Evaluating a Grassroots PageRank Model for Predicting Elections from Political Contributions

Ben Lasley, Ben Shapero, and Charlton Soesanto

CS 224W: Social and Information Networks

1 Aim

We wish to use a modified version of the PageRank Algorithm to rank candidate support based on contributions received and number of supporters.

Such an algorithm can quantify the importance of money versus the number of donors in predicting the winner of an election.

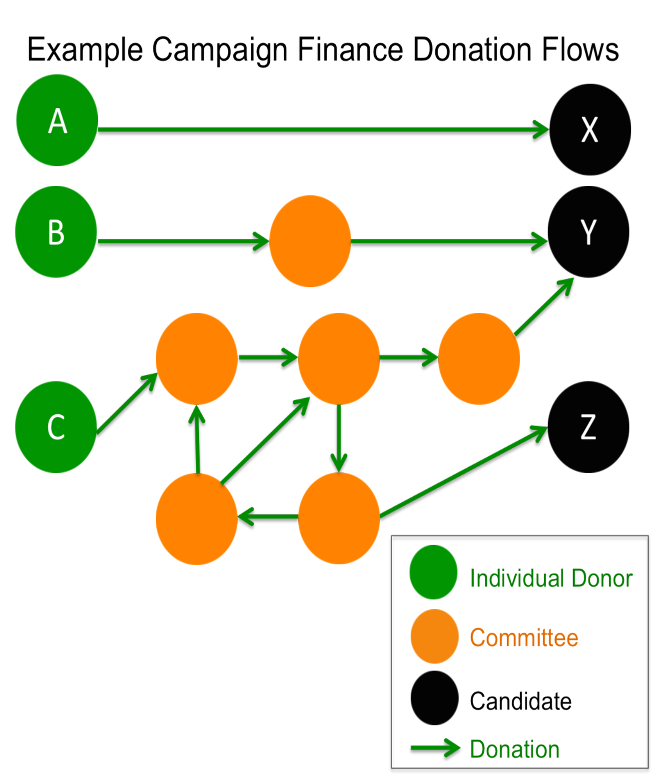

Figure 1: Example ways in which an individual's money can reach a candidate.

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.

2 Political Donation Network Structure

	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

We compared the structure of the donation network to $G_{n,m}$ random and power-law generated graphs of the same size.

- The average shortest path length demonstrates that on average, **individual donations go through many intermediary groups before reaching an individual candidate.**
- **Both 2008 and 2010 resemble power-law graphs for degree distribution**, including in and out degree distributions.
- Both **had strongly connected components of over 1,000 nodes**, which can only happen amongst committees, as individuals are sources and candidates are sinks.
- Both had weakly connected components of sizes exceeding the random distribution, suggesting that **a large number of donors feed into the strongly connected component of committees**.

3 Model and Algorithm

We developed a model, the *PageRank Grassroots Prediction Model,* in which we represent *directed support* as a function on the amount of money given.

We quantify support as both a function of number of donors and dollar amounts received.

For example, 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.

PageRank Grassroots Model: Applying a power function only on initial support scores and 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.

Initial Support score: $$d_u = \sum dollars_{outgoing} - \sum dollars_{incoming}$$

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

Function applied over edges: $f(x) = x^a \; where \; -0.2 < a < 2$

Calculating the probability adjacency matrix: $p_{i \to j} = \dfrac{e_{i \to j}}{\sum_k e_{i \to k}}$

We then use the PageRank power iteration method to calculate the "support score."

The candidate in a given race with the highest support score is the predicted winner.

We compared results with the following 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.

4 Results

2008 ELECTION RESULTS	Republican Accuracy	Democrat Accuracy	Overall Accuracy
Model A: **f(x) = x^0**	**.8238**	**.9291**	**.8866**
Model B: **f(x) = x^1**	**.8238**	**.9291**	**.8866**
Comparison Algorithms			
Original PageRank	.9171	.8120	.8550
Naive Greedy (money only)	.5564	.8714	.6939
Naïve Grassroots (in-degree only)	.0489	.0190	.0356

2010 ELECTION RESULTS	Republican Accuracy	Democrat Accuracy	Overall Accuracy
Model A: **f(x) = x^0**	**.7895**	**.8714**	**.8260**
Model B: **f(x) = x^0**	**.8120**	**.8905**	**.8470**
Comparison Algorithms			
Original PageRank	.8158	.8810	.8449
Naive Greedy (money only)	.7782	.9429	.8512
Naïve Grassroots (in-degree only)	.0263	.0333	.0294

5 Discussion and Future Work

Discussion:
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.

Outperforming all versions of PageRank in 2010 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.

Future Work:
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*.