Utilizing Political Campaign Contribution Networks to Identify Political Affiliations and Important Actors in US Elections

Leah Balter[a], **William Cortes**[b], **Jing Lu**[a], **Aviva Prins**[a], **and Oladimeji Salako**[c]

[a]Department of Mathematics, University of California, Los Angeles, CA 90095; [b]Department of Physics and Astronomy, University of California, Los Angeles, CA, 90095; [c]Department of Computer Science, University of California, Los Angeles, CA, 90095

This manuscript was compiled on June 15, 2018

We seek to identify important political committees and states in recent election cycles by applying centrality measures to networks formed from political campaign data. One network we focus on is the senatorial bipartite committee-state network, which denotes relationships between committees and senatorial candidates from a specific state. We find that a state's weighted PageRank score correlates to the contentiousness of that state's senatorial election, and that the important committees of a given senatorial election cycle tend to reflect contemporary trends in national political discourse. The political action committee associated with the National Rifle Association is found to be the single-most important committee by weighted PageRank. By analyzing the community that contains this committee as detected by the Louvain algorithm, we find that network community structure corresponds to real-life political relationships. Furthermore, we utilize another community detection algorithm to split this network into two groups, and find that a state's community tends to correlate to the party that won that state's senatorial election.

networks | elections | PageRank | community detection

In recent years, United States politics has become increasingly more partisan. As bipartisan compromise has become less frequent and more politically charged, the Democratic and Republican parties have tried to win as many Senate and House seats as possible in order to pass legislation congruent to their ideals without depending on support from members of the opposing party. In order to garner the voter support necessary to win elections, candidates often rely on financial backing from various entities, such as the political party itself or Politcal Action Committees (PACs).

In this paper, we look at two networks formed from political campaign data. The first is a committee-committee network that traces how money moves between committees. The second is a bipartite committee-state network, wherein an edge is defined between a committee and a state if a committee donates money to a candidate running for a seat in that state. Since PACs tend to support those candidates that they believe will enact policies that support their interests, we posit that the structure of the latter network may be influenced by contentiousness of elections and by contemporary political issues. As such, we try to identify the important nodes of these networks by utilizing centrality measures, and aim to see if we can detect communities in the networks that reflect political affiliation.

Data Acquisition and Cleaning

The United States Federal Election Commission (FEC) is the federal agency tasked with enforcing campaign finance law. All candidates for US House of Representatives, US Senate, and President must register with the FEC in order to perform election-related financial transactions, as must labor unions, PACs, and candidates' political campaign committees. Each such entity is assigned a unique FEC code.

Andrew Waugh compiled a list of transactions filed with the FEC for presidential, Senate, and House elections between 1980 and 2010, organized by two-year election cycle (1). He also compiled a list of all House, senatorial, and presidential candidates for each election cycle, with associated meta-data, as well as a list of all entities with an FEC code for each election cycle. We obtain this dataset courtesy of Professor Porter (2). Each transaction in the dataset represents money received or disbursed by an entity with an FEC code. The FEC tags each transaction with a code that describes the nature of the transaction, like whether the transaction represents the candidate loaning money to his campaign, or an individual donating money to a PAC, or a PAC contributing money to a candidate, or one of a multitude of other scenarios.

The Committee-Committee Network. A group of codes – 24C, 24E, 24F, 24H, 24K, 24R, and 24Z – delineate money being disbursed by an entity with an FEC code to another entity with an FEC code. Because we are interested in studying the network comprised of committees transferring money to other committees, we filter the dataset to only include transactions tagged with these codes. Thus, we form a directed edgelist between committee-nodes, where a transaction between two committees represents a directed edge from the sending com-

Significance Statement

United States politics forms a complex ecosystem wherein voters, political parties, and interest groups jockey for the power to define and decide public policy by backing the winning candidate in an election. We utilize network centrality measures and community detection to analyze political campaign contributions by identifying important and influential campaign donors and to discover trends in voting patterns and political affiliations.

