# Spheres of Legislation: Polarization and Most Influential Nodes in Behavioral Context

Andrew C. Phillips, Mohammad T. Irfan(✉), and Luca Ostertag-Hill

Bowdoin College, Brunswick, ME 04011, USA
andrewphillips182@gmail.com, {mirfan,ldostert}@bowdoin.edu
http://www.bowdoin.edu/~mirfan

**Abstract.** Game-theoretic models of influence in networks often assume the network structure to be static. In this paper, we allow the network structure to vary according to the underlying behavioral context. This leads to several interesting questions on two fronts. First, how do we identify different contexts and learn the corresponding network structures using real-world data? We focus on the U.S. Senate and apply unsupervised machine learning techniques, such as fuzzy clustering algorithms and generative models, to identify different spheres of legislation as context and learn an influence network for each sphere. Second, how do we analyze these networks in order to gain an insight into the role played by the spheres of legislation in various interesting constructs like polarization and most influential nodes? To this end, we apply both game-theoretic and social network analysis techniques. In particular, we show that game-theoretic notion of most influential nodes brings out the strategic aspects of interactions like bipartisan grouping, which typical centrality measures fail to capture. We also show that for the same set of senators, some spheres of legislation are more polarizing than others.

**Keywords:** Influence in networks · Machine learning and networks · Computational game theory · Graphical games

## 1 Introduction

In recent times, the study of social influence has extended beyond mathematical sociology [12,31] and has entered the realm of computation [1,3–5,16,18,21–24]. A computational study of "influence"– however we define it – is key to understanding the behavior of individuals embedded in networks. In this paper, we model and analyze social influence in a strategic setting where one's behavior depends on others' behavior. Since game theory reliably captures such interdependence of behavior in a population, we ground our computational approach

ACP worked on this research as an undergraduate student at Bowdoin. MTI is the corresponding author. LOH is currently an undergraduate student at Bowdoin.


H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2019, SCI 882, pp. 66–80, 2020.
https://doi.org/10.1007/978-3-030-36683-4_6

in game theory. The strategic setting of our interest here is the U.S. Senate. We model the influence structure among the senators by taking into account the relevant context, which we call the spheres of legislation. We learn these models of influence from the real-world behavioral data on Senate bills and voting records. Our particular focus is on analyzing machine learned influence networks to answer various questions on polarization and most influential nodes.

Interestingly, most computational models of influence assume a fixed network structure among individuals. We relax this simplifying assumption, allowing the network of influence to vary according to the spheres of legislation. For example, bills on finance may induce a very different influence network among senators than bills on defense, which may in turn have different impacts on inference problems like polarization and most influential nodes. One central question in this regard is: How do we identify different spheres of legislation that may have different implications on these inference problems? We address this in Sect. 2.

After identifying spheres of legislation, we can learn an influence network among the senators for each sphere by adopting game-theoretic models of strategic behavior. Broadly speaking, modeling and analyzing congressional voting behavior has been a trending topic in both political science and computer science [6,10,16,18,29], in part due to the availability of data. In particular, we use the linear influence game (LIG) model of strategic behavior proposed by Irfan and Ortiz [17,18] and its recent extension [16]. We learn these models using data from the spheres of legislation. In LIG, each senator exerts influence upon (and is subject to influences from) other senators in a network-structured way. The model focuses on interdependence among the senators and adopts the game-theoretic solution concept of *Nash equilibrium* to predict stable outcomes from a complex system of influences. This notion of Nash equilibrium leads to a definition of the most influential senators, where a group of senators is called most influential with respect to a desirable outcome if their support for that outcome influences enough other individuals to achieve that outcome. The LIG model will be elaborated in Sect. 3 and machine learning of this model using the spheres of legislation will be detailed in Sect. 4.

While game-theoretic *prediction* of congressional votes has been well studied using the LIG model and its extensions [16–18], an *analysis* of the machine learned networks of influence did not get much attention, which we address here. Similarly, algorithms for computing most influential nodes in a strategic setting have been studied before (e.g., [18]), but their structural analogs like centrality measures have not been explored in a comparative fashion. In other words, what do we gain by using a game-theoretic definition of most influential nodes as opposed to a structural definition? We address questions like this.

Furthermore, polarization in social networks has been well studied [1,13,27], especially in the political arena [7,9,26,32,33]. Three salient points distinguish our approach from the rich body of literature: (1) Ours is a model-based approach, where networks are central to predicting collective outcomes, (2) we learn the networks using behavioral data, because the networks are not observable, and (3) we seek to show that polarization in Senate varies according to the spheres of legislation. We do not touch on the rising polarization in Senate over time, which by now is a well-settled matter [8].

The past two terms of Congress are especially interesting for analyzing network behavior and polarization. The 114 th Congress ran from January 2015 to January 2017 and the 115 th Congress ran from January 2017 to January 2019. In both terms, Republicans controlled the Senate, but the executive power was different. In the 114 th Congress, Barack Obama (D) held the presidency; in the 115 th, Donald Trump (R) held the presidency. Despite the two opposing parties holding presidency, both terms are perceived to be deeply polarized. Interestingly, when we study different influence networks *among the same group of senators* arising from different spheres of legislation we find that polarization is not really equally applicable. It very much depends on the sphere under consideration. Our aim is to put polarization and other inference questions like most influential nodes in context. Before continuing, we note that all supplementary materials, including detailed literature reviews, visualizations, and technical details, are included in the Appendix.

