Empirical Analysis of Networks of Contributors in American Electoral Campaigns in the 1990

Jakub Langr

April 21, 2013

Contents

List of Figures

1 Introduction

American politics was the first one to run such huge budgets merely on campaigning. A lot of interesting literature has been written on the matter, everything from how Barack Obama has recently used social networks to his advantage to lack of empirical testing in the use of the accumulated money. One of the aspects, that are tremendously interesting is partisanship hidden in the networks. While politicians frequently, on the outside, purport their strong adherence to the core values, these actions might be guided by someone completely different – contributors. This paper will examine their influence.
Just a note to the reviewer, however, I would very much appreciate any additional comments you might have as the rubric, although meaningful for grading purposes, does not give me any feedback. If I happen to grade your paper, I would be delighted to reciprocate. So feel free to send any thoughts to: james.langr@gmail.com; thank you!

2 Statistical methods

2.1 Data

The data was downloaded from

$$http://datacommons.s3.amazonaws.com/subsets/td-20120424/contributions.fec.1990.csv.zip$$

and unzipped into a 223 MB file, which turned out to be a bigger issue than it previously seemed. This was a nice, pre-packaged record of all contributions to politicians in the year 1990: everything from an individual contributors to large corporations.
I chose the 1990 dataset, because of it (a) smaller size, (b) meaningful insights, as the newer datasets are more frequently discussed. The data consistent of almost 800,000 datapoints and had 42 variables, some of which were empty, so the data had to be cleaned. This is an overview of the variables:

Overview: All variables in the original dataset

[1]	"*id*"	"*import_reference_id*"
[3]	"*cycle*"	"*transaction_namespace*"
[5]	"*transaction_id*"	"*transaction_type*"
[7]	"*filing_id*"	"*is_amendment*"
[9]	"*amount*"	"*date*"
[11]	"*contributor_name*"	"*contributor_ext_id*"
[13]	"*contributor_type*"	"*contributor_occupation*"
[15]	"*contributor_employer*"	"*contributor_gender*"
[17]	"*contributor_address*"	"*contributor_city*"
[19]	"*contributor_state*"	"*contributor_zipcode*"
[21]	"*contributor_category*"	"*organization_name*"
[23]	"*organization_ext_id*"	"*parent_organization_name*"
[25]	"*parent_organization_ext_id*"	"*recipient_name*"
[27]	"*recipient_ext_id*"	"*recipient_party*"
[29]	"*recipient_type*"	"*recipient_state*"
[31]	"*recipient_state_held*"	"*recipient_category*"
[33]	"*committee_name*"	"*committee_ext_id*"
[35]	"*committee_party*"	"*candidacy_status*"
[37]	"*district*"	"*district_held*"
[39]	"*seat*"	"*seat_held*"
[41]	"*seat_status*"	"*seat_result*"

2.2 Data processing

The data was loaded to GoogleRefine, after numerous problems with memory, only to find out that the file was, for the most part, cleaned. The only major transformation necessary was to remove non-positive donations, which is an odd artifact of the dataset. So I returned to `R 2.15.1`, where the file was cleaned of unnecessary columns or those that were too incomplete to provide any meaningful insight. Next, the data was converted to graph, so that it can be further exported to Gephi `0.8.2 beta`, which has better tools for network analysis. As for creation of the graph, the nodes were the contributors and the edges were contributions. The transformation was done thanks to `igraph`. Raw code for the transformation can be found in the appendix. Typesetting was done in LaTeX.

2.3 Exploratory Analysis

The file was loaded into Gephi. The graph file produced in `R` was already reduced to only approximately 85 MB in size by (a) removing unnecessary variables (b) removing heavily incomplete variables or (c) removing small and medium contributors. Small contributors were those who made “less than 100 contributions". It was interpreted as follows: nodes: candidates/contributors; edges: contributions. It was loaded as a graphML file with attributes.

This dataset was then colored by party, nodes differentiated by size by degree so that the larger ones can be easily spotted. Layout ForceAtlas2 was used to visualize the results better.

2.4 Hypotheses and Explanation

Let me firstly say the whole hypothesis and then give full explanation: **Contributors will tend to donate to candidates of both parties as to hedge against the risk of their favorite losing and hence will give money both to Republicans and Democrats; hence these donors gain influence over the candidate regardless of income.**
Now, there are number of obvious issues with this:

1. What if these large donors genuinely want to donate to both parties, because few have preferences aligned perfectly along party lines and so it is normal to give money to both since you agree with both on some points? If that is the case, then this supports the hypothesis, because donors are not so focused on the party divide as they are focused on the policy.

