Partisan Scale

Sedat Gokalp
Computer Science
Arizona State University
Sedat.Gokalp@asu.edu

Hasan Davulcu
Computer Science
Arizona State University
Hasan.Davulcu@asu.edu

ABSTRACT

US Senate is the venue of political debates where the federal bills are formed and voted. Senators show their support/opposition along the bills with their votes. This information makes it possible to extract the polarity of the senators. We use signed bipartite graphs for modeling debates, and we propose an algorithm for partitioning both the senators, and the bills comprising the debate into binary opposing camps. Simultaneously, our algorithm scales both the senators and the bills on a univariate scale. Using this scale, a researcher can identify moderate and partisan senators within each camp, and polarizing vs. unifying bills. We applied our algorithm on all the terms of the US Senate to the date for longitudinal analysis and developed a web based interactive user interface **www.PartisanScale.com** to visualize the analysis.

Categories and Subject Descriptors

H.2.8 [**Database applications**]: Data mining;
H.3.3 [**Information Search and Retrieval**]: Clustering

General Terms

Algorithms

Keywords

Community discovery, Link Analysis, Partitioning, Ranking, Scaling, HITS, Signed Bipartite Graphs, Spectral Clustering

1. INTRODUCTION

The United States has a bicameral legislature that comprises the US Senate as the upper house, and the US House of Representatives. The terms of the US Senate last for two years, and the senators serve three terms (six years) each. The terms are staggered in such a way that approximately one-third of the seats are up for election every two years.

The Senate meets in the United States Capitol in Washington, D.C. to form and debate on motions, or bills. When debates conclude, the bill in question is put to a vote, where senators respond either 'Yea' (in favor of the bill) or 'Nay' (against the bill). For most of the bills, only the total number of 'Yea' and 'Nay' votes are recorded, except for the roll call votes. According to The Library of Congress[1],

> A roll call vote guarantees that every Member's vote is recorded, but only a minority of bills receive a roll call vote.

The current political party system in the United States is a two-party system, which suggests a bipolar nature for both the senators and the bills; such that, there exists two polarized camps of senators that oppose each others views, and two sets of bills that polarize the senators. It can be presumed that these camps would purely split according to the political parties of the senators, or the political parties of the sponsors of the bills. Although this is true to a certain extent, our analysis show that the actual behaviours can be different for a minority.

Senators show their support/opposition along the bills with their votes. This information makes it possible to extract the polarity of the senators. We use signed bipartite graphs for modeling the opposition, and we used our previous work ANCO-HITS algorithm for partitioning both the senators, and the bills into two polarized camps. Simultaneously, our algorithm scales both the senators and the bills on a univariate scale. Using this scale, a researcher can identify moderate and partisan[2] senators within each camp, and polarizing vs. unifying bills.

Partitioning and scaling help a researcher to better understand the structure of political debates in the Senate. While partisan ends of a scale may represent senators with irreconcilable viewpoints, moderate senators may represent viewpoints that are more amenable to engage in a constructive dialog through a set of unifying issues. Moderates may sympathize with some of the claims and grievances of the other side. Longitudinal analysis using our proposed algorithms could reveal interesting dynamics, such as, moderates from opposing camps could be in the process of forming a coalition by making the necessary compromises to reach a consensus.

Major contributions of this paper are: (1) a modification of our previous algorithm ANCO-HITS, to propagate the scores on a signed bipartite graph to solve the partitioning and scaling problems described above; (2) applying the algorithm on 112 terms of the US Senate for longitudinal analysis; (3) developing a web based interactive user interface to visualize the analysis.

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

[1] `http://thomas.loc.gov/home/rollcallvotes.html`
[2] *Partisanship* can be defined as being devoted to or biased in support of a party.