Leah cleaned the dataset and created edgelists for both the committee-committee network and committee-state network in R with help from Jing and Oladimeji. Leah performed all analysis on the committee-state network in R and MATLAB, wrote MATLAB scripts to aid visualization of these networks, and performed all relevant political research. Oladimeji, Jing, and Aviva investigated the committee-committee network. Oladimeji studied the Walktrap algorithm. William investigated a configuration model (see Supplementary Information). Aviva organized biweekly meetings, created all visualizations in Tableau, and created the presentation. Leah wrote the paper with help from Aviva.

The authors declare no conflict of interest.

To whom correspondence should be addressed. leahmbalter@gmail.com

mittee to the receiving committee. Each edge is assigned a weight equal to the amount of money transferred in that transaction. We refer to this weighted, directed network as the committee-committee network.

For this network, we primarily focus on the 2010 election cycle. After selecting out only the transactions with the appropriate 24__ tags, we obtain a network of 7,883 nodes and 356,004 edges, with a total of $885,266,972 transferred between committees. Since the weight of each edge represents the amount of money moving from the sending committee to the receiving committee, we expect it to be positive. However, 7,828 transactions – representing $15,574,220 and 1.8% of the money in the original network – have non-positive amounts. Given the scope of this project, we simply remove these edges and create a final weighted, directed network of 7,757 nodes and 348,176 edges. Further analysis would try to determine if these non-positive monetary amounts were a result of an error in how the data was compiled, or if they in fact represent money moving in the opposite direction, from the node usually designated the "receiving" node to the node usually designated the "sending" node.

The Bipartite Committee-State Network. We create a second type of network from this dataset in order to study how political campaign contributions relate to voting patterns and election results. To do this, we focus only on senatorial elections, which occur every two years. To create this network, we select out those transactions from the committee-committee network which transfer money to general election senatorial candidates or their associated campaign committees. Since each candidate is associated with a state, we can collapse this committee-senatorial candidate edgelist by state, thus producing the committee-state edgelist that defines the bipartite committee-state network. Each edge between a committee and a state indicates that the committee donated money to a general-election senatorial candidate in that state in that election cycle. The weight associated with each edge represents the sum total of money a committee donated to candidates from that state.

We primarily study the networks associated with the 2008 and 2010 general senatorial elections. Even though only 33 seats were up in 2008, and 34 in 2010, all 50 states appear in both networks. We suspect that candidates campaigning early for seats available in those other states in future election cycles received some money during the election cycles in question, and thus states associated with such candidates appeared in our networks. The 2008 network has 14,567 edges and 2,371 nodes, with 2,321 committees and 50 states, while the 2010 network has 15,769 edges and 2,677 nodes, with 2,627 committees and 50 states.

Methods

Centrality Measures. We want to see if we can determine important committees and states in a given election cycle utilizing centrality measures, which attempt to identify important nodes in a network. One type of centrality measure is eigenvector centrality, which considers a node important if it itself is connected to other important nodes. Eigenvector centrality is simply the vector x that satisfies

$$A\mathbf{x} = \lambda \mathbf{x}$$

where A is the adjacency matrix associated with the network, and λ is A's largest eigenvalue. The Perron-Frobrenius theorem guarantees that if the network is strongly connected, λ's associated eigenvector can be chosen to have strictly positive components (3, p. 170).

However, our committee-committee network is not strongly connected; the largest connected component of the network is just nine nodes. This makes sense since once money reaches a particular candidate, that individual is more likely to utilize those funds to aid his own campaign instead of transferring the money to other committees or campaigns. Since this network is perhaps more similar to a directed acyclic graph than to a strongly connected component, eigenvector centrality is not the appropriate tool to study it.

