# Legislative Vote Prediction using Campaign Donations and Fuzzy Hierarchical Communities

1st Scott Wahl
*Gianforte School of Computing*
*Montana State University*
Bozeman, MT, USA
scott.wahl@msu.montana.edu

2nd John Sheppard
*Gianforte School of Computing*
*Montana State University*
Bozeman, MT, USA
john.sheppard@montana.edu

3rd Elizabeth Shanahan
*Department of Political Science*
*Montana State University*
Bozeman, MT, USA
shanahan@montana.edu

*Abstract*—An important aspect of social networks is the discovery and partitioning of the complex graphs into dense sub-networks referred to as communities. The goal of such partitioning is to find groups who have similar attributes or behaviors. In the realm of politics, it is possible to group individuals with similar political behavior by analyzing campaign finance records. In this paper we use fuzzy hierarchical spectral clustering to find communities with campaign finance networks. Multiple experiments were performed using varying edge weighting, number and type of communities, as well as analyzing multiple different years of voting data. The results show that using the full hierarchy of community assignments for legislators is highly predictive of voting behavior in the US House of Representatives and Senate.

*Index Terms*—social network, campaign finance, fuzzy clustering

## I. Introduction

Based on data from the National Institute on Money in Politics[1], contributions to candidates and committees in 2016 reached $5.5 billion for federal elections and $3.6 billion for state level politics. Roughly $1.7 billion were for federal House and Senate candidates. Approximately half of the federal legislative money came from only 1.6% of the donors. State level politics tell a similar tale where approximately 12% of donors gave 95% of the total money. An important question is how money may impact legislators and legislation. One apparent benefit to donating to a politician is that it provides access. While there is no clear evidence that political donations directly influence legislation, those who donate can more easily get their legislators to listen (1).

Based on an economic game theory model, contribution limits reduced aggregate contributions (2). A lobbyist targeting a politician who prefers the same policies does not need to spend as much as it is impossible for an opposing lobbyist to spend enough in opposition. Another model shows how legislators may elect to adopt policy choices based on how interest groups donate (3). This money can have an impact even without the expectation of *quid pro quo*. Akey notes most large campaign donations come from areas where political opinion does not necessarily reflect a majority of the population (4). Changes in legislative committees can also impact donations as groups focus on short term access (5).

We focus our analysis of campaign contributions through social networks. Social networks have shown up in other fields such as genetics (6), neuroscience (7), and many more. Some prior work on social-political networks looks at social interaction and its effect on political participation (8; 9). Other work showed donations follow a preferential attachment network, showing how shares of a donor pool can affect the probability of gaining more donors (10). Gubernatorial data showed self financing a campaign is not as effective as getting external funds when tested against the results of elections (11). Prior work used community assignments and identifying features of the edges of the network to provide context to the communities (12). Association rule mining on the transactions provided frequent patterns and rules within each community, showing how underlying transactions differed between communities.

Bonica used campaign finance records to create an ideological estimate for both donors as well as candidates (13). Common-space campaign finance score (CFscore) has the advantage of applying to both types of entities, where prior research focused solely on legislators or recipients. CFscores were calculated using correspondence analysis over the set of donations to recipients. The first step calculated the federal ideal points. Those values were used as bridges to state data to inform the ideological ideal points for state level candidates and recipients. Results yielded similar ideological measures to prior work and from which accurate voting records for legislators could be predicted.

The goal of finding communities is to locate shared areas of possible influence. Hierarchical communities can assist in that they highlight differences between communities at multiple levels. In our work, we use a hierarchical fuzzy spectral clustering algorithm for social networks to find communities within contribution networks. Instead of correspondence analysis, we use the entire donation network and find community structure. A concrete result can be obtained by using communities to predict behavior. Within this paper, the communities are used directly to predict voting behavior of legislators in the House of Representatives and Senate of the United States government. Hierarchical communities were shown to be more robust than non-hierarchical clustering, while providing additional context to the communities over eigenvectors.

---

[1] Based on numbers provided by www.followthemoney.org

## II. BACKGROUND

Community discovery is a much researched topic in social networks. Crisp communities comprised much of the early work wherein groups of nodes only belong to a single community (14; 15; 16; 17). A limitation of early clustering techniques is that they did not handle non-convex clusters well. $K$-means clustering, as an example, is biased towards spherical clusters. Spectral clustering is one alternative since it is relatively easy to implement and it can find non-convex clusters (18; 19).

Early approaches limited nodes to belong to a single community. However, people can belong to more than one community at a time. Some methods also do not account for sub-groups within a community. Such sub-groups can consist of smaller groups of individuals within a larger community forming a hierarchy of communities. There are more recent approaches that attempt to improve on the older algorithms by allowing fuzzy clusters as well as creating a hierarchical structure for the communities (20; 21; 22; 23; 24; 25). The work presented in this paper may be of benefit to socio-political research. The fuzzy community values found at each hierarchy may help determine if there are specific groups who are having a strong impact.