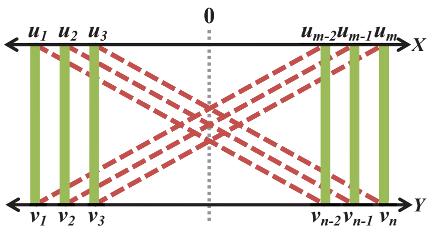

Figure 1: Perfectly polarized bipartite graph

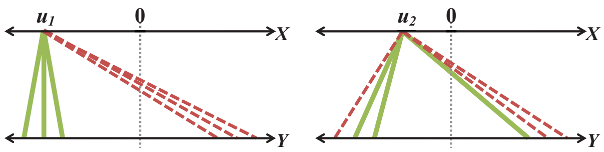

Figure 2: Partisan vs. Moderate senators

2. PROBLEM FORMULATION

There are many applications [5, 4, 7, 1, 3] for recognizing political orientation, and bipartite graphs [2, 6, 8] have been widely used to represent relationships between two sets of entities. We use bipartite graphs to model the relationships between the senators and the bills. We use signed edges to represent the votes, where positive edges denote support, and negative edges denote opposition on a bill by a senator. Given

- $G = (U \cup V, A)$ is a bipartite graph consisting of senators U and bills V, and a signed vote matrix A
- $U = \{u_1, u_2, \ldots, u_m\}$, a set of m senators
- $V = \{v_1, v_2, \ldots, v_n\}$, a set of n bills
- $A \in \mathbb{R}^{m \times n}$, where a_{ij} represents the vote of senator u_i on bill v_j

Find

- $X = (x_1, x_2, \ldots, x_m)$, where $x_i \in \mathbb{R}$ is the assigned value of the senator u_i
- $Y = (y_1, y_2, \ldots, y_n)$, where $y_j \in \mathbb{R}$ is the assigned value of the bill v_j

such that

- x_i value for a senator u_i should be closer to the y_j values of the bills that he supports, and further away from the y_k values of the bills that he opposes. The magnitude of x_i denotes the *partisanship* of the senators u_i, and the magnitude of y_j denote how polarizing the bill v_j is. i.e. magnitudes closer to 0 meaning *more moderate* and larger magnitudes meaning *more partisan.*

Figure 1 depicts a perfectly polarized bipartite graph. The two axes X and Y represent the univariate scale for the senators and bills. The vertices to the right of zero have positive values, and the vertices to the left have negative values on the scale. A green solid line between a senator u_i and a bill v_j represents support, and a red dashed line represents opposition.

Figure 2 shows an example of two senators; u_1 being extreme and u_2 being more moderate. u_1 supports the bills of same polarity, and opposes the vertices of the opposite polarity. However, u_2 has mixed support and opposition. Same relation holds between polarizing and unifying bills.

Although partitioning algorithms can be utilized to detect the polarity of senators and bills, it is not possible to distinguish partisans from moderates. Scaling overcomes this problem and makes it possible to compare two senators of same polarity. In this paper, we are not only able to compare pairs of senators, but also provide the exact locations on the scale, therefore providing valuable information about the shape of the distribution as well.

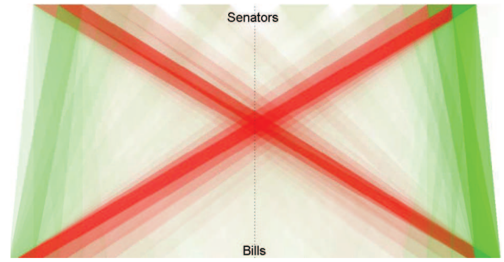

Figure 3: Vote matrix for the 111$^{\text{th}}$ US Senate after scaling with ANCO-HITS

3. ANCO-HITS

In this study, we used a modified version of our previous work ANCO-HITS. Algorithm 1 describes the steps of the ANCO-HITS algorithm for the co-scaling problem.