Instead, we turn to PageRank, a variant of eigenvector centrality particularly well-suited to directed acyclic graphs because of its teleportation parameter α. PageRank can be expressed as the leading eigenvector solution to

$$(\alpha A + (1-\alpha)\mathbf{v}\mathbf{1}^T)\mathbf{x} = \mathbf{x}$$

where $\alpha \in [0, 1]$, A is the weighted adjacency matrix, $\mathbf{1}$ is a column vector of 1's, and v describes the teleportation strategy, the probability of jumping to some random node in the network instead of moving to a neighboring node when traveling from a given node. When we set α to be 1, we recover eigenvector centrality. By default, v_j is taken as $\frac{1}{N}$ for all j, where N is the number of nodes in the network, but can be customized to describe other teleportation strategies (4). Because of this teleportation parameter, PageRank can be used to study the importance of nodes in networks that are not strongly connected, like the committee-committee network.

Community Detection. We also want to apply community detection to these networks to see if communities found by such algorithms correspond to actual political relationships. One common method of community detection is to organize nodes in groups in such a way that maximize modularity, a quantity that is defined as

$$Q = \frac{1}{2m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})\ \delta(c_i, c_j)$$

where m is the number of edges in the network, A_{ij} is the ijth entry of the adjacency matrix associated with the network, k_i is the degree of node i, and c_i represents the community that node i belongs to. $\delta(c_i, c_j)$ is 1 if nodes i and j are in the same community, and 0 otherwise (3, p. 224).

One popular implementation of modularity-based community detection is the Louvain algorithm, which is well-suited for weighted networks like the committee-state network (5). This algorithm is split into two phases. In the first, each node is assigned to its own community. The algorithm places i in the community of each of it's neighbors, and calculates the change in modularity for the network as a whole associated with each change of community. After it tests each neighboring node, the algorithm places i in the community that yields the largest increase in modularity for the network as a whole, and if no increase is possible, i remains in its original community. The second phase of the algorithm creates a second network from the communities that were created in the first phase, with each first-phase community representing a node of this new network. Intra-community edges are represented as self-edges for a given node, and inter-community edges from nodes in one community to a given node in another community are represented as multiedges between community-nodes. The first phase of the algorithm is once again applied to this new network until no further modularity increases occur.

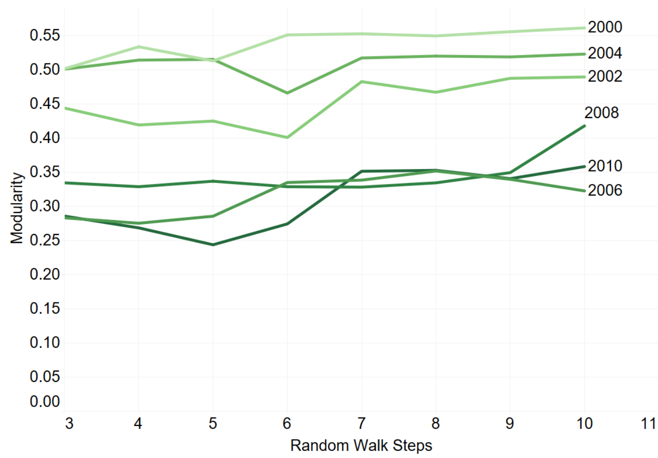

Fig. 1. The modularity of the committee-committee network for different step sizes of the Walktrap algorithm for community detection.

The Walktrap Algorithm. Since the committee-committee network is weighed and directed, we use the Pons and Latapy Walktrap algorithm (6) for community detection, which maximizes the modularity score and is efficient for complex networks. The hierarchial agglomerative method uses random walks to calculate the distances between nodes. Clusters are computed by using Markov chains on the random walks. If two nodes are in the same community, the probability to get to a third node located in the same community through a random walk should not be very different for i and j. The distance is constructed by summing these differences over all nodes, with a correction for the degree. The Walktrap algorithm uses the property that a random walker traversing the network will likely get stuck within a community. In Figure 1 we see that as we increase the step size of a random walker on our committee-committee network, the modularity of the network stabilizes.

