ANALYZING AND VISUALIZING POLARIZATION AND BALANCE WITH SIGNED NETWORKS: THE U.S. CONGRESS CASE STUDY

Arthur Capozzi Department of Computer Science Università degli Studi di Torino Via Pessinetto, 12, 10149, Torino, Italy Corresponding author: arthurtomasedward.capozzi@unito.it

Alfonso Semeraro Department of Computer Science Università degli Studi di Torino Via Pessinetto, 12, 10149, Torino, Italy **Giancarlo Ruffo**

Department of Science and Technological Innovation (DISIT) Università del Piemonte Orientale "A. Avogadro" Viale Teresa Michel, 11, 15121 Alessandria, Italy

ABSTRACT

Signed networks and balance theory provide a natural setting for real-world scenarios that show polarization dynamics, positive/negative relationships, and political partisanship [1]. For example, they have been proven effective in studying the increasing polarization of the votes in the two chambers of the U.S. Congress from World War II on [1, 2].

To provide further insights into this particular case study, we propose the application of a pipeline to analyze and visualize a signed graph's configuration based on the exploitation of the corresponding Laplacian matrix' spectral properties [3]. The overall methodology is comparable with others based on the frustration index [4], but it has at least two main advantages: first, it requires a much lower computational cost; second, it allows for a quantitative and visual assessment of how arbitrarily small subgraphs (even single nodes) contribute to the overall balance (or unbalance) of the network.

The proposed pipeline allows the exploration of polarization dynamics shown by the U.S. Congress from 1945 to 2020 at different resolution scales. In fact, we are able to spot and point out the influence of some (groups of) congressmen in the overall balance, as well as to observe and explore polarization's evolution of both chambers across the years.

Keywords signed networks · balance theory · network visualization · polarization

1 Introduction

Studying opinion and political polarization dynamics in social networks is considered crucial to understand how individuals that belong to different communities can be exposed to a plurality of information and beliefs [5]. The consequences of strong polarization, such as the emergence of the so-called "echo-chambers" [6] and increased partisanship in politics [1], can represent a risk for modern democracies. They can lead to "filter bubbles" undermining social media recommendation algorithms [7], accelerate the spreading of misinformation [8], among other things. There is a general agreement that polarization is a negative consequence of network's homophily, and explicit relationships between individuals have been intensively used to study related phenomena in complex social systems. Friendship is usually represented by means of positive relationships, even if in many different domains it is important to consider also negative ties to better understand polarization dynamics [9].

Signed networks can effectively represent patterns of collaboration, friendship, or membership in a group, by means of positive edges, as well as explicit disagreements, different opinions, antagonisms, by means of negative edges.

In this type of networks, the coexistence of two forces inevitably leads to the formation of cohesive groups of individuals, also called clusters. A *balanced network* represents a perfectly polarized system, where there are only positive relationships between nodes of the same group and negative edges between nodes belonging to different clusters. Perfectly balanced systems rarely manifest themselves in practice; however, we can have several shades of unbalance.

In many analyses, we may want to check both the presence of homophily/heterophily due to the node attribute and how far a network is from being balanced: these measures could help to understand the polarization dynamics of the system from different points of view. Political systems, as roll call votes in the U.S. Congress, are paradigmatic: we can label nodes as belonging to a party (e.g., Democrats or Republicans), and we expect that positive links are more frequent than negative ones between intra parties representatives, as well as the opposite between inter parties nodes. Observing how balance changes over time can help us understand polarization dynamics; monitoring homophily may support the analyst to reveal signals of partisanship or anti-partisanship behaviors among (group of) nodes in the network.

1.1 Our contribution

Although there are several algorithms and measures in the literature that would support signed network analysis, a pipeline that combines these techniques is still missing.

In this paper, we present an analytical pipeline that leverages techniques and metrics from the literature to efficiently analyze and visualize signed networks in terms of both homophily and balance. We present the potential of this pipeline to dissect parts of the network and assess their contribution to the overall structural balance, allowing for a fine-grained analysis of the actors (and the parties they belong to) that make the network less balanced and unstable in its configuration.

We showcase this analytical pipeline by studying the polarization of the votes in the two chambers of the U.S. Congress from 1945 to 2020, where nodes have explicit attributes denoting their membership to a specific party. The results are consistent with others in the literature, and a fine-grained analysis of Senator Manchin's political transformation demonstrates the new potential of this pipeline.

The measures adopted to support the analysis and visualization of the signed graph are based on the spectral analysis of the Laplacian matrix. The computation of the smallest eigenvalue λ_n can be used to understand whether the network is balanced and, if not, how far it is from being balanced. We compare the results and the computational cost to another metric based on the frustration index [4]. Furthermore, the eigenvector v_n corresponding to the smallest eigenvalue λ_n , as suggested by Kunegis et al. [10] and Galimberti et. al. [3], can be used to i) study the influence of arbitrarily large subgroups of nodes on the overall balance, and ii) display nodes in a Cartesian plane depending on their individual contribution to the network's balance.

Given the great opportunities of applying signed networks in real-world scenarios, a Python library containing the analytics and visualizations underlying the pipeline presented in this study is being built and released. In this library, called Sygno¹, we intend to implement further functions to support the study of signed networks and balance theory.

Related work is reviewed in Section 2, while in Section 3: i) we summarize some basic definitions of signed networks, spectral analysis, and structural balance; ii) we give a mathematical definition of the metrics used in this paper to compute the structural balance. In Section 4, we describe the construction of signed networks starting from the voting data of the U.S. Congress.

In Section 5, we present a comparative analysis of two structural balance computation metrics; we show how the spectral analysis of the Laplacian matrix and the data visualization can support the study of the influence of individual nodes on the network balance. We conclude by presenting a computational cost comparison of structural balancing computation metrics.

