Open Secrets Analysis

Nick Greene, William Lifferth, Denizhan Pak, Christian Payne, Zachary Trzil

Abstract—The United States spent 3.15 billion dollars on lobbying in 2016[1]. The trend of lobbying in the United States has only risen over the years. Lobbying is intended to give the stakeholders of certain policies a direct method of communicating with their representatives. The failing of this system is that it gives disproportionate amount of influence to those with more financial resources. This creates an undesirable scenario where those with more economic resources will pursue goals that reinforce their control over those resources, without regard for the well-being of others. To understand the impact this has on governance, we analyze how impactful these donations are. What is the relationship between lobbying and the support for policies that are beneficial to those entities that fund said lobbying? To answer this question, we will collect and analyze data quantifying amounts spent on lobbying for various politicians and respectively the policies supported by those politicians. We will look for correlations between industries that spend money on lobbying and policies passed affecting said industries. We will then look at the types of policies that are more commonly implemented based on the industry's lobbying capacity. The data will be acquired through an open database, *opensecrets.org*, where donations made to politicians are stored and made freely available. We will then collect federal policy from an open Congressional database *govtrack.us* where we will analyze trends and look for statistical associations.

Index Terms—Politics, corruption, lobbying, data retrieval, public policy.

I. Background

Many in the United States are concerned that the current structure of congressional campaign funding is the cause of unequal representation. Intuitively, this idea is based on the concept that politicians have a direct incentive to please individuals who have the ability to contribute larger amounts to their campaigns. The inherent conflict with the United States governace is the expectation that few are selected to represent the many. This requires that those few take decisions to support their constituency and affect policy which will meet this burden. The issue arises when the means of acquiring the power necessary to affect change occurs through the process of concession to those with a larger capability to donate. While the complexity of modern government necessitates the existence of a system of representation in any fundamentally democratic nation-state, it is important to recognize its limitations. However inevitable, this system is a textbook example of the principal-agent problem, generally defined to be the differential of incentives between a principal (the citizens of a nation state) and a selected agent (elected officials in a representative democracy) when it is unfeasible to evaluate in extreme granularity the decision-making process of the agent. While ideally elected officials should be made easily accessible to their constituents such that they

Department of Computer Science, University of Tennessee - Knoxville

can actively align their objectives with the desires of their respective principals, access extends to interest groups and private entities. These private entities can use their access to lobby an elected official and effectively incentivize them to pursue their own goals over the goals of their principals–namely the citizenry they were elected to represent. While theoretically the natural realigning of incentives should be gradually achieved in an efficient electoral system with regular opportunities to replace representatives, this is rarely the case in modern American political life. Alternatively, making available a better means of evaluating the decision-making process of agents (elected officials) can provide a separate path to alleviating the failings of the principal-agent problem. The central question to alleviating the principal-agent problem is, for a given politician, to what degree is their decision making on matters of public policy associated with desires of the interest groups which provide financial resources, particularly in the case where these desires may be contrary to those of their electorate. Using records on the financial relationships between lobbying entities and elected officials, we can derive the groups and entities an elected official may feel potentially beholden to. Although there are many valid ways of going about this, a simple mapping of magnitude of financial contribution to potential for corrupt rent-seeking behavior may provide a robust mechanism for identifying the third party-agent pairings of most interest. Additionally, an objective way of determining whether the behavior of an elected official exhibits corrupt rent-seeking potential is necessary. A preliminary method to establish this potential is simply clustering groups of elected officials together based on the nature of their received financial contributions, and verifying whether or not these clusters hold any predictive power over certain actions (i.e. votes on public policy). If this is the case, we can at least show that lobbying behavior is not independent of decision of public policy. However the above stated method is limited in that it can only show trends in aggregate, as it depends on the clustering of many agents (elected officials). Ideally, we would be able to evaluate a single elected official on the grounds of corruption. For this purpose, it is necessary to build a model capable of labeling a certain public policy decision as either beneficial or disadvantageous for a given third-party lobbying entity. If this can be done with relative confidence, we can address to what degree different elected officials are affected by financial contributions to behave in ways that subjugate the well-being of their principals to the well-being of those third-party entities from which they have received financial resources.

