CS224W Final Project: Super-PAC Donor Networks

Rush Moody — rmoody@stanford.edu

December 9, 2015

1 Introduction

In a landmark case decided in January of 2010, *Citizens United v. Federal Election Commission*, the United States Supreme Court created a new class of nominally independent political action committees (P.A.C.s) with the ability to raise unlimited sums of money from individuals, corporations, and unions, sidestepping the previous $2,500 per-contribution limit. Proponents of this decision hailed it as a victory for free speech over government regulation, while critics claim that it has opened a floodgate of spending by wealthy donors and special interest groups that threatens the foundations of our democracy. Regardless, the Federal Election Commission (F.E.C.) requires that all contributions to these committees (which have come to be known as "Super PACs") be disclosed on a quarterly basis, and the disclosure data is made freely available on the FECs website. The volume and availability of the contribution data allows us to explore the donor networks underpinning this new phenomenon in the political landscape, in order to gain deeper insight into the nature and structure of the interrelated individuals and groups who are shaping the country's political future. Given the innately partisan nature of the debate around the Citizens United decision, I believe that network analysis tools could help provide an objective and minimally biased framework from which to approach this phenomenon.

2 Related Work

Although Super-PACs are a relatively new class of entity, significant research has been performed on both political donation network structure and on other forms of politically charged networks. Many of these previous papers focus on attempting to look for meaningful differences between the structure of liberal and conservative networks, which I will also attempt at various points.

In Koger, et al's 2009 paper, social network analysis is used to reveal the subtle interactions between official partisan organizations and nominally unaffiliated interest groups and media outlets, with the goal of "revealing the communities that drive modern politics" [1]. To obtain data on partisan information exchange they used a technique called snowball sampling, wherein they gave political donations under a number of unique names to various political organizations, and then tracked the spread of that information as they received donation solicitations from various other groups. Their subsequent analysis reveals that the Democratic National Committee (DNC) has a much more densely connected egonet than its Republican counterpart (RNC), along with a stark division of these outside groups into polarized camps

In another paper by Matt Grossman and Casey Dominguez, network analysis tools are utilized to help reveal the structure of various interest groups communities and determine how they interact with party-aligned political organizations [2]. They used a mix of several signals to connect interest groups in the network: campaign endorsements, legislative ties, and financial contributions, from which they constructed single-type and multiplex networks of the involved interest groups. Interestingly, this paper also reaches a similar conclusion to the previous paper in finding that Democratic networks are much denser than their Republican counterparts

Finally, the 2004 paper by Lada Adamic and Natalie Glance explores the cross-linking patterns and discussion topics of U.S. political blogs leading up to the 2004 presidential election, with the goal of measuring the level of interaction between the liberal and conservative blogging communities and analyzing the unique characteristics of these respective networks [3]. In a depature from the previous papers, their analysis revealed the conservative blogging community to be more densely linked than liberal blogging community, and additionally provided evidence to dispel the myth of the conservative "echo chamber" by noting that post content within the liberal and conservative blogging communities shared a similar level of similarity.

3 Model and Methods

3.1 Data

For this project I am using the complete FEC contribution filings from the 2012 U.S. election cycle. The raw data is contained in two database tables; the first table contains a row for a single contribution by an individual to some committee, and the second table contains a row with information about each of the committees. Sanitizing this data involved removing contributions to non-Super-PAC committees and clustering together all of the individual contributions by a single person. The final cleaned data contains 128,506 individual contributions to 1,397 Super PACs, totaling $848,090,207.

3.2 Graph Models

This FEC data lends itself to several graph formulations. The primary and most intuitive model is a bipartite graph between individuals and PACs, with weighted edges representing the total contributions from an individual to a committee. I use this model for the first part of my analysis where I attempt to find quantitative differences between the donor profiles of conservative-leaning versus liberal-leaning Super PACs, as well as for examining the formation and growth of these networks over time. The second formulation is an individual co-donation graph, where each individual is a node and edges occur between individuals that donated to the same PAC; this formulation is used in the second part of my analysis, where I attempt to detect community structures in the donor networks and to assess the overall degree of partisan polarization.

