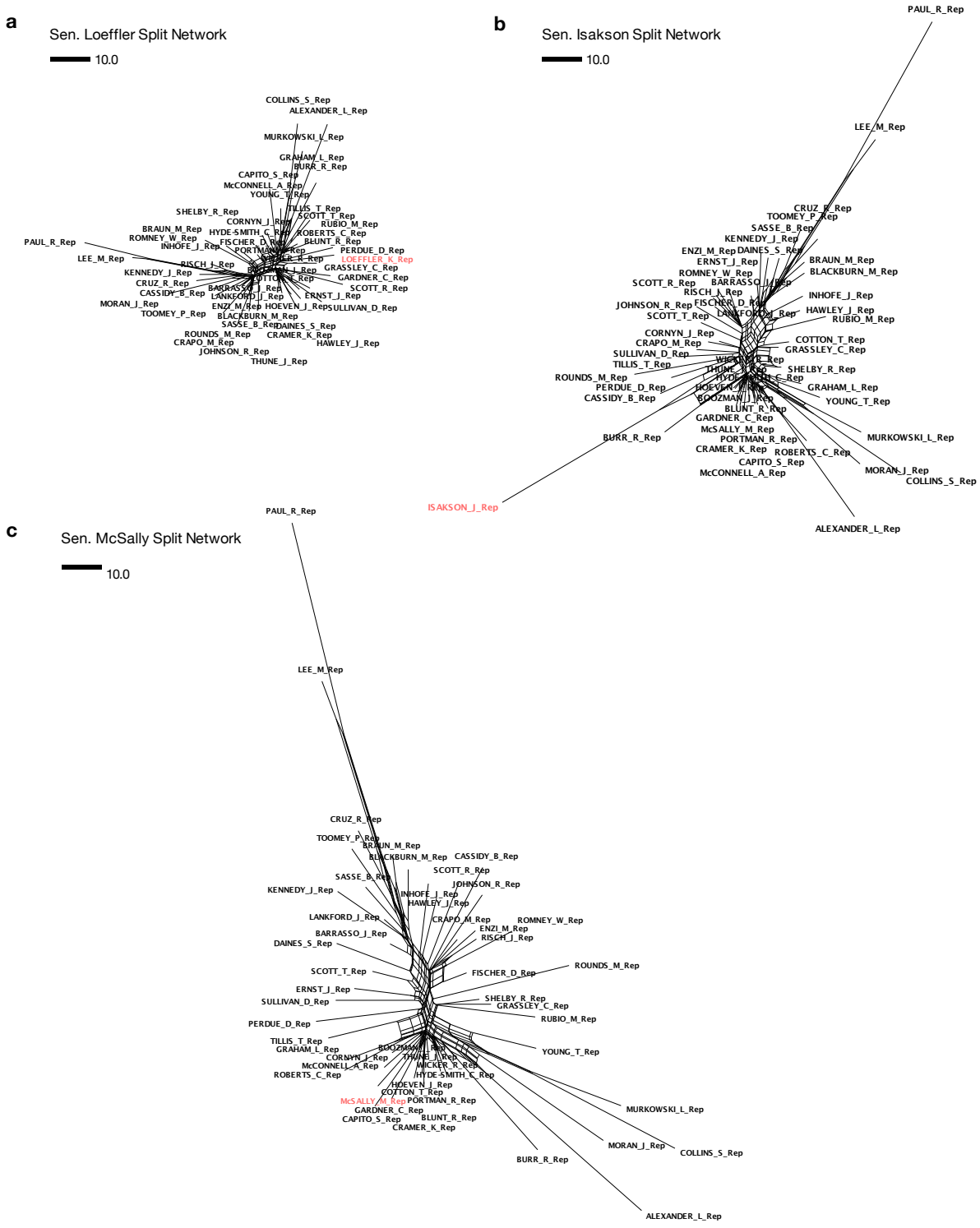

Figure 7: **a)** Splits graph of the Republican party (in the 116th Senate) for votes Sen. Loeffler was present for **b)** Splits graph of the Republican party for votes Sen. Isakson was present for **c)** Splits graph of the Republican party for votes Sen. McSally was present for. [Code]

Split System Visualizations for Partially Active Senators

For Senate members who were not active for the full term of the 116th Congress, and thus not included in the main Senate analysis, we created split-network visuals for the voting period in which they were present (Fig. 7). From this, we can view for these senators who their nearest political neighbors are given the votes they did cast. This includes a split-network for Sen. Loeffler (Rep.)(Fig. 7a) and her predecessor Sen. Isakson (Rep.) (Fig. 7b), as well as Sen. McSally. Sen. Kelly (Dem.)(Fig. 7c), who recently succeeded Sen. McSally, did not have enough votes to create a representative splits graph. While Sen. Loeffler's nearest neighbor is her fellow Georgia senator, Sen. Perdue, Sen. Isakson appears more discordant in his voting behavior during the 116th Senate, visible by the long split separating him from the rest of the party. Sen. McSally's positioning is similar to that of Sen. Loeffler, in that she is clustered near other members (neighbors with Senators McConnell and Capito), lying within the denser region of the party's split-network.