2. Does money actually mean influence? Now, this is hard to measure, but it is the assumption of the model. After all, we always hear "follow the moneyrq', right?

3. Why only the large contributors? Because only the people who have made more than 100 donations are (a) wealthy enough to actually influence any given candidate – because only they are important enough so that candidates care, (b) they have provided us with sufficient data to actually show some partisanship or lack thereof.

Hence the variables I am going to be interested in are only: (a) candidate-donor connections, (b) party of the candidate, (c) InDegree of the candidate, (d) donor's consistency in giving to candidates.

2.5 Metrics

Now that I have set out what we want to know, how does this hypothesis actually translate into measurable $a - priori$ metrics that we could check by testing the dataset? I have identified the following:

1. High degree of clustering: we would expect the average clustering coefficient to be at least above 0.02, otherwise the null-hypothesis (i.e. contributors hold party line) is true. Average clustering coefficient for directed graphs, such as ours is defined as:

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

For a directed graph, e_{ij} is distinct from e_{ji}, and therefore for each neighbourhood N_i there are $k_i(k_i - 1)$ links that could exist among the vertices within the neighbourhood (k_i is the number of neighbors of a vertex). This definition as well as further resources can be found at Wikipedia. (See: Clustering Coefficient)

2. We would expect a small average shortest path (ASP) to be below 3.5, as we expect this network to perfectly mimic a small world network, as contributions should, in theory, mimic population. Hence we would expect a low ASP as is in the general population.

3. We would expect the connected component to be over 95%: there will be no niche donors who only want to donate to a small set of people, but will donate to most.

4. Relatively high average degree: again, we would expect that the degree would be over 20.

The last condition almost follows from the previous three, but we can leave it there, just as another benchmark.

If all of the above hypotheses are true, then we would proclaim the initial hypothesis true and conclude that donors genuinely consider the influence they can get from a candidate over party considerations.

3 Results

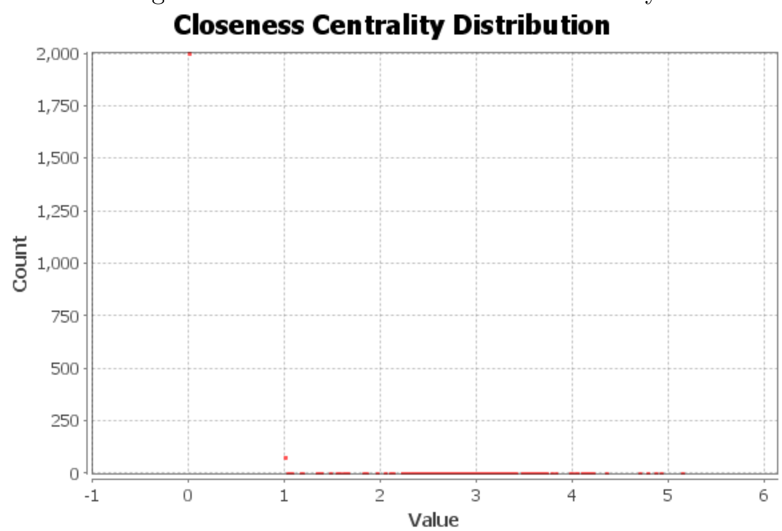
3.1 Visual

This section includes interesting results, not just related to the hypothesis:

Figure 1 is the distribution of closeness centrality: on the x-axis is the number of nodes with corresponding number of . It seems generally hard to interpret, however, whether it supports or goes against the hypothesis, since it actually hard to compare. A computer model to simulate the two scenarios for closeness centrality would be necessary in order to judge this. I am just adding it for completeness.

Figure 2 is only a PNG screen-shot added for convenience, because the full vector graphics image takes long to load; however, it is added at the end of this PDF for you to explore. Figure 2 is the map of the network with contributions to Democrats in blue, contributions to Republicans in red and contributions to Independents in yellow. To visualize it I used ForceAtlas 2; yet, there are no visible clusters–with the notable exception of "lonesome" anonymous donors. Those cannot be identified as individual entities and hence have contributions to both Democrats and Republicans (more often the latter than the former). One interesting snippet is that highest InDegree goes to $MaxBaucus(D)$ and highest outdegree goes to $AT\&T$.