Results

The Committee-Committee Network. We use weighted PageRank to identify the important nodes in this network. As shown in Figure 2, the two most important nodes, by a significant margin, are the Democratic Congressional Campaign Committee (DCCC)and the National Republican Congressional Committee (NRCC), the entities responsible for raising money to support House candidates from the Democratic and Republican parties, respectively. We also use Walktrap community detection to cluster communities. In Figure 1 we see that more recent years have a slightly lower modularity. Perhaps this is an indication that networks are more insular in recent years. We also see a slight variation in presidential election years, which needs to be confirmed with further analysis.

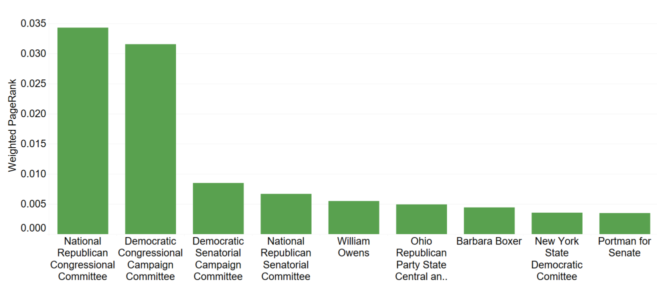

Fig. 2. Top nine weighted PageRank results on the committee-committee network in 2010.

States and Weighted PageRank. In order to identify the important nodes of the committee-state network, we compute the weighted PageRank centrality scores for each of the nodes. Unsurprisingly, as shown in Table 1, states tend to have relatively high weighted PageRank scores when compared to most of the committees in the network, as individual states usually receive a lot more money from committees than individual committees donate to candidates running in a particular state. Two exceptions in the 2010 senatorial election year are the National Rifle Association Political Victory Fund (NRA-PVF), which has the highest weighted PageRank score in the entire network across states and committees alike, and the American Crossroads PAC, which has a score between that of Washington and Illinois.

Table 1. 2010 Committee-State Network by Weighted PageRank

Name	Weighted PageRank
National Rifle Association PVF	0.037716369
Pennsylvania	0.030960336
Colorado	0.028719673
Nevada	0.025300206
Missouri	0.023812033
Ohio	0.022040072
Florida	0.021850998
California	0.021757912
Washington	0.021361687
American Crossroads	0.019941478
Illinois	0.017467654
Arkansas	0.016379646
Alaska	0.015171672
North Carolina	0.014467764
New York	0.014176559

When we rank the states by weighted PageRank in the 2010 network, we find an interesting correlation between a state's centrality to the network by weighted PageRank and the contentiousness of a state's Senate election, as shown in Figure 3 (7). In general, states with close election races tend to have higher PageRank scores since committees and interested parties will often pour money into vulnerable seats in an attempt to win those elections.

For example, the outcome of the 2010 Pennsylvania Senate race between Republican Pat Toomey and Democrat Joe Sestak was categorized as a toss-up by many analysts in the months leading up to the election (8). Toomey ultimately defeated Sestak 51-49. Pennsylvania has a PageRank score of .039, the highest score of any state in the network. In contrast, states like Arizona - where 2008 Republican presidential candidate and four-term senator John McCain handily beat a relatively unknown Democratic challenger 58.7-35.4 - have

PageRank scores an entire order of magnitude smaller than that of Pennsylvania.

We find similar results in the 2008 committee-state network, as seen in Figure 3.

