Controlling Elections through Social Influence

Bryan Wilder
Center for Artificial Intelligence in Society
University of Southern California
bwilder@usc.edu

Yevgeniy Vorobeychik
Electrical Engineering and Computer Science
Vanderbilt University
yevgeniy.vorobeychik@vanderbilt.edu

ABSTRACT

Election control considers the problem of an adversary who attempts to tamper with a voting process, in order to either ensure that their favored candidate wins (constructive control) or another candidate loses (destructive control). As online social networks have become significant sources of information for potential voters, a new tool in an attacker's arsenal is to effect control by harnessing social influence, for example, by spreading fake news and other forms of misinformation through online social media.

We consider the computational problem of election control via social influence, studying the conditions under which finding good adversarial strategies is computationally feasible. We consider two objectives for the adversary in both the constructive and destructive control settings: probability and margin of victory (POV and MOV, respectively). We present several strong negative results, showing, for example, that the problem of maximizing POV is inapproximable for any constant factor. On the other hand, we present approximation algorithms which provide somewhat weaker approximation guarantees, such as bicriteria approximations for the POV objective and constant-factor approximations for MOV. Finally, we present mixed integer programming formulations for these problems. Experimental results show that our approximation algorithms often find near-optimal control strategies, indicating that election control through social influence is a salient threat to election integrity.

KEYWORDS

Election control; influence maximization

1 INTRODUCTION

The integrity of elections is crucial to the functioning of democratic institutions. As a result, a large body of work has focused on the robustness of elections to various forms of control, where a malicious party attempts to manipulate election results, for example, by bribing voters and adding or removing votes. While it is important to understand the vulnerability of elections to such control, there are many countries where blatant tampering is (fortunately) uncommon. For instance, outright voter fraud is very rare in the United States federal and state elections [1, 13].

However, more subtle forms of election control may attempt to subvert legitimate information channels towards malicious means. For example, political advertising and news (in the editorial form) are common legitimate means for convincing prospective voters. Such communication, when sufficiently transparent, is often critical

to the effective functioning of democracy, and can exert considerable influence on voting behavior [5, 17, 25]. Malicious control over information promulgated through these channels can thus have considerable impact, but is also difficult to achieve due to the relative transparency of traditional media sources.

The increasing importance of social media, such as Facebook and Twitter, for propagating information, including about political issues [12, 29, 44], is a game changer. Both the decentralized nature of information sources on social media, and their associated lack of transparency, present malicious parties with an unprecedented opportunity to influence a democratic political process. Recent evidence of deliberate election tampering in the 2016 US presidential election through *fake news*—deliberately falsified news content—suggests that this issue is a major concern for election integerity for years to come [2]. For example, it is estimated that the typical American adult saw at least one fake news story during the election cycle [2], and such stories have been shown to impact voters' judgment [39].

Motivated by these concerns, we initiate the first algorithmic study of the problem of *election control through social influence*. In our setting, there is a social network of voters who elect a single winner by plurality vote. An outside party may select a subset of nodes as *seed nodes* for a news story or advertisement. Each of these seed nodes shares the story with their friends. Each friend has some probability of being influenced in their voting preferences, as well as sharing the story further. The question is whether, given a limited budget, the attacker can influence enough voters to ensure that a target candidate wins or loses the election. This is similar to the election bribery setting considered by previous work, but with the added consideration of social influence from the bribed nodes[1].

This problem is closely related to *influence maximization*, which has been studied primarily in the context of viral advertising. There, the objective is simply to maximize the expected number of people who receive a message. While influence maximization admits a simple $(1 - 1/e)$-approximation algorithm, election control through social influence presents a number of new algorithmic challenges. We study both constructive and destructive control for two different objectives: 1) maximizing the expected margin by which a target candidate wins (loses) the election (*margin of victory*, or MOV), and 2) maximizing the probability that a target candidate wins (loses) election (*probability of victory*, or POV).