## 2 Spheres of Legislation

We use an unsupervised machine learning technique, namely fuzzy clustering, to assign bills to different spheres of legislation based on the bill subjects. The clustering algorithm uses data obtained from the @unitedstates project (https://github.com/unitedstates/congress). In particular, we use bill data and roll-call data. The latter contains each senator's "yea," "nay," or abstaining votes, while bill data includes a list of subjects incident to the bill, among other attributes. There are 820 subjects ranging from "Abortion" to "Zimbabwe," and multiple subjects describe each bill. Additionally, each bill is assigned a single "top term," the broad subject which best describes the bill out of 23 possible top-level subjects. We use the roll-call data to learn strategic interactions and bill data to extract bill topics. For the 114 th and 115 th Senates, we have a total of 103 senators and 722 bills (details are in the Appendix).

We seek to split the bills into a small number of broad categories, each of which encompasses many bills. On their own, the "top terms" are too specific to be used as clusters of their own. Making each top term its own cluster would result in some clusters containing only one bill and others containing a hundred. Due to the exponential "outcome space" of LIGs, learning LIGs requires a relatively large amount of data. Therefore, small clusters would be unusable.

Rather than manually re-categorizing bills, we took a statistical clustering approach to grouping, based on a bill's assigned "top term" in addition to all subjects it contains. For each data point, we assigned each possible subject a weight: 0 if missing, 1 if present, or 10 if it is the "top term." By including both measures of subjects (top and regular), we produce more meaningful categories than using top terms or bill subjects lists alone.

In data science, K-Means (KM) is often used as a simple yet effective clustering algorithm [25]. Cluster membership in KM is crisp, meaning each data point belongs to one and only one cluster. While effective at producing distinct clusters, KM is not ideal for our purposes because bills often belong to multiple clusters. For example, a bill about increasing defense spending is about national security as well as economics. The Fuzzy C-Means (FCM) clustering algorithm addresses this problem. FCM is an extension of KM which allows for overlaps in clusters [2,30]. The objective function in FCM is largely the same as in KM, with the addition of membership values $w_{ij}$ and a fuzzifier $m$. Membership values describe how closely each data point $i$ belongs to cluster $j$. The fuzzifier changes membership values: $m = 1$ results in crisp clusters ($w_{ij} \in \{0, 1\}$), and higher values of $m$ result in fuzzier clusters.

Iterating over a range of values, we found that number of clusters, $c = 4$ and $m = 1.3$ resulted in clusters which were relatively distinct, had intuitive descriptions and also contained an adequate number of bills for machine learning. Additionally, we experimented with the threshold values for cluster membership and settled on 0.15. That is, a bill is considered a member of a cluster if it's membership value is above 0.15. Table 1 describes the results of our chosen FCM parameters. Each cluster is assigned a shorthand name describing its contents and is called a sphere of legislation in this paper. We next describe the model.

**Table 1.** Shorthand names and descriptions for each of the spheres of legislation identified by the FCM algorithm. Spheres 1 and 2 are relatively distinct from the rest, while Spheres 3 and 4 share a large number of bills.

| Sphere# | Size | Name of sphere | Sampling of bill subjects | Ovlp. 1 | Ovlp. 2 | Ovlp. 3 | Ovlp. 4 |
|---|---|---|---|---|---|---|---|
| 1 | 105 | Security & Armed Forces | Armed forces and national security (77), Emergency management (11), Transportation and public works (10) | | 7% | 20% | 20% |
| 2 | 263 | Economics & Finance | Economics and public finance (263) | 3% | | 0% | 0% |
| 3 | 284 | Energy & Infrastructure | Energy (69), Education (31), Taxation (28), Transportation and public works (27) | 7% | 0% | | 76% |
| 4 | 313 | Public Welfare | Health (52), Crime and law enforcement (43), Taxation (38), Education (31) | 7% | 0% | 69% | |

## 3 The LIG Model

We represent the senate influence network as a linear influence game (LIG) [17,18], one type of 2-action graphical game [20]. Nodes represent senators, or *players*, and are connected by directed edges. Edge weights represent the influence exerted by the source node upon the target. Influence weights can be negative, positive, or zero. The directed edges are allowed to be asymmetric, meaning nodes A and B may exert different levels of influences on each other. Additionally, nodes have a threshold level, which represents "stubbornness." Nodes with thresholds further from zero are more resistant to change. Absent influences, a node with negative threshold is predisposed to adopting action +1 (*yea* vote), and a node with positive threshold is predisposed to $-1$ (*nay* vote). The matrix of influence weights $\mathbf{W} \in \mathbf{R}^{n \times n}$ and the threshold vector $\mathbf{b} \in \mathbf{R}^n$ constitute the LIG model. The action $x_i \in \{+1, -1\}$ chosen by each node $i$ is the outcome of the model, as described below in game-theoretic terms.

Each node's *best response* to other nodes' actions depends on the net incoming influence and the node's threshold. When the total incoming influence from nodes playing $+1$ minus the total incoming influence from nodes playing $-1$ exceeds the node's threshold level, that node's best response is $+1$. If below, it is $-1$; in the case of a tie, the node is indifferent and can play either. Note that the best responses of the nodes are interdependent. A vector of *mutual best responses* of all the nodes is a stable outcome of the model, formally known as a *Nash equilibrium*. It is stable because no node has any incentive to deviate from it. The LIG model adopts Nash equilibria to represent stable collective outcomes from a complex network of influence. Below is a formal description, using the same notation as [18]. An example is provided in the Appendix.

**Definition 1 (Linear Influence Game (LIG) [18]).** *In* LIG*, the influence function of each individual $i$, given others' actions $\mathbf{x}_{-i}$, is defined as $f_i(\mathbf{x}_{-i}) \equiv \sum_{j \neq i} w_{ij} x_j - b_i$ where for any other individual $j$, $w_{ij} \in \mathbb{R}$ is a* weight *parameter quantifying the "influence factor" that $j$ has on $i$, and $b_i \in \mathbb{R}$ is a* threshold *parameter for $i$'s level of "tolerance."*

Here, individuals receive influences from other players and have an influence threshold of their own, which accounts for their own resistance to external influence. The influence function $f_i$ calculates the weighted sum of incoming influences on $i$, as described in the paragraph above Definition 1, and subtracts $i$'s threshold from it. The payoff of each player is defined next.