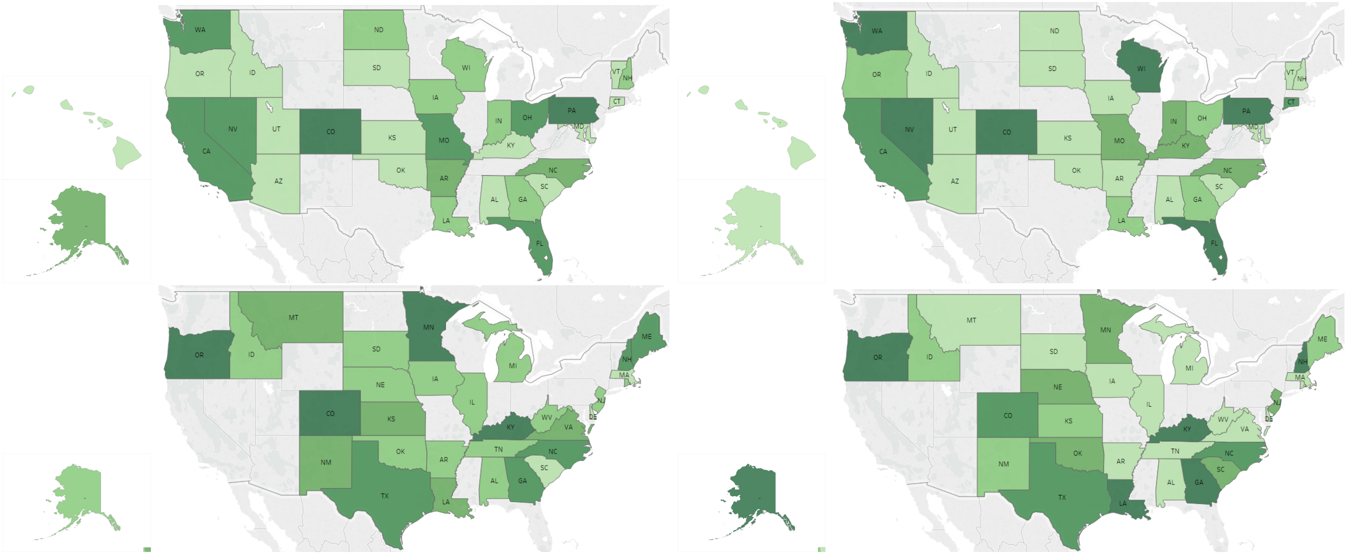

Fig. 3. Comparison of weighted PageRank of state nodes in the committee-state network (left) versus the winning margin of senatorial elections (right), in 2010 and 2008 (top and bottom respectively). Higher values of weighted PageRank are shaded darkly, as is lower winning margin. The winning margin is calculated using $|\mathrm{win}(\%) - \sum \mathrm{loss}(\%)|$, provided the candidate received at least 10% of the vote, and ranges from 1.02% to 25% in 2010, and from 1.25% to 30% in 2008.

Committees and Weighted PageRank. As seen in Table 1, the single most important committee in the 2010 network is the NRA-PVF. Other important nodes in this network include American Crossroads, a PAC that promotes conservative Republican candidates (9); the National Republican Senate Committee, the arm of the Republican party that raises money to support Republican senatorial candidates nationwide; the National Right to Life PAC, which supports candidates that are anti-abortion (10); the Senate Conservatives Fund, a PAC that promotes conservative Republican Senatorial candidates in an attempt to unseat more moderate Republicans (11), and Alaskans Standing Together, a super PAC that supported establishment Republican write-in senatorial candidate Lisa Murkowski, who narrowly won the three-way 2010 Alaskan Senate race as a write-in candidate against a Democratic candidate and the conservative Tea Party Republican candidate that had defeated the incumbent Murkowski in the primary election (12).

Although the Republicans failed to take a majority in the Senate in 2010, the efforts of these Republican-leaning PACs and others like them yielded the Republican Party a six-seat gain in the Senate, and also installed a conservative wing in both the US House and Senate that influenced, and continues to influence, American politics and legislation. The influence these PACs exerted on American politics in the 2010 election cycle is evident in the network structure.

