Ben Lasley
Ben Shapero
Charlton Soesanto
CS224 Final Project Write-Up

Google (Campaign) Finance:
Evaluating a Grassroots PageRank Model for Predicting Election Results from Political Contributions

1. Introduction
1.1 Overview
Our project determines if the outcome of an election can be predicted using campaign finance data adjusting for grassroots contributions, weighing the number of donors more than simply the total amount of money a candidate receives. Campaign finance data involves individuals, PACs, anonymous donations, candidates, and incomplete information about where the money came from. The campaign finance donation network is difficult to understand, obscured, and has incomplete information. By first analyzing the structure of the network and comparing it with the random and power law graphs to understand main differences, we then propose an algorithm to predict election winners that compensates for unique structural attributes.

While our PageRank Grassroots Prediction Model uses campaign donations to predict the winners of elections, the ultimate result of the election is dependent on many factors. Some of these factors include how that money is spent, the regional demographics of the election, or the relative competitiveness of a race. Campaign contribution data alone cannot model all of these different dimensions, but our hypothesis is that money and grassroots support (as a proxy of literal votes in an election) account for a significant amount.

We propose and test the model of political gifting that weighs the number of political donors against the total amount of money donated. The initial "rank" vector is calculated by taking the net sum positive donations and dividing it by the total unique net positive donations in the graph. This allows anonymous, unreported donations to be counted in the PACs, because if they donate more than they receive we assume that money was not reported.

Donations will be given weights based on functions applied to the donation amounts. We will use the PageRank random surfer model and power method iteration to calculate a "support score" for each candidate. This model describes the value of grassroots support as an indicator of electoral success, and is valuable toward answering the question "What is the importance between number of donors and the total amount of money received for prediction election results?"

1.2 Relevant Prior Work
Much research has been done on the nature of political donations, but most studies focus on only a single mode of political gifting. Koger and Victor only look at relationships between legislators and lobbyists [5]. Dominguez studies the overlap in donor lists between the political parties and PACs [1]. Studies on the network of small donors are much less common. This is because donations under $200 are not required to be disclosed to the FEC. The Campaign Finance

Institute conducted a wide survey of small donors as part of The CFI Small Donor Project [9], the first large-scale attempt to answer questions about the impact of small donor participation. Even so, The CFI Small Donor Project only surveyed in seven states across only a sample of small donors, and only looked at political races at the state level. The Small Donor Project compared donors using a variety of criteria, including race, income level, education level, and political ideology. Our project will use different techniques to analyze the data and make inferences based on the network structure.

Prior research has been done on making inferences on probabilistic relational models. Kaelin and Precup describe an approach using block sampling [4]. Probabilistic relational models describe relationships between classes and variables, encoding these relationships as dependencies. The probabilistic relational model builds off of the Directed Acyclic Probabilistic Entity-Relationship (DAPER) model described in Heckerman et al. (2004) [3]. DAPER models describe the network in a first order logic like structure, requiring the construction of large Ground Bayesian Networks (GBN). Computing these networks can be computationally expensive. One technique to address this is aggregation, when the values of nodes are computed by their parents, without knowing the number of parents before hand. This technique works very well in our model, as our weighted edges lend themselves to aggregation easily. In their study, Kaelin and Precup use aggregation and construct the GBN incrementally on an as needed basis [4]. They describe a Lazy Aggregation Gibbs Block algorithm that scales well with the size of the data. As a proof of their algorithm and its relevance to this proposal, they showcase their algorithm on political contribution data. Their tests involve predicting the political affiliation of donors. Our tests will instead test how accurately the model can predict electoral victors.

2. Structural Analysis

2.1 Data Collection Methods

All data has been retrieved from The Center for Responsive Politics at opensecrets.org [6]. The data has been pulled from the Federal Election Commission (FEC) website, cleaned and compiled into database tables. The Center for Responsive Politics provides multiple APIs and widgets in various programming languages. This project will utilize the Python APIs, converting the raw relational data into a NetworkX graph for analysis and visualization. We include every provided category of political giving, including individual donations, PACs, 527 groups, and political parties. Contributions and expenditures are listed per candidate, allowing for fast constructions of models.

We downloaded the CSV files for Campaign Finance data, specifically from Individuals to Candidates, PACs to Candidates, and PACs to PACs. We converted the CSV files to GraphML files and then loaded them into NetworkX, with each node having a person or group ID and corresponding attributes party and type. The directed edges go from the donors to the donees with the edge weight being the amount of the contribution. Records with negative contributions (indicating refunds) or records where either the donor or the recipient is unknown are ignored.