### A. Spectral Clustering

Spectral graph partitioning methods already existed that relied on repeatedly cutting a network into smaller partitions (16). The second eigenvector of a graph's Laplacian could be used to find an approximation of an optimal partition based on the sign of the entries in the vector. Each new partition would then be divided by isolating its nodes and performing the split again. Spectral clustering instead uses the top $k$ eigenvectors of the Laplacian of an affinity matrix instead of doing an iterative split. This calculation is a common complaint of users of spectral clustering as it can be costly for large datasets. Approximation algorithms can help as not all eigenvectors are necessary for the calculation (26).

Consider points within a set of data $S = \{s_1, \ldots, s_n\} \in \mathbb{R}^l$. Affinity matrix $\mathbf{A}$ is created by

$$a_{ij} = \exp\left(-\|s_i - s_j\|^2 / 2\sigma^2\right)$$

where $\sigma$ is a parameter to control how quickly affinity drops off as the distance increases. In the case of networks, the adjacency matrix can be used as the affinity matrix. Using the adjacency matrix, spectral clustering proceeds by performing the following steps.

1) Define $\mathbf{D}$ as a diagonal matrix where $d_{ii} = \sum_j a_{ij}$ and $d_{ij} = 0$ for all $i \neq j$.
2) Construct Laplacian $\mathbf{L} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$.
3) Find the $k$ largest eigenvectors of $\mathbf{L}$ as $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$.
4) Form matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]$ by stacking eigenvectors in columns.
5) Form matrix $\mathbf{Y}$ by normalizing each row in $X$ to have unit length.
6) Cluster $\mathbf{Y}$ using a clustering algorithm, typically $K$-means.

A common method for doing the final clustering is to use $K$-means (27).

### B. Fuzzy c-Means

Crisp clustering is not always ideal, and fuzzy clusters can be an improvement. A common method is to use fuzzy $c$-means clustering (FCM) (28; 29). Like $K$-means, FCM begins by identifying the number of clusters. Initial coefficients of belonging to each cluster are assigned randomly to each data point $w_i, j$. Like $K$-means, FCM then proceeds by repeating two steps. The centroid for each cluster is calculated using $c_k = \sum_s w_k(s)^m s / \sum_s w_k(s)^m$ where $m$ determines the amount of fuzzy overlap between the clusters. A higher value results in fuzzier clusters. The assignments are then updated with the new centroids using

$$w_{ij} = \frac{1}{\sum_k \left(\frac{\|s_i - c_j\|^2}{\|s_i - c_k\|^2}\right)^{\frac{2}{m-1}}}$$

until the algorithm converges or the maximum number of iterations is reached. The resulting weights and centroids define the discovered fuzzy clusters.

### C. Random Forests

For the purposes of this research, once the data is separated into clusters, or communities, that information is used to predict behavior of the members of the network. By combining the cluster data with voting history, the goal is to predict behavior. Decision trees perform classification by generating a tree from repeatedly splitting training data based on testing features (30). The appeal of most decision trees are they are simple to create and also simple to interpret. However, one issue decision trees can have is that they overfit the data.

Random forests were created as an ensemble method to avoid overfitting and increase generalization (31), (32) using use an ensemble of decision trees that each vote on the predicted class. Each of the decision trees uses a bootstrap aggregated (bagged) sample of the dataset. The bagged method samples with replacement from the original data to create an equal sized dataset for training, resulting in approximately $\frac{1}{3}$ of the data being left out for each tree. Each tree randomly selects a subset of the features for splitting at each point. The criteria for choosing a feature and value to partition the data can vary greatly. Common techniques utilizes information theory to determine partitions. Iterative Dichotomizer 3 (ID3) is one well known decision tree algorithm which uses entropy and information gain (33). Consider a set of classes $c \in C$. For these classes, the entropy of any partition of the data $D$ is defined as

$$H(D) = \sum_{c \in C} -p(c) \log_2 p(c)$$

where $p(c)$ is a proportion of the data with class $c$, or

$$p(c) = \frac{|d_c \in D|}{|D|}.$$

ID3 works by selecting the test $t$ that maximizes the information gained when partitioning the data into sets $D_t$ using

$$H\left(D\right) - \sum_{D_t} p\left(D_t\right) H\left(D_t\right).$$

Since the entropy of $D$ is fixed for dataset $D$, this is equivalent to minimizing the entropy of partitions $D_t$. Beginning with the root of the tree, the attribute maximizing information gain is selected to partition the data. The same procedure is applied to the resulting partitions until a stopping criteria is met.

Another method for determining how to split data at each point is Gini impurity, which is used in the Classification and Regression Tree (CART) algorithm (30). The principle behind this metric is to minimize the impurity at each split. The impurity is defined by the probability of a data point being associated with the wrong class when randomly assigning a class to each member in the partition based on the class distribution (34). If the partition contains only one class, then the impurity would be zero as the class distribution would allow only that class. The measure is worst when there is an even split of classes in the partition. More formally, the probability of misclassification is defined as

$$\phi = 1 - \sum_{j=1}^{n} \left(p\left(c_j|t\right)\right)^2$$

where $p\left(c_j \mid t\right)$ is the probability of assigning the incorrect class based on the class distribution in partition $t$.