**Definition 2 (Payoff Function [18]).** *For an LIG, we define the* payoff *function $u_i : \{-1, 1\}^n \to \mathbb{R}$ as $u_i(x_i, \mathbf{x}_{-i}) \equiv x_i f_i(\mathbf{x}_{-i})$, where $\mathbf{x}_{-i}$ denotes the vector of a joint-action of all players except $i$ and $f_i$ is defined in Definition 1.*

The payoff function quantifies the preferences of the players based on the actions of other players. Given the action of all other individuals $\mathbf{x}_{-i}$ and influence function $f_i(\mathbf{x}_{-i})$, an individual will prefer to choose either $+1$ or $-1$ as follows. When $f_i(\mathbf{x}_{-i})$ is negative, $x_i = -1$ will result in a positive payoff; when $f_i(\mathbf{x}_{-i})$ is positive, $x_i = +1$ will result in a positive payoff. Actions chosen in this fashion in order to result in a positive payoff (i.e., to maximize payoff) is defined as the *best response*. When everyone is playing their best responses simultaneously, we get a pure-strategy Nash Equilibrium (PSNE) as defined below.

**Definition 3 (Pure-Strategy Nash Equilibrium [18]).** *A* pure-strategy Nash equilibrium (PSNE) *of an LIG $\mathcal{G}$ is an action assignment $\mathbf{x}^* \in \{-1, 1\}^n$ that satisfies the following condition. Every player $i$'s action $x_i^*$ is a simultaneous best-response to the actions $\mathbf{x}_{-i}^*$ of the rest.*

We adopt PSNE as the notion of stable outcomes arising from a network of influence. We are interested in questions like how the network changes based on the spheres of legislation and what impact the spheres have on polarization and most influential nodes. For these, we learn the networks using the spheres data.

## 4 Machine Learning

We use Honorio and Ortiz's machine learning algorithm in order to instantiate an LIG from raw roll-call data [15]. The goal of the algorithm is to capture as much of the ground-truth data as possible as PSNE (the *empirical proportion of equilibria*), without having so many total PSNE (the *true proportion of equilibria*) that the model is meaningless. For example, if all influence weights and threshold levels are 0 ($\mathbf{W} = 0$, $\mathbf{b} = 0$), then all $2^n$ possible joint actions among $n$ players would be PSNE, trivially covering all observed voting data. However, this is undesirable as it has no predictive power at all. Therefore, we would like to maximize the empirical proportion of equilibria while minimizing the true proportion. To balance the true and empirical proportions of equilibria, the learning algorithm uses a generative mixture model that picks a joint action which is either a PSNE or non-PSNE with probability $q$ and $1 - q$, respectively. Maximizing the empirical proportion of equilibria relative to the true proportion can be framed as a maximum likelihood estimation problem in this generative model. Under mild conditions, the final optimization problem is the following: $\min_{\mathbf{W},\mathbf{b}} \frac{1}{m} \sum_l \max_i \ell\big[x_i^{(l)}(\mathbf{w}_{i,-i}^T \mathbf{x}_{-i}^{(l)} - b_i)\big] + \rho||\mathbf{w}||_1$. Here, $m$ is the number of bills, $\ell$ is the typical logistic loss function, and $\rho$ is an $l_1$ regularization parameter controlling the number of edges $||\mathbf{w}||_1$. This is a gist of Honorio and Ortiz's machine learning algorithm resulting from a very lengthy proof [15].

We solve the above optimization for each sphere of legislation and obtain an influence network. While doing this, we rigorously cross validate to avoid overfitting or underfitting as follows. In the model selection phase, we wish to choose an "appropriate" value of the $l_1$ regularization parameter $\rho$. Since $\rho$ penalizes the number of edges, high values of $\rho$ result in sparser graphs at the risk of underfitting, and low values of $\rho$ lead to denser graphs at the risk of overfitting. The number of edges is important, because the problem of computing equilibria is NP-hard [17,18], and an extremely complex model would have so many edges that equilibrium computation would not finish within days. We use 10-fold cross-validation, track multiple metrics, and choose $\rho = 0.002728$, 0.003888, 0.003070, 0.003888 for the four spheres, respectively. Details are in the Appendix.

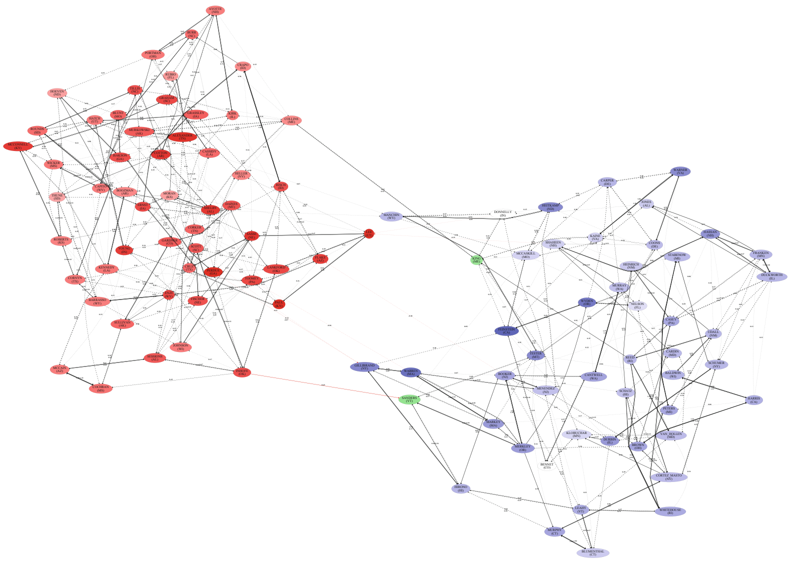


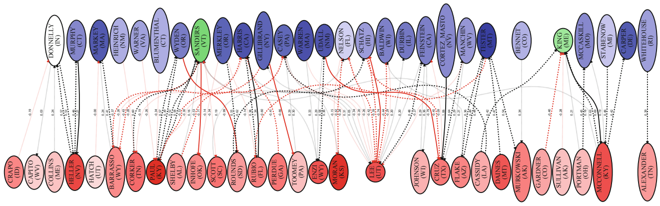
**Fig. 1.** A bird's eye view of the LIG network for Sphere 2 (Economics & Finance). Red nodes are Republicans, blue Democrats, green Independents. Darker nodes have higher threshold and thicker edges have more influence weights. The strongest 40% incoming and outgoing edges for each node are shown.

