# Predicting Congressional Votes Based on Campaign Finance Data

**Samuel Smith, Jae Yeon (Claire) Baek, Zhaoyi Kang, Dawn Song, Laurent El Ghaoui, Mario Frank**
Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

## Abstract

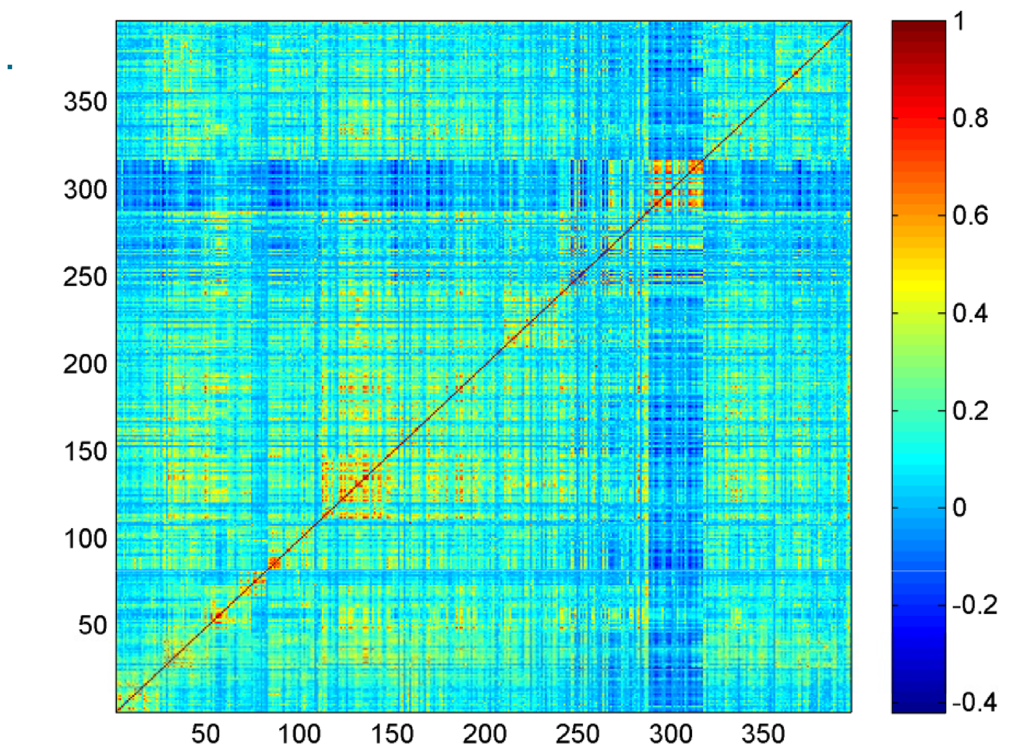
- Analysis of how campaign contributions influence voting in Congress.
- High accuracies achievable for predicting Congress members votes by their received donations.
- However, party line is even better predictor.
- Party is a variable that influences both voting behavior and donations sources.

## Background

- Political campaign contributions for Congress members are heavily disputed.
- Nearly limitless corporate funding permitted thought the *Citizens United* Supreme Court decision.
- Nearly $6 billion spent on the 2012 US federal election, over $2.5 billion on the Congressional races alone.