### D. DW-NOMINATE

The classification task of predicting votes requires a representation of socio-economic factors of legislation. Prior work analyzes ideological estimates of legislators and the bills upon which they vote. A widely known tool is DW-NOMINATE: dynamic, weighted, nominal three-step estimation (35), (36). DW-NOMINATE is built upon the idea of a random utility model where legislator $i$'s utility for an outcome (Yea) on a bill $j$ is given by $U_{ij}^y = u_{ij}^y + \varepsilon_{ij}^y$ where $u_{ij}^y$ is a utility function and $\varepsilon_{ij}^y$ is a random error (37). Early DW-NOMINATE work was based on a normal distribution utility function. The utility of a legislator's choice of voting Yea or Nay is centered around an estimated ideal point. The more distant an option is to a legislator's ideology, the less utility is gained by voting for that option. As examples, the authors' refer to the concepts of alienation and indifference. Alienation represents where the set of choices are far removed from the ideal point but on the same side of the ideological space. Indifference is where the choices are far away but on either side of the space, as in a moderate politician faced with voting for two extremes, one on each side of the political spectrum.

Encapsulating this information requires determining the ideal points of legislation and legislators. DW-NOMINATE does so by maximizing a likelihood function based around the probability of a legislator's choices regarding bills. The utility of a legislator $i$ voting Yea in the $k$th dimension is defined by

$$u_{ij}^y = \beta \exp\left(-\frac{1}{2}\sum_{k=1}^{s} w_k (d_{ijk}^y)^2\right)$$

where $(d_{ijk}^y)^2$ is the squared distance of legislator $i$ to a Yea outcome in dimension $k$, $w_k$ are salience weights, and $\beta$ is an adjustment for overall noise, which is proportional to the variance of the error distribution. From this, the probability of voting Yea in the normal utility model is based upon the relative utilities of voting Yea or Nay. Using those values, this can be written as

$$P_{ij}^y = P\left(u_{ij}^y > u_{ij}^n\right) = P\left(\varepsilon_{ij}^n - \varepsilon_{ij}^y < u_{ij}^y - u_{ij}^n\right)$$
$$P_{ij}^y = \Phi\left[u_{ij}^y - u_{ij}^n\right]$$

In the normal model, this is given by

$$P_{ij}^y = \Phi\left[\beta\left\{\hat{u}_{ij}^y - \hat{u}_{ij}^n\right\}\right].$$

where

$$\hat{u}_{ij}^\tau = \exp\left(-\frac{1}{2}\sum_{k=1}^{s} w_k (d_{ijk}^\tau)^2\right)$$

Using these equations, the model is estimated by maximizing the likelihood as defined by

$$L = \prod_{i=1}^{p}\prod_{j=1}^{q}\prod_{\tau\in\{y,n\}} (P_{ij}^\tau)^{C_{ij}^\tau}$$

where $\tau$ is an index for choices Yea and Nay, $P_{ij}^\tau$ is the probability of voting $\tau$ as defined above, and $C_{ij}^\tau = 1$ if the actual choice was $\tau$ and zero otherwise. In learning the model, Rosenthal and Poole estimated a single parameter at a time, holding the others fixed. The authors claim two dimensions worked well in practice.

The result of the model gives estimates of ideal points for both legislator and legislation in two dimensions. In practice, the first of the two dimensions is considered to correspond to a liberal and conservative economic spectrum. The second dimension corresponds to social issues. In the experiments below, the estimates of legislation are used in order to show the expressive power of the communities discovered in campaign finance networks. The discovered communities perform well in predicting votes in those two dimensions without explicitly using the legislators' ideal points as determined in the model.

### III. DATA & APPROACH

The donation data is provided by Bonica and Stanford's Social Science Data Collection (38; 13). Contributions are given in two year election cycles spanning 1979 through 2012.[2] While accompanying donor data is sparse, recipient data includes information such as the state, district, seat, party, and other pertinent information regarding a candidate running for office. Unique identifiers were assigned to the candidates and donors across states and years. Candidates are attached via a separate identifier to legislative voting data provided by Voteview (39). This legislative data provides two estimated DW-NOMINATE measures of the socioeconomic space of legislation (36). We use these socioeconomic values, along with the community assignments, to generate a classifier to determine voting behavior in the legislature.

---

[2]The dataset has been updated to include more recent years after originally downloading the data; however, we continue to use the 1979–2012 subset.

## A. Data Preprocessing

The Stanford dataset uses two separate identifiers for candidates and donors. Using them separately would cause duplication of nodes in the network as an individual in the network may both donate and receive money. The candidate (or recipient) data includes the donor identifier so that it is possible to bridge the two nodes and create a single unique identifier. This new identifier is used to populate the nodes and edges in the induced social network.

Some restrictions are applied to types of edges used to create the networks. Loans and similar records are removed from consideration as they do not necessarily indicate support of a candidate. The list of transaction types kept are as follows: 10, 11, 12, 13, 15, 15C, 15E, 15F, 15I, 15J, 15L, 15PD, 15S, 15T, 15Z, 18G, 18H, 18J, 18K, 18S, and 18U. An initial network is created out of all donations where edge $a_{ij} = \sum amt$ is the sum of all transactions between entities $i$ and $j$. Following the procedure outlined in prior work using this data, any node $i$ with degree $d_i = 1$ is removed from the network. The largest connected component of the remaining network is then used as the network of contributions. Using all years from 1979 to 2012, the resulting network consists of over 5.26 million nodes and 29.85 million edges.