Summary of main results: We provide a mix of negative (hardness and inapproximability) and positive (algorithmic) results for the problem of election control through social influence. Our main contributions are the following:

[1]Our use of the term "election control" deviates slightly from common terminology. Here, control refers generically to the objective of changing election outcomes, not the means of doing so. The means considered in this paper are closest to the bribery setting in previous work, e.g., we do not study adding or removing candidates or voters.

- We show that the MOV objective in the two candidate case is monotone submodular and hence admits a $(1 - 1/e)$ greedy approximation algorithm.
- We prove that the POV objective is hard to approximate to within *any* multiplicative factor for both constructive and destructive control, even in elections with only two candidates.
- We provide a bicriteria approximation algorithm for the POV objective in the two-candidate case. In fact, our algorithm applies to the more general problem of maximizing the probability that a submodular function exceeds a given threshold value and may be of general interest.
- In the multicandidate case, we provide algorithms which achieve similar guarantees as the two-candidate case up to the loss of a constant factor (independent of the number of candidates). Such guarantees hold for both constructive and destructive control, for both the MOV and POV objectives.
- We give mixed integer linear programming (MILP) formulations for all of the above settings which can be used to find optimal solutions.
- We experimentally compare our approximation algorithms to the optimal strategies produced by the MILP. Despite formal hardness results, our approximation algorithms often find near-optimal solutions, particularly for the MOV objective. This suggests that computational hardness may not always be a practical barrier to controlling elections via social influence.

2 RELATED WORK

Our work is closely related to two research areas: election control and influence maximization. These bodies of work are separate: to our knowledge there is no previous work which considers election control *using* social influence (though there is a sizable literature studying opinion dynamics and social choice [15, 36, 40, 43]). The computational study of election control was started by Bartholdi et al. [3], who considered constructive control. The destructive control setting was introduced by Hemaspaandra et al. [28]. A large body of work has studied election control under different settings and voting rules [16, 20] including bribing voters [4, 21, 22, 24, 46], adding or deleting voters [19, 23, 33, 34], and adding or deleting candidates [8, 23, 33]. The main difference between our work and previous work on election control is that we introduce and analyze social influence as a novel mechanism for both constructive and destructive control. Perhaps the closest related work is that of Bredereck and Elkind [6], who study the manipulation of opinion dynamics in a different model (though not in the context of elections).

There is a large, parallel literature devoted to influence maximization in social networks. This line of work was introduced by Kempe et al. [30] who introduced influence maximization in the independent cascade model and proposed a greedy algorithm based on submodularity. Since then, a number of newer algorithms have been proposed, mostly attempting to scale up the greedy algorithm to very large graphs [10, 14, 26, 42]. Subsequent work has also introduced several related settings, e.g., continuous time dynamics [18, 41], bandit settings where dynamics are learned over time [7, 11, 32], or robust problems where influence probabilities are uncertain [9, 27, 35, 45]. None of this work considers using social influence to control an election, and our setting brings a range of new technical challenges.

Almost all work on influence maximization is founded on submodularity of the objective function. However, even though we use the same model of influence spread, objectives related to election control often violate submodularity, and we need to develop new algorithmic techniques. We mention here work by Krause et al. [31] on robust submodular optimization. For optimizing the POV objective, we use a similar form of surrogate objective. However, our objective is to maximize the *probability* of a desired outcome, not the worst-case value, so both our final algorithm and analysis are novel.

3 PROBLEM FORMULATION

We consider an election with candidates $C = \{c_*, c_1, c_2, ...c_\ell\}$. c_* is a special target candidate, and the objective of the election control problem is to make c_* either win the election (constructive control) or lose (destructive control). The voters are the nodes of a graph $G = (V, E)$. Each voter v has an ordering π_v over the candidates and casts a vote for $\pi_v(1)$, i.e., their first ranked candidate. We assume that voters do not behave strategically. The winner is decided via the plurality rule (the candidate with the most votes wins the election). If there is a tie, we say that the attacker fails. This tie-breaking does not impact any of our results. Let $V_{c_i}^j = \{v \in V : \pi_v(j) = c_i\}$ be the set of voters who rank candidate c_i in place j. Initially, c_i has $|V_{c_i}^1|$ votes.