## 5 Polarization in Context

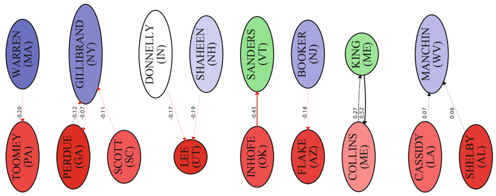
Visualization of the machine learned networks clearly shows that the network structure varies according to the spheres of legislation. In all spheres, however, the force-directed drawing algorithm automatically distinguishes Republicans from Democrats. Figure 1 depicts this LIG visualization for Sphere 2 (Economics & Finance) as a representative example. The visualizations for the remaining spheres can be found in the Appendix.

The boundary between the two parties is interesting for studying polarization. Even though negative edges more often occur at the boundary, the connectivity between the two parties varies a lot according to the spheres of legislation. These are depicted in Fig. 2 for Spheres 1 and 2 (others are in the Appendix).

Figure 2b shows the cross-border edges in Sphere 2 (Economics & Finance), which starkly contrasts those of Sphere 1 (Security & Armed Forces). In Sphere 2, only 12 of the strongest 40% of edges are between members of different parties. Of these, 2/3 are negative, suggesting a very polarized network. Aside from two positive influences between Maine senators King (a left-leaning Independent) and Collins (a center-leaning Republican), the remaining two positive connections are the weakest of all connections shown for this sphere.



(a) Sphere 1 (Security & Armed Forces): 52 boundary edges are positive, 37 negative.




(b) Sphere 2 (Economics & Finance): 4 cross-border edges are positive, 8 negative.

**Fig. 2.** *Graphviz* visualization of cross-border edges connecting members of the opposing parties within the strongest 40% of all edges.

Similarly, examining inter-party edges reveals that Sphere 3 (Energy & Infrastructure) is also very polarized. While there are many edges between both parties in this network, about 70% of them are negative. Positive influences come from a few sources, again including the centrist Senator Collins. Incongruously, prominent right-wing senator Tom Cotton (R-AR) also exhibits positive influences with democratic senators. However, most other far-left or far-right leaning senators, including Sanders (I-VT) and Cruz (R-TX) only exhibit negative influences with the opposite party.

Sphere 4 (Public Welfare)'s inter-party edges strike a balance between the polarities exhibited by the previous three spheres. There are slightly more positive edges (9) than negative edges (7), but still a low number of edges overall. Again, there are positive influences between Maine senators King (I-ME) and Collins (R-ME), but also positive influences between Senator McConnell and Democratic senators King (D-ME) and Tester (D-MT).

Overall, each sphere exhibits some level of polarization, but influences within some spheres are far less polarizing than others. Some senators are present in every sphere's inter-party boundary, whether for positive or negative influences. Senators Collins and King often share positive influences with each other, as well as other senators. Senator Lee (R-UT), a conservative libertarian, always exhibits negative edges with members of the other party, although in Sphere 1, he also shares positive influences with senators Harris (D-CA) and Feinstein (D-CA). Meanwhile, left-wing icon Bernie Sanders (I-VT) exhibits the equivalent behavior, with only negative cross-border edges in all spheres *except* Sphere 1. These results suggest that Sphere 1 (Security & Armed Forces) is least polarized, whereas Spheres 2 (Economics & Finance) is highly polarized.

Furthermore, a formal study of polarization rooted in network science produces similar results. Modularity [11,27,28] has been widely used as a measure of polarization in networks. We apply the following definition of modularity derived for directed networks with signed weights [14].

$$Q = \frac{1}{2w^{+} + 2w^{-}} \sum_{i} \sum_{j} \left[ w_{ij} - \left( \frac{w_i^{+,out} w_j^{+,in}}{2w^{+}} - \frac{w_i^{-,out} w_j^{-,in}}{2w^{-}} \right) \right] \times \delta\left(C_i, C_j\right).$$

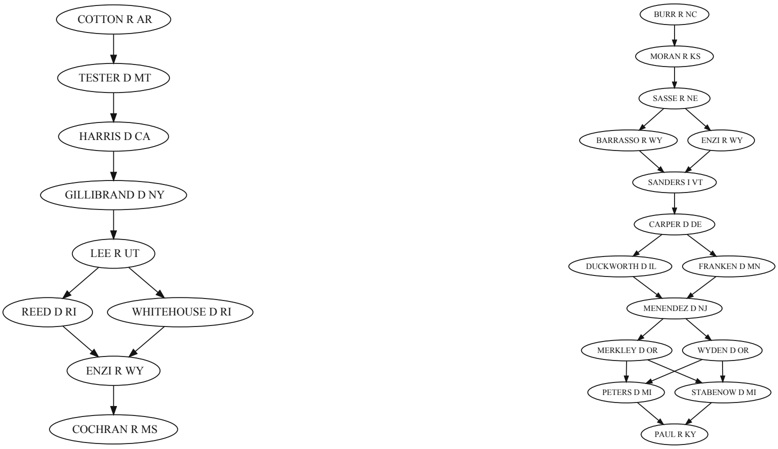
Here, $w_{ij}$ is the weight of edge $i$ to $j$, $w_{ij}^{+} = max\{0, w_{ij}\}$, $w_{ij}^{-} = max\{0, -w_{ij}\}$, and $2w^{\pm}$ is the total weight of all positive or negative edges, expressed by $\sum_i \sum_j w_{ij}^{\pm}$. Furthermore, $w_i^{\pm,out}$ is the weighted out-degree $\sum_k w_{ik}^{\pm}$ and $w_j^{\pm,in}$ is the weighted in-degree $\sum_k w_{kj}^{\pm}$. The Kronecker delta function $\delta\left(C_i, C_j\right)$ is 1 if $i$ and $j$ belong to the same party; it is 0 otherwise.