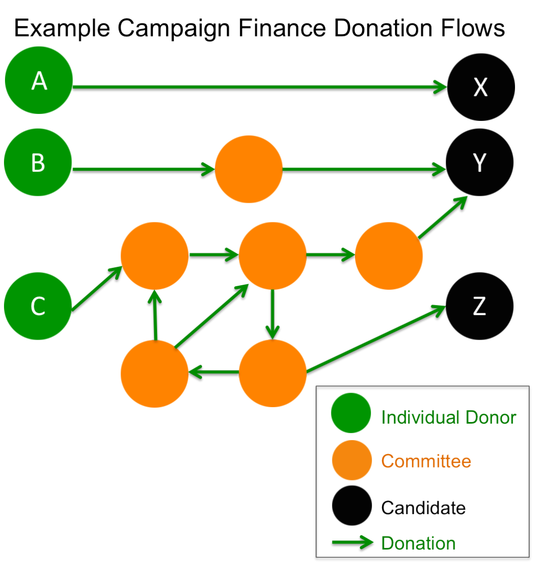
2.2 Initial Findings and Summary Statistics

Before running our models on the data, we identified structural information about our graphs. Our hypothesis is that the donation network follows a power-law distribution. We generated a $G_{n,m}$ random graph and a power-law distribution for comparison. The statistics are presented with both election years in parallel below.

	2008	2010
Number of Nodes	1,389,033	842,584
Number of Edges	2,439,734	1,548,008
Maximum Degree	362,553	48,237
Average Degree	3.513	3.674
Maximum In Degree	362,553	48,091
Maximum Out Degree	679	763
Average Shortest Path Length	4.226	4.192

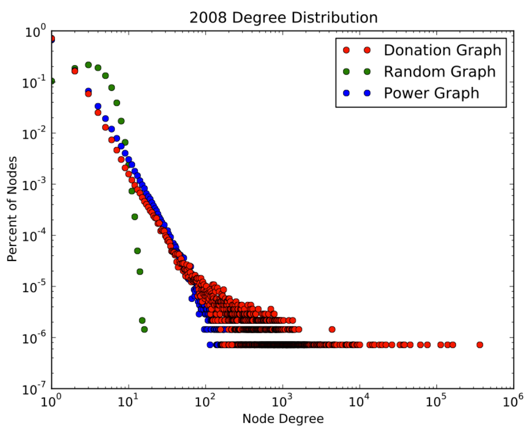
Table 1 – Structural Features of the Donation Graphs

Unique nodes are comprised of individuals, candidates, and political committees including PACs, 527 groups, and political parties (among others). The edges were calculated disregarding refund edges and edges without two known edge points, and aggregating multiple donations over the campaign finance period into a single edge. In this data, it makes sense that the maximum in degree is similar to the maximum overall degree, as there are relatively few sinks (candidates) compared to sources (individuals). The average shortest path length seems high relative to the presumed structure of this network, and suggests that a significant number of individuals donate to at least one political committee, possibly connecting them with a large strongly connected component. An average shortest path length near 1 would suggest that the majority of donations are directly from individuals to candidates. The computed values, therefore, show that there are often many steps between donor and recipient, who may not even know about the other. The breakdown of strongly connected and weakly connected components below aids this hypothesis.



This figure explains three types of donations that individuals can make. Individual Donor A donates directly to Candidate X. Individual Donor B chooses to donate only to a candidate's PAC which donates directly to Y. Individual Donor C donates only to a PAC which donates to other PACs, which donate to each other, creating cycles. Since PACs do not always support only one candidate, it possible for an individual who donates to a single PAC to have the money go to two different candidates. Note that A, B, and C are three *types* of ways that individuals can donate money, but in the actual graph individuals can donate a combination of the three ways. For example, a person can donate directly to a candidate *and* donate to a candidate's PAC to funnel more money to that candidate.

To test the distribution, we plotted the degree distribution of our graphs on a logarithmic scale, as compared to a randomly generated graph with the same number of nodes and edges, and a generated power-law graph with the same number of nodes. Figures 1a-b show the 3 distributions for both 2008 and 2010.



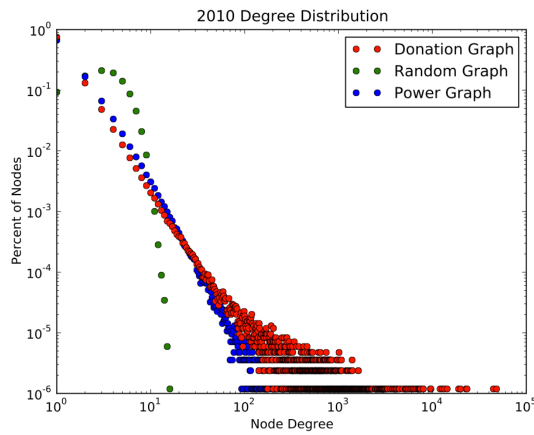
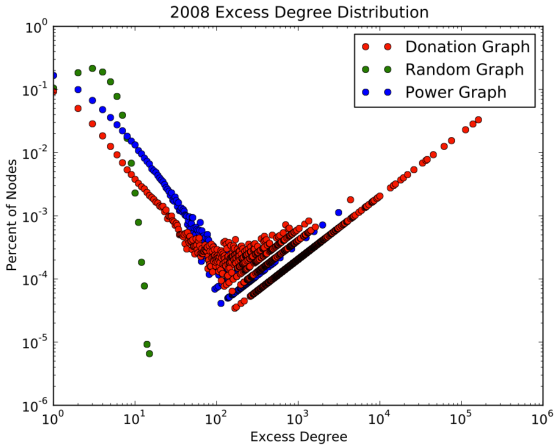