Social influence: There is an attacker who wishes to change the results of the election by spreading messages which cause voters to change their ordering over candidates. In *constructive control*, the attacker can spread a message which causes any voter v who becomes influenced to promote c_* by one place in π_v (exchanging c_* with the candidate previously ranked above them). If $\pi_v(c_*) = 1$, the message has no effect on v, but v may still decide to share the message with their neighbors. In *destructive control*, a voter who is influenced demotes c_* by one place in π_v. Influence spreads via the independent cascade model (ICM), the most common model in the influence maximization literature. Each edge $(u, v) \in E$ has a propagation probability $p_{u,v}$. If u is influenced, it makes one attempt to influence each neighbor v. Each attempt succeeds independently with probability $p_{u,v}$. The attacker may select a set of k *seed nodes* who are influenced at the start of the process. The diffusion then proceeds in discrete time steps until no new activations are made.

We also introduce a useful alternate view of the ICM, the *live-graph* model. We can equivalently see the ICM as removing each edge (u, v) from the graph with probability $1 - p_{u,v}$. A node is influenced if it is reachable from any seed node via the edges that remain. Call any specific setting of present/absent edges a scenario y, with induced graph G_y. Let $m = 2^{|E|}$ be the total number of scenarios. Let $f(S, y)$ denote the number of nodes which are reachable from any seed node in S on graph G_y. The expected number of nodes influenced under the ICM is just $f(S) = \mathbb{E}_y [f(S, y)]$. Similarly, the probability that the number of influenced nodes exceeds any threshold value Δ is $\Pr_y [f(S, y) \geq \Delta]$. At times, we will want to specifically reason about the probability some subset of V is influenced. For any $A \subseteq V$, let $f(S, y, A)$ denote the number of nodes in A reachable from S in scenario y. Analogously, $f(S, A) = \mathbb{E}_y[f(S, y, A)]$. We remark that such functions can be evaluated up to arbitrary precision by averaging over random samples for y. For simplicity, we ignore such issues here since they are well understood [10, 14, 42].

Objectives: We now formally introduce our two objectives, starting with the simpler two-candidate case. In a two-candidate election, constructive and destructive control are clearly equivalent since maximizing the probability that c_* loses is the same as maximizing the probability that the other candidate wins (and vice versa). Hence, we study only constructive control without loss of generality.

In the margin of victory (MOV) objective, we want to maximize the expected number of votes by which c_* wins the election. We define our objective as the *change* in the expected margin:

$$\mathrm{MOV}(S) = 2 \mathop{\mathbb{E}}_{y} \left[f(S, y, V_{c_*}^2) \right].$$

The factor 2 is present since reaching a voter in $V_{c_*}^2$ both adds a vote for c_* and removes a vote for the opponent. We study the expected change in the margin (not the margin itself) so that approximation ratios are well defined even when the margin is negative.

In the probability of victory (POV) objective, we want to maximize the probability that c_* wins the election. Let $\Delta = \frac{1}{2} \left(|V_{c_1}^1| - |V_{c_*}^1| \right) + 1$ be the number of voters that c_* needs to reach in order to win the election. The POV objective is

$$\mathrm{POV}(S) = \mathop{\mathrm{Pr}}_{y} \left[f(S, y, V_{c_*}^2) \geq \Delta \right]$$

which is just the probability that at least Δ of the voters who have c_* in second place are reached.

In the multicandidate case, constructive and destructive control are no longer equivalent. Further, the impact of messages is more complex than before. E.g., in constructive control not only does c_* gain a vote, but another candidates loses a vote; we need to keep track of the number of votes lost by each other candidate.