3.3 Bipartite Network Graph Methods

To lay the groundwork for my analysis, I supplemented the raw FEC contribution data by manually annotating the top 40 PACs (as determined by donation totals, which roughly corresponds to all PACs with more than $1 million in funding) with their known partisan affiliation. Using this information I attempt to assess any qualitative differences between the contributions to conservative versus liberal PACs by examining the distribution of donation totals. To do so I created a number of buckets for ranges of contribution totals and plot the resulting contribution totals as a cumulative distribution function. To aid in analysis I divided these plots into both low-end and high-end contributors, using $500,000 as the threshold value between the two.

Next, I attempt to gain insight into the formation of these Super-PAC networks by examining distribution of total funds available to all the PACs in the ecosystem; there is clearly a long tail of smaller committees, but can we fit a power law to their distribution (both unweighted, i.e. number of donors, and weighted, i.e. total donation amount) to determine whether their formation is governed by a rich-get-richer phenomenon? To perform this analysis I place the PACs into fixed-size buckets based on their total funding base and number of donors, and then attempt to fit a power law to the resulting distributions and plot the resulting ccdf. From there, I attempt to track edge formation over time using the contribution timestamps to gather further evidence for such a phenomenon, i.e. by tracking whether donors change their loyalties over time to favor better-funded Super-PACs as opposed to only switching when the candidate back by their preferred PAC is withdrawing from the race.

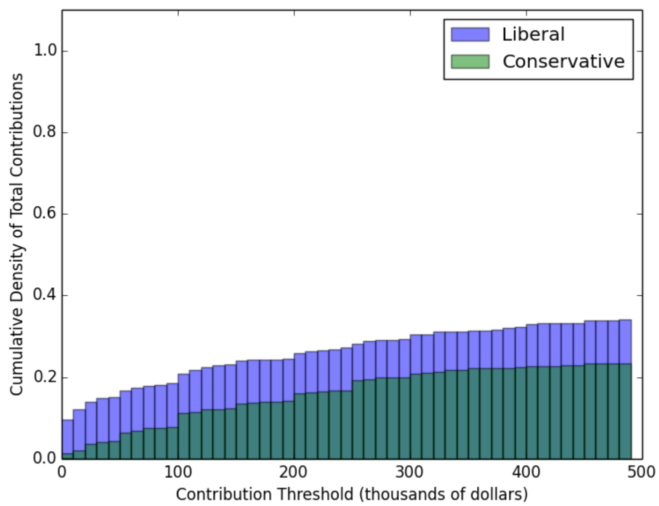
3.4 Co-Donation Graph Methods

To analyze the donor co-donation network, I start by again subdividing the graph into liberal and conservative donors based on the affiliations of the groups that they contribute to, and then perform some fairly rudimentary analysis to determine some basic properties of the resulting networks, including their connectivity and their Freeman Centralization using various centrality measures. From there I attempt to determine more sophisticated underlying structure within the networks, such as the presence of densely connected sub-communities, and then use the Girvan and Newman algorithm [4] to perform hierarchical clustering on the donors. Finally, I analyze the full graph of both liberal and conservative donors to determine whether and where any connection between the partisan extremes occurs, with the goal of assessing the overall polarization of the political donor landscape.

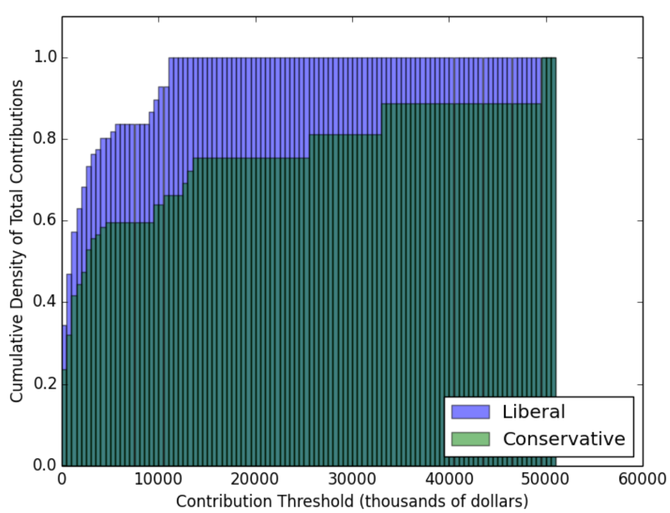
4 Results and Findings

4.1 Bipartite Graph Results

The following plots show the cumulative density of donation totals for liberal and conservative Super-PACs, graphed by total donation threshold. The first plot shows the cumulative contribution of relatively small donors (less than $500,000 total each):