2 Related works

2.1 Signed networks

Network science has many applications in social media and social networks analysis [11, 12]. Networks that only represent relationships of friendships or following in social media can be seen as signed networks with only positive edges. Many scholars have studied how to deal with the increased complexity of the models that also represent the relations of antagonism through negative edges, Tang et al. [13] describe the process of mining social media data to

¹https://github.com/alfonsosemeraro/draw_signed_networkx

build signed networks based models. The increase in complexity makes modeling the network more difficult, but it also provides many opportunities for further analyses.

Balance theory is central to many studies involving signed networks. The theory's fundamental observation is that three nodes linked to each other can show both stable and unstable configurations. For instance, the enmity between two nodes in a friendly relationship with a third node is a source of contrast, and the triangle is defined unbalanced. Thus, the *structural balance* of a signed network is the overall property of the network to be built upon stable triangles. In social networks, the concept of balance was first formulated by Heider in 1946 [14]. He theorized that the relationships that form the structure of social networks may be characterized by the dualism of friendship and hostility. Heider's studies influenced the social psychology of the years to come, but it was only in 1956 that the balance was modeled in terms of signed graphs [15]; in fact, Cartwright and Harary's intent was to mathematically formalize balance theory and to extend it to other domains, such as communication networks, power systems, sociometric structures, systems of orientations, or neural networks.

Nowadays many applications of this theory to signed networks can be found in the literature; e.g., it has been used to analyze the interactions and polarization of users in social media [16, 17], or to study international relations between countries [18], or biotic interactions in real ecosystems [19].

In particular, Leskovec et al. [20] tried to evaluate social-psychological theories on the balancing of social systems by modeling large-scale datasets with signed networks. They also proved that positive link prediction can be improved by using both positive and negative edges [21]. Other studies have focused on the application of signed networks with the aim of improving the performance of recommendation systems [22, 23].

Some studies have modeled recommendation systems [24] or social networks [25] with time-varying signed networks. In fact, real-world signed networks can change over time in many aspects, such as structure, node clustering, or structural balance.

Another relevant field of application of signed networks concerns the study of the voting behavior of politicians and voters. Porco et al. [26] tried to predict polarity and voting behavior by using a recommendation system-based approach. Intal and Yasseri [27] analyzed the individual interactions of House of Commons Members of Parliament (MPs) in the voting process. The purpose of their study was to predict how MPs would have voted in a forthcoming Brexit deal. They computed pairwise similarity scores and calculated "rebellion" metrics based on eigenvector centrality obtaining an accuracy of 90% in the prediction task. Eigenvector centrality has been shown to be an effective method for quantifying the power and overall strength of a node in a signed network [28].

Several aspects of American politics and the U.S. Congress have been the subject of network science studies and analyses. In 2015, Andris et al. [1] analyzed the evolution of partisanship in the United States Congress from 1949 to 2012. From roll-call voting they have created an undirected network of over 5 million pairs of representatives, and showed that partisanship has increased exponentially over the past 60 years. In 2020, Neal [29] analyzed the polarization in the U.S. Congress by applying the Stochastic Degree Sequence Model and creating signed networks of political relationships among legislators. An increase in polarization in the U.S. Congress has been shown by using the triangle index, a metric for computing partial structural balance of signed networks.

In 2021 Aref and Neal [2] studied polarization on political networks proposing new models for optimal clustering of signed networks. They have confirmed the increase in polarization in the US Congress since 1979. They also found that partisanship may positively influence the effectiveness of passing bills. In their study, they used two metrics for measuring partial balance: triangle index, i.e., the fraction of positive cycles of length 3, and the frustration index, based on the minimum number of edges that should be removed to get a balanced network.

Similar analyzes regarding political votes and polarization, involve among others the study of the European Parliament [30, 31], the U.S. Congress [32, 33] and the Brazilian House of Representatives [34].

2.2 Compute structural balance

The first formal definition of structural balance in terms of signed graphs is due to Cartwright and Harary [15]: a signed network Γ is balanced if and only if its nodes can be partitioned into two disjoint subsets V_1 , V_2 in such a way that each positive edge of Γ joins two nodes of the same subset and each negative edge joins two nodes of different subsets.

It is possible to determine whether a signed network is balanced or not in polynomial time [35]. When a signed network is not balanced, it is possible to compute the partial structural balance, which is the distance of the network from being balanced. Calculating how far a network is from being balanced is much more complex. Approaches based on the computation of the least number of edges that must be dropped to obtain a balanced network, as the frustration index,

are nondeterministic polynomial-time hard problems [35]. Other approaches are based on the spectral analysis of the Laplacian matrix of the signed graph, and have a complexity of $O(n^2)$, like the algebraic conflict.

Some of the most common approaches are walk-based [36, 37, 38]. A walk of length k in Γ is a sequence of connected nodes v_0, \ldots, v_k . The sign of the walk is the product of the signs of the edges that connect the nodes v_0, \ldots, v_k [39]. The walk is defined as closed (or as a cycle) if $v_0 = v_k$ and if its length is equal to k. A signed network is walk-balanced if every closed walk is positive [36]. Hence, a network is balanced if every closed walk has an even number of negative edges [35].

2.3 Visualize structural balance

Many network visualization techniques are proposed in the literature, but very few approaches are designed to visualize specifically signed networks. Kunegis et al. [10] showed how the properties of spectral analysis can be used to draw signed graphs. In order to obtain an embedding of the nodes of a signed graph into a plane, they used the eigenvectors of the two smallest eigenvalues as coordinates for drawing the nodes. Using this method, when the graph is balanced $(\lambda_{|V|} = 0)$, all nodes are placed on two parallel lines, reflecting the perfect bi-partitioning present in the graph.

In our study we use another spectral analysis based algorithm called *Structural-balance-viz* [3]. If compared with a generic force-based graph drawing layout where edges can be shown as positive/negative (for example as solid/dashed lines), this visualization scheme has multiple advantages: i) it shows clearly whether the network is balanced or not; ii) it allows to visually represent the partial balance; iii) it is capable of providing information on the contribution of individual nodes to the balance of the entire network; iv) even a slight change to the graph would correspond to a visual modification, allowing to spot and compare differences between different graphs. In Section 5.2 we provide an accurate description of this method and we apply it to the U.S. Congress roll call votes dataset.