Figure 1a - 2008 Degree Distribution **Figure 1b - 2010 Degree Distribution**

The degree distribution of the donation graph is linear on the logarithmic scale while the degree of a node is small, but fans out into a wide cone at very high (> 100) degrees. This is what we expect to happen if a graph follows a power-law distribution. Being linear with a slope of -α on a log-log axis means that the actual distribution is $k^{-\alpha}$. The generated power-law graph shows a similar shape, with a slightly different coefficient. This confirms our hypothesis that the donation graph follows a power-law distribution. The 2008 graph appears to have a longer tail proportionately than the 2010 graph, which corresponds to the high degree of nodes involved at the presidential level.

Another observation relates the high expected degree of a node in the donation graph (about 3.5 for both 2008 and 2010). The majority of the nodes in the graph correspond to individuals, with candidates and committees comprising fewer, but more centralized nodes. Presumably, most individuals do not give money to many groups. As both Figures 1a-b show, the probability of a node having degree of 1 is quite high, close to 0.5. An expected degree so much higher than 1 indicates that there are some nodes with extremely high degree.

In addition to the degree distribution, we also calculated the excess degree distribution for the three graphs. This distribution was plotted on log-log axes as Figures 2a-b.



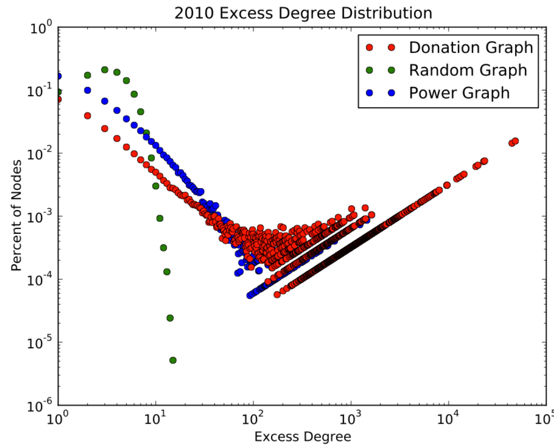
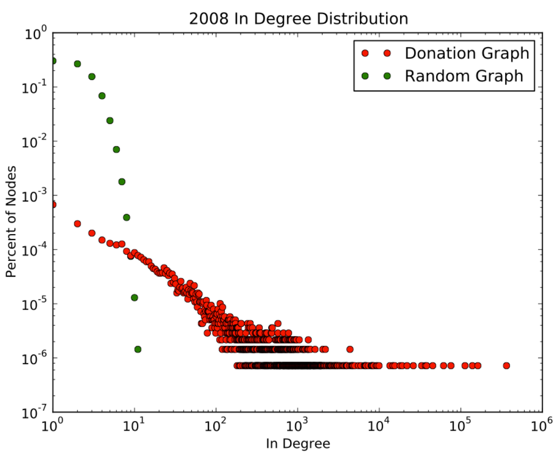


Figure 2a - 2008 Excess Degree Dist. **Figure 2b - 2010 Excess Degree Dist.**

The excess degree distribution in Figures 2a-b shows a very distinctive V shape. The point of the V is around degree K = 100. This is the same point where the degree distribution shown in Figure 1 begins to fan out.

We also calculated the degree distributions for in-degree and out-degree (Figures 3a-b and 4a-b). All clearly hold the power-law distribution shape, although it appears that the out-degree distribution fits the power-law model better. The in-degree distribution has a very long tail in 2008, which is probably associated with the presidential election year, where fundraising campaigns start earlier and have a wider reach. The random graph is shown for comparison - the generated power-law graph was undirected, so we did not include it.



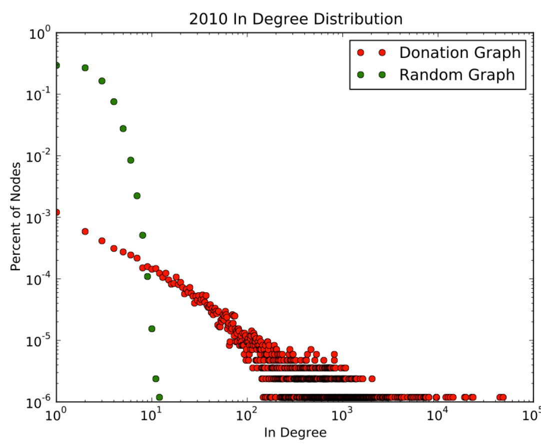


Figure 3a - 2008 In Degree Distribution **Figure 3b - 2010 In Degree Distribution**