If the above claims are true and it is the case that the missaligned incentives of our representatives have a relationship with the perpetuation of their power through contributions,

it would follow that there will be a relationship between the contributions made and the policy that was supported by these individual politicians. For this analysis, we looked at the 2014 election cycle and the

legislative session immediately following.

II. METHODOLOGY

A. Sources

We pulled data on financial contributions from the site OpenSecrets.org, which is maintained by The Center for Responsive Politics, to inform our initial assessment of potential for corrupt rent-seeking behavior between a given elected official and any third-party interest group. These financial contributions are structured as individual donation amounts made by a numeric donor identification number associated with an individual PAC to an associated candidate identity. The donations are strictly numeric values that are the total quantity donated. From this data only donations that were not refunded were kept. Additionally all donations made to candidates that did not win their respective election cycles were dropped. Furthermore, we will utilize the documents made available by the U.S. Government Publishing Office via the site govtrack.us to evaluate the nature of specific behavior by agents. This data is a list of legislation proposed and the corresponding votes by each individual representative on those legislative pieces. All pieces of proposed legislation were analyzed and representatives absent from voting or abstaining were accounted for.

B. Computing Tools

The project was done utilizing a python3 based jupyter notebook. The following libraries were used:

- Numpy: Used for numerical and trend analysis.
- Pandas: used for the purpose of reading and cleaning source data.
- MatPlotLib: Used for visualizing the resulting data for analysis.

C. Methods

1) Data Collection: The donation data was collected from the opensecrets.org data base and the voting data was scraped from its html format on the government database. The data was read in and unnecessary features were dropped.

2) Feature Selection: Since the variation between donation data is so vast, we applied feature selection as a collection of five categorical values into which each donation could fit. To determine these values, we first placed all donations on a logarithmic scale as it made sense that larger contributions would have a wider range but should have similar impacts based on our hypothesis. Once the scale was decreased, we then found the mean contribution and defined the categories in relation to standard deviations removed from the mean with the intent of maintaining the core features of each donation. This lead to the creation of the following five donation categories:

- Up to $580.49
- $580.49 to $1526.17

- $1526.17 to $4012.47
- $4012.47 to $10549.24
- Over $10549.24

There does not seem to be major correlation between party and amount of donations per category[1].
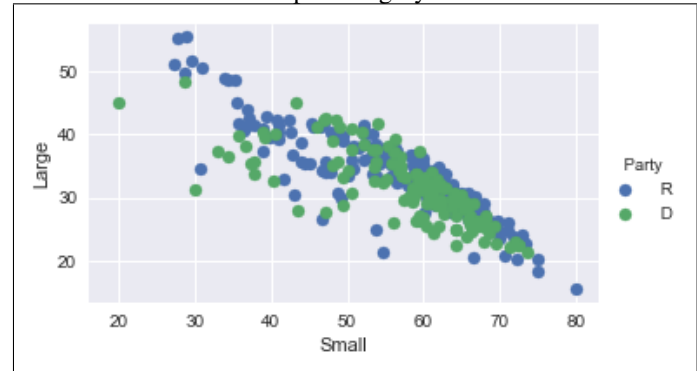
3) Representative Profiles: Each representative was defined in terms of the number of donations they had received in each of the above categories. Once they were defined each representative was the clustered based on their donation profiles. To determine how to best cluster the data we conducted a k-means analysis looking at the accuracy over the possible numbers of clusters. [2]
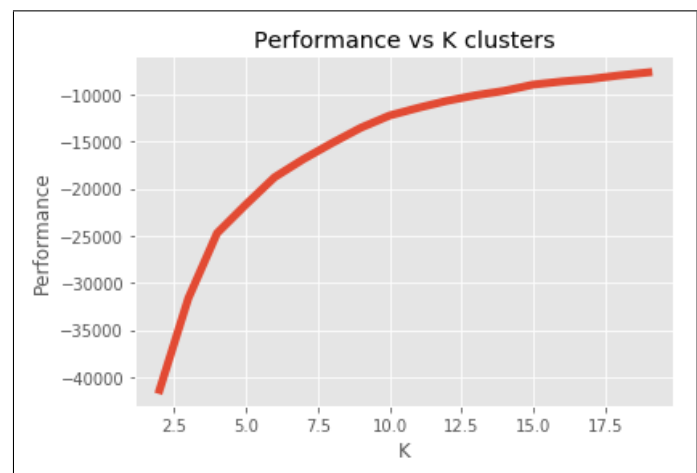