Figure 1: Distribution of Closeness Centrality



More interesting facts are that $119,631$ of these 160,000+ contributions were to a winning candidate and most generous district in 1990 was $OR-05$. Democrats, however, overall received more donations from these "elite" donors: $95,626$, as compared to Republican's $68,230$. (The rest went to independents).

Figure 3 is a vectorized version of dollar amounts of donations. Republican (in red) and Democrat (in blue) donations are overlayed in a rough form of points. Be warned however, the figure is misleading, for instance, because not all donations in the right part of the curve are Republican! The color just dominates so the blue points seem to be missing.

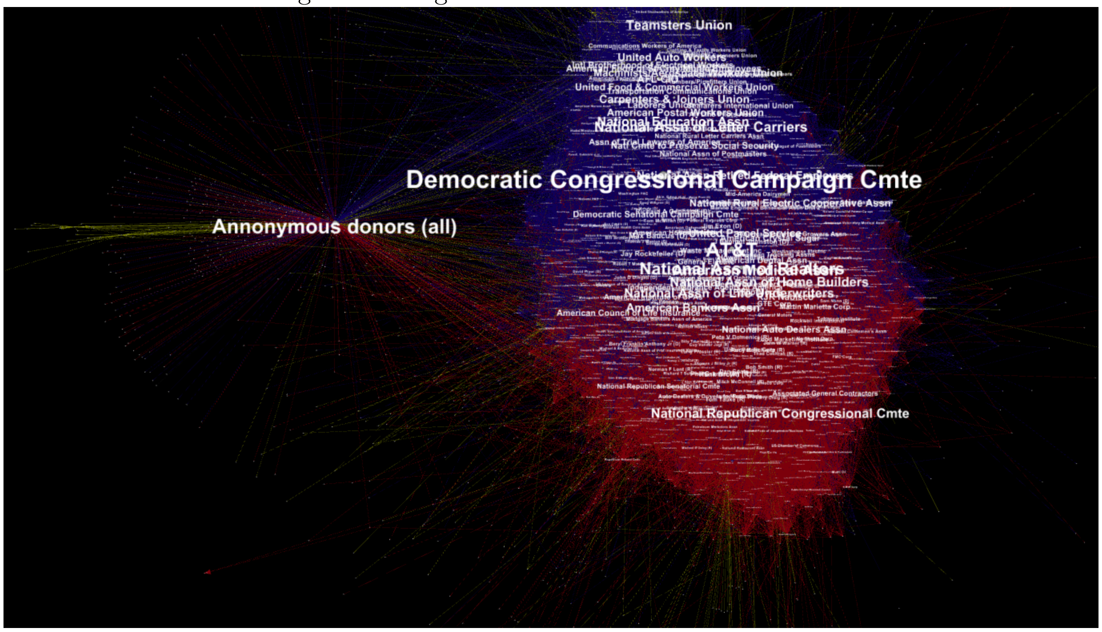
3.2 Numeric

For convenience, the results are summarized in the following table:

Table 1: Summary of results

Variable	Result	Fulfilled criteria?
Number of nodes	2530	–
Average clustering coefficient	0.032	Yes
Average shortest path	2.879	Yes
Connected component, rel	0.977	Yes
Connected component, abs	2472	Yes
Average degree	31.243	Yes