Discussion

Our findings highlight the utility of the NNet-SplitsTree algorithms in creating representations and visualizations of voting data that facilitate exploratory analysis and facilitate identification of voting patterns that may not be readily apparent. This non-model-based approach minimizes assumptions on the structure of the input data, though the circular nature of the split-system can limit which relationships are accurately recapitulated in the visualization. However, as mentioned previously, these discrepancies can be utilized to detect members of the network who display more discordant or 'maverick' behavior. The analysis framework we have proposed is additionally limited to political relationships and structures visible at the level of voting behavior, and it is important to keep in mind that there may be other factors and behaviors which may influence the relationships between political members.

With the NNet-based approach, we determined relationships between pairs of senators within and across their respective parties, highlighting the impact of inter versus intra-party voting behaviors on the stability of those relationships. The relative lengths of the generated splits additionally provided a quantitative visualization of both the strengths of these relationships and the level of divergence those shared behaviors represented. This gave rise to visible coalitions of senators within the greater split system, notably five of the 2020 Democratic Party presidential candidates and a separate group of Democrat 'centrists'. Utilizing the direct relationship of the defined splits to the input voting data, we recovered the contributing votes to each coalition and verified the existence of shared voting behaviors unique to the primary candidate coalition beyond the abstentions common during presidential bids. The split-system also provided an interpretable framework for the development of a statistical procedure to rank contributions of each vote to any given split of interest. Given these ranked votes, we extracted the defining votes underlying the unique voting behaviors of both the primary candidates and the noted 'centrists' as well as the content of these votes, connecting shared behaviors to their comprising features in a statistically rigorous manner.

To take further advantage of the quantitative and qualitative aspects of the split-networks we generated, we expanded our analyses to the dynamics of inter- and intra-party voting agreement over Senate sessions. Given the natural representation of a Senate 'center' inherent in the split-

network, we measured relative levels of polarization in each of the main parties with respect to each member's voting patterns. This provided a representation which not only reflected the ideal point placements of senators within a given Senate session, but also highlighted periods of greater or lesser party polarization and the possible effects of majority party's influence on these noted changes in party-specific voting behaviors. Thus, the generated split-system enables comparative, temporal analysis across the represented social structures. Together, this work demonstrates the power of the NNet-SplitsTree approach, a biologically-minded methodology, to address common questions in exploratory political data analysis. Moreover, we believe that the methods we have described here will be powerful for the structural analysis of voting patterns in other domains where it is of interest to understand the nature of coalitions within social structures.

Methods

Defining the Roll-Call Based Dissimilarity Matrix

To assess the behaviors of the senators we chose to use their roll-call votes [27], a popular choice for determining similarities, differences, and coalitions between political members [28, 29, 43, 46]. We initially investigated the structure of the current Senate in the 116th Congress. We placed the vote records of the senators for the given Congress in an $n \times v$ matrix R with n senators, and v votes. Senators not present for the entirety of a given Congress are not used in this evaluation. Values for each entry in matrix R are defined by each of the n senator's votes:

$$R_{ij} = \begin{cases} 0 & \text{if Nay} \\ 1 & \text{if Yea} \\ 1/2 & \text{if Abstain*} \end{cases}$$

for $i = 1...n, j = 1...v$. *Abstain incorporates abstentions and 'Present' votes.

A distance (dissimilarity) matrix δ was created by computing pairwise L_1 distances between pairs of senators to obtain an $n \times n$ matrix with entries:

$$\delta_{ij} = \|R_i - R_j\|_1 \tag{1}$$

for $i, j = 1...n$. This matrix δ was provided as the distance matrix input to the NNet implementation in SplitsTree4 [19], to calculate the ordering π and splits Σ for a graphical, non-hierarchical visualization, and to extract the split weights λ for the system.

