Election Manipulation: The Average Case

ELCHANAN MOSSEL
University of California, Berkeley and Weizmann Institute of Science
and
MIKLÓS Z. RÁCZ
University of California, Berkeley

We review recent research on quantitative versions of the Gibbard-Satterthwaite theorem, which analyze the average-case manipulability of elections. The main message of these results is that computational hardness cannot hide manipulations completely. We conclude with open problems.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]: Economics

General Terms: Algorithms, Economics, Theory

Additional Key Words and Phrases: Voting, computational social choice, manipulation, Gibbard-Satterthwaite, average-case

1. INTRODUCTION

A naturally desirable property of a voting system is *strategyproofness* (a.k.a. nonmanipulability): no voter should benefit from voting strategically, i.e., voting not according to her true preferences. However, the classical result of Gibbard [1973] and Satterthwaite [1975] says that no reasonable voting system can be strategyproof: if voters rank three or more alternatives and all of them can be elected, then the only strategyproof voting systems are dictatorships.

This has contributed to the realization that it is unlikely to expect truthfulness in voting. But is there a way of circumventing the negative results? What is the extent of manipulability of voting systems? This problem is increasingly relevant not only in social choice theory, but also in artificial intelligence and computer science, where virtual elections are now an established tool for preference aggregation (see the survey by Faliszewski and Procaccia [2010]).

Bartholdi, Tovey and Trick [1989] suggest computational complexity as a barrier against manipulation: if it is computationally hard for a voter to manipulate, then she would just tell the truth (see [Faliszewski and Procaccia 2010] for a detailed history of the surrounding literature). However, this is a worst-case approach and does not tell us anything about *typical* instances of the problem—is it easy or hard to manipulate *on average*?

2. AVERAGE-CASE MANIPULABILITY

A natural approach was taken by Friedgut, Kalai, Keller and Nisan [2008; 2011], who looked at the fraction of ranking profiles that are manipulable. To put it differently: assuming each voter votes independently and uniformly at random

This work is supported by NSF grant DMS 1106999 and DOD ONR grant N000141110140. Authors' addresses: `mossel@stat.berkeley.edu`, `racz@stat.berkeley.edu`.

(known as the *impartial culture assumption*), what is the probability that a ranking profile is manipulable? Is it perhaps exponentially small (in the number of voters n and the number of alternatives k), or is it nonnegligible?

Of course, if the social choice function (the function which maps the votes of the voters to the winner of the election, abbreviated as SCF) is nonmanipulable (a dictatorship or a monotone function on two alternatives) then this probability is zero. Similarly, if the SCF is "close" to being nonmanipulable, then this probability can be small. We say that a SCF f is ε-far from the family of nonmanipulable functions, if one must change the outcome of f on at least an ε-fraction of the ranking profiles in order to transform f into a nonmanipulable function. Friedgut et al. conjectured that if $k \geq 3$ and the SCF f is ε-far from the family of nonmanipulable functions, then the probability of a ranking profile being manipulable is bounded from below by a polynomial in $1/n$, $1/k$, and ε. Moreover, they conjectured that a random manipulation will succeed with nonnegligible probability, suggesting that manipulation by computational agents in this setting is easy.

Friedgut et al. proved their conjecture in the case of $k = 3$ alternatives. Note that this result does not have any computational consequences, since when there are only $k = 3$ alternatives, a computational agent may easily try all possible permutations of the alternatives to find a manipulation (if one exists).

Several follow-up papers have since extended this result. Xia and Conitzer [2008] extended the result to a constant number of alternatives, assuming several additional technical assumptions. Dobzinski and Procaccia [2008] proved the conjecture in the case of two voters under the assumption that the SCF is Pareto optimal. Isaksson, Kindler and Mossel [2012] then proved the conjecture in the case of $k \geq 4$ alternatives with only the added assumption of neutrality. Moreover, they showed that a random manipulation which replaces four adjacent alternatives in the preference order of the manipulating voter by a random permutation of them succeeds with nonnegligible probability. Since this result is valid for any number of ($k \geq 4$) alternatives, it does have computational consequences, implying that for neutral SCFs, manipulation by computational agents is easy on average.