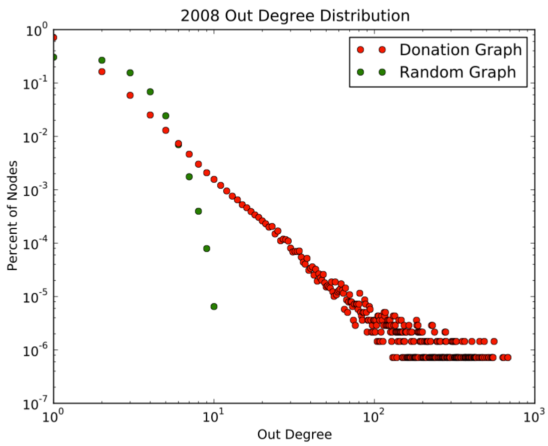


Figure 4a - 2008 Out Degree Distribution

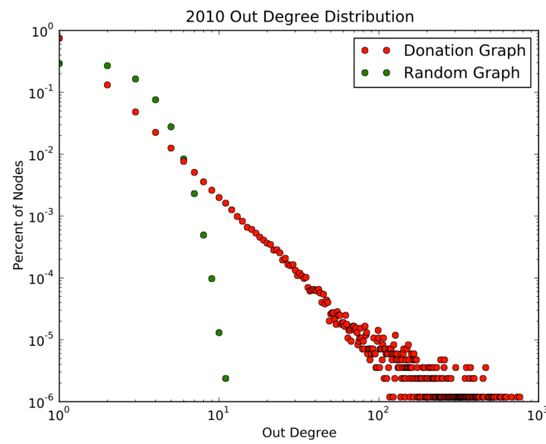


Figure 4b - 2010 Out Degree Distribution

We next determined the distribution of the sizes of connected components, both strongly and weakly connected. This is where our data diverges from the generated power-law graph, because that graph is generated as a single connected component.

Figures 5a-b show the distribution of strongly connected components for all three graphs for both 2008 and 2010. The high number of single nodes represents both sources and sinks, which in a donation graph are expected to be individual donors and candidates, respectively. PACs, 527s, and other committees are likely the nodes involved in any SCC of size greater than 1, with the largest having 1,432 nodes, in 2008, and 1,606 nodes in 2010.

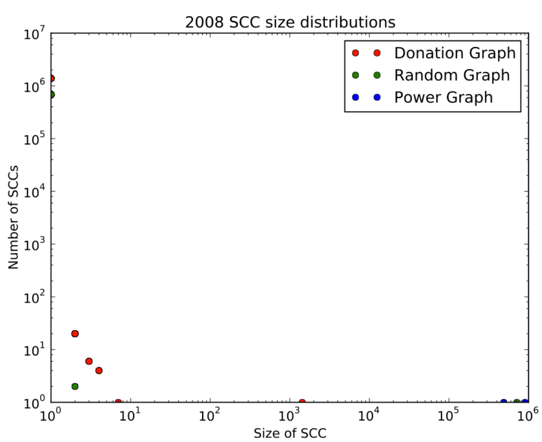


Figure 5a - 2008 SCC Size Distribution

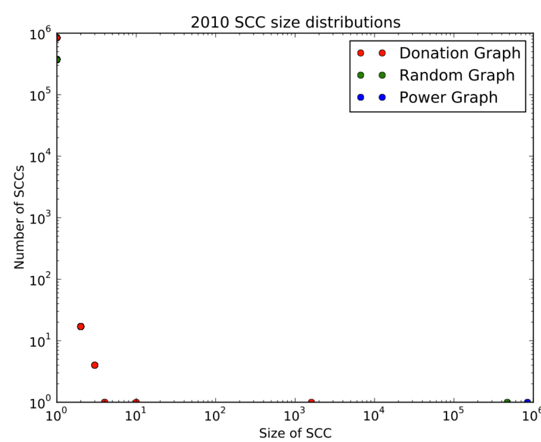


Figure 5b - 2010 SCC Size Distribution

Figures 6a-b show the distribution of weakly connected components. The power-law graph is not included because it is undirected.

Weakly connected components can display independent components that do not overlap at all. Two large weakly connected components would suggest that there are two main components that don't have any contribution edges between them. Although it is difficult to see in the graph, the donation graph's largest weakly connected component is just larger than the random, at 1.39 million nodes for 2008 and 841 thousand in 2010 (versus 1.34 million and 819 thousand for the random graph, respectively). In comparison, there are just a relative handful of nodes that do not touch the rest of the graph – these are likely PACs or candidates that did not enjoy widespread networked support with their party or other organizations.

