

Master Thesis proposal

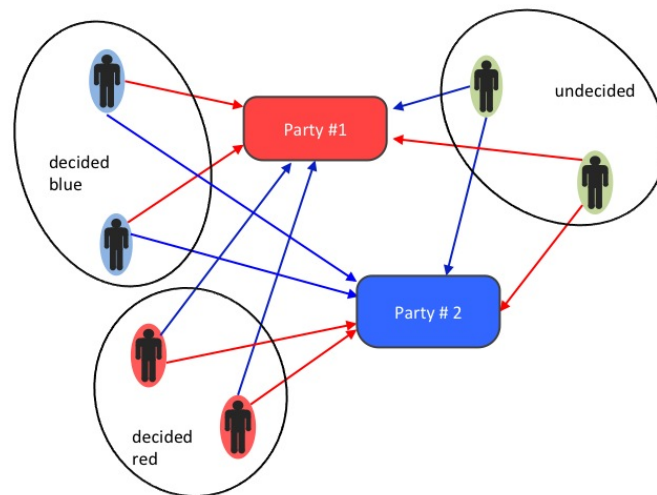
Extrapolating voting prediction from microblogging and studying its “structural balance”

Claudio Altafini

Division of Automatic Control, Dept. of Electrical Engineering,
Linköping University, SE-58183, Sweden.
email: `claudio.altafini@liu.se`

April 14, 2014

This proposal is for a Master thesis in the field of **Data mining and Social Network Analysis**.



In bi-partite (or bi-coalition) political systems, the fraction of undecided voters tends to decrease, and the political opinion of electors tend to polarize, as election day approaches.

The questions that this thesis aims to investigate are:

1. if it is possible to use simple sentiment analysis to “see” this polarization of opinions on microblogging sources such as Twitter.
2. if the “social network” extrapolated from the polarized sentiments converges to a so-called structurally balanced state.

In order to accomplish the first part, suitable tools for lexicon-based sentiment analysis must be developed, able to extrapolate positive and negative “sentiments” towards candidates/parties from tweets. There is already a consistent literature on using the aggregate of these tweets-based sentiments as a predictor of electoral vote, see [2] for a survey. The

approach that will be developed in the thesis is different, more inspired by the field of recommendation systems (or rating systems, used for books, movies, etc.) [3] and by the notion of “disorder” of Statistical Physics. We will look at the user-party map, where edges corresponds to tweets having positive or negative opinions on a party. This is a bipartite signed graph like the one in the figure (where red edges correspond to negative opinions and blue edges to positive opinions) Voters that are “decided” are those that give opinions on both parties coherent with their decision, while voters that are “undecided” are not showing any specific preference. From a Statistical Physics perspective, our signed bipartite (party-user) graph is similar to a so-called spin glass. Disorder corresponds to the presence of negative (undirected) cycles like, i.e., cycles that contain an odd number of undirected negative edges (example: decided - party #1 - undecided - party #2 - decided), see [1] for an application to social networks of similar type. If all users are “decided” then all cycles are positive. If Tweeter is a faithful source for opinion polls, then the “undecided” should decrease as election day approaches (and then maybe we could even predict elections...).

The ideal candidate for the thesis will have a curriculum in Engineering or Computer Science. As the nature of the work is essentially computational, experience in programming and in data mining is highly recommended.

References

- [1] G. Facchetti, G. Iacono, and C. Altafini. Computing global structural balance in large-scale signed social networks. *Proc. Nat. Ac. Sci.*, 108:20953–8, 2011.
- [2] D. Gayo-Avello. A meta-analysis of state-of-the-art electoral prediction from twitter data. *Social Science Computer Review*, 31(6):649–679, 2013.
- [3] F. Ricci, L. Rokach, B. Shapira, and P. Kantor. *Recommender Systems Handbook*. Springer, 2010.