Finally, in recent work [Mossel and Rácz 2012] we removed the neutrality condition and resolved the conjecture of Friedgut et al. Removing this assumption is important because in many settings neutrality is not a natural assumption. Consider a search engine aggregating rankings of webpages; if one searches in child-safe mode, then the top-ranked webpage cannot have adult content, and so the aggregating function cannot be neutral.

The main message of these results is thus that computational hardness cannot hide manipulations completely, because manipulation is easy on average unless the SCF is close to being a dictatorship or taking on at most two values.

In addition, the interplay between various techniques used in the proofs is interesting. Friedgut et al. used combinatorial techniques together with discrete harmonic analysis, reducing the problem to a quantitative version of Arrow's theorem. The techniques of Isaksson et al. are more geometric in nature, using canonical path arguments to show isoperimetric results about the interface of more than two bodies. Our recent result combines tools from both of these proofs, in addition to crucially using reverse hypercontractivity.

3. OPEN PROBLEMS

We conclude with a few open problems that arise naturally.

—There are various ways to measure the manipulability of a function: in terms of the probability of having manipulating voters, in terms of the expected number of manipulating voters, etc. How are these related, and which of these is the most relevant for applications?

—Our techniques do not lead to tight bounds and it would be interesting to find the correct tight bounds (in terms of one of the quantities above).

—A related question is: what is the "least manipulable" function? More precisely, in some natural subsets of functions (e.g., anonymous and such that each of the k alternatives are chosen with probability at least $1/k^2$), find the one that minimizes manipulation. Without the naturality constraint, related questions are addressed in [Maus et al. 2007].

—Voting profiles typically have some structure. Can we prove similar results if the underlying distribution over rankings is not i.i.d. uniform? It would be interesting to consider the questions asked above in this setting as well.

—Studying incentives to manipulate is also important in order to understand the difference between when manipulation is possible and when it actually occurs [Carroll 2011].

REFERENCES

Bartholdi III, J., Tovey, C., and Trick, M. 1989. The Computational Difficulty of Manipulating an Election. *Social Choice and Welfare 6,* 3, 227–241.

Carroll, G. 2011. A Quantitative Approach to Incentives: Application to Voting Rules. Working paper.

Dobzinski, S. and Procaccia, A. 2008. Frequent Manipulability of Elections: The Case of Two Voters. In *Proceedings of the 4th International Workshop on Internet and Network Economics*. Springer, 653–664.

Faliszewski, P. and Procaccia, A. 2010. AI's War on Manipulation: Are We Winning? *AI Magazine 31,* 4, 53–64.

Friedgut, E., Kalai, G., Keller, N., and Nisan, N. 2011. A Quantitative Version of the Gibbard-Satterthwaite Theorem for Three Alternatives. *SIAM J. Comput. 40,* 3, 934–952.

Friedgut, E., Kalai, G., and Nisan, N. 2008. Elections can be manipulated often. In *Proceedings of the 49th Annual Symposium on Foundations of Computer Science*. IEEE, 243–249.

Gibbard, A. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica: Journal of the Econometric Society*, 587–601.

Isaksson, M., Kindler, G., and Mossel, E. 2012. The Geometry of Manipulation: A Quantitative Proof of the Gibbard-Satterthwaite Theorem. *Combinatorica 32,* 2, 221–250.

Maus, S., Peters, H., and Storcken, T. 2007. Anonymous voting and minimal manipulability. *Journal of Economic Theory 135,* 1, 533–544.

Mossel, E. and Rácz, M. 2012. A quantitative Gibbard-Satterthwaite theorem without neutrality. In *Proceedings of the 44th ACM Symposium on Theory of Computing (STOC)*. ACM. Full version available as arXiv preprint at arXiv:1110.5888.

Satterthwaite, M. 1975. Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory 10,* 2, 187–217.

Xia, L. and Conitzer, V. 2008. A Sufficient Condition for Voting Rules to be Frequently Manipulable. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*. ACM, 99–108.