Here we see an obvious increase in accuracy of prediction based upon an increase of clusters. However, although accuracy increases clustering causes a trade off with accuracy and generalizability as with most machine learning algorithms.

Since the sample size is not too large, we also looked at the size of the smallest cluster for each number of clusters.

---

[1] Large donations are the two largest categories, while small donations are the two smallest

[2] Accuracy is defined in terms of predictability of a vote on a given bill using a k-means clustering.
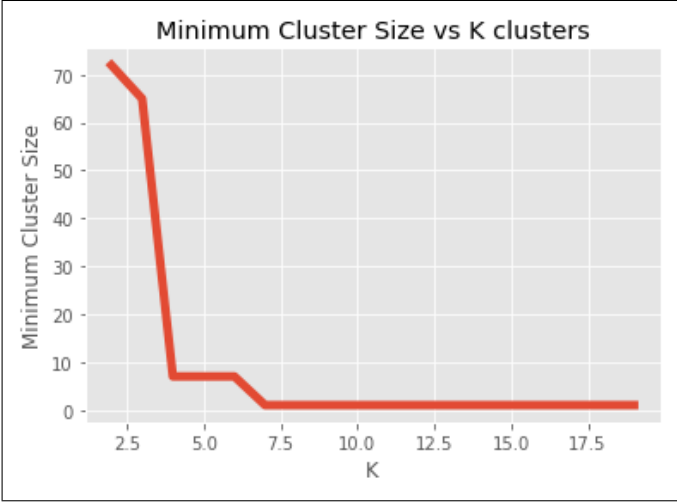
Three clusters demonstrates a relatively decent elbow in accurracy and prevents any one cluster from being too small. As such the distribution of the resulting clusters roughly correlates to Large, Medium, and Small donation recievers with some outliers.
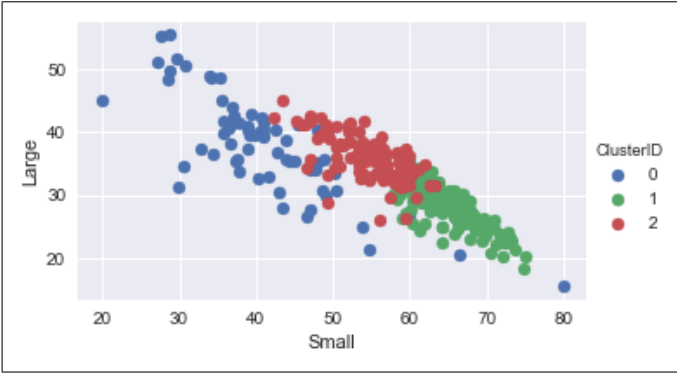
From here we see that the clusters are more correlated to donation information than party information. This is important to note as it alleviates the concern that the analysis will only find party based trends.