As seen in Figure 4, in addition to the NRA-PVF and National Right to Life PACs, most of the other important committees in the 2008 network are in some way involved with the healthcare industry, including the American Medical Association; the Service Employee International Union, a labor union that largely represents employees of the healthcare industry (13); The International Association of Firefighters, a labor union representing firefighters and paramedics; the American Academy of Orthopedic Surgeons; and the American Hospital Association. In the 2008 presidential election, one of the fundamental promises of Democratic candidate Barack Obama's platform was universal healthcare for all Americans. From this network, we can see that in 2008, the national debate over healthcare filtered down to the Senate races and clearly impacted the structure of the network.

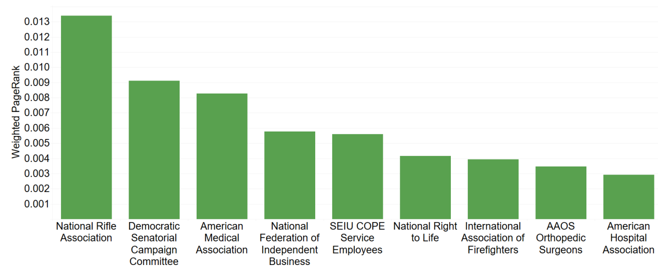

Fig. 4. Top nodes in the committee-state network by weighted PageRank in 2008. Only PACs with weighted PageRank > 0.002 are shown.

Modularity Community Detection and Political Party. We use undirected modularity to group the committee-state network nodes in communities. Using the Brain Connectivity Toolbox's `modularity_und` algorithm, we detect two communities in each of the committee-state networks. Since United States politics is largely a two-party system, with most candidates identifying as belonging to one or the other of the two major political parties, we suspect that this partition may have perhaps split the network along party lines.

In order to compare this to real-life phenomena, we decided to assign each state and each committee a label. Each state was assigned the party of the candidate who won the general senatorial election in that state; states that appeared in our network but did not have a Senate election in that cycle were labeled 'other'. We assigned the label 'Democrat'