We start out by defining functions which give the change in the margin between c_* and another candidate c_i in a given scenario y when seed set S is chosen. Let $\chi(v, S, y)$ be 1 if node v is reachable from seed set S in the graph G_y. The change in margin (in constructive and destructive control, respectively) is given by

$$g_C(S, y, c_i) = \sum_{v \in V_{c_*}^2 \setminus V_{c_i}^1} \chi(v, S, y) + 2 \sum_{v \in V_{c_*}^2 \cap V_{c_i}^1} \chi(v, S, y)$$

$$g_D(S, y, c_i) = \sum_{v \in V_{c_*}^1 \setminus V_{c_i}^2} \chi(v, S, y) + 2 \sum_{v \in V_{c_*}^1 \cap V_{c_i}^2} \chi(v, S, y)$$

which gives value 2 for every node that is flipped from c_i to c_* (or vice versa) and hence count double towards the margin, and 1 for other nodes. Based on this, we now give expressions for the change in margin given any fixed scenario y and seed set S. We start with constructive control. Note that before any intervention, the margin is just $\max_{c_i} |V_{c_i}^1| - |V_{c_*}^1|$. Afterwards, the margin is $\max_{c_j} |V_{c_j}^1| - g_C(S, y, c_j) - |V_{c_*}^1|$. Hence, the change in margin is

$$m_C(S, y) = \left[\max_{c_i} |V_{c_i}^1| - |V_{c_*}^1| \right] - \left[\max_{c_j} |V_{c_j}^1| - g_C(S, y, c_j) - |V_{c_*}^1| \right]$$

$$= \min_{c_j} \left(g_C(S, y, c_j) + \max_{c_i} |V_{c_i}^1| - \left| V_{c_j}^1 \right| \right).$$

That is, the change in margin is driven by candidate with largest starting vote ($|V_{c_j}^1|$) and smallest loss in vote ($g_C(S, y, c_j)$). Now considering all scenarios y, the constructive control objectives are

$$\mathrm{MOV}_C(S) = \mathop{\mathbb{E}}_{y} \left[m_C(S, y) \right] \quad \mathrm{POV}_C(S) = \mathop{\mathrm{Pr}}_{y} \left[m_C(S, y) \geq \Delta_C \right].$$

where $\Delta_C = \max_{c_i} |V_{c_i}^1| - |V_{c_*}^1| + 1$ is the necessary change in margin for c_* to win. For destructive control, we can similarly write the change in margin and corresponding objectives as

$$m_D(S, y) = \max_{c_i} \left(g_D(S, y, c_i) + |V_{c_i}^1| \right) - \max_{c_j} |V_{c_j}^1|$$

$$\mathrm{MOV}_D(S) = \mathop{\mathbb{E}}_{y} \left[m_D(S, y) \right] \quad \mathrm{POV}_D(S) = \mathop{\mathrm{Pr}}_{y} \left[m_D(S, y) \geq \Delta_D \right].$$

where $\Delta_D = |V_{c_*}^1| - \max_{c_i} |V_{c_i}^1| + 1$.

4 ELECTIONS WITH TWO CANDIDATES

We start with elections with only two candidates. Recall that in this setting, constructive and destructive control are equivalent, so our results are stated only for constructive control (trying to ensure c_* wins the election). In order to state our algorithmic results, we first introduce some background on submodular optimization and influence maximization. A set function $f : V \to R$ is submodular if for all $A \subseteq B \subseteq V$ and all $x \notin B$, $f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$. Intuitively, submodularity formalizes the property of diminishing returns. The function $f(S)$ which gives the expected number of nodes reached by S under the independent cascade model is known to be monotone submodular. It is well known that whenever f is a monotone submodular function, the greedy algorithm gives a $(1 - 1/e)$-approximation to the problem $\max_{|S| \leq k} f(S)$.

It is natural to hope that submodularity would transfer to our election control objectives MOV and POV. Our first result is that submodularity does in fact hold for the MOV objective with two candidates. Previous results for influence maximization do not directly apply because the MOV objective only counts nodes who have c_* in second place. Nevertheless, similar reasoning applies.