Figure 2: Image of the network of contributions



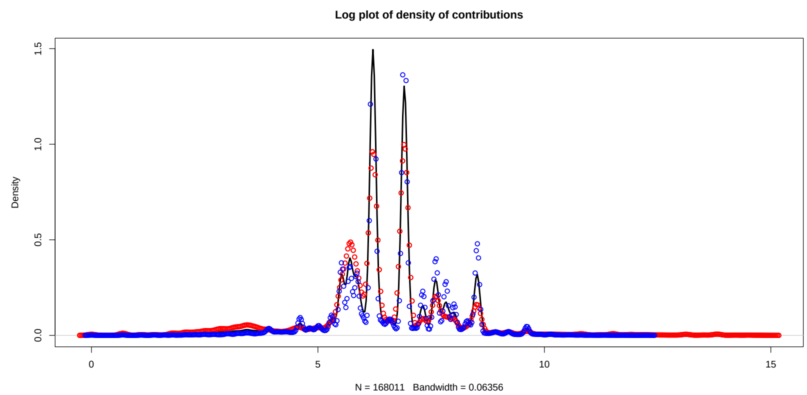
3.3 Hypothesis

From information above, we can conclude, that the hypothesis seems to be true. Admittedly, this was very superficial analysis, but the predictions were tested and it seems that it would be valid. Yet, I have to admit that choice of some of them was rather arbitrary–clustering coefficient could be much different. Again, this is because the metric is not very intuitive and to test our intuition an agent-based model would have to give us the respective coefficients for models that would be pre-programmed with or without hedging. Other than that, I think that it is clear from the picture that there are no two distinct communities–nor was it possible to separate the two parties in any meaningful way. Ergo, there is some overlap, though, obviously, it is not a seamless transition, simply because we would expect some degree of partisanship even with hedging.

4 Conclusion

In the end, the hypothesis seemed to be confirmed and it seems that even as far back as in 1990s contributors were looking to be always on the winning side. Obviously, the benefits are not clearly visible, but simply giving a lot of money to a candidate, I would imagine would surely secure you at least one meeting with him when he finally reaches office to plead your case for whatever might be troubling you. This sort of political quid-pro-quo was obviously happening ever since the beginning of politics and hence is hardly scandalous. It just

Figure 3: Log distribution of donations - USD amounts

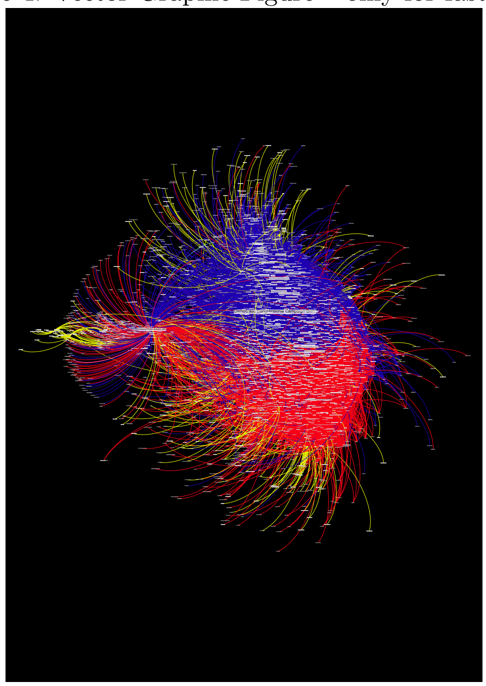


interesting to note that people are aware that perhaps money does buy you friends and those friends then can help you. The unsurprising, but perhaps more interesting fact is that companies are purely calculating (as well as the fact that they can contribute at all!, but that is for another discussion), for instance, *AT&T* seems to have no partisan allegiance: of its 1,578 independent contributions, 886 went to Republicans and 673 went to Democrats. A perfect example of hedging.

Also note that bellow the conclusion is the vector-graphic figure, but be warned–only fast PCs will display it reasonably.

5 Appendix A: Vector Graphic Figure

Figure 4: Vector Graphic Figure – only for fast PCs!



6 Appendix B: Unrefined R Code

Note that the data were loaded as "cont".

```
FreqCon <- table(cont$contributor_name)
FMI <- names(FreqCon)[FreqCon > 100]
LC <- subset(cont, contributor_name %in% FMI)

LC[,1] <- LC$contributor_name
LC[,2] <- LC$recipient_name

LC = LC[,c(1,2,3,5,13,14,15,17,20,22,24,25)]
GLC <- graph.data.frame(LC)

write.graph(GLC,"Large_donations.graphml",format="graphml")

plot(density(log(LC$amount)),lwd=2.5, main="Log plot of density of contributions")
lines(density(log(LC$amount[LC$recipient_party=="R"])),type='p',pch=21,lwd=1.5,col="red")
lines(density(log(LC$amount[LC$recipient_party=="R"])),type='p',pch=21,lwd=1.5,col="red")
lines(density(log(LC$amount[LC$recipient_party=="D"])),type='p',pch=21,lwd=1.5,col="blue")
dev.copy2pdf(file="Density_plot.pdf")
```

The rest was done in Gephi, using standard GUI.