Each node in the network is attached to the voting data by use of the ICPSR id (Inter-university Consortium for Political and Social Research). The voting data includes 9.2 million observations of 'Yea' or 'Nay' votes on various bills from the years 1979 through 2012. Each bill also contains two features that are the spatial estimates of the socioeconomic measure: DW-NOMINATE midpoints $mid1$ and $mid2$. These are the ideological estimates generated from the voting behavior of legislatures as given by Voteview. The final voting dataset is created by combining the discovered communities assignments, the DW-NOMINATE ideological scores, and the predictive class variable of 'Yea' or 'Nay'.

## B. Network Community Assignment

The entire process of performing the prediction requires two primary steps. First is hierarchical fuzzy spectral clustering on the campaign contributions network in order to find the community assignments for the legislators (40). These values are then attached to the corresponding entries in roll call data.

Finding fuzzy clusters uses prior spectral clustering work in (18; 41; 40). Clustering is performed on the adjacency matrix $\mathbf{A}$ corresponding to data matrix $\mathbf{D}$ for cluster numbers $K = 2, 3, \ldots, k-1, k$, as shown in Algorithm 1. Using the adjacency matrix $\mathbf{A}$ and the number of communities desired $k$, find the Laplacian $\mathbf{L}$. The top $k$ eigenvalues of the Laplacian are calculated, where $k = c$, the number of communities to be found. In practice, the eigenvectors are calculated once using the largest number of communities desired $\max(k)$. These $k$ eigenvectors then become the columns for matrix $\mathbf{X}$. After normalizing each row to form a new data point, fuzzy c-means clustering (FCM) is used on the rows of $\mathbf{X}$ to obtain $\mathbf{U}$. This $n \times k$ matrix $\mathbf{U}$ contains the community assignments where $\mathbf{U}_{i,j}$ is the assignment of $i$ to community $j$.

---

**Algorithm 1** Fuzzy Spectral Clustering

1: **function** FSC($\mathbf{A}, k$)
2: $\quad \mathbf{D} \leftarrow \{d_i = \sum_1^n a_{ij}\}$
3: $\quad \mathbf{L} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$
4: $\quad \mathbf{V} = eigenvectors(\mathbf{L}, k)$
5: $\quad$ **for all** $\mathbf{v}_i \in \mathbf{V}$ **do**
6: $\quad\quad$ **for all** $v_{ij} \in \mathbf{v}_i$ **do**
7: $\quad\quad\quad x_{ij} = v_{ij}$
8: $\quad$ **for all** rows $\mathbf{x}_i \in \mathbf{X}$ **do**
9: $\quad\quad \mathbf{x}_i = \mathbf{x}_i / \|\mathbf{x}_i\|$
10: $\quad \mathbf{U} = \text{FCM}(\mathbf{X}, k)$

---

**Algorithm 2** Hierarchical Generation

1: **function** HFSC($\mathbf{A}, k$)
2: $\quad$ **for** $i = 2$ to $k$ **do**
3: $\quad\quad \mathbf{C}_i = \text{FSC}(\mathbf{A}, i)$
4: $\quad$ **for all** $C_{i,m} \in \mathbf{C}_i : i > 2$ **do**
5: $\quad\quad$ **for all** $C_{i-1,n} \in \mathbf{C}_{i-1}$ **do**
6: $\quad\quad\quad s_{i,m,n} = sim(\mathbf{C}_{i,m}, \mathbf{C}_{i-1,n})$

---

The process is repeated with a varying $k$ corresponding to the number of clusters in each hierarchical level. Each level is connected to its previous by calculating the fuzzy Jaccard similarity measure of two communities $\mathbf{C}_i$ and $\mathbf{C}_j$ given by

$$sim(\mathbf{C}_i, \mathbf{C}_j) = \sum_{e \in \mathbf{C}_i \cup \mathbf{C}_j} \frac{\min(c_i^e, c_j^e)}{\max(c_i^e, c_j^e)}$$

where $c_i^e$ the community assignment of element $e$ in community $\mathbf{C}_i$, or 0 if $e$ is not in the set. In Algorithm 2, $s_{i,m,n}$ is the similarity between child community $m$ at level $i$ and a possible parent community $n$ at $i-1$. The above procedure gives fuzzy cluster values for varying $k$ which can be used in additional analysis.

At the top level of a hierarchy is the entire network. Level two splits the network into two overlapping communities. At each increasing level one more community is added. There are two important sets of community assignment values within the tree. $T_i$ includes all assignments up to level $i$ in the tree. $L_i$ is defined only by those assignments on level $i$ itself. These communities are used to predict new behavior of the actors within the network, specifically legislative voting. Classification relies on additional data provided by Voteview in the form of socio-economic estimates of bills and voting records of United States legislators (36). The resulting dataset contains a record for each recorded vote in the legislature. Each record also contains the fuzzy assignment values for that legislator and the two DW-NOMINATE dimensions for the bill being voted upon. Each record also has the class to be predicted: a 'Yea' or 'Nay' vote.