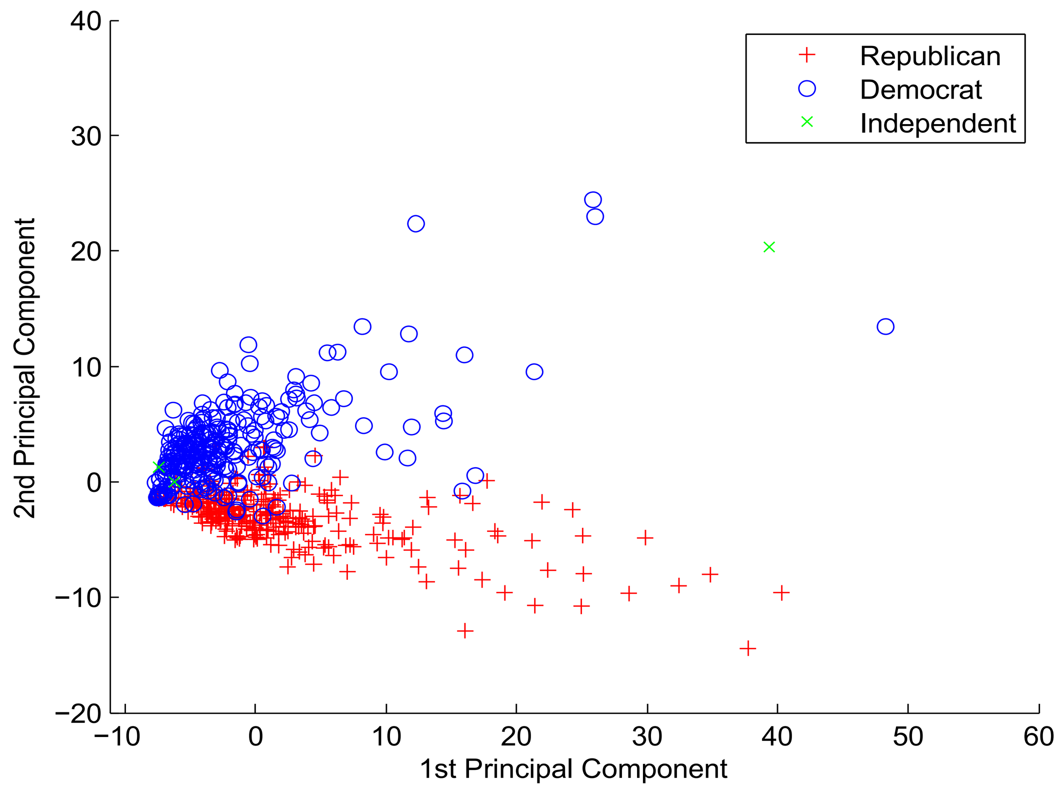
## Data

- Source: MapLight, a nonprofit that collects information about corporations and special interest groups that contribute to campaigns.
- Datasets:
  - Votes on 1262 measures from Congress between 2006 and 2012.
  - Positions held by various interest groups on those bills.
  - Individual and corporate contributions to campaigns from FEC filings.
  - A list of politicians, their district and party.



## Principal Component Analysis

- Add up all the money given to each politician from each subsector and compute the correlation among the subsectors.
- Project the politicians along the first two principal components of the donation matrix.
- Second principal component provides clear separation by party.

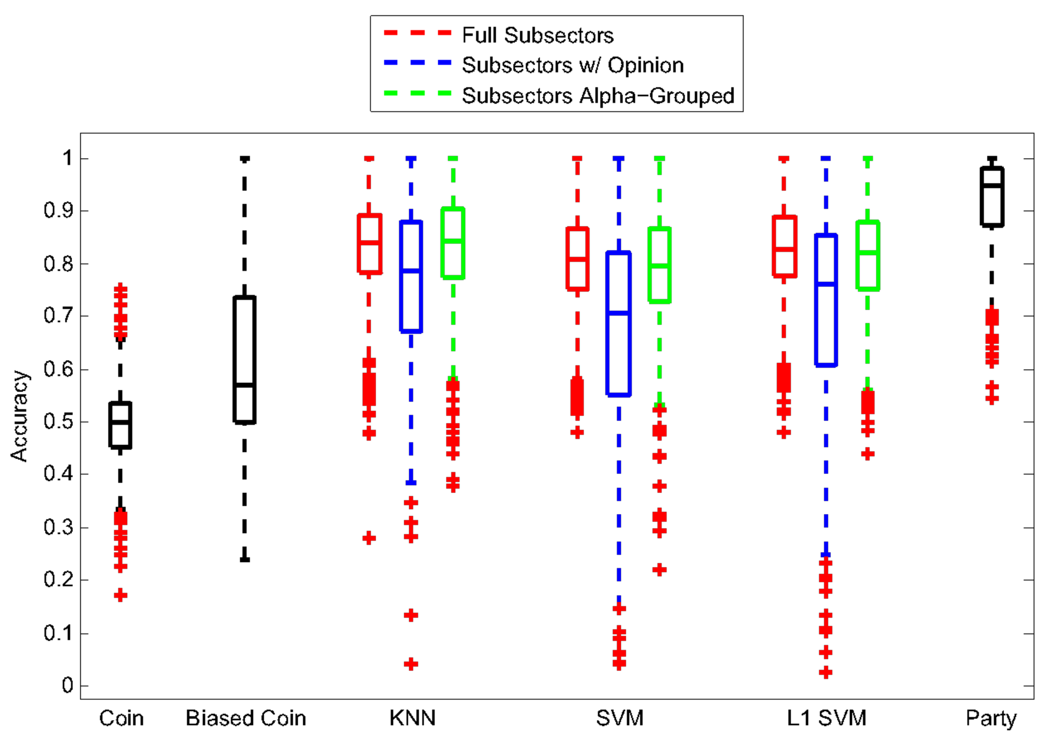



## Classification Methods

- Classify how a politician votes based on campaign contributions.
- Baselines: coin toss and an empirically biased coin toss.
- Methods used:
  - *k*-Nearest neighbors (*k*NN),
  - linear support vector machine (SVM), and
  - $L_1$-regularized SVM.
- Party line classifier to assess significance of political party.
- Classifiers were run for each bill for three different donation matrices:
  1. All subsectors
  2. Subsectors which expressed and opinion on a measure.
  3. Similar to 2, but with the addition of related subsectors.

## Results

- All methods used significantly outperform the randomized baseline
- Most accurate: party line
- Given the PCA results, political party likely a significant latent variable in the analysis.
- Experiment conditioned on political party:
  - Tested only one party on bills with high disagreement within party.
  - 62% accuracy for *k*NN method.
  - 52% for a biased coin.




## Conclusions and Outlook

- The *k*NN method was found to have the highest accuracy and lowest variance of all classification schemes tested.
- But the party line is a better predictor.
- Conclusion that money influences votes to first order is not strongly supported by evidence.
- Money is usually funneled through lobbyists, political parties, and political action committees.
- This is usually not transparent. In particular: no link between donations and individual bills.
- Complexities cannot be captured in a simple model containing only information about direct campaign contributions.

## References

[1] "Maplight: Revealing money's influence on politics." [Online]. Available: http://maplight.org/

[2] "Center for responsive politics." [Online]. Available: http://www.opensecrets.org/

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[4] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995, 10.1007/BF00994018. [Online]. Available: http://dx.doi.org/10.1007/BF00994018

[5] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 58, pp. 267–288, 1996.

[6] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," *in ICML'04: Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 78.

[7] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

International Conference on Machine Learning Applications
December 12-15, 2012, Boca Raton, FL