Statistical Analysis of Vote Contribution to Split Definition

We then used this representation of the matrix δ to identify outstanding structures or relationships and traced their origin to the original input votes, utilizing the inherent properties of split-systems highlighted previously i.e. the feature representation underlying the construction of the splits. To extract votes which contribute to particular splits of interest we applied the split to the original voting input R, and selected for features (votes) which characterized that split (where the votes of that individual or group of individuals differ from the other members).

We additionally ranked votes by their likelihood of being associated with any given split of interest. To do so we defined p-values for a vote using the Fisher's exact test [16] for a 2×3 contingency tables between a split $\{A|B\}$ and counts of the vote types $\{0, 1/2, 1\}$. The table, shown below, denotes the counts for the intersections between members in $\{A, B\}$ and $\{C, D, E\}$, where C, D, and E are the sets of members whose votes were 0, 1/2, and 1 respectively.

	C (0)	D (1/2)	E (1)
A	$\lvert A \cap C \rvert$	$\lvert A \cap D \rvert$	$\lvert A \cap E \rvert$
B	$\lvert B \cap C \rvert$	$\lvert B \cap D \rvert$	$\lvert B \cap E \rvert$

We used a two-sided Fisher's exact test to determine p-values for assessing how likely a more 'extreme' contingency table for (how far from random) a particular vote's table would be. We were thereby ranking, for a given split of Senate members, the likelihood of the voting behaviors in each vote being associated with that split. For ranking purposes, we report the raw p-values of each vote.

Voting Agreement as a Measure of Distance from the Senate Center

In order to quantitatively compare agreement of senators within and across parties we used the circular split-system to define a Senate 'center' against which all senators can be compared. For the given Congress, the center can be defined by the exact split which delineates the two main divisions within the split-network i.e. the Republican party members and the Democrat members (inclusive of Independents) [35]. From this split, distances of any individuals (or groups) can be obtained by summing the appropriate split weights from the calculated λ. We define the distance $d(S)$ between some subset S of all members M as the sum of the split weights of all splits $A|B$ for which the elements (members) of S are separated [31]:

$$d(S) = \sum_{\substack{A|B \in \Sigma \\ A \cap S \neq \emptyset \\ B \cap S \neq \emptyset}} \lambda_{A|B}. \tag{2}$$

Center distances are then calculated for individuals over their time in the Senate by summing the weights of all splits which separate the member from all members of the opposing party.

Data Availability

All data was downloaded from https://voteview.com/data, and is also available in https://github.com/pachterlab/CP_2021. The data was last accessed on 12/13/2020.

Code Availability

All code for analysis and figure generation is available at https://github.com/pachterlab/CP_2021. The SplitsTree4 program was used to run Neighbor-Net and to create the split networks displayed in Figures 1, 2 and 4. Python code is also provided to implement all Neighbor-Net calculations and vote analyses without use of the SplitsTree4 GUI.

Acknowledgements

This project started with exploratory data analysis by Tonje Fredriksen and Corina Ray, and was inspired by a lecture of David Bryant where he discussed the possibility of using Neighbor-Net and SplitsTree to analyze United Nations votes. We also thank Daniel Ebanks for his helpful feedback on the manuscript.