3 Methods

In this section we define the partial balance indices that we compute and compare in section 5 to analyze the temporal dynamics of political polarization from 1941 to 2020 in the U.S. Congress.

3.1 Notations and metrics

Let $\Gamma = (G, \sigma)$ be an undirected signed graph, where G = (V(G), E(G)) is a graph, and σ is the sign function $\sigma : E(G) \to \{-1, +1\}$. We define E_+ as the set of positive edges, and E_- as the set of negative edges. A is the signed adjacency matrix of Γ , i.e., for each pair of nodes $u, v \in V$, A[u, v] = 1 if $(u, v) \in E_+$, A[u, v] = -1 if $(u, v) \in E_-$. Let $\overline{D} = diag(\overline{d}_{u1}, \ldots, \overline{d}_{u|V|})$ be the signed degree matrix of Γ , where $\overline{d}_u = \sum_{v \in V} |A[u, v]|$. Finally, we define the signed Laplacian matrix $\overline{L} \in \mathbb{R}^{V \times V}$ of Γ as:

$$\bar{L}(\Gamma) = \bar{D} - A \tag{1}$$

The Laplacian matrix spectrum captures some characteristics of the relationship between the nodes. It is composed by the |V| Laplacian eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{|V|}$. Laplacian eigenvalues are all real and non-negative, and contained in the interval $[0, min\{|V|, 2d_{max}\}]$, where d_{max} is the maximum degree of Γ [40]. The second smallest eigenvalue $\lambda_{|V|-1}$ is equal to zero only if the graph is disconnected. If $\lambda_{|V|-1} > 0$, then its multiplicity indicates how many connected components exist in Γ [41]. Furthermore, as we report below, the smallest eigenvalue captures some structural balance properties of signed graphs.

In this study we apply and compare the following two approaches for computing partial structural balance:

• Frustration index ($\epsilon(\Gamma)$): it is equal to the absolute value of the *minimum detection set* ($|E^*|$), that is the lowest number of edges to be deleted from a signed network to obtain a balanced network. Computing this metric is very expensive from a computational point of view: in the worst case, when all edges are negative, the frustration index is NP-hard (it is equivalent to the MAX-CUT problem [42]). In this study we use the *normalized frustration index* suggested in Aref [4]:

$$F(\Gamma) = 1 - \frac{\epsilon(\Gamma)}{m/2} \tag{2}$$

Where m is the number of edges of Γ .

• Algebraic conflict $(\lambda_{|V|}(\Gamma))$: it is the smallest eigenvalue of the signed Laplacian matrix. In this study we use the *normalized algebraic conflict* suggested by Aref [4]:

$$A(\Gamma) = 1 - \frac{\lambda_{|V|}(\Gamma)}{\bar{d}_{max} - 1}, \qquad \bar{d}_{max} = \max_{(u,v)\in E} (d_u + d_v)/2$$
(3)



Figure 1: A balanced signed network is represented on the left. The first cluster is composed by nodes 1, 2, and 3. The second cluster consist of nodes 4 and 5. On the right there is the visualization of the same network created with the *Structural-balance-viz* algorithm. In both representations, positive edges are blue, and negative edges are red.

In positively weighted graphs, when the smallest eigenvalue $\lambda_{|V|}(\Gamma)$ is nonzero, the corresponding eigenvector $u_{|V|}$ provides the embedding coordinates of each vertex [43]. In signed networks, the sign of the elements in $u_{|V|}$ can be used for partitioning the nodes into two clusters [44, 3], and the value can be used as coordinates [17].

The normalized frustration index equals to one if the signed network is perfectly balanced; otherwise, a value less than one measures the partial structural balance.

If a signed graph is balanced, the corresponding Laplacian matrix is proved to be positive semidefinite (all its eigenvalues are non-negative), and $\lambda_{|V|}(\Gamma) = \epsilon(\Gamma) = 0$ [17, 45]. If a signed graph is unbalanced, the corresponding Laplacian matrix is proved to be positive-definite, and an algebraic conflict greater than zero estimates how much the signed network is far from being balanced [46]. Furthermore, it has been proved that $\lambda_n(\Gamma) \leq \epsilon(\Gamma)$ [45].

3.2 Spectral analysis and structural balance

A balanced signed network is shown in Figure 1. The vertices can be divided into two subsets, with only negative edges between them and only positive edges within them. The first subset consists of vertices 1, 2 and 3, the second of vertices 4 and 5. The equation in 4 shows the computation of L as D - A. The vector of the eigenvalues of L, and the eigenvector $u_{|V|}$ corresponding to the smallest eigenvalue are shown in equation 5. The smallest eigenvalue $(\lambda_{|V|}(\Gamma))$ of the signed Laplacian matrix L is equal to zero and this confirms that the network is perfectly balanced. The equation shows also the computation of the partial structural balance using the normalized algebraic conflict (Eq. 3). The network is balanced, so $A(\Gamma)$ is equal to 1.

$$\begin{bmatrix} 3. & 0. & 0. & 0. & 0. \\ 0. & 3. & 0. & 0. & 0. \\ 0. & 0. & 4. & 0. & 0. \\ 0. & 0. & 0. & 3. & 0. \\ 0. & 0. & 0. & 0. & 3. \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 & -1 & 0 \\ 1 & 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ -1 & 0 & -1 & 0 & 1 \\ 0 & -1 & -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3. & -1. & -1. & 1. & 0. \\ -1. & 3. & -1. & 0. & 1. \\ -1. & -1. & 4. & 1. & 1. \\ 1. & 0. & 1. & 3. & -1. \\ 0. & 1. & 1. & -1. & 3. \end{bmatrix}$$
(4)

$$\lambda(\Gamma) = \begin{bmatrix} 0. & 3. & 5. & 3. & 5. \end{bmatrix}, u_{|V|} = \begin{bmatrix} 0.4472\\ 0.4472\\ 0.4472\\ -0.4472\\ -0.4472\\ -0.4472 \end{bmatrix}$$
(5)

$$A(\Gamma) = 1 - \frac{0}{\bar{d}_{max} - 1} = 1, \qquad \bar{d}_{max} = 3.5$$
(6)

The visualization on the right of Figure 1 is created with the *Structural-balance-viz* (SBV) algorithm [3]. SBV creates reproducible visualizations that show the degree of partial network balancing and clearly identify clusters of nodes. The eigenvector $u_{|V|}$ is used to show the contribution of each single node to the structural balance of the network. In *Structural-balance-viz* the nodes of the network are drawn on a Cartesian plane and the coordinates are calculated as it follows:

- x of node u is equivalent to the element of the eigenvector corresponding to u.
- y is calculated in such a way as not to overlap the nodes with the same x coordinate.