Applying this definition, We obtain the following modularity scores for the four spheres of legislation respectively: 0.7861, 0.8904, 0.8724, and 0.8857. This shows that Sphere 1 (Security & Armed Forces) is least polarized and Spheres 2 (Economics & Finance), 3 (Energy & Infrastructure), and 4 (Public Welfare) are much more polarized.

## 6 Most Influential Nodes in Context

There exists a number of centrality measures that are derived from a structural analysis of networks [19]. However, our model is behavioral where nodes adopt their best responses to each other. In a strictly game-theoretic model of behavior, a set of nodes will be called most influential *with respect to achieving a desirable stable outcome* if their choice of actions leads the whole system of influence to that desirable outcome [17,18]. Here, a crucial aspect is a desirable stable outcome, represented by a PSNE. For example, let us say that our desirable outcome is to pass a bill by a 100-0 vote. A set of senators will be called most influential if their voting together influences every other senator to also vote for the bill, thereby having the desirable outcome as the *unique* PSNE outcome. This concept can be extended to other types of desirable outcomes like passing a bill with at least 60 votes, forcing/avoiding a filibuster, etc.

An approximation algorithm for computing most influential senators was given by Irfan and Ortiz [18], which produces a directed acyclic graph (DAG). The algorithm requires precomputation of all PSNE, which is a provably hard problem [18]. We apply Irfan and Ortiz's PSNE computation algorithm to the LIG for each sphere of legislation. We elaborate this in the Appendix. Having computed all the PSNE, we then compute the DAG representing most influential sets of nodes. Figure 3 shows the results of the most influential nodes algorithm for Spheres 1 and 3, where the desirable outcome is set as passing a bill unanimously. The way to read Fig. 3 is to inspect each DAG and find a top to bottom path. Each of these paths gives a most influential set.




(a) Sphere 1 (Security & Armed Forces): 4 Republicans and 4 Democrats are most influential.

(b) Sphere 3 (Energy & Infrastructure): 5 Republicans and 6 Democrats are most influential.

**Fig. 3.** Directed acyclic graphs (DAGs) representing sets of most influential nodes for Spheres 1 and 3. Any top to bottom path gives a most influential set.

The sets of most influential senators in each sphere support the inferences gained from analyzing the LIG networks. As illustrated in Fig. 3, in Sphere 1 (Security & Armed Forces), 4 Republicans and 4 Democrats comprise a set of 8 most influential senators. In other words, 8 senators are sufficient to guarantee enough support that a bill will pass unanimously. In Sphere 3 (Energy & Infrastructure), 5 Republicans and 6 Democrats comprise a set of 11. This suggests that Sphere 3 is more polarized than Sphere 1, since it requires a larger body of influencing senators. The DAGs for the other spheres are in the Appendix.

**Game-Theoretic vs. Structural Centrality Measures.** In the above game-theoretic formulation of most influential nodes, we find that each set of most influential senators across all spheres is comprised of an (almost) equal number of Democrats and Republicans. This signifies the need for bipartisan support to guarantee passing a bill unanimously. As we show next, this also happens to be a distinguishing feature between game-theoretic and structural measures. Table 2 shows various centrality measures and other quantities computed for each sphere. For each sphere, we show the top 10 most central senators with respect to four centrality measures: degree, closeness, betweenness, and eigenvector. Most notably, these centrality measures do not capture the strategic aspects of behavior. Throughout most measures, Republican senators are over-represented, comprising the majority of the top ten most central nodes. In contrast, the game-theoretic measure gives a balanced coalition between Democrats and Republicans. This is important because when networks are polarized, achieving a desirable outcome requires support from both sides.

**Table 2.** Network analysis of learned influence networks for different spheres of legislation. Various centrality measures and network-level properties are shown.