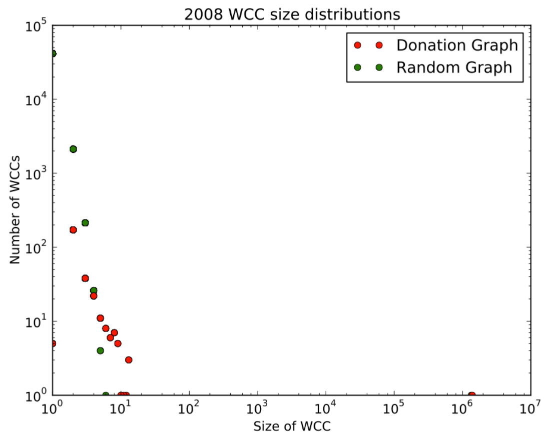


Figure 6a - 2008 WCC Size Distribution

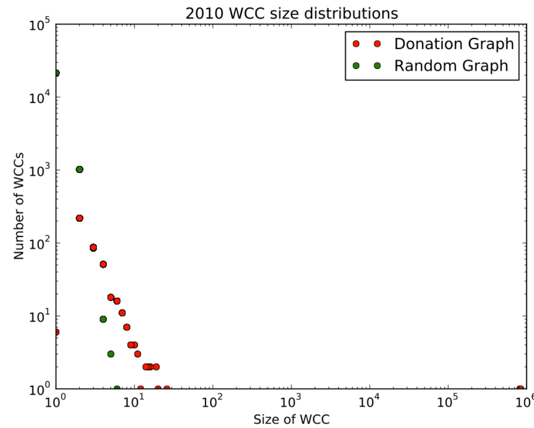


Figure 6b - 2010 WCC Size Distribution

3. Detailed Algorithm Description and Implementation

3.1 The Grassroots Election Prediction Model for Campaign Donor Networks

We developed a model, the *PageRank Grassroots Prediction Model,* in which we represent *directed support* as a function on the amount of money given. The most basic version of this prediction model would weigh all edges with the value 1, and predict that the candidate with the most number of incoming edges, or the highest in-degree would win the election.

A more nuanced, realistic version of the grassroots prediction model involves applying a function over all the edges in the graph to represent diminishing marginal increases of support per donated dollar, which edits PageRank adjacency matrix. That is, grassroots support is not the only important factor, but neither is only money.

Concretely, a donation of $1 to a candidate may be worth 1 unit of support, and a donation of $100 may be worth only 10 units of support. We then use the PageRank algorithm over the modified graph, where the initial rank, hereby referred to as the "support score," is determined by the amount of positive net outgoing dollars. For example, individuals who are not candidates do not "receive" donations, so their initial support score is the sum of their outgoing donations. Similarly, PACs that have larger outgoing donations then total incoming also have a non-zero initial support score. The nodes that merely transfer money have an initial support score of 0.

Since we use the power iteration method, we normalize the outgoing support score for each node and enter that to our probability matrix. To account for total differences in donation for each node (i.e, someone who gives $1000 to each candidate as opposed to someone who gives $1 to each candidate), we initialize the support score for individuals as their total outgoing scores divided by the total unique outgoing dollars by all individuals and PACs with unaccounted for donations.

Then the winner of an election is predicted based on the number and weights of political donations using each model. By applying this model on multiple data sets (from 2008, 2010) we can compare how our model generalizes to different Presidential elections and years without a Presidential election.

3.2 Modifying PageRank to Implement The Grassroots Prediction Model
The PageRank algorithm is used and described in the original PageRank paper [8].

At a high level, PageRank works with an adjacency matrix *A* which represents the probability that one node hops to another node and an initialized support score vector s_0 which is multiplied by the adjacency matrix repeatedly until the vector s_0 converges. We used the random walk model of PageRank and the power-iteration method to determine the support score for the candidates. In order to get our algorithm to converge, we needed to focus on making the graph irreducible.

Due to the tree-like structure of our graph, the graph edges tend to flow in one direction toward the candidate nodes which are sink nodes or "dead ends." To fix this, we had to modify the "random jumps" hereby referred to as "teleporting" so that teleports happened only to individuals and PACs with unaccounted-for donations. We also had to consider how to apply the grassroots model over all the edges, or over just the initialization of the support score vector.

We used a modified version of PageRank called TrustRank [2], where the seed nodes in our graph (the ones to which others can teleport) are individual nodes and candidates with a positive net outgoing dollar amount since they received this money from anonymous individual donors who were not listed.

We made two major modifications to the TrustRank algorithm:
1) We set the initialization vector s_0 based on the total of outgoing donations for individuals and PAC's "leftover money."
2) We applied different functions over all the edges to represent diminishing marginal "support" for each dollar, which we used to evaluate our Grassroots Support hypothesis.

In order to set the initialization vector, we need to calculate the *total unique outgoing dollars* t_d and the *unique outgoing dollars of node i,* d_{u_i}. For each node *i* in the graph, individual or PAC), if its net outgoing dollars d_{u_i} is positive, their initialized "support" is $\frac{d_{u_i}}{t_d}$.