## IV. EXPERIMENTAL DESIGN

The experiments were performed with Python implementations of the eigen-decomposition, fuzzy $c$-means, and random

forests. An approximate eigen-solver was used to find the eigenvectors of the network to obtain only the vectors and values of interest. A value of $m = 10$ for fuzzy $c$-means was used to find the clusters. This was set to ensure highly fuzzy communities instead of closer to crisp clusters. In this experiment, communities are found for $k \in [2, 12]$. Early experiments using random forests showed the community features were not equally expressive when trying to predict votes and limiting the number of features available during a split to $\sqrt{(n_{features})}$ resulted in poor performance. Thus, the algorithms were allowed to use any feature when determining how to perform a split. The random forests use the Gini impurity metric when performing splits. During training for each random forest, the class weights were inversely proportional to the number of instances of that class. The final parameter to note is that the random forest used 50 trees.

Three different primary experiments were performed to investigate properties of the data: different weighting of the network edges, using all data combined and individual years separately, and varying community types and numbers. All experiments were evaluated using 10-fold cross-validation.

### A. Edge Weighting

Three different edge weighting schemes were tested. The first used 1 to indicate an edge between two entities in the network regardless of the donation amounts. The second used base-10 logarithm of total contributions between two entities as $a_{ij} = \log_{10}(amount)$, with a minimum value of 1. Finally, the raw amount $a_{ij} = amount$ of total donations between two entities was used, again with a minimum value of 1. By analyzing these measures, we can infer the impact of connectivity versus donation on predicting voting behavior.

### B. All vs Yearly Data

Along with the edge weighting, a test is performed to compare the results of testing the entire dataset against individual cycles. The full network utilized the entirety of the data from all available years. Separate networks were also created for each individual 2-year cycle from 1979 through 2012 as provided in the data. This created 17 different networks. The sizes of the networks greatly increase over time due to the increase in the amount of money in politics. After performing the same processing step for each cycle, the number of nodes ranged from approximately 34 thousand in 1980 to 1.4 million in 2012. Similarly, the number of edges ranged from 169 thousand in 1980 to 6.2 million in 2012.

### C. Communities

Three different applications of the spectral clustering were applied. The first of these methods used a single level of the hierarchical tree (shown in figures as FCM-L). This represents the process of using the best single performing clustering as would be typical in many techniques. The second included all parents of the level in question, effectively pruning anything further down the tree (shown in figures as FCM-A). Finally, the eigenvectors themselves were used directly in performing

prediction (shown in figures as EV). This was chosen as a baseline to show any loss in predictive power.

## V. RESULTS AND DISCUSSION

Fig. 1 shows the results of predicting votes over the entire set of data. For just the adjacency matrix, performance began by middling but improved until about $k = 10$ communities. At this point, the performance of the single level (FCM-L) began to drop. However, the eigenvector (EV) and full tree (FCM-T) held steady due to not losing the information provided by fewer communities.

Raw contributions show almost all of the predictive power of this schema is held within the first few levels of communities. The large donations appear to dominate the community analysis at that stage, and additional communities do not provide much in additional predictive ability. This can be shown in that after $k = 7$, the performance of the single level drops rapidly. Unlike the other two, scaled contributions performed poorly at first for this particular set of data. As shown by EV, this was not due to the performance of FCM, but integral to the structure of the network. As the tree grew the performance improved and matched the other two weightings.

Using votes by year showed additional evidence of polarization within the legislature. Beginning with 1980, the accuracy of prediction was low. Accuracy increased over time, hitting a peak in 2010 with votes being predicted at roughly 94% for each of the weighting methods. This fell to $\sim 90\%$ in 2012, which was more in line with 2008. The performance of the differing number of communities also flattened over time, implying fewer communities needed to define splits.

### A. Communities

Increasing the number of communities improved classification accuracy of the vector (EV) and whole tree models (FCM-T), as shown in Fig. 1. Both of these showed consistent performance where more features helped with only small deviations. This was not true when using single levels (FCM-L). In two weighting schemes, the accuracy increased with the number of communities until a certain point at which it sharply fell. It is expected that as the number of communities grow in the scaled weighting, a similar pattern will hold and the accuracy will begin to drop.

As a more extreme example, consider the results from the experiments using only data in 1998 from Fig. 2. While the performance of FCM-L was consistent in the case of raw contributions, the performance in both the adjacency matrix and the scaled values were highly erratic. For certain numbers of communities the accuracy improved to nearly match that of EV and FCM-T. However, since both EV and FCM-T have access to the entirety of the data up to that point, those two methods are far more robust to those changes. Due to these issues, the selection of $k$ for FCM-L is critically important when using discovered communities to inform predictions.