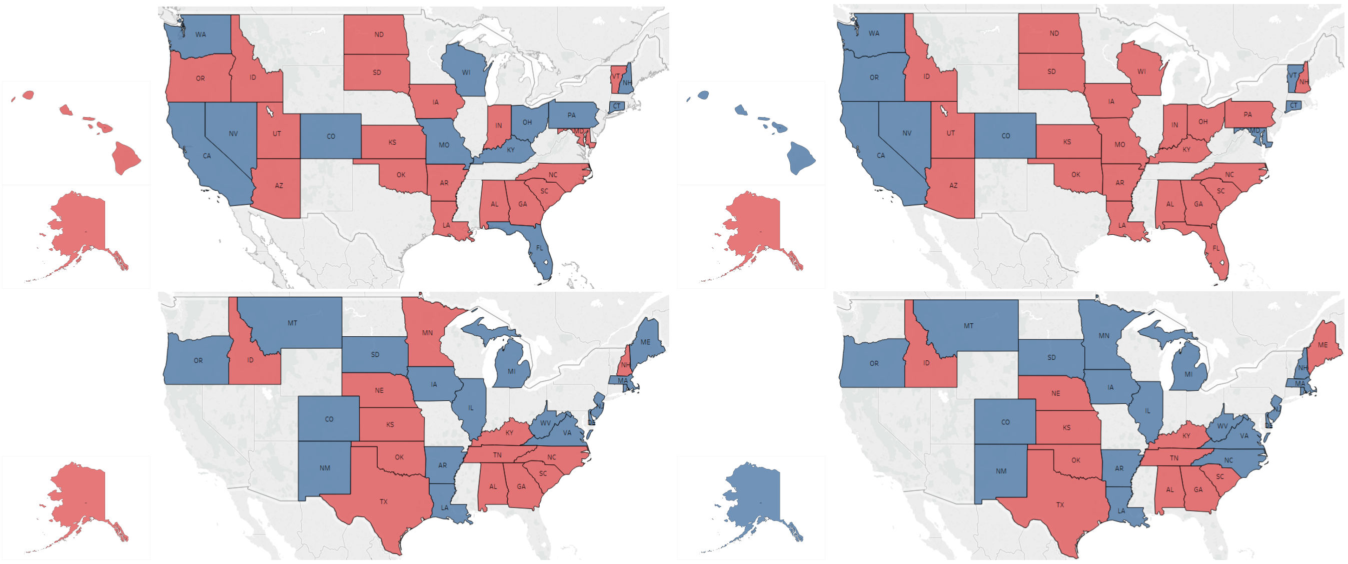

Fig. 5. Comparison of the modularity community detection (left) to winning party (right) in the 2010 and 2008, (top and bottom, respectively) senatorial elections.

to committees if they, in sum, donated more money to candidates affiliated with Democratic party than to candidates affiliated with the Republican party, and labeled committees 'Republican' if the opposite was true.

Figures 5 and 6 compare the partitions created by both community detection and party affiliation. We suspect the fact that 2008 was a presidential election year has some relationship with the fact that our results for 2008 were more accurate than our results for 2010; further research would try to analyze the difference between presidential election years and non-presidential election years to see if this is a general trend if either year is an outlier.

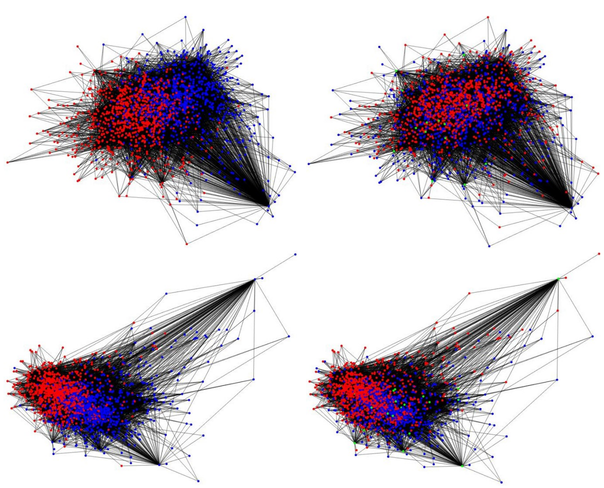

Fig. 6. Comparison of the modularity community detection algorithm on all nodes and party associated with that node (left and right, respectively). The top row is 2010, which has an accuracy of 53.7%. The bottom row is 2008, which has an accuracy of 65.2%. Green nodes in the figures on the left denote states that did not have senatorial elections that year.

Louvain Community Detection and the NRA. In both the 2008 and 2010 committee-state networks, the NRA-PVF had one of the highest weighted PageRank scores. As seen in Table 1, in 2010 the NRA-VCF is the single-most important node in the network by weighted PageRank, outscoring all states and all other committees, including even the National Republican Senate Committee and the Democratic Senatorial Campaign Committee. In 2008, the NRA is outscored by several states but is still the most important committee in the network by weighted PageRank.

Because the NRA features so prominently in each of these networks, we decide to analyze this node further by studying its community in each year. We apply the `community_louvain` algorithm from the Brain Connectivity Toolbox in MATLAB to both networks, and find that there are 398 nodes in the NRA community in 2010, and 433 in 2008; Figure 7 highlights the nodes in the NRA Community in each year. There are 72 committees that appeared in the NRA community in both years. We utilize the Center for Responsive Politics' OpenSecrets website to analyze four nodes chosen randomly from this list: the Virginian-Carolinian Peanut Ownership Membership PAC, which has clear geographic ties to the region in which the NRA tends to operate; the National Ocean Industries Association, an oil lobbying PAC that tends to donate to Republican candidates, like the NRA; the Conservative Victory Fund, a PAC that supports conservative Republican House and senatorial candidates; and the Association of Kentucky Fried Chicken Franchisees Inc PAC.