Specifically,

$$d_u = \sum dollars_{outgoing} - \sum dollars_{incoming}$$

$$t_d = \sum_{i \in N} d_{u_i}$$

If the sum was negative (or if the PAC gave less money than it received), we set d_u to 0.

To modify A, we applied a function $f(x) = x^a$ where we varied a between -1 and 2 over all the edges. Then, to calculate the probability that a node would get from node *i* to node *j*, we calculated the weight of the edge from node *i* to node *j* divided by the sum of the weights from node *i* to node *k* where *k* is a successor of *i*. In other words,

$$p_{i \to j} = \frac{e_{i \to j}}{\sum_k e_{i \to k}}$$

where $e_{i \to j}$ represents the dollar amount from node *i* to node *j*.

After modifying the adjacency matrix and initializing s_0, we used the power iteration method to calculate the rank, hereby referred to as a "support score." The power iteration method does the following until convergence:

$$As_i = s_{i+1}$$
$$then$$
$$As_{i+1} = s_{i+2}$$

Once PageRank converges, we have a final set of support scores. For each race, we predict the winner is the one with the highest support score.

3.3 Programming Implementation Details

To implement the modified PageRank algorithm using NetworkX, we modified the source code for the method used in the NetworkX library for PageRank and the creation of a right stochastic graph.

4. Prediction Model Results

4.1 How to Evaluate the Success of our Algorithm

To evaluate the success of our algorithm, we ran the following two versions of our prediction model:

> **PageRank Grassroots Model A:** Applying a power function only on initial support scores, then running PageRank to determine final support score and comparing that with other candidates in the same race to determine the winner.
> **PageRank Grassroots Model B:** Applying a power function modifying all edges and initial support scores, then running PageRank to determine final support score and comparing that with other candidates in the same race to determine the winner.
> Recall that PageRank Grassroots Model A applies the function to all unique outgoing edges and thus recalculates the total unique outgoing dollars and the initial importance and support of individuals and PACs.

In addition to what PageRank Grassroots Model A does to initial support scores, PageRank Grassroots Model B uses the applies the same function on edges to change the probability adjacency matrix that determines possible-next-hops.

The PageRank Grassroots Models both use the TrustRank method of teleporting, where every node in the graph sets restarts to either individuals (the beginning of a donation flow) or a PAC that has unaccounted money (representing anonymous individuals).

We compared the with the results of the following prediction models:

> **Classic PageRank Model:** Using a standard implementation of PageRank to determine support scores, without adjusting teleportation or taking any of the edge weights into account.
> **Naïve Greedy Model:** Using only total received donations to determine the winner of an election.
> **Naïve Grassroots Model:** Using a candidate's in-degree relative to other candidates in the race to determine the winner of an election.

The classic PageRank model serves as a baseline to see if using the donations is at all a necessary measure to determine prediction accuracy.

The Naïve Models represent two simple, quick approaches that focus solely on either only money or only number of contributors as measured by in-degree. If a simple, fast algorithm can make predictions with a success rate approaching that of PageRank's, then there is little incentive to spend the extra time running the PageRank algorithm. We tested two such simple algorithms, one that predicted winners by whoever had received the most money, and one that predicted winners by whoever had the most different donors (only counting direct donations, so PACs only counted as 1 donor).

4.2 Results in a Table (Page 11):
The highlighted grey percentages are those with the highest accurate prediction percentage. "Republican Accuracy" denotes that given a Republican won the seat, our algorithm correctly predicted the winner. "Democratic Accuracy" does the same, except given a Democrat won the seat. Overall accuracy evaluates the percentage of candidates we predicted correctly.

2008 ELECTION RESULTS

PageRank Grassroots Prediction Model A	Republican Accuracy	Democrat Accuracy	Overall Accuracy
f(x) = x^0	.8238	.9291	.8866

PageRank Grassroots Prediction Model B (edge)	Republican Accuracy	Democrat Accuracy	Overall Accuracy
f(x) = x^{-2}	.9016	.8582	.8761
f(x) = $x^{-1.5}$	.8187	.9220	.8803
f(x) = x^{-1}	.8187	.9220	.8803
f(x) = $x^{-0.5}$	.9482	.8227	.8739
f(x) = x^0	.9120	.8121	.8529
f(x) = $x^{0.5}$	.8549	.8830	.8718
f(x) = x^1	.8238	.9291	.8866
f(x) = $x^{1.5}$	.8135	.9255	.8803
f(x) = x^2	.8031	.9255	.8761
f(x) = log (x)	.9067	.8156	.8592

Comparison Algorithms	Republican Accuracy	Democrat Accuracy	Overall Accuracy
Original PageRank	.9171	.8120	.8550
Naive Greedy (money only)	.5564	.8714	.6939
Naïve Grassroots (in-degree only)	.0489	.0190	.0356

2010 ELECTION RESULTS

Pagerank Grassroots Prediction Model A (rank)	Republican Accuracy	Democrat Accuracy	Overall Accuracy
f(x) = x^{-2}	.7895	.8714	.8260