Another notable difference between FCM-L and FCM-T can be seen in the behavior in the low number of communities.
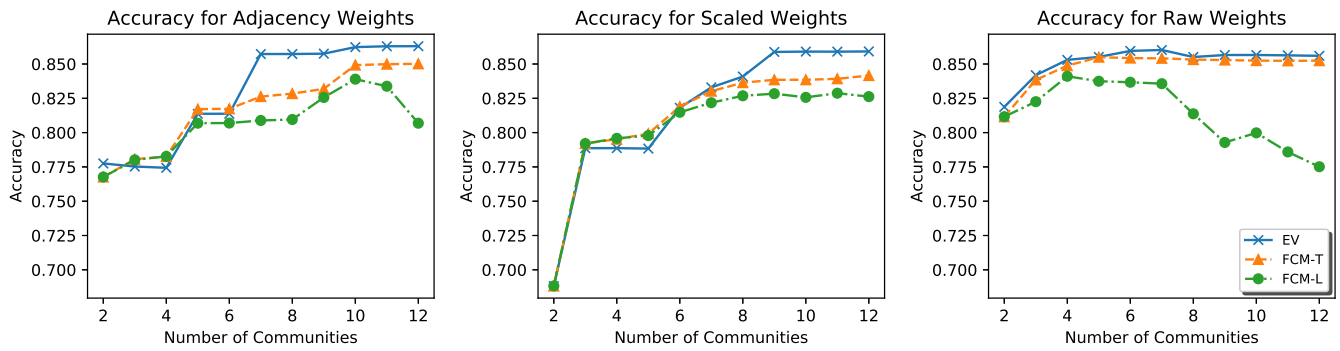
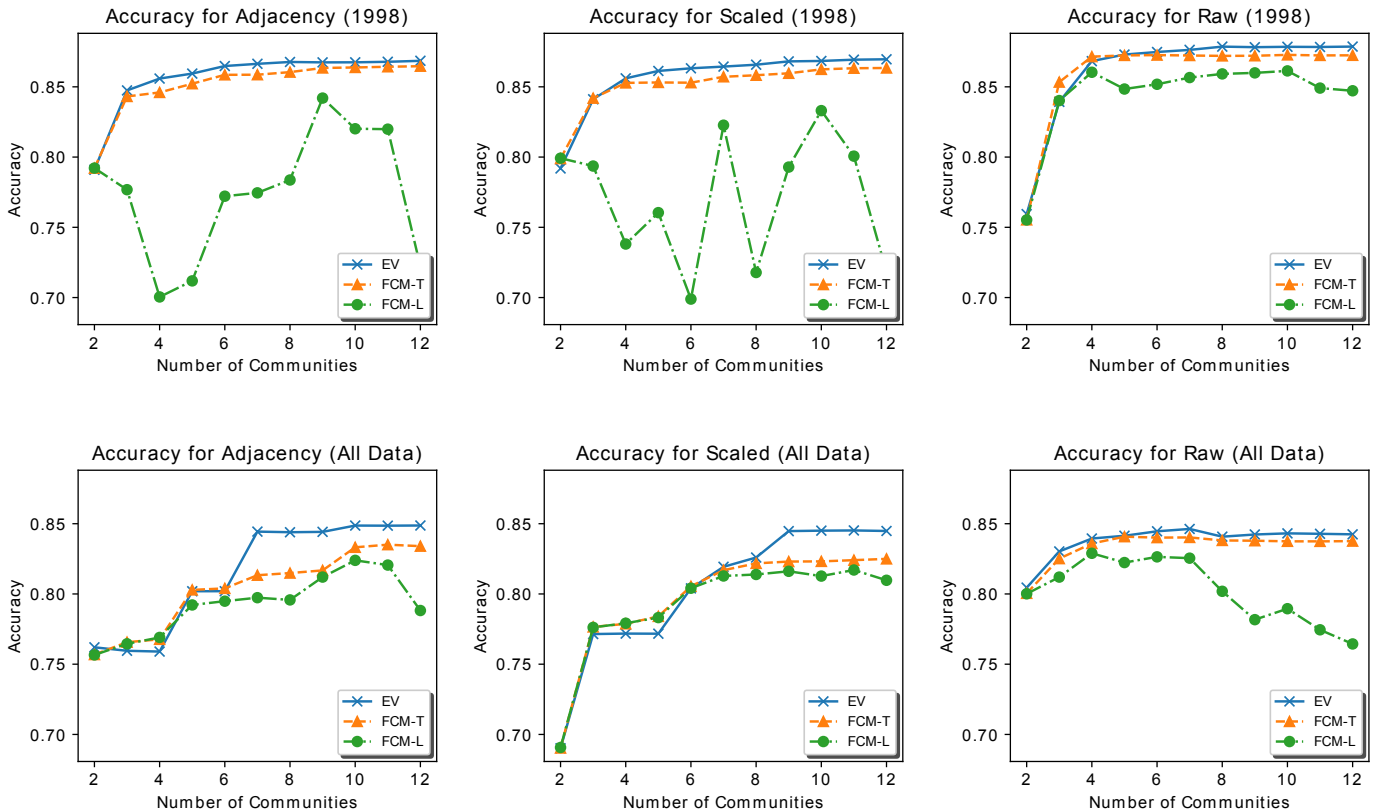Figure 1. Experimental results for differing communities using all years of data.