Data: Signed vote matrix A
Result: Scale vectors X and Y
Initiate $X^{<0>} = (1, 1, \ldots, 1)$;
Initiate $Y^{<0>} = (1, 1, \ldots, 1)$;
repeat
  Update X;
  Update Y;
until X *vector converges*;

Algorithm 1: Iterative update procedure for ANCO-HITS

This research uses a different normalization scheme than the original ANCO-HITS algorithm. The update functions for X and Y are modified such that the vectors X and Y would converge not only in direction, but also in value.

$$x_i^{<k>} = \frac{\sum_{j=1}^{n} a_{ij} y_j^{<k-1>}}{\sum_{j=1}^{n} |a_{ij} y_j^{<k-1>}|} \qquad y_j^{<k>} = \frac{\sum_{i=1}^{m} a_{ij} x_i^{<k>}}{\sum_{i=1}^{m} |a_{ij} x_i^{<k>}|} \tag{1}$$

The convergence values for X and Y vectors will satisfy $-1 \leq x_i, y_j \leq +1$.

Figure 3 represents the bipartite graph of the 111$^{\text{th}}$ US Senate data after scaling both the senate and the bills with ANCO-HITS. The light green colored edges represent 'Yea' votes, and dark red represents 'Nay' votes. Similar to our motivating Figure 1, this figure also shows partisan behavior in the 111$^{\text{th}}$ US Senate.

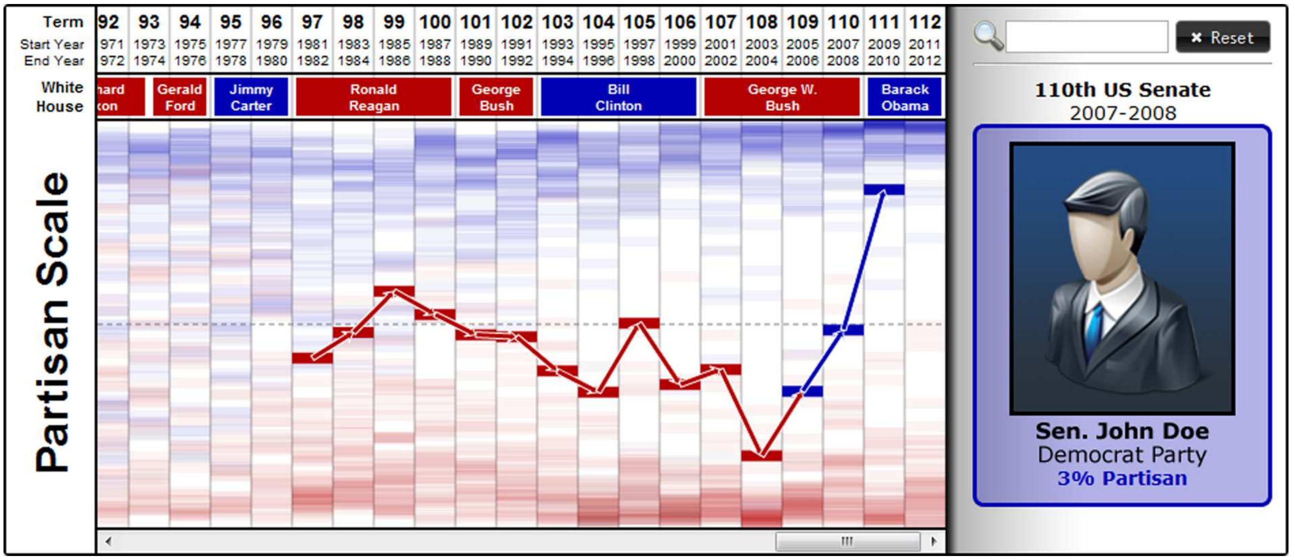

Figure 4: A screenshot from PartisanScale.com showing the partisanship history for a senator

4. INTERACTIVE USER INTERFACE

The US Congress has been collecting data since the very first congress of the US history. This data has been encoded as XML files and publicly shared through the govtrack.us project[3]. We collected the *roll call votes* of the US Senate for the terms 1 through 112, covering the years 1789-2011. We ran the ANCO-HITS algorithm for each individual term. The sign of the ANCO-HITS values are arbitrary; therefore, we aligned consecutive terms by mirroring the scale if necessary. By analyzing more than 3,000,000 votes, we produced the web based interactive user interface **www.PartisanScale.com** that allows the users to navigate through the history of the US Senate.