Structural-balance-viz shows two types of edges:

- *Frustrated edges* decrease the structural balance of the network. These edges can be negative (in red) if they connect nodes in the same cluster (i.e., in the same quadrant of the Cartesian plane); or positive (in blue) if they connect two nodes of different clusters (i.e., in different quadrants of the Cartesian plane).
- *Non-frustrated edges* represent an agreement between two nodes of the same cluster (positive edge), or a disagreement between two nodes of different clusters (negative edge). Positive and negative non-frustrated edges increase the structural balancing of the network.

The computational time required to execute the SBV algorithm is the same as that required to compute the algebraic conflict.

Since the network in Figure 1 is perfectly balanced, the nodes are represented in two distinct clusters and there are no positive edges between the two clusters nor negative edges within the clusters, so there are no frustrated edges.



Figure 2: An unbalanced signed network is represented on the left. This network is obtained by adding to the network in Fgure 1 a positive edge between nodes 2 and 5. On the right there is the visualization of the same network created with the *Structural-balance-viz* algorithm.

In Figure 2 we show a variation of the same network introduced above, but with a negative edge between vertices 2 and 5. In this case, the network is no longer balanced as shown in Eq. 8 by a value of $A(\Gamma)$ less than 1. In Figure 2 it is shown the visualization created with the *Structural-balance-viz* algorithm. The positive frustrated edge between 2 and 5 is colored in blue and connects the two clusters, and it actually crosses the y-axis, instead of increasing inter-group clustering. Even a small variation in the network (and its balance) produces a significant difference in the corresponding visualization, making comparisons self-explanatory.

$$\lambda(\Gamma) = \begin{bmatrix} 0.5505 & 1.5857 & 4. & 4.4142 & 5.4494 \end{bmatrix}, u_{|V|} = \begin{bmatrix} -0.5411 \\ -0.2979 \\ -0.4865 \\ 0.5411 \\ 0.2979 \end{bmatrix}$$
(7)

$$A(\Gamma) = 1 - \frac{0.5505}{\bar{d}_{max} - 1} = 0.7797, \qquad \bar{d}_{max} = 3.5$$
(8)

This visualization paradigm allows for spotting intra-groups and inter-groups dynamics as well; for example, nodes' attributes could be represented by means of different colors or shapes. Intra-groups and inter-groups edges can be compared to each other to measure homophily. For example, if negative frustrated edges are established between nodes belonging to the same group, or, conversely, if positive frustrated edges connect nodes belonging to different groups, we can detect signals of inverse homophily that contribute to reduce the polarization in the system, making the network less balanced. In the next two sections we show how we can consider inter-groups and intra-groups positive/negative edges to observe both homophily and balance.

4 Creation of the Congress network

In this study, we use congressional roll call votes to analyze the dynamics of political polarization in the U.S. Congress from 1945 (77th Congress) to 2020 (117th Congress). The polarization of the U.S. Congress has been extensively studied in the literature, both through the paradigm of unsigned networks [1] and through the paradigm of signed networks [29, 2]. Our paper proposes a quantitative and visual analysis at a new level of detail, namely of individual nodes. Furthermore, using political voting as an emblematic case study, we show how it is possible to analyze polarization in terms of both homophily and structural balance. In this section, we present the vote processing and network construction pipeline for both the House of Representatives and U.S. Senate.

The data are available by the Voteview Project². In a roll call vote, a representative can vote 'yay' or 'nay', and we consider non-attendances as 'nay'. A complete signed network (G_v) is created for each vote. A node in the network represents a politician, and there is a positive edge between two nodes if they voted similarly, otherwise there is a negative edge. Trivially, the network representing a single roll call vote is always balanced by construction, since the 'yay' voters are grouped together in one cluster and the 'nay' voters are grouped together in another cluster. Within a cluster, there are only positive edges. Between the clusters there are only negative edges.

For each two-year Congress, we collapse the networks representing each vote into a single weighted network (G_c), where the weight of an edge represents the number of collapsed edges. The sum of the weights of the edges connecting two politicians is equal to the number of times they have participated in the same roll call votes. Furthermore, in a similar way to [1], an edge is labeled as "inter-party" if it connects two representatives of two different parties, and as "intra-party" if they are from the same party.

 G_c is not always a complete network: in fact, if two politicians have never taken part in the same vote, then there is neither a positive nor a negative edge between them. Such cases occur, for example, when a politician joins the House of Representatives or the U.S. Senate at the end of a Congress.

For each pair of nodes of G_c we compute the percentage of votes in agreement (p_+) as the weight of the positive edges divided by the sum of the weights of the positive and negative edges (see equations 9). The same process is used to compute the percentage of disagreements (p_-) .

$$p_{+} = e_{+}/(e_{+} + e_{-})$$

$$p_{-} = e_{-}/(e_{+} + e_{-})$$
(9)

Where e_+ and e_- are respectively the positive and negative edges between two nodes. For each pair of nodes we keep only the edge with the highest percentage p_+ or p_- .

In order to keep only the edges that represent a clear positive or negative relationship between two politicians, two thresholds are computed for each Congress: one for the positive edges and one for the negative edges. The threshold for the positive edges is defined as the crossing point between the kernel density estimation (KDE) of the percentage of votes in agreement of the "intra-party" and "inter-party" edges. The threshold for negative edges is computed similarly. An example is shown in Figure 3. In [1], the authors apply a similar approach to the same dataset in terms of network homophily, without considering structural balance.