Figure 2. Experimental results for differing communities using only 1998 data vs all data.

Moving down the hierarchy, the performance of FCM-L decreased for $k = 3$ and $k = 4$ for both the adjacency and scaled weights. However, despite using those same values from the hierarchy of community assignments, the performance of FCM-T increased during that period. This shows that FCM-T is more robust to individually poor community assignments and that the combination of data from the levels is more useful.

In general, the best performance was obtained using the eigenvectors directly. This was not always the case, however, especially for small numbers $k$ of communities. FCM-T was typically comparable in performance and did not suffer the same issues as FCM-L. EV was more easily computable due to fewer features required. With EV, $k$ indicates the number

of eigenvectors to use. For FCM-T, the entire tree structure is used up to the level with $k$ communities. This results in $k(k-1)/2$ features. This increase in the number of features, along with the large amount of voting data, resulted in FCM-T taking longer to compute models than EV. Interpretability is an important factor in the selection and use of models, and FCM-T has the benefit of being easier to interpret. Two vectors representing two different entities does not have meaning out of the larger context. However, two lists of community assignments for those same entities can be more easily interpreted for anyone using these data or methods.

## B. Edge Weighting

Raw values outperformed the other two measures. Out of the possible comparisons, raw weighting was better statistically than the adjacency matrix 53.1% of the time and better than scaled 50.8% of the time. Adjacency was better than raw 34.3% of the time and scaled was better than raw 36.9% of the time. From these results, The best communities for predicting votes were more often those generated by large, raw dollar amounts. This impact is small, however. Even though the different weights had an impact in any individual test, overall they were quite similar in their performance.

## C. All vs Yearly Data

A little more information was necessary in order to compare the models learned from all of the data to those trained on two year cycles. When calculating the performance of the all-data models, the accuracy of each individual year was calculated as well as that of the entire dataset. This shows the results of prediction in a specific cycle that can then be compared with the model trained only on that cycle. The notable result from this is that using all of the data at once did not typically do better than using data in a specific year. Instead, out of the different comparisons, the models built from individual cycle data were statistically better in 91.7% cases. Fig. 2 shows this trend from examples of these experiments.

The results of the experiments show another issue related to polarization of the legislature. This can be seen from two different aspects of the results from individual cycles. First, the overall accuracy of prediction increased over time. This hit a peak in year 2010, though it was still quite high in the following 2012 cycle. Additionally, the impact of the number of communities is less pronounced in those later years. The result is nearly a flat line for all the experiments in later years. This is true even for FCM-L, which mostly stops behaving as in prior years where there are peaks and valleys in performance based on the number of communities.

## VI. CONCLUSION

As shown, fuzzy hierarchical spectral clustering is effective in predicting voting behavior of legislators based on campaign finance records. The results show using the full hierarchy improves classification performance relative to a single level in the tree while also showing less volatility in predictive power. The full hierarchy also performs similarly to using eigenvectors directly while providing more context for those who use the data. While different weightings of the edges had an impact on the results of the predictions, overall the best performing weighting was based on raw dollar amounts. However, both the adjacency matrix and scaled weights were promising. The results also again highlight the growing issue of partisanship within the legislature, where communities become more accurate in more recent years.

Future work includes improving the temporal analysis and creation of communities as they change over time–both in the communities and how those changing communities may indicate shifts in ideology over time. Additional analysis could better examine groups who differ in their type of donations beyond ideology. These datasets could also be augmented with additional lobbying or expenditure data to improve the definition of the network. We are also looking to use use deep neural networks to analyze the social network to allow for improvements to the speed of the experiments by making use of better approximations as well as GPU processing.

## REFERENCES

[1] J. L. Kalla and D. E. Broockman, "Campaign contributions facilitate access to congressional officials: A randomized field experiment," *American Journal of Political Science*, vol. 60, no. 3, pp. 545–558, 2016.

[2] I. Pastine and T. Pastine, "Politician preferences, law-abiding lobbyists and caps on political contributions," *Public Choice*, vol. 145, no. 1, pp. 81–101, Oct 2010.

[3] J. Fox and L. Rothenberg, "Influence without bribes: A noncontracting model of campaign giving and policy-making," *Political Analysis*, vol. 19, no. 3, pp. 325–341, 2011.

[4] P. Akey, "Valuing changes in political networks: Evidence from campaign contributions to close congressional elections," *Review of Financial Studies*, vol. 28, no. 11, pp. 3188–3223, 2015.

[5] E. N. Powell and J. Grimmer, "Money in exile: Campaign contributions and committee access," *The Journal of Politics*, vol. 78, no. 4, pp. 974–988, 2016.

[6] B. Zhang and S. Horvath, "A general framework for weighted gene Co-Expression network analysis," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, pp. Article 17+, 2005.

[7] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen, "Functional network organization of the human brain," *Neuron*, vol. 72, no. 4, pp. 665–678, 2011.

[8] J. H. Aldrich, R. K. Gibson, M. Cantijoch, and T. Konitzer, "Getting out the vote in the social media era: Are digital tools changing the extent, nature and impact of party contacting in elections?" *Party Politics*, vol. 22, no. 2, pp. 165–178, 2016.