| | Sphere 1 | Sphere 2 | Sphere 3 | Sphere 4 |
|---|---|---|---|---|
| Number of Edges | 1191 | 1071 | 1280 | 1076 |
| Network Diameter | 5 | 5 | 4 | 5 |
| Modularity | 0.7861 | 0.8904 | 0.8724 | 0.8857 |
| Avg. (Shortest) Path Length | 2.2295 | 2.5132 | 2.1506 | 2.5476 |
| Avg. Clustering Coefficient | 0.1867 | 0.2057 | 0.174 | 0.2176 |
| Degree Centrality | | | | |
| Degree (1) | 0.5784: LEE R-UT | 0.3529: TOOMEY R-PA | 0.3725: LANKFORD R-OK | 0.3725: COTTON R-AR |
| Degree (2) | 0.4216: PAUL R-KY | 0.3333: PERDUE R-GA | 0.3627: SASSE R-NE | 0.2941: LEAHY D-VT |
| Degree (3) | 0.4118: SANDERS I-VT | 0.3235: ENZI R-WY | 0.3529: WARNER D-VA | 0.2843: CAPITO R-WV |
| Degree (4) | 0.3824: MORAN R-KS | 0.3137: LANKFORD R-OK | 0.3333: CAPITO R-WV | 0.2843: MURKOWSKI R-AK |
| Degree (5) | 0.3725: MANCHIN D-WV | 0.3137: YOUNG R-IN | 0.3235: TOOMEY R-PA | 0.2745: AYOTTE R-NH |
| Degree (6) | 0.3627: RUBIO R-FL | 0.3039: COTTON R-AR | 0.3235: MURKOWSKI R-AK | 0.2745: SHELBY R-AL |
| Degree (7) | 0.3529: CRUZ R-TX | 0.2941: CASEY D-PA | 0.3235: COTTON R-AR | 0.2647: PERDUE R-GA |
| Degree (8) | 0.3333: ALEXANDER R-TN | 0.2843: CASSIDY R-LA | 0.3137: BROWN D-OH | 0.2549: ALEXANDER R-TN |
| Degree (9) | 0.3333: ENZI R-WY | 0.2843: WICKER R-MS | 0.3137: FEINSTEIN D-CA | 0.2549: PETERS D-MI |
| Degree (10) | 0.3235: LEAHY D-VT | 0.2745: CORKER R-TN | 0.3137: PAUL R-KY | 0.2549: PAUL R-KY |
| Closeness Centrality | | | | |
| Closeness (1) | 0.5862: LEE R-UT | 0.5126: PERDUE R-GA | 0.5514: WARNER D-VA | 0.5204: COTTON R-AR |
| Closeness (2) | 0.5635: RUBIO R-FL | 0.4951: COTTON R-AR | 0.5426: LANKFORD R-OK | 0.5178: KIRK R-IL |
| Closeness (3) | 0.5574: PAUL R-KY | 0.4766: COLLINS R-ME | 0.5368: SASSE R-NE | 0.4951: MURKOWSKI R-AK |
| Closeness (4) | 0.5455: SANDERS I-VT | 0.47: ENZI R-WY | 0.5368: BENNET D-CO | 0.4928: AYOTTE R-NH |
| Closeness (5) | 0.5455: BALDWIN D-WI | 0.4636: SASSE R-NE | 0.534: KING I-ME | 0.4766: SULLIVAN R-AK |
| Closeness (6) | 0.5397: ENZI R-WY | 0.4636: MANCHIN D-WV | 0.5231: MURKOWSKI R-AK | 0.4744: MCCONNELL R-KY |
| Closeness (7) | 0.5368: MORAN R-KS | 0.4615: FLAKE R-AZ | 0.5178: COTTON R-AR | 0.4722: PAUL R-KY |
| Closeness (8) | 0.5285: CORKER R-TN | 0.4595: SHELBY R-AL | 0.5152: BROWN D-OH | 0.4636: COLLINS R-ME |
| Closeness (9) | 0.5258: CASEY D-PA | 0.4595: YOUNG R-IN | 0.5126: CASEY D-PA | 0.4554: CAPITO R-WV |
| Closeness (10) | 0.5231: DURBIN D-IL | 0.4595: HEITKAMP D-ND | 0.51: CARPER D-DE | 0.4554: SANDERS I-VT |
| Betweenness Centrality | | | | |
| Betweenness (1) | 0.0696: LEE R-UT | 0.0538: PERDUE R-GA | 0.0278: LANKFORD R-OK | 0.0685: SANDERS I-VT |
| Betweenness (2) | 0.0362: PERDUE R-GA | 0.0468: HEITKAMP D-ND | 0.0272: SASSE R-NE | 0.0641: WYDEN D-OR |
| Betweenness (3) | 0.033: MORAN R-KS | 0.0452: GILLIBRAND D-NY | 0.0265: WARNER D-VA | 0.0559: COTTON R-AR |
| Betweenness (4) | 0.0314: KING I-ME | 0.045: ENZI R-WY | 0.0226: BENNET D-CO | 0.0533: MARKEY D-MA |
| Betweenness (5) | 0.0301: MANCHIN D-WV | 0.0434: COLLINS R-ME | 0.0206: MURKOWSKI R-AK | 0.0532: ALEXANDER R-TN |
| Betweenness (6) | 0.0288: DURBIN D-IL | 0.0383: MERKLEY D-OR | 0.0196: BROWN D-OH | 0.0421: MURKOWSKI R-AK |
| Betweenness (7) | 0.0283: RUBIO R-FL | 0.0382: COTTON R-AR | 0.0194: CORNYN R-TX | 0.0381: PAUL R-KY |
| Betweenness (8) | 0.0283: PAUL R-KY | 0.0381: SANDERS I-VT | 0.0193: SCHATZ D-HI | 0.0376: SASSE R-NE |
| Betweenness (9) | 0.0253: SANDERS I-VT | 0.0344: LEE R-UT | 0.0187: SANDERS I-VT | 0.0347: HARRIS D-CA |
| Betweenness (10) | 0.0249: CRUZ R-TX | 0.034: TESTER D-MT | 0.0185: WICKER R-MS | 0.032: AYOTTE R-NH |
| Eigenvector Centrality | | | | |
| Eigenvector (1) | 0.271: LEE R-UT | 0.2114: COTTON R-AR | 0.181: WARNER D-VA | 0.2029: KIRK R-IL |
| Eigenvector (2) | 0.2291: SANDERS I-VT | 0.2061: PERDUE R-GA | 0.1703: CORNYN R-TX | 0.2017: HOEVEN R-ND |
| Eigenvector (3) | 0.228: PAUL R-KY | 0.1866: SULLIVAN R-AK | 0.1675: LANKFORD R-OK | 0.1727: GARDNER R-CO |
| Eigenvector (4) | 0.1894: BALDWIN D-WI | 0.1865: ENZI R-WY | 0.1567: BENNET D-CO | 0.1724: PORTMAN R-OH |
| Eigenvector (5) | 0.1892: RUBIO R-FL | 0.183: YOUNG R-IN | 0.1551: JOHNSON R-WI | 0.1715: CAPITO R-WV |
| Eigenvector (6) | 0.1696: ENZI R-WY | 0.1758: THUNE R-SD | 0.1541: SASSE R-NE | 0.1706: COTTON R-AR |
| Eigenvector (7) | 0.1681: BARRASSO R-WY | 0.1734: WICKER R-MS | 0.1525: MURKOWSKI R-AK | 0.1706: ROBERTS R-KS |
| Eigenvector (8) | 0.161: CASEY D-PA | 0.1674: MORAN R-KS | 0.1454: KING I-ME | 0.1683: FISCHER R-NE |
| Eigenvector (9) | 0.1607: MORAN R-KS | 0.1653: JOHNSON R-WI | 0.1426: COTTON R-AR | 0.1682: MURKOWSKI R-AK |
| Eigenvector (10) | 0.1602: MANCHIN D-WV | 0.163: GARDNER R-CO | 0.1379: BOOZMAN R-AR | 0.1646: ISAKSON R-GA |