III. RESULTS

For each of the clusters we normalized defined a aggregate vote value for each bill as the average of all votes in that cluster regarding the given bill, with an affirmative vote coded as a 1 and a negative vote coded as a 0. Since each cluster votes on a bill we can define them as paired events. To check for statistically significant differences we can conduct a paired T-Test. This will indicate the probability that differences in paired measurements or events from two populations occurred as a result from random variation in a single population following a normal distribution. Thus the results can be interpreted as showing a relationship between the cluster in which a representative is located and their likelihood of voting similarly to another cluster.

```
In [6]:  1  t_test_pv = np.empty((3, 3))
         2  for i, cvl1 in enumerate(cluster_voting):
         3      for j, cvl2 in enumerate(cluster_voting):
         4          t_test_pv[i][j] = stats.ttest_rel(cvl1, cvl2)[1]
         5
         6  t_test_pv

Out[6]:  array([[          nan,  5.10779906e-05,  4.11646508e-14],
                [ 5.10779906e-05,          nan,  3.73281923e-12],
                [ 4.11646508e-14,  3.73281923e-12,          nan]])
```

The resulting T-Test shows the highest p-value between clusters to be close to 0.000051. This implies a relatively small chance of these observations being made of fundamentally similar entities. With some interpretation this means that candidates in different donation profile clusters are unlikely to vote similarly. However, this test assumes a normal distribution among voting records–and by definition our coding of votes into the set $\{0, 1\}$ means our observations cannot follow a normal distribution. To further probe the potential dissimilarity between donation profile clusters' voting records we apply a logistical regression model which uses the cluster ID as a categorical variable.

We can query the coefficients of the second and third variables to establish the degree to which our model organizes them as categorically different; we established an alpha value of 0.05 such that a coefficient with an absolute value greater than 0.05 is statistically significant. 93% of the bills has statistically significant differences in our logistic regression model. This evidences the original claim that there is significance in the relationship between donation cluster and voting record.

```
[[ 0.          0.93019943  0.9031339 ]
 [ 0.93019943  0.          0.96153846]
 [ 0.9031339   0.96153846  0.        ]]
```

IV. CONCLUSION

From our results it seems there is a clear relationship between the clustering of a representatives donation profile and their voting record. What this implies is that a representative is indeed likely to be skewed in their votes by the quantities of donations they receive. Further analysis also showed that there was little correlation between the party of a representative and the size of the donations they received. This implies that there may be an impact of campaign contributions that is independent of party affiliation which would further support the hypothesis that representatives can be influenced by the size of the donations they receive. Although this is correlative and not causative it indicates that future research may be able to determine if lobbying has become the deciding factor in United States policy. between the types of donations a representative recieves in terms of amount and the legislative decisions they make. Future research could be done to analyze the impact of these donations on both a PAC and Super PAC level. Additionally, if more terms were to be accounted, for it might show this analysis to simply be an outlier; although some of the statistical correlations may be too strong for that to simply be the case. Finally, it would further inform the conversation to be had if research was performed on the

actual contents of the legislation and information about the donating parties. Such an analysis would imply if there is a causative relationship between the donations and the resulting votes. The current state of the United States electoral and campaign finance system may be undermining the intent of a representative

democracy. Such an issue should be addressed promptly using analysis and facts supported by empirical data. The goal of this project is to further this idea and serve in determining if the systems in place are meeting their underlying goals. Alleviating the principal-agent problem will make the American democratic process stronger. Accurately representing the citizens of the nation, whether or not they can financially support an elected official, is key to accountability in the political process. Unfortunately, there is no accountability in American politics. This is an effort to restore some faith back into the process by shedding light on the parts that Washington aims to keep hidden, rather than sitting idly by while interest groups silently push their own agendas.

V. References

1) GovTrack.us. (2017). GovTrack.us. [online] Available at: https://www.govtrack.us/ [Accessed 14 Dec. 2017].
2) Politics, T. (2017). Data on Campaign Finance, Super PACs, Industries, and Lobbying. [online] OpenSecrets. Available at: http://www.opensecrets.org/ [Accessed 14 Dec. 2017].