A fair portion of the committees that appear in the NRA community in both years seem to have geographical, political, or ideological ties to the same conservative wing of US politics that the NRA is known to support (14); communities detected by the Louvain algorithm seem to correlate to actual political factions.

Discussion

The committee-committee network gives expected results: the DCCC and NRCC have by far the highest weighted PageRank scores of the network. However, we note in Figure 1 that the modularity of each year varies considerably. This may be due to presidential elections, which occur every four years, or it may be due to a changing political landscape over time. Six

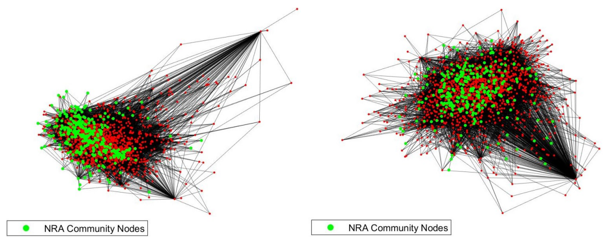

Fig. 7. The green nodes are nodes detected to be in the same community as the NRA by the Louvain algorithm in the committee-state networks for 2008 (left) and 2010 (right).

years is not enough data points to identify a trend, so we suggest that further research is necessary to form a conclusion.

The bipartite committee-state network yields far more interesting results. We believe that money is a proxy for political power in that candidates utilize donor money to influence the electorate in order to win elections. Our findings justify this belief. We find a correlation between donation activity and contentious elections as measured by weighted PageRank on states, a correlation between unweighted modularity community detection and the outcome of those elections, and a correlation between unweighted community detection and the associated party of a PAC. Finally, we identify influential PACs using weighted PageRank.

Further Research. Since the dataset we were given only recorded transactions through the 2010 elections, we would be interested in getting data from more recent election cycles in order to study more recent trends in American politics. Furthermore, we believe that the methodology used to create the bipartite committee-state network can be extended to study other sorts of elections, including presidential elections, primaries, and House races. In particular, we suspect that studying the committee-House district bipartite network of a particular state over a period of several election cycles could be used to analyze the effects of gerrymandering - the practice of manipulating district boundaries in order to benefit a particular political party in elections.

Additionally, we speculate that there is some element of preferential attachment that underlies the structure of these geographic-committee networks: candidates who win a seat are probably more likely to get more money in future election cycles than new or failed candidates. Geographic-committee networks could probably thus be studied as an application of the preferential attachment model.

ACKNOWLEDGMENTS. We would like to express our gratitude to Professor Mason Porter for his guidance and for providing the dataset, which was originally compiled by Andrew Waugh.

1. Waugh AS (2011) Community Structure in Federal Election Donation Networks, 1980-2008.
2. Telang N (2012) An Investigation of Federal Election Donation Networks from 1980 to 2010.
3. Newman M (2010) *Networks: An Introduction.* (Oxford University Press).
4. Haveliwala T, Kamvar S, Jeh G (2003) An analytical comparison of approaches to personalizing pagerank, (Stanford InfoLab), Technical Report 2003-35.
5. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10):P10008.
6. Pons P, Latapy M (2005) Computing communities in large networks using random walks in *International symposium on computer and information sciences.* (Springer), pp. 284–293.
7. Federal Election Commission (2010) Federal Elections 2010: Election Results for the US Senate and the US House of Representatives (online).
8. (2010) Pennsylvania senate profile, election 2010 (online).
9. Willis D (2011) Super PACs Report Strong Early Fund-Raising Numbers.
10. (2018) About the National Right to Life Political Action Committee (online).
11. Raju M, Palmer A (2013) Senate Conservatives Fund Roils GOP.
12. Wing N (2010) Super PAC 'Alaskans Standing Together' Using Unlimited Corporate Donations To Help Keep Murkowski In Office.
13. SEIU (2018) These fast facts will tell you how we're organized and what we do.
14. Enton H (2018) The nra used to be much more bipartisan. now it's mostly just a wing of the gop (online).