## 7 Richer Models: Ideal Point Models with Social Interactions

We also apply a richer model of influence recently proposed by Irfan and Gordon [16] that extends the LIG model by incorporating ideal points of the senators and polarities of bills [29]. Their work showed the value of combining game-theoretic and statistical models for studying strategic interactions in context, but they assume the network to be fixed, regardless of the bill context. We use their model and allow the network to change based on the spheres of legislation. We also perform an analysis of the networks learned.

As a cautionary note, the way Irfan and Gordon's model [16] combines networks with ideal points makes it difficult to disentangle the two. Analyzing the networks alone may be inconclusive, because ideal points also supply the model with predictive power. Moreover, the machine learning algorithm also learns these two components simultaneously. Figure 4 shows the learned network for Sphere 2 (Economics & Finance) under this richer model. It is evident that the two parties are not as clustered as they are in the LIG model (see Fig. 1 for Sphere 2 under LIG). This is because the ideal points are now being used in addition to social interactions to discriminate the behaviors of opposing senators. More interestingly, an analysis of the cross-party edges show that there are a lot more negative edges between the two parties under this richer model than there are under the LIG model (details are in the Appendix).

Nevertheless, we still study the ideal point distributions and influence networks under this model. We apply ideal point-based polarization metrics [26] as well as network modularity metrics [14] to calculate polarization levels across the four spheres. The ideal point distributions for two of the spheres are depicted in Fig. 5 (others are in the Appendix).

Applying the well-known ideal point-based polarization metric (i.e., distance between the means of the two parties) [26], we obtain values of 0.754, 1.235, 1.126, and 0.889 for Spheres 1 to 4 respectively. Evidently, Sphere 1 is least polarizing with respect to the ideal point distributions alone.

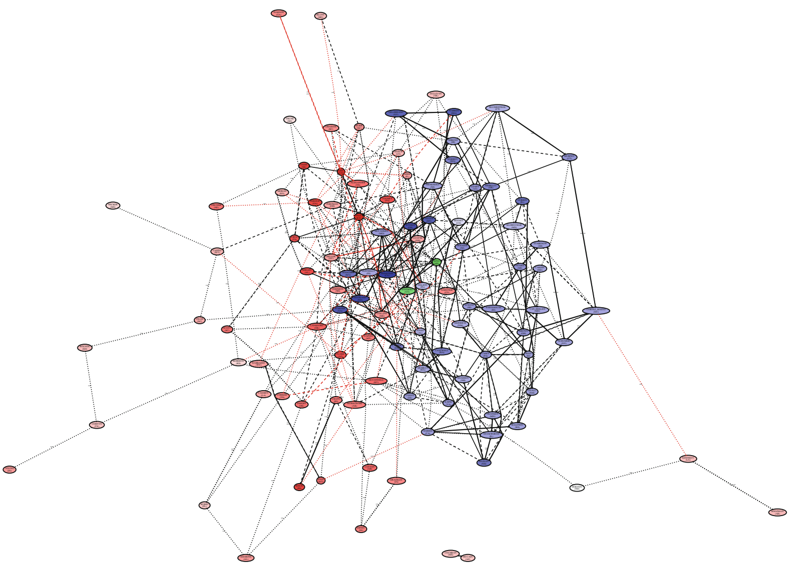
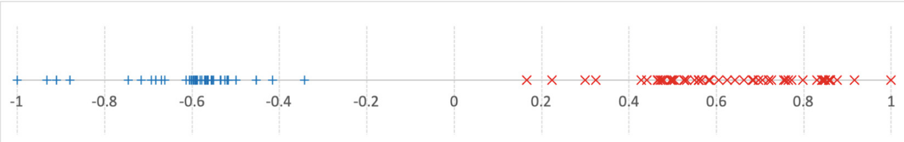


**Fig. 4.** A bird's eye view of the influence network for Sphere 2 (Economics & Finance) learned under the ideal point model with social interactions. The strongest 33% of the edges are shown. Contrast this with Fig. 1 where the two parties were relatively well separated.

We also analyze the influence networks for the four spheres. The modularity framework discussed in Sect. 5 yields scores of 0.5392, 0.6801, 0.6887, and 0.6229, respectively. Both the ideal point metric and modularity scores indicate that Spheres 2 (Economics & Finance) and 3 (Energy & Infrastructure) are most polarizing, whereas Sphere 1 (Security & Armed Forces) is least polarizing.


(a) Sphere 1 (Security & Armed Forces): The distance between the mean ideal points of the two parties is 0.754. It shows that Democratic and Republican senators are ideologically close to each other when it comes to national security.




(b) Sphere 2 (Economics & Finance): The distance between the mean ideal points of the two parties is 1.235, which shows more polarization compared to Sphere 1.

**Fig. 5.** The ideal point distributions of Democratic (blue +) and Republican (red x) senators, scaled linearly between $-1$ and 1.