The yellow areas of Figure 3 are respectively the KDE of the distributions of negative "inter-party" edges (on the left) and positive "inter-party" edges (on the right). A positive "inter-party" edge represents the case when a politician votes in the same way as another politician of a different party. This is a type of frustrated edge, i.e., an edge that reduces polarization by reducing the partial structural balance of the network.

²https://voteview.com/



Figure 3: For each node pair, we compute a percentage of agreement and a percentage of disagreements as shown in 9. The plots show the kernel density estimation (KDE) of each percentage for "inter-party" and "intra-party" edges. The figure on the right refers to the percentage of disagreements (negative edges) of the House of Representatives of 78th U.S. Congress, and the figure on the right refers to the percentage of agreement (positive edges) of the House of Representatives of 82nd U.S. Congress.

The blue areas in Figure 3 are respectively the KDE of distributions of negative "intra-party" edges and positive "intraparty" edges. The edges labeled as "intra-party" connect two politicians of the same party, so a negative "intra-party" edge is an example of a frustrated edge, while a negative one is not.

The red vertical lines in Figure 3 represent the thresholds for positive and negative edges of the House of Representatives of the 78th and 82th U.S. Congresses.

Considering all the Congresses, for the House of Representatives there are a total of 3, 436 unique nodes, 1, 984, 618 "inter-party" edges, and 2, 055, 660 "intra-party" edges. For the U.S. Senate, there are a total of 707 unique nodes, 107, 843 "inter-party" edges, and 108, 513 "intra-party" edges.

5 Results

The first step in the study of the dynamics of polarization in the U.S. Congress is the analysis of network homophily. By construction, each node has a membership cluster, i.e., its party; in a signed network, network homophily is manifested by positive edges between nodes of the same party, and negative edges between nodes of different parties. The dynamics of homophily over time can be understood by comparing the frequency distributions of agreement and disagreement percentages. The overlap between the two kernel density estimation (red area in Figure 3) shows that i) the number of votes in agreement with the affiliation party and the number of votes in agreement with the opposing party are comparable (Figure 3 on the left); ii) the number of votes in disagreement with the opposing party are also comparable (Figure 3 on the right). This overlap can be interpreted as a tendency of politicians to reduce polarization by challenging their own party and cooperating with the opposing party, thus reducing the network homophily.

Figure 4 shows the variation of this overlap over time, from Congress 77 to Congress 116, for the House of Representatives and the U.S. Senate. The overlap between the frequency distribution of "inter-party" edges and "intra-party" edges tends to decrease dramatically over time. This suggests an increase in polarization and a decrease in cooperation among politicians. In the next Section we verify that the increase in network homophily over time is consistent with an increase in structural balance.

Sections 5.1 and 5.2 address polarization in terms of balance, both quantitatively and visually. In Section 5.3, we show how decomposing the eigenvalues of the Laplacian matrix can bring the analysis to a level of detail not previously available in the U.S. Congressional polarization literature. Finally, Section 5.4 shows the computational advantages of algebraic conflict, both with respect to the case study of the American Congress and with respect to artificially created networks. Emphasis is placed on the relationship between the percentage of frustrating edges and the computational cost of both the frustration index and the algebraic conflict.



Figure 4: Overlap between "inter-party" and "intra-party" positive edges frequency distributions over the congresses (U.S. Senate on the left and House of the Representatives on the right).



Figure 5: Comparison across Congress of frustration index, algebraic conflict, and the values of the eigenvector corresponding to the least eigenvalue. The plot on the left refers to the U.S. Senate, the plot on the right to the House of the Representatives.

5.1 Frustration index and algebraic conflict

In order to study the dynamics of polarization through the analysis of partial structural balance, we focus on two metrics: frustration index and algebraic conflict (see section 3).

Figure 5 shows the partial structural balance of the Congresses from 1941 to 2021 computed with frustration index (red line) and algebraic conflict (green line). The blue line represents the mean of the values of the eigenvector $u_{|V|}$ corresponding to the smallest eigenvalue $\lambda_{|V|}(\Gamma)$. The blue area represents the confidence interval computed as the standard deviation of the mean of the eigenvector $u_{|V|}$. Each element of the eigenvector $u_{|V|}$ refers to a node of the network. Nodes can be divided into two clusters using the sign of the elements of the eigenvector. The influence of each node on the structural balance of the whole network can be quantified by means of its corresponding value in $u_{|V|}$. A node not involved in frustrated edges has a corresponding eigenvector value of one. Conversely, a node with a corresponding value close to zero is involved in a process of balance reduction. Therefore, the higher the mean of the eigenvector, the more cohesive are the node clusters. A larger standard deviation indicates a greater tendency of nodes to autonomously interact with the rest of the network.

Furthermore, in Section 3 we discuss how the eigenvector of the smallest eigenvalue can be used for partitioning and positioning the nodes in a signed network visualization. In the visualizations based on the *Structural-balance-viz* algorithm in Figure 11, the reduction in the standard deviation of the eigenvector $u_{|V|}$ over time is clearly represented by the fact that the nodes become less scattered along the x-axis over time.



Figure 6: Network created with the *Structural-balance-viz* algorithm and representing the votes of the House of Representatives of Congress 82 (1951). Non-frustrated edges (82440) are colored in gray, positive frustrated edges in blue (6836), and negative frustrated edges in red (10605). Nodes are color-coded according to the corresponding party: red nodes are Republicans, and blue nodes are Democrats. It is striking that a tendency to polarization is slowed down by nodes belonging to different parties whose position in the x-axis is mixed, no matter the clusters determined by the structure of the underlying signed network. The partial balance of the network partial balance computed with the algebraic conflict is 0.832, and computed with the frustration index is 0.765.

Pearson's correlation is computed for all the combinations of frustration index, algebraic conflict, and the mean of the eigenvector $u_{|V|}$. For both the U.S. Senate and the House of the Representatives, Pearson's correlation coefficients are between 0.92 and 0.98, with all p-values below 0.005.