Figure 4 shows a screenshot of the user interface. Each term of the senate is shown as a column in the figure. The top row shows the terms and the years for each senate with the incumbent US president shown below. The senators are represented by boxes which are colored according to their political parties.

The vertical axis of the scale represents the bipolar nature of the US Senate. The polarity of each senator is represented by the location of each box. The dashed line shows the zero point. Senators around this point are calculated to be moderate, and the senators away from the dashed line are calculated to be more polarized. Hovering along these boxes will show the picture, the political party, and the amount of partisanship for the senator in focus. Clicking on the scale will further filter the figure to show the partisanship history. This filtering can also be done with the quick search tool on the top right corner. The auto-completion feature will help the users easily select the senator.

For example, Figure 4 shows a senator that is calculated to be moderate for the 110th term. It can be seen that this senator was first elected in 1981 and served for 15 terms until the year 2010. It also shows us that after 12 terms of service as a republican, he switches membership to the Democratic Party for the last 3 terms of his service.

[3]`http://www.govtrack.us/data`

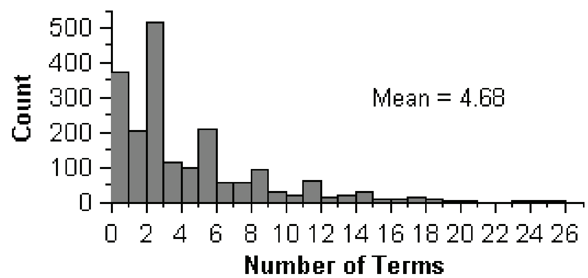

Figure 5: Longevity of service

An introductory screencast video that shows the usage of the system can be found on the website.

5. STATISTICS

Figure 5 shows the histogram for the number of terms each senator served. The average number of terms the senators served is 4.68, and the longest run is 26 terms.

Figure 6 shows the partisanship displacement distribution for three ΔT values on a semi-log scale. Partisanship displacement is defined as the absolute distance of partisan scale values for a senator between two terms T_1 and T_2. $C_{\Delta T}(d)$ is the number of displacements $\geq d$ between any two terms T_1 and T_2 satisfying $\Delta T = T_1 - T_2$.

This figure shows three plots of $C_{\Delta T}$ values for $\Delta T = 1$, $\Delta T = 2$ and $\Delta T = 3$. It can be clearly seen that the plots on the semi-log scale form a linear function, which suggests an exponential distribution.

Figure 7 aggregates the party polarities. The mean partisanship values of the senators from each party is shown as a solid line. The shaded areas show 1 standard deviation along the mean for each term. This figure is helpful to identify the times of partisan politics within the US Senate.