Sphere 4 (Public Welfare) sits in between. These results are somewhat similar to our earlier conclusions based on LIG without ideal points.

As mentioned above, investigating the influence networks and ideal points separately does not give us the complete picture since the model combines these two components together to make predictions. Therefore, we should also combine them in a meaningful way to infer polarization. We leave this as future work.

**Acknowledgement.** This research was partially supported by NSF grant IIS-1910203. We sincerely thank Dr. Stephen Majercik (Bowdoin College) for reading an earlier draft of this paper and giving us many valuable suggestions. We are also thankful to Drs. Honorio and Ortiz for letting us use their codes [15] and to the anonymous reviewers for their suggestions.

## Appendix

Link to the Appendix: http://bit.ly/appendix-spheres-of-legislation.

## References

1. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on facebook. Science **348**(6239), 1130–1132 (2015)
2. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
3. Chen, W., Collins, A., Cummings, R., Ke, T., Liu, Z., Rincon, D., Sun, X., Wang, Y., Wei, W., Yuan, Y.: Influence maximization in social networks when negative opinion may emerge and propagate. In: Proceedings of the Eleventh SIAM International Conference on Data Mining, pp. 379–390 (2011)

4. Christakis, N.A., Fowler, J.H.: The spread of obesity in a large social network over 32 years. N. Engl. J. Med. **357**(4), 370–379 (2007). http://content.nejm.org/cgi/content/abstract/357/4/370
5. Christakis, N.A., Fowler, J.H.: The collective dynamics of smoking in a large social network. N. Engl. J. Med. **358**(21), 2249–2258 (2008). http://content.nejm.org/cgi/content/abstract/358/21/2249
6. Clinton, J., Jackman, S., Rivers, D.: The statistical analysis of roll call data. Am. Polit. Sci. Rev. **98**(2), 355–370 (2004)
7. Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on Twitter. In: Fifth International AAAI Conference on Weblogs and Social Media (2011)
8. Farina, C.R.: Congressional polarization: terminal constitutional dysfuction. Colum. L. Rev. **115**, 1689 (2015)
9. Garcia, D., Abisheva, A., Schweighofer, S., Serdült, U., Schweitzer, F.: Ideological and temporal components of network polarization in online political participatory media. Policy Internet **7**(1), 46–79 (2015)
10. Gerrish, S.M., Blei, D.M.: How they vote: issue-adjusted models of legislative behavior. Adv. Neural Inf. Process. Syst. **25**(1), 2762–2770 (2012)
11. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)
12. Granovetter, M.: Threshold models of collective behavior source. Am. J. Sociol. **83**(6), 1420–1443 (1978). https://www.jstor.org/stable/27781
13. Guerra, P.C., Meira Jr., W., Cardie, C., Kleinberg, R.: A measure of polarization on social media networks based on community boundaries. In: 7th International AAAI Conference on Weblogs and Social Media (2013)
14. Gómez, S., Jensen, P., Arenasl, A.: Analysis of community structure in networks of correlated data. Institut des Systèmes Complexes **1**(1), 2–3 (2009)
15. Honorio, J., Ortiz, L.: Learning the structure and parameters of large-population graphical games from behavioral data. J. Mach. Learn. Res. **16**, 1157–1210 (2015). http://arxiv.org/abs/1206.3713
16. Irfan, M.T., Gordon, T.: The power of context in networks: ideal point models with social interactions. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, pp. 910–918. International Foundation for Autonomous Agents and Multiagent Systems (2018)
17. Irfan, M.T., Ortiz, L.E.: A game-theoretic approach to influence in networks. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, pp. 688–694 (2011). http://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3746
18. Irfan, M.T., Ortiz, L.E.: On influence, stable behavior, and the most influential individuals in networks: a game-theoretic approach. Artif. Intell. **215**, 79–119 (2014). http://www.sciencedirect.com/science/article/pii/S0004370214000812
19. Jackson, M.O.: Social and Economic Networks. Princeton University Press, Princeton (2010)
20. Kearns, M., Littman, M., Singh, S.: Graphical models for game theory. In: Proceedings of the Conference on Uncertainty in Artificial Intelligence, pp. 253–260 (2001)
21. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2003)

22. Kempe, D., Kleinberg, J., Tardos, É.: Influential nodes in a diffusion model for social networks. In: Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP) (2005)
23. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. ACM Trans. Web (TWEB) **1**(1), 5 (2007)
24. Li, Y., Chen, W., Wang, Y., Zhang, Z.L.: Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 657–666. ACM (2013)
25. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1: Statist, pp. 281–297 (1967). https://projecteuclid.org/euclid.bsmsp/1200512992
26. McCarty, N., Poole, K.T., Rosenthal, H.: Polarized America: The Dance of Ideology and Unequal Riches. MIT Press, Cambridge (2016)
27. Newman, M.E.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. **103**(23), 8577–8582 (2006)
28. Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**(2), 026113 (2004)
29. Poole, K.T., Rosenthal, H.: A spatial model for legislative roll call analysis. Am. J. Polit. Sci. **29**(2), 357–384 (1985). Published by: Midwest Political Science Association St. Political Science 29(2), 357–384 (2008)
30. Ross, T.J.: Fuzzy Logic with Engineering Applications, 2nd edn. Wiley, Hoboken (2004)
31. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, New York (1994)
32. Waugh, A.S., Pei, L., Fowler, J.H., Mucha, P.J., Porter, M.A.: Party polarization in congress: a network science approach. arXiv preprint arXiv:0907.3509 (2009)
33. Zhang, Y., Friend, A.J., Traud, A.L., Porter, M.A., Fowler, J.H., Mucha, P.J.: Community structure in congressional cosponsorship networks. Phys. A Stat. Mech. Appl. **387**(7), 1705–1712 (2008)