References

1. C. Andris *et al.*, en, *PLoS One* **10**, e0123507 (Apr. 2015).
2. C. Boudreaux, R. M. Coats, B. Walia, *Voting and Abstaining in the US* (2012).
3. T. J. Brazill, B. Grofman, *Soc. Networks* **24**, 201–229 (July 2002).
4. D. Bryant, F. Filimon, R. D. Gray, *The evolution of cultural diversity: A phylogenetic approach*, 67–84 (2005).
5. D. Bryant, V. Moulton, en, *Mol. Biol. Evol.* **21**, 255–265 (Feb. 2004).
6. A. Burns, M. Flegenheimer, J. C. Lee, L. Lerer, J. Martin, en, *The New York Times* (Jan. 2019).
7. A. Bycoffe, *Tracking congress in the age of trump*, https://projects.fivethirtyeight.com/congress-trump-score/, Accessed: 2020-12-30, Jan. 2017.
8. R. Carroll, J. B. Lewis, J. Lo, K. T. Poole, H. Rosenthal, *Polit. Anal.* **17**, 261–275 (2009).
9. J. L. Carson, A. J. Madonna, *et al.*, *Am. Polit. Q.* (2016).
10. J. De Leeuw, *Real Data Analysis*, 405–411 (2006).
11. F. G. de Borja, C. M. D. S. Freitas, presented at the 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, pp. 257–264.
12. A. Dress, K. T. Huber, J. Koolen, V. Moulton, A. Spillner, *Basic Phylogenetic Combinatorics*, en (Cambridge University Press, 2012).
13. M. Dunn, A. Terrill, G. Reesink, R. A. Foley, S. C. Levinson, en, *Science* **309**, 2072–2075 (Sept. 2005).
14. P. Everson, R. Valelly, A. Vishwanath, J. Wiseman, *Studies in American Political Development* **30**, 97–115 (Oct. 2016).
15. D. P. Faith, *Biol. Conserv.* **61**, 1–10 (Jan. 1992).
16. R. A. Fisher, *J. R. Stat. Soc.* **85**, 87 (Jan. 1922).
17. S. Holloway, *Canadian Journal of Political Science/Revue canadienne de science politique* **23**, 279–296 (1990).
18. D. H. Huson, en, *Bioinformatics* **14**, 68–73 (1998).
19. D. H. Huson, T. Kloepper, D. Bryant, *Bioinformatics* **14**, 68–73 (2008).
20. D. Jahn, *Party Politics* **17**, 745–765 (Nov. 2011).
21. A. Jakulin, W. Buntine, T. M. La Pira, H. Brasher, *Polit. Anal.* **17**, 291–310 (2009).
22. I. S. Kim, D. Kunisky, *Code Ocean. https://doi. org/10. 24433/CO* **776811**, v1 (2020).
23. A. Kreppel, G. Tsebelis, *Comp. Polit. Stud.* **32**, 933–966 (Dec. 1999).
24. F. E. Lee, *J. Polit.* **80**, 1464–1473 (Oct. 2018).
25. S. D. Levitt, *Am. Econ. Rev.* **86**, 425–441 (1996).
26. D. Levy, L. Pachter, *Adv. Appl. Math.* **47**, 240–258 (Aug. 2011).
27. J. B. Lewis *et al.*, *See https://voteview. com/(accessed 27 July 2018)* (2018).

28. A. J. Madonna, *Am. J. Pol. Sci.* **55**, 276–288 (Apr. 2011).
29. G. B. Markus, *Am. J. Pol. Sci.* **18**, 595–607 (1974).
30. W. G. Mayer, *The Divided Democrats: Ideological Unity, Party Reform, And Presidential Elections*, en (Routledge, Feb. 2018).
31. B. Q. Minh, S. Klaere, A. von Haeseler, en, *Syst. Biol.* **58**, 586–594 (Dec. 2009).
32. J. Moody, P. J. Mucha, *Network Science* **1**, 119–121 (Apr. 2013).
33. D. A. Morrison, *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **4**, 296–312 (July 2014).
34. R. Nunes Moni da Silva, A. Spritzer, C. Dal Sasso Freitas, presented at the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 150–157.
35. *Party Division*, https://www.senate.gov/history/partydiv.htm, Accessed: 2020-12-30, Jan. 2019.
36. A. Perer, B. Shneiderman, presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 265–274.
37. B. Plechanovová, en, *Czech Economic Review* **5**, 249–266 (2011).
38. K. T. Poole, H. Rosenthal, *Am. J. Pol. Sci.* **29**, 357–384 (1985).
39. C. Possieri, C. Ravazzi, F. Dabbene, G. C. Calafiore, en, *PLoS One* **15**, e0238481 (Sept. 2020).
40. *Report cards for 2019 - ideology score - all senators - GovTrack.Us*, https://www.govtrack.us/congress/members/report-cards/2019/senate/ideology, Accessed: 2020-12-30.
41. J. M. Sabucedo, C. Arce, en, *Eur. J. Polit. Res.* **20**, 93–102 (July 1991).
42. D. Schoch, U. Brandes, en, *Sci. Rep.* **10**, 17369 (Oct. 2020).
43. J. A. Segal, C. M. Cameron, A. D. Cover, *Am. J. Pol. Sci.* **36**, 96–121 (1992).
44. R. Shikiar, en, *Multivariate Behav. Res.* **9**, 461–477 (Oct. 1974).
45. J. M. Snyder, T. Groseclose, *Am. J. Pol. Sci.* **44**, 193–211 (2000).
46. T. Stratmann, *Am. Polit. Sci. Rev.* **94**, 665–676 (2000).
47. S. Wahl, J. Sheppard, E. Shanahan, presented at the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), pp. 718–725.
48. A. Wald, J. Wolfowitz, *Ann. Math. Stat.* **11**, 147–162 (1940).
49. T. Zhang *et al.*, en, *Sci Adv* **6** (Nov. 2020).