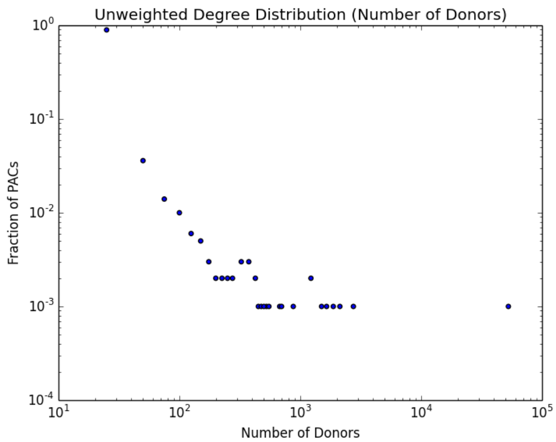


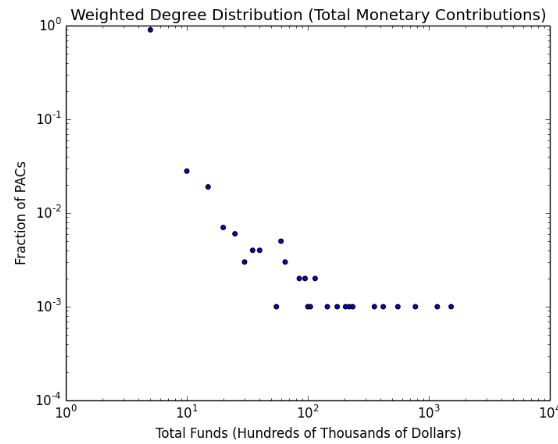
While the shape of these distributions is quite similar, the liberal distribution is clearly more densely distributed at the lower end, implying a greater reliance on smaller donors. Next we examine the entire distribution to get an idea of the impact of high-rollers:



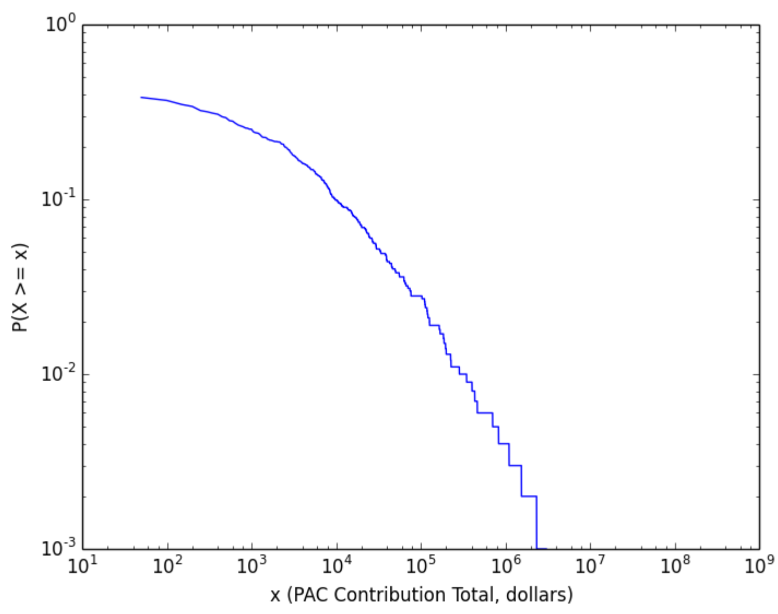
While the liberal Super-PACs seem to derive a significant proportion of their funding from these high-end donors, their conservative counterparts have a clear advantage here, raising more than 75% of funding from high-rollers (as opposed to slightly less than 65% for the liberal PACs). Additionally, these high-end donors only compose 9% of the conservative and around 1% of the liberal donors by number, lending credence to the criticism that the Citizens United decision has allowed these mega-donors to have an outsize influence over the election process.