The frustration index and the algebraic conflict increase significantly since Congress 90 (1967) for the U.S. Senate and since Congress 91 (1967) for the House of Representatives. The progressive increase in political polarization, computed in terms of both homophily and structural balance, is consistent with the literature [29, 32, 34, 2, 47].

5.2 Visualize the U.S. Congress

The analysis of the polarization in terms of homophilia and balance can be supported by the visualization of the structural properties of the network. *Structural-balance-viz* [3] algorithm, is not only computationally efficient, but also provides an at-a-glance view of network balance and which nodes are most interested in frustrated edges.

Figure 6 shows the *Structural-balance-viz* applied to the voting network of the House of Representatives of Congress 82 (1951). A large number of frustrated edges (17441, while e.g. Congress 117 has only 118 frustrated edges), both positive (blue edges) and negative (red edges), are clearly visible in the visualization. Since we are interested in highlighting the edges that reduce the partial balance of the network, we color all non-frustrated edges light gray. In addiction, the two clusters, represented by the two upper quadrants of the Cartesian plane, are heterogeneous. The nodes are highly scattered along the x-axis, although the right quadrant cluster has more blue nodes (Democratic politicians) and the left quadrant cluster has more red nodes (Republican politicians). The visualization of a highly unbalanced network is characterized by nodes scattered along the x-axis, and a significant number of frustrated edges.

Figure 7 shows the voting network of the House of Representatives of Congress 108 (2003). In this case, it is immediately apparent that the situation is very different from that shown in Figure 6: there are few positive (in blue) and negative (in red) frustrated edges, and the nodes are represented in two cohesive and homogeneous clusters. These clusters almost exactly represent the two parties, the Democrats in the right quadrant and the Republicans in the left quadrant. These are the characteristics of a highly balanced network view; the partial balance is equal to 0.995, computed with the algebraic conflict, and to 0.992 computed with the frustration index (see Figure 5). Homophily emerges more clearly in



Figure 7: Network created with the *Structural-balance-viz* algorithm and representing the votes of the House of Representatives of Congress 108 (2003). Non-frustrated edges are colored in gray (96435), positive frustrated edges in blue (247), and negative frustrated edges in red (319). Nodes are color-coded according to the corresponding party: red representatives are republicans, and blue ones are democrats. The strong signal of polarization is shown by the almost perfect match between nodes' colors and their positions at opposing sides of the x-axis.

this scenario: nodes in the same party agree more often than nodes in different parties. Analogously, nodes in different parties disagree more likely than nodes in the same party.

The network would be almost totally balanced if it were not for the two nodes (a Republican and a Democrat) in the right quadrant which are spatially located closer to the origin of the x-axis, w.r.t. the nodes of their own parties.

Figure 11 shows an overview of the Structural-balance-viz algorithm applied to the voting networks of 20 congresses of the U.S. Senate between the years 1943 and 2019. Over time there is a clear progressive reduction in frustrated edges and a separation of the two clusters into increasingly polarized positions.

In the next section, we propose an analysis of the political behavior of Democratic Senator Joe Manchin over 3 Congresses (from 2015 to 2021). This fine-grained analysis is supported by the networks' visualizations created with the *Structural-balance-viz* algorithm.

5.3 Linking visualization, unbalance metric, and intra/inter-parties dynamics

The possibility to study the influence of each node on the partial network balance and to follow the intra- and inter-group dynamics is one of the main advantages of studying the network balance through the spectral analysis of the Laplacian matrix. By exploiting the eigenvector $v_{|V|}$, the algorithm *Structural-balance-viz* shows which are the nodes that most influence the network structure. As an example, below we show how it is possible to trace the political behavior of a Senator through the analysis of his position in three different visualizations.

The networks in Figure 8 represent the votes of the U.S. Senate in Congresses 114, 115, and 116. The node representing Democratic Senator Joe Manchin is highlighted with a red rectangle. His political stance has changed dramatically over the years. According to the Senate Vote Network for the 114th Congress, on average, Senator Manchin has taken positions very close to his party, with a slight tendency to cooperate with Republicans. In the next Congress, the 115th (2017-2019), the Senator moved closer to the positions of the Republican Party. In fact, the node representing Joe Manchin, while remaining in the quadrant containing the Democrat cluster, is located closer to the origin of the x-axis.

In the last Congress shown in Figure 8, the 116th (2019-2021), the highlighted node is in the right quadrant, meaning it is closer to the cluster of Republican nodes than to the Democratic nodes.



Figure 8: Networks created with the *Structural-balance-viz* algorithm and representing the votes of the Senate from Congress 114 (2015) to Congress 116 (2019). The node representing Democratic Senator Joe Manchin is highlighted with a red rectangle. In only 3 congresses (6 years) he comes closer to the Republican cluster than to his party cluster.



Figure 9: The y-axis represents the computation time in seconds to compute the frustration index and the smallest eigenvalue for each congress. The networks that represent the Congresses are sorted on the x-axis according to the number of frustrated edges. The Figure on the left refers to the U.S. Senate, the one on the right to the House of Representatives.

Senator Manchin's political transformation toward greater openness to Republicans has been a source of political debate in the newspapers since 2020^{34} . The analysis of the eigenvector $v_{|V|}$ of the two previous Congresses (114th and 115th) shows that this process is not unexpected, and results from a slow approach over the years to the political counterpart. Some of the interpretations justify this approach to the Republican Party with the electoral need to be re-elected. Elected in the red state of West Virginia, the Senator seems to be moving closer and closer to the demands of his potential electorate, even if they are not in line with the rest of the Democratic Party. It is also interesting that in the less polarized system observed in 1951, this behavior would have been interpreted as normal, whereas in 2020 it is clearly seen as an outlier. (Figure 6).

5.4 Unbalance metrics computational costs

The time and space complexity of computing the algebraic conflict is governed by the complexity required by the decomposition of the eigenvalues of the Laplacian matrix. In this study, for the decomposition of the eigenvalues, we use the implementation in the Python library *Scipy*⁵. The algebraic conflict is normalized as in 3.