Pagerank Grassroots Prediction Model B (edge)	Republican Accuracy	Democrat Accuracy	Overall Accuracy
f(x) = x^{-2}	.7857	.8809	.8260
f(x) = $x^{-1.5}$	.7970	.8762	.8302
f(x) = x^{-1}	.8083	.8952	.8449
f(x) = $x^{-0.5}$	.8008	.9000	.8428
f(x) = x^0	.8120	.8905	.8470
f(x) = $x^{0.5}$	.7932	.8857	.8344
f(x) = x^1	.7895	.8714	.8260
f(x) = $x^{1.5}$	.7744	.8524	.8092
f(x) = x^2	.7857	.8238	.8029
f(x) = log (x)	.8045	.8810	.8386

Comparison Algorithms	Republican Accuracy	Democrat Accuracy	Overall Accuracy
Original PageRank	.8158	.8810	.8449
Naive Greedy (money only)	.7782	.9429	.8512
Naïve Grassroots (in-degree only)	.0263	.0333	.0294

4.3 Explaining the Results

The results from running our modified PageRank algorithm and other algorithms showed that the PageRank algorithm was very good at predicting the winners of elections. Both in 2008 and 2010, PageRank correctly predicted over 80% of all elections correctly, but the results are much more interesting upon closer inspection of the success of our algorithm with respect to each party's candidates. Model A outperformed Model B in 2008, but underperformed Model B in 2010. In both years, unmodified PageRank was roughly as good as either Model A or Model B. The biggest change between elections was the naive greedy strategy was much worse than PageRank in 2008, but was the best predictor of election results in 2010. This difference could be the result of 2008 being a general election versus the 2010 midterms. Another factor could be the change in campaign finance law after the *Citizens United v. Federal Election Commission* case in early 2010.

We tested two differing models of PageRank at various parameter values. PageRank Grassroots Model A always had the same rate of success, no matter what function was applied to the initial scores. This makes sense because the initialization vectors should not matter so much, as the adjacency matrix will predominantly determine the support scores.

For the 2008 election, Model A correctly predicted 82.38% of all Republican victories and 92.91% of all Democratic victories, an overall success rate of 88.66%. PageRank Grassroots Model B had slight variations in success depending on the function applied to the donation weights and initial scores. When $f(x) = x$, these results were exactly the same as those from Model A. For other function values, Model B had a consistently slightly lower success rate overall than Model A. However, unlike in Model A, Model B was sometimes more successful at predicting Republican victories than it was at predicting Democratic victories.

The unmodified version of PageRank had a slightly lower success rate than either one of our models. The unmodified PageRank was more successful at predicting Republican wins than Democratic ones. The greedy approach was not nearly as accurate as our PageRank algorithms, only correctly predicting 69.39% of all elections correctly for 2008. However, most of the inaccuracy comes from only correctly predicting 55.64% of all Republican wins. Democratic wins are still correctly predicted 87.14% of the time, a rate comparable to PageRank's success rate. The success rate of the second simple algorithm was incredibly poor, correctly predicting less than 5% of all Republican wins and under 2% of all Democratic wins.

While the results from 2008 indicated that PageRank was a much better predictor of election results than simply choosing the candidate with the most money, the results from 2010 show the opposite. Model A correctly predicted 78.95% of all Republican victories and 87.14% of all Democratic victories, for a total success rate of 82.60%. As in 2008, Model B when $f(x) = x$ produced the same results as Model A, with small variations in success rate across the different functions. However, unlike in 2008, Model B in 2010 always had a higher success rate with Democrats than with Republicans. Also unlike the results for 2008, Model B in 2010 was more successful at predicting election results than Model A, except for when $f(x) > x$. In 2008, the maximum success rate of Model B was when $f(x) = x^{\pm 1.5}$. In 2010, Model B's success rate declined as $f(x)$ increased.

The unmodified PageRank algorithm had a higher success rate in 2010 than either of our models, with the single exception of our Model B algorithm when donation sizes were ignored (i.e. $f(x) = 1$). Even then, Model B outperformed the unmodified PageRank algorithm by less than 0.25%. Outperforming all versions of PageRank was the greedy approach that selected the candidate with the most money. This greedy approach correctly predicted 77.82% of all Republican winners and an amazing 94.29% of all Democratic winners. The second simple algorithm that only counted the number of incoming edges was once again a very poor predictor, getting fewer than 3% of all races correct, although this time it was slightly better with Democrats than with Republicans.

The reasons behind these mixed results could lie in the nature of the election years we tested. 2008 was an election where Democrats won the presidency as well as large majorities in both the House and the Senate. 2010 was a year where conservative Republicans retook control of the House of Representatives and almost won a majority in the Senate. In between these two elections, the United States Supreme Court ruled that corporations and unions could donate unlimited amounts of money to political campaigns in the landmark case of *Citizens United*. This controversial decision enabled the creation of many new political campaigning entities designed to take advantage of the new laws regarding donation limitations and disclosure. Many Democrats blame this decision and the resulting political spending by groups such as American Crossroads for their losses in the 2010 election. In the upcoming election in 2012, both political parties have vowed to increase their use of "Super PACs" and unlimited campaign funding.