Next we examine the distribution of contribution totals over all of the Super-PACs in the FEC database, both by unweighted degree (number of donors) and weighted degree (total amount donated). For the former we discretize into buckets of 25 donors, while for the later we place them into buckets of size $500,000 and graph the resulting distributions:





Both the weighted and unweighted degree distribution seems to display signs of a power-law fit. For the unweighted graph we obtain that xmin equals 25 and alpha equals 5.98, while for the weighted degree distribution we obtain xmin equal to 500,000 and alpha equal to 4.56. To further look for evidence of a power-law phenomenon, I plotted the empirical ccdf (complementary cumulative distribution function) for the weighted network:



While there is some curvature to the initial portion of the graph, the majority seems to exhibit the straight-line appearance of a power law phenomenon. To delve deeper into the dynamics of the formation of these networks, we exploited the donation timestamps in the individual contribution database to follow the behavior of donor edge formation over time. To help validate the hypothesis that donors display preferential attachment, as suggested by the power law fit, we examine the incidence of donors changing their initial PAC loyalty to a PAC that is currently more successful. To keep this analysis tractable, we restrict our inquiry to a subset of the 2012 election Super-PACs representing the top Republican primary candidates (as the Democrats had an incubant, Barack Obama, the same dynamics likely don't apply to them). These candidates and their PACs are summarized in the table below:

Candidate	Super-PAC	PAC Contribution Total ($)	Date Suspended Campaign
Mitt Romney	Restore Our Future, Inc	152,680,805	N/A
Ron Paul	Endorse Liberty	3,831,483	N/A
Newt Gingrich	Winning Our Future	23,314,935	May 2, 2012
Rick Santorum	Red, White, and Blue Fund	8,462,221	Apr 10, 2012
Rick Perry	Make Us Great Again, Inc	5,605,174	Jan 19, 2012
Jon Huntsman	Our Destiny PAC	3,165,544	Jan 16, 2012

We propose three different models of donor decision-making. In all cases, we only consider changes in donor loyalty made by donors whose original donation choice remains in the race when they decide to switch. The first model is our null model, where the donor makes a uniformly random choice about where to donate. In our second, location-based model, a donor decides where to donate proportional to where other donors from a similar location have donated (based on donors' home state). Our final model is a rich-gets-richer model, where a donor decides to change their donation target with probability proportional to total funds raised by each PAC. We found approximately 500 individuals who made multiple donations to the above PACs, of whom approximately 100 decided to change their choice from the original donation. For each valid donor switch (i.e., their original candidate is still in the running) and for each model we derive the probability of the donor making that particular choice to donate to PAC p based on the model at date t and the donor d using a function $P_t^d(p)$ in the table below, and describe as follows: for the null model the probability is uniform between the PACs of those candidates still in the running (which we define for date t to be X_{run}^t; for the location-based model the probability of choosing a PAC is proportional to the percentage of donors from the same state who have given to that PAC out of all the PACs still in the running (defined as $frac(pac, state, running\_pacs)$); and for the rich-gets-richer model the probability of choosing a PAC is proportional to that PAC's share of funding raised to date out of PACs that are still in the running, which we define as f_p^t. Assuming that each individual choice is made independently, we derive the following log likelihood values for the sequence of changes in donor loyalty over time:

Model	Log Likelihood of All Decisions	Probability Function
Uniformly Random	-40.779	$P_t^d(p) = 1[p \in X_{run}^t] \frac{1}{\lvert X_{run}^t \rvert}$
Location-Based	-46.894	$P_t^d(p) \propto frac(p, d.state, X_{run}^t)$
Rich-get-Richer	-38.178	$P_t^d(p) \propto f_p^t$

Here we can see that the rich-gets-richer model does a marginally better job of explaining the data than the null model, and significantly better than the location-based model, lending further evidence that some such phenomenon helps shape formation of the donor networks. To strength this claim further we would need to control for other factors in the race, such as current poll numbers or some other quantitative measurement of how well a candidate is doing, to derive a more sophisticated understanding of how Super-PAC fundraising compares in importance to these other factors. The ongoing 2016 Republican Primary will also provide a solid testing group for these hypothesis once the full contribution data is available for the election cycle.