The theoretical formulation of the frustration index is NP-hard. An approximation based on simulated annealing implemented in the R library $Signnet^6$ is used in this study.

³https://www.theguardian.com/us-news/2021/oct/29/senate-manchin-sinema-biden-democrats-outsizeinfluence

⁴https://www.washingtonpost.com/politics/2022/01/26/manchin-sinema-democrats-legislation/

⁵https://docs.scipy.org/

⁶https://cran.r-project.org/web/packages/signnet/



Figure 10: Each cell of the heatmap shows the computation time in seconds required for computing the frustration index (on the left) and the eigenvalue decomposition of the Laplacian matrix (on the right). Note the difference in scale between the two heatmaps.

Figure 9 shows the computation time required to compute the frustration index and the algebraic conflict for the U.S. Senate and the House of Representatives. The computations are performed on a computer with 8 GB of RAM and the Intel Core i5-6200U as CPU. The computation time for the frustration index is orders of magnitude greater than the computation time for the algebraic conflict.

Furthermore, the computation time required to compute the frustration index increases as the number of frustrated edges in a network increases (on the x-axis the networks are sorted by the number of frustrated edges). Since the frustration index is defined as the minimum number of edges whose removal results in balance, it is clear therefore that its computation time depends on the number of frustrated edges to remove. The same phenomenon is not observed for the computation of the smallest eigenvalue: eigenvalue decomposition of the Laplacian matrix is independent of the edge types.

In order to generalize this concept, we generate ten random signed networks each composed of an increasing number of nodes, from 100 to 1000. For each network we vary the number of frustrated edges, from 10% of the total number of edges, up to 90%.

The process of creating a random signed network is shown below:

- 1. Two different Erdős-Rényi networks are created (named A and B respectively).
- 2. Each node of the two networks is assigned an attribute based on the network it belongs to.
- 3. Networks A and B are merged, and edges are added between the nodes of networks A and B.
- 4. Each edge connecting two nodes of the same network is assigned the positive sign. Each edge between network A and network B is assigned the negative sign.
- 5. The network is now perfectly balanced. Based on the percentage of frustrated edges out of the total, some edges between A and B have their sign changed to positive (positive frustrated edges), and some edges inside A and inside B have their sign changed to negative (negative frustrated edges).

We compute both the frustration index and the smallest eigenvalue to verify that only the former has a computation time influenced by the number of frustrated edges. The results of the computation times required for each of the network configurations are shown in Figure 10, with the frustration index on the left and the eigenvalue decomposition of the Laplacian matrix on the right. For both indices, the number of nodes in the network has an effect on the computation time. However, only the computation of the frustration index is significantly affected by the percentage of frustrated edges in the network.

6 Conclusion and future work

Signed networks are a powerful model for representing relationships between individuals. Community detection, as a consequence of network homophily by attribute, and partial balance computation offer the possibility to analyze

conflicts and polarization between communities in a signed network of relationships. The analytical pipeline we present is based on the synergy of methods known in the literature for the study of partial structural balance. The advantages of applying this pipeline are numerous and significant: i) it offers results comparable to those obtainable with an approach based on the frustration index (Figure 5); ii) it allows to visualize and analyze in fine grain the influence of every single node on the structural balance of the entire network (Figures 8); iii) it is very advantageous from a computational point of view (Figures 9).

The application of this approach to the study of political polarization in the U.S. Congress has revealed implications missing from previous studies in the literature [29, 1]. This is exemplified by the case of Senator Manchin. Although it has only been in the newspapers in recent years, we show and quantify that his much-discussed political transformation began much earlier. We examine the intra-group dynamics of the senators and how this affected the partial balance of the networks over three Congresses.

In addition, since many researchers in the field of complex networks use Python as a programming language, a library for the analysis and visualization of the structural balance of the signed network is being built and released.⁷.

References

- [1] Clio Andris, David Lee, Marcus J. Hamilton, Mauro Martino, Christian E. Gunning, and John Armistead Selden. The rise of partisanship and super-cooperators in the u.s. house of representatives. *PLOS ONE*, 10:1–14, 04 2015.
- [2] Samin Aref and Zachary Neal. Detecting coalitions by optimally partitioning signed networks of political collaboration. *Scientific reports*, 10(1):1–10, 2020.
- [3] Edoardo Galimberti, Chiara Madeddu, Francesco Bonchi, and Giancarlo Ruffo. Visualizing structural balance in signed networks. In Hocine Cherifi, Sabrina Gaito, José Fernendo Mendes, Esteban Moro, and Luis Mateus Rocha, editors, *Complex Networks and Their Applications VIII*, pages 53–65, Cham, 2020. Springer International Publishing.
- [4] Samin Aref and Mark C Wilson. Measuring partial balance in signed networks. *Journal of Complex Networks*, 6(4):566–595, 09 2017.
- [5] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 89–96, Aug. 2011.
- [6] Cass R. Sunstein. Echo chambers : Bush v. Gore, impeachment, and beyond. Princeton University Press, 2001.
- [7] Eli Pariser. *The Filter bubble: What the Internet is hiding from you.* Penguin, 2011.
- [8] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018.
- [9] Anna Keuchenius, Petter Törnberg, and Justus Uitermark. Why it is important to consider negative ties when studying polarized debates: A signed network analysis of a dutch cultural controversy on twitter. *PLOS ONE*, 16(8):1–23, 08 2021.
- [10] Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto W. De Luca, and Sahin Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In Proc. of the 2010 SIAM International Conference on Data Mining (SDM), pages 559–570, 2010.
- [11] David Knoke and Song Yang. *Social Network Analysis*. SAGE Publications, Inc., Thousand Oaks, California, 2008.
- [12] Charu C. Aggarwal. An introduction to social network data analytics. In Social Network Data Analytics, 2011.
- [13] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. A survey of signed network mining in social media. *ACM Computing Survey*, 49(3), aug 2016.
- [14] Fritz Heider. Attitudes and cognitive organization. The Journal of Psychology, 21(1):107–112, 1946.
- [15] Dorwin Cartwright and Frank Harary. Structural balance: A generalization of heider's theory. *Psychological Review*, 63(5):277–293, 1956.
- [16] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, page 1361–1370, New York, NY, USA, 2010. Association for Computing Machinery.