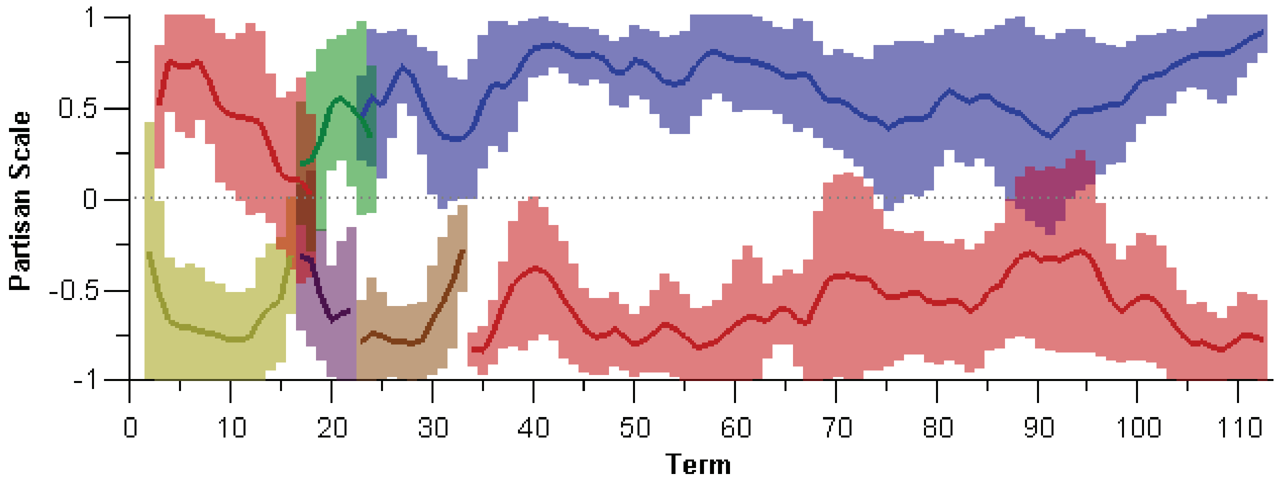

Figure 7: Aggregated Party Partisanship

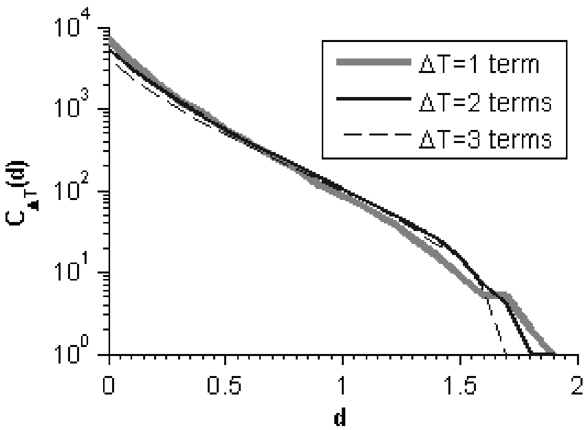

Figure 6: Partisanship displacement distribution

6. CONCLUSIONS

In this paper, we introduced a measure for partisanship, and applied it on 112 terms of the US Senate for longitudinal analysis. We further developed an interactive user interface **www.PartisanScale.com** to visualize the analysis. The data set and the algorithm in source code are available online.

7. ACKNOWLEDGMENTS

This research was supported in part by US DOD Minerva Research Initiative grant N00014-09-1-0815.

8. REFERENCES

[1] M. Bansal, C. Cardie, and L. Lee. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. *Proceedings of COLING: Companion volume: Posters*, pages 13–16, 2008.

[2] H. Deng, M. Lyu, and I. King. A generalized co-hits algorithm and its application to bipartite graphs. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 239–248. ACM, 2009.

[3] W. Lin and A. Hauptmann. Are these documents written from different perspectives?: a test of different perspectives based on statistical distribution divergence. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 1057–1064. Association for Computational Linguistics, 2006.

[4] R. Malouf and T. Mullen. *Graph-based user classification for informal online political discourse.* 2007.

[5] T. Mullen and R. Malouf. A preliminary investigation into sentiment analysis of informal political discourse. In *AAAI symposium on computational approaches to analysing weblogs (AAAI-CAAW)*, pages 159–162, 2006.

[6] M. Rege, M. Dong, and F. Fotouhi. Co-clustering documents and words using bipartite isoperimetric graph partitioning. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pages 532–541. IEEE, 2006.

[7] M. Thomas, B. Pang, and L. Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *In Proceedings of EMNLP*, pages 327–335, 2006.

[8] H. Zha, X. He, C. Ding, H. Simon, and M. Gu. Bipartite graph partitioning and data clustering. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 25–32. ACM, 2001.