4.2 Large Co-Donation Graph Results

To make the datasets more tractable, we restricted the co-donor network to donors who contributed at least $5,000 in total; this value was chosen because it is twice the maximum value of traditional political donations to non-Super-PAC entities. The properties of the two resulting networks are summarized in the tables below:

Orientation	Liberal	Conservative
Nodes	2270	1590
Total Donations ($)	289,611,140	434,572,971
Avg Donation ($)	127,582	273,833
LWCC Fraction	0.9958	0.998
Avg Norm. Degree	0.619	0.29
Freeman Centrality (closeness)	0.513	0.498
Freeman Centrality (betweeness)	0.0069	0.013

As exhibited above the two networks have some notable differences in structure. The liberal network is larger and more densely connected, but has a much lower average donation and slightly higher betweenness centralization compared to its conservative counterpart. These differences mirror both Koger and Grossman's finding that liberal political networks tend to be denser than their conservative counterparts [2][1], lending further credence to such a conclusion. Interestingly, both networks appear to be quite highly connected, as evidenced by the fact that in both cases the largest connected component constitutes the vast majority of the graph. Although the co-donors to each Super PAC represent natural communities in the graph, to help look for deeper network structure I used the Girvan and Newman hierarchical clustering algorithm [4] to locate other densely-connected communities within the graphs. This alogorithm runs in $O(n^3)$ time, which unfortunately was not tractable to run on the full networks, so here we further refine our domain to donors contributing more than $175,000, leaving us with 272 conservative and 208 liberal donors. The results of the clustering analysis are summarized below:

Orientation	Liberal	Conservative
Num. Nodes	208	272
Num. Communities	13	16
Median Community Size	5	4
Largest Community Size	72	154
Final Clustering Modularity	0.363	0.215

The broader community structure for both groups looks similar, as each has a singular largest community significantly larger than the others, and then a collection of peripheral communities ranging in size from (1, 59) nodes on the conservative side and (1, 40) on the liberal side. The evidence also suggests the liberal network is slightly better-developed and contains denser communities, based on the higher modularity value and the larger median community size despite the fewer overall nodes in the network.

In our analysis we now examine several representative communities from each network. On the liberal side one of the mid-size communities consists of 19 different organized labor associates, including giant unions such as the larger AFL-CIO (American Federation of Labor) and smaller unions like the Plasterers' Cement Masons' and Shop. This finding make intuitive sense, as organized labor constitutes a significant fraction of the Democratic support base that are likely to share common interests and contribute to similar candidates. A smaller community in the liberal network is composed of four environment-focused organizations: Environment California, Environment Oregon, Environment Washington, and Environment Illinois. These loosely-affiliated organizations all hail from relatively liberal states and share the common goal of fighting for environmental preservation, again providing a good intuition for their placement in the same community.

On the conservative side there appear to be mainly clusters of individual donors, although one smaller cluster contains two large conservative advocacy groups, Americans for Limited Government and American Conservative Union, as well as a secretive conservative political organization called New Models, and an individual named Rex Sinquefield who espouses very conservative views (including ending the income tax!). While these four different organizations might not be otherwise overtly linked, they clearly support a similar conservative political philosophy; using such analysis techniques could help groups identify individuals to target for donation solicitations.

Finally, I examined the combined liberal and conservative network to attempt to find any key connecting points bridging the liberal and conservative donor communities. The combined graphs consisted of 6863 nodes, of which 6856 where in the largest connected component, which implies that at most around 10 nodes donated to both liberal and conservative candidates. I used betweenness centrality to attempt to locate these key connectors; the top three nodes are listed in the table below, along with their total contributions to Liberal and Conservative Super PACs:

Name	Betweenness C.	Liberal Total ($)	Conservative Total ($)
Fidelity Information Services	533,891	75,000	75,000
Williams, John E.	111,295	750,000	75,000
Shenker, Scott	42,478	730,000	0