While *Citizens United* was generally seen as more beneficial to Republicans, our results show that the electoral landscape as a whole was much more determined by money after the decision than in the previous election cycle. Predicting the candidate with the most money as the winner was more successful than any PageRank algorithm we tested in 2010. For Republicans, the success rate increased over 22% from 55% to 77%, for Democrats, the success rate increased over 7% from 87% to 94%. This may suggest that Republicans did indeed benefit from the additional political contributions. Despite this, the greedy strategy successfully predicted Democratic wins with a higher success rate than it predicted Republican wins. Partially this is because there were so many Democrats incumbents. Incumbents typically have much larger campaign war chests, particularly in the Senate, where reelections are only every six years instead of every two years. Also, because of how few Democratic candidates won elections, those that did were primarily in safe districts, where they faced weak, poorly funded opposition.

In both election years, the strategy of picking the most well-financed candidate was more successful at predicting Democratic wins than Republican wins. In 2008, this strategy still underperformed PageRank in predicting Democratic victories. By 2010, this strategy outperformed PageRank in predicting Democratic winners, even though it still slightly underperformed PageRank in predicting Republican victories.

5. Conclusion

PageRank algorithms worked very well in both 2008 and 2010, with our modified PageRank algorithms slightly outperforming unmodified PageRank in 2008 and slightly underperforming basic PageRank in 2010. The simple strategy of picking the candidate with the most money was far worse than PageRank in 2008, but was better than PageRank in 2010. Political cynics and campaign finance reformers will see this as confirmation of their worst fears about American politics, that votes and elections can be bought.

It is still unknown how well PageRank algorithms will predict election results against a cynical strategy in the future. It has been less than two years since the *Citizens United* decision, and trends such as the Occupy movement pushing for campaign finance reform are even more recent. Future work on the appropriateness of PageRank for predicting election results should test PageRank on many more elections. Comparing the results from 2010 to 2008 alone is not enough to say that money is now all that matters. 2008 was a presidential election year, and one that saw the election of America's first African American president, whose campaign brought in unprecedented amounts of young and minority voters. In contrast, 2010 was a more or less typical midterm election year, where the amount of donations and the size of the electorate is much smaller than in presidential elections. PageRank performed well in 2008, and testing on additional elections preceding that year will confirm if PageRank was a good algorithm before *Citizens United* or whether 2008 was an outlier year when PageRank performed exceptionally well compared to the greedy strategy. As campaign donation data for the 2012 campaign becomes available, testing on the new data will be a better comparison to 2008 when determining the impact of *Citizens United*.

Other areas for future study is refining of the dataset. Comparing PageRank and the greedy strategy in races where they are almost certain to make the same prediction, such as a safe seat where the incumbent always wins easily is not efficient. Research into the performance of these strategies in highly competitive races would reveal to what extent our political system is "one dollar, one vote."

References

1. Dominguez, C. Groups and the Party Coalitions: *A Network Analysis of Overlapping Donor Lists.* Meeting of the American Political Science Association, 2005. <http://home.sandiego.edu/~caseydominguez/groupspartycoalitions.pdf>.
2. Gyöngyi, Z., Garcia-Molina, H., and Pedersen, Jan. 2004. *Combating Web Spam with TrustRank*. Proceedings of the Thirtieth international conference on Very Large Data Bases. Volume 30.
3. Heckerman, D., Meek, C., and Koller, D. Probabilistic models for relational data. Technical Report MSR-TR-2004-30, Microsoft Research, 2004.
4. Kaelin, F. and Precup, D. An Approach to Inference in Probabilistic Relational Models using Block Sampling. JMLR: Workshop and Conference Proceedings, 13: 315-330. http://jmlr.csail.mit.edu/proceedings/papers/v13/kaelin10a/kaelin10a.pdf
5. Koger, G. and J. Victor. *The Beltway Network: A Network Analysis of Lobbyists' Donations to Members of Congress.* Meeting of the American Political Science Association, 2009. <http://opensiuc.lib.siu.edu/pn_wp/32/>.
6. OpenSecrets.org: Center for Responsive Politics. <http://www.opensecrets.org/>
7. PageRank Analysis of Graph Structure. <http://networkx.lanl.gov/svn/networkx/trunk/networkx/algorithms/link_analysis/pagerank_alg.py>
8. Page, L., Brin, S., Motwani, R., and Winograd, T. 1999. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report. Stanford InfoLab.
9. Small Donors and Online Giving: A Study of Donors to the 2004 Presidential Campaigns. <http://www.cfinst.org/pdf/federal/president/IPDI_SmallDonors.pdf>.