[9] R. La Due Lake and R. Huckfeldt, "Social capital, social networks, and political participation," *Political Psychology*, vol. 19, no. 3, pp. 567–584, 1998.

[10] G. Vonnahme, "A preferential attachment model of campaign contributions in state legislative elections," *Public Choice*, vol. 159, no. 1, pp. 235–249, Apr 2014.

[11] A. R. Brown, "Does money buy votes? The case of self-financed gubernatorial candidates, 1998–2008," *Political Behavior*, vol. 35, no. 1, pp. 21–41, Mar 2013.

[12] S. Wahl and J. Sheppard, "Association rule mining in fuzzy political donor communities," in *Proceedings of the 14th International Conference on Machine Learning and Data Mining in Pattern Recognition*, 2018, pp. 231–245.

[13] A. Bonica, "Mapping the ideological marketplace," *American Journal of Political Science*, vol. 58, no. 2, pp. 367–386, 2014.

[14] V. Blondel, J. Guillaume, R. Lambiotte, and E. Mech, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10, p. P10008, 2008.

[15] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[16] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review*, vol. E 69, no. 026113, p. 026113, 2004.

[17] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *J. of Graph Alg. and App. bf*, vol. 10, pp. 284–293, 2004.

[18] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*. MIT Press, 2001, pp. 849–856.

[19] A. Pothen, H. Simon, and K. Liou, "Partitioning sparse matrices with eigenvectors of graphs," *SIAM Journal on Matrix Analysis and Applications*, vol. 11, no. 3, pp. 430–452, 1990.

[20] S. Bandyopadhyay, "Automatic determination of the number of fuzzy clusters using simulated annealing with variable representation," in *Foundations of Intelligent Systems*, ser. Lecture Notes in Computer Science, M.-S. Hacid, N. Murray, Z. Ras, and S. Tsumoto, Eds. Springer Berlin Heidelberg, 2005, vol. 3488, pp. 594–602.

[21] A. Devillez, P. Billaudel, and G. V. Lecolier, "A fuzzy hybrid hierarchical clustering method with a new criterion able to find the optimal partition," *Fuzzy Sets and Systems*, vol. 128, no. 3, pp. 323 – 338, 2002.

[22] J. Liu, "Fuzzy modularity and fuzzy community structure in networks," *The European Physical Journal B*, vol. 77, no. 4, pp. 547–557, 2010.

[23] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–8, Jun. 2005.

[24] V. Torra, "Fuzzy c-means for fuzzy hierarchical clustering," in *Proceedings of the 14th IEEE International Conference on Fuzzy Systems (FUZZ)*, May 2005, pp. 646–651.

[25] J. Xie, B. Szymanski, and X. Liu, "Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process," in *Proceedings of the 11th International Conference on Data Mining Workshops (ICDMW)*, Dec 2011, pp. 344–349.

[26] R. Lehoucq and D. Sorensen, "Deflation techniques for an implicitly restarted arnoldi iteration," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, no. 4, pp. 789–821, 1996.

[27] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, 1967, pp. 281–297.

[28] J. C. Bezdek, "Cluster validity with fuzzy sets," *Journal of Cybernetics*, vol. 3, no. 3, pp. 58–73, 1973.

[29] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[30] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Wadsworth, 1984.

[31] T. K. Ho, "Random decision forests," in *Proceedings of the Third International Conference on Document Analysis and Recognition*, ser. ICDAR '95, vol. 1. Washington, DC, USA: IEEE Computer Society, 1995, pp. 278–282.

[32] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001.

[33] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar 1986.

[34] L. Raileanu and K. Stoffel, "Theoretical comparison between the gini index and information gain criteria," *Annals of Mathematics and Artificial Intelligence*, vol. 41, pp. 77–93, 05 2004.

[35] R. Carroll, J. B. Lewis, J. Lo, K. T. Poole, and H. Rosenthal, "Measuring bias and uncertainty in dw-nominate ideal point estimates via the parametric bootstrap," *Political Analysis*, vol. 17, no. 3, p. 261275, 2009.

[36] K. Poole, *Spatial Models of Parliamentary Voting*. Cambridge University Press, 01 2005.

[37] D. L. McFadden, *Quantal Choice Analaysis: A Survey*. NBER, October 1976, pp. 363–390.

[38] A. Bonica, "Database on ideology, money in politics, and elections: Public version 1.0 [computer file]," 2013. [Online]. Available: http://data.stanford.edu/dime

[39] J. B. Lewis, K. Poole, H. Rosenthal, A. Boche, A. Rudkin, and L. Sonnet, *Voteview: Congressional Roll-Call Votes Database*, Voteview, 2019, https://voteview.com/. [Online]. Available: https://voteview.com/

[40] S. Wahl and J. Sheppard, "Hierarchical fuzzy spectral clustering in social networks using spectral characterization," in *Proceedings of the 28th International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, 2015, pp. 305–310.

[41] S. Zhang, R.-S. Wang, and X.-S. Zhang, "Identification of overlapping community structure in complex networks using fuzzy c-means clustering," *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 1, pp. 483–490, 2007.