From examining these nodes, it is clear that the first two play a key role in connecting the graph. The first, Fidelity Information Services, donated an equal amount to the Super PAC supporting Mitt Romney, the Republican Candidate, and Hilary Clinton, the Democratic candidate, providing a clear link between two of the largest communities in the network and giving it by far the highest betweenness centrality. This finding mirrors the results from the Grossman-Dominguez paper, which identified professional associations and business interests as key bridging components between the otherwise disjoint liberal and conservative networks. The second node donated mainly to PACs supporting a liberal national agenda, but also gave money to the Texas Conservative Fund, perhaps suggesting a rare combination of a liberal national political bent with a loyalty to local conservative causes. Although the conservative PAC here is relatively minor, it still provides enough of a link to clearly increase the number of shortest paths between other nodes that travel through this node, and thus increase the betweenness centrality. The final node donated to only a large number of liberal PACs; note that the betweenness centrality for this node is significantly below the ones above it. The rest of the nodes in the top 20 in betweenness centrality all have values in the range (16,000, 41,000), implying that they do not play nearly as important a role in connecting the graph. On the whole, the presence of only a tiny number of nodes that connecting the two ideological communities lends credence to the thesis that the political landscape is currently extremely polarized, again mirroring a finding from the Koger paper [1].

5 Conclusion and Future Work

Our methods of analyzing Super-PAC donor networks both lend some weight to the previously existing work on partisan networks, and raises several new questions worthy of investigation. From analyzing the complete network, we can see that Democratic Super-PACs rely much more heavily on small donors than do their Republican counterparts, which further implies they are less-well equipped to take advantage of the unlimited donation limits that differentiate Super-PACs from traditional PACs or political campaigns. The Republicans also rely heavily on a small number of super donors giving more than $10,000,000 each to provide almost a quarter of their overall funding, with a single donor providing more than 10% of the funds raised; without these donors, the gap between Democratic and Republican fundraising totals decreases from around $150,000,000 to only around $35,000,000. The distribution of funding totals for the PACs themselves also seems to mirror a power-law distribution, implying that rich-get-richer dynamics help shape their formation. Our use of several competing models helps support this claim, but there is significant room for improving the model to control for other factors about the state of the campaign (such as polling strength) to help determine whether there is a substantial causitive relationship between PAC fundraising totals and individual donation decisions. Finally, our analysis of the co-donation network mirrors previous findings about the denser structure of liberal versus conservative partisan networks, and the small number of links between the liberal and conservative communities lends further credence to the idea that the political landscape is extremely polarized. Although some basic work has been done to derive the underlying structure of these donor networks, different methods of clustering (such as using an agglomerative approach instead of divisive method used by the Girvan-Newman algorithm) could help reveal other interesting sub-communities.

Although this paper merely presents a starting point for further analysis, I feel that using using graph analysis techniques provide a useful framework to approach the novel political phenomenon presented by these Super-PACs. These organizations are likely to play an important role in the U.S. political landscape for the near future, so understanding how they impact the fundraising dynamics of political campaigns is critical for gauging whether the intense criticism focused on them is valid. A final useful approach to this larger question would involve comparing the Super-PAC networks to traditional PACs to see whether they have different structures or whether Super-PACs are just an extension of previous techniques.

References

[1] Gregory Koger, Seth Masket, and Hans Noel. *Partisan Webs: Information Exchange and Party Networks.* Cambridge University Press, 2009
`http://works.bepress.com/cgi/viewcontent.cgi?article=1010&context=gregorykoger`

[2] Matt Grossman and Casey B. K. Dominguez. *Party Coalitions and Interest Group Networks.* American Politics Research, 2009
`http://apr.sagepub.com/content/37/5/767.refs`

[3] Lana Adamic and Natalie Glance. *The Political Blogosphere and the 2004 U.S. Election: Divided They Blog.* Proceedings of the 3rd International Workshop on Link Discovery, 2005
`http://dl.acm.org/citation.cfm?id=1134277`

[4] Santo Fortunato. *Community Detection in Graphs.* Arxiv, 2009.
`http://snap.stanford.edu/class/cs224w-readings/fortunato10community.pdf`