⁷https://github.com/alfonsosemeraro/draw_signed_networkx



Figure 11: Structural-balance-viz algorithm applied to the voting networks of 20 congresses of the U.S. Senate between the years 1943 and 2019.

- [17] Jérôme Kunegis. Applications of structural balance in signed social networks, 2014.
- [18] Patrick Doreian and Andrej Mrvar. Structural balance and signed international relations: Joss. *Journal of Social Structure*, 16:1–49, 2015.
- [19] Hugo Saiz, Jesús Gómez-Gardeñes, Paloma Nuche, Andrea Girón, Yolanda Pueyo, and Concepción L. Alados. Evidence of structural balance in spatial ecological networks. *Ecography*, 40(6):733–741, 2017.
- [20] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, page 1361–1370, New York, NY, USA, 2010. Association for Computing Machinery.
- [21] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In WWW '10: Proceedings of the 19th international conference on World wide web, WWW '10, page 641–650, New York, NY, USA, 2010. Association for Computing Machinery.
- [22] Patricia Victor, Chris Cornelis, Martine De Cock, and Ankur Teredesai. Trust- and distrust-based recommendations for controversial reviews. *IEEE Intelligent Systems*, 26:48–55, 01 2011.
- [23] Hao Ma, Michael R. Lyu, and Irwin King. Learning to recommend with trust and distrust relationships. In Proceedings of the Third ACM Conference on Recommender Systems, RecSys '09, page 189–196, New York, NY, USA, 2009. Association for Computing Machinery.
- [24] Jianrui Chen, Lidan Wei, Liji U, and Fei Hao. A temporal recommendation mechanism based on signed network of user interest changes. *IEEE Systems Journal*, 14(1):244–252, 2020.
- [25] Hua Liu, Cunquan Qu, Yawei Niu, and Guanghui Wang. The evolution of structural balance in time-varying signed networks. *Future Generation Computer Systems*, 102:403–408, 2020.
- [26] Aldo Porco, Andreas Kaltenbrunner, and Vicenç Gómez. Low-rank approximations for predicting voting behaviour. In *Proc. of NIPS 2015 Workshop: Networks in the Social and Information Sciences*, 01 2015.
- [27] Carla Intal and Taha Yasseri. Dissent and rebellion in the house of commons: a social network analysis of brexit-related divisions in the 57th parliament. *Applied Network Science*, 6:1–12, 2021.
- [28] Phillip Bonacich. Some unique properties of eigenvector centrality. Social Networks, 29(4):555–564, 2007.
- [29] Zachary P. Neal. A sign of the times? weak and strong polarization in the u.s. congress, 1973–2016. Social Networks, 60:103–112, 2020. Social Network Research on Negative Ties and Signed Graphs.
- [30] Nejat Arinik, Rosa Figueiredo, and Vincent Labatut. Signed graph analysis for the interpretation of voting behavior. In *Proceedings of the Workshop Papers of i-Know 2017*, 10 2017.
- [31] Israel Mendonça, Rosa Figueiredo, Vincent Labatut, and Philippe Michelon. Relevance of negative links in graph partitioning: A case study using votes from the european parliament. In 2015 Second European Network Intelligence Conference, pages 122–129, 2015.
- [32] Andrew Scott Waugh, Liuyi Pei, James H. Fowler, Peter J. Mucha, and Mason A. Porter. Party polarization in congress: A network science approach, 2009.
- [33] Dana R. Fisher, Joseph Waggle, and Philip Leifeld. Where does political polarization come from? locating polarization within the u.s. climate change debate. *American Behavioral Scientist*, 57(1):70–92, 2013.
- [34] Carlos Ferreira, Breno Matos, and Jussara Almeira. Analyzing dynamic ideological communities in congressional voting networks. In Social Informatics. SocInfo 2018. LNCS Vol. 11185, pages 257–273. Springer, 09 2018.
- [35] Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini. Computing global structural balance in large-scale signed social networks. *Proceedings of the National Academy of Sciences*, 108(52):20953–20958, 2011.
- [36] Ernesto Estrada and Michele Benzi. Walk-based measure of balance in signed networks: Detecting lack of balance in social networks. *Physical Review E*, 90:042802, Oct 2014.
- [37] Ranveer Singh and Bibhas Adhikari. Measuring the balance of signed networks and its application to sign prediction. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(6):063302, jun 2017.
- [38] Alec Kirkley, George T. Cantwell, and M. E. J. Newman. Balance in signed networks. *Physical Review E*, 99:012320, Jan 2019.
- [39] Thomas Zaslavsky. Matrices in the theory of signed simple graphs. arXiv: Combinatorics, 2013.
- [40] K.Ch. Das. The laplacian spectrum of a graph. *Computers & Mathematics with Applications*, 48(5):715–724, 2004.
- [41] Ann Marsden. Eigenvalues of the laplacian and their relationship to the connectedness, 2013.

- [42] Francisco Barahona. The max-cut problem on graphs not contractible to k5. *Operations Research Letters*, 2(3):107–111, 1983.
- [43] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2001.
- [44] Tom Coleman, James Saunderson, and Anthony Wirth. Spectral clustering with inconsistent advice. In *Proceedings* of the 25th International Conference on Machine Learning, ICML '08, page 152–159, New York, NY, USA, 2008. Association for Computing Machinery.
- [45] Francesco Belardo. Balancedness and the least eigenvalue of laplacian of signed graphs. *Linear Algebra and its Applications*, 446:133–147, 2014.
- [46] Yaoping Hou. Bounds for the least laplacian eigenvalue of a signed graph. *Acta Mathematica Sinica*, 21:955–960, 2005.
- [47] Jonathan Haidt. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Vintage Books, New York, New York, United States, 2008.