THEOREM 4.1. *In an election with two candidates, MOV is a monotone submodular function.*

PROOF. We first fix a particular scenario y and show that the function $f(\cdot, y, V_{c_*}^2)$ is submodular. This suffices to show that $\mathbb{E}_y[f(\cdot, y, V_{c_*}^2)]$ is submodular since a nonnegative linear combination of submodular functions remains submodular. Monotonicity is clear since adding additional seeds to A can only make more nodes reachable. To show submodularity, we can write the marginal gain as

$$f(A \cup \{x\}, y, V_{c_*}^2) - f(A, y, V_{c_*}^2) = \sum_{v \in V_{c_*}^2} (1 - \chi(v, A, y)) \chi(v, \{x\}, y).$$

Compare the above expression for a set A and any $B \supseteq A$. For any single node v, $\chi(v, B, y) = 1$ whenever $\chi(v, A, y) = 1$. Hence, the term in the above summation for each node v can only be smaller for $f(B \cup \{x\}, y, V_{c_*}^2) - f(B, y, V_{c_*}^2)$ than for $f(A \cup \{x\}, y, V_{c_*}^2) - f(A, y, V_{c_*}^2)$. We conclude that $f(A \cup \{x\}, y, V_{c_*}^2) - f(A, y, V_{c_*}^2) \geq f(B \cup \{x\}, y, V_{c_*}^2) - f(B, y, V_{c_*}^2)$ and submodularity now follows by taking the expectation over y. □

Algorithm 1 Algorithms for MOV objective

1: **function** GREEDY(h, k)
2:      $S \leftarrow \emptyset$
3:      **while** $|S| < K$ **do**
4:          $v \leftarrow \arg\max_{v \in V \setminus S} h(S \cup \{v\}) - h(S)$
5:          $S \leftarrow S \cup \{v\}$
6:      **return** S
7: **function** MOVCONSTRUCTIVE(k)
8:      $h(S) := \mathbb{E}_y \left[f(S, y, V_{c_*}^2) \right]$
9:      **return** GREEDY(h, k)
10: **function** MOVDESTRUCTIVE(k)
11:      $h(S) := \mathbb{E}_y \left[f(S, y, V_{c_*}^1) \right]$
12:      **return** GREEDY(h, k)

Algorithm 2 Algorithms for POV objective

1: **function** ENUMERATETHRESHOLD(h, Δ, k)
2:      **for** $\beta = \Delta \dots n$ **do**
3:          $h'(S) := \mathbb{E}_y \left[\min(\beta, h(S, y)) \right]$
4:          $S_\beta = $ GREEDY(h', k)
5:      **return** $\arg\max_{S_\beta, \beta = \Delta \dots n} \Pr_y \left[h(S_\beta, y) \geq \Delta \right]$
6: **function** POVCONSTRUCTIVE(k)
7:      //recall $m_C(S, y) = \mathbb{E}_y \left[f(S, y, V_{c_*}^2) \right]$ for 2-candidate case
8:      **return** ENUMERATETHRESHOLD(m_C, Δ_C, k)
9: **function** POVDESTRUCTIVE(k)
10:      **return** ENUMERATETHRESHOLD(m_D, Δ_D, k)

Hence, we can apply a greedy algorithm to MOV to obtain a $(1 - 1/e)$-approximation (MOVCONSTRUCTIVE in Algorithm 1). Moreover, this ratio is tight since MOV contains regular influence maximization as a special case when all nodes have c_* in second place ($V_{c_*}^2 = V$). It is NP-hard to approximate influence maximization with ratio better than $1 - 1/e$ [30]. Hence, two-candidate MOV is computationally intractable with respect to exact optimization but high quality and efficient approximation algorithms exist.

We now turn to the POV objective, where we maximize the *probability* that c_* wins the election. It is natural to think that submodularity may also carry over to this setting. However, this is not the case; we can provide a simple counterexample where POV violates submodularity. Consider n isolated nodes, where $\frac{n}{2} - k + 1$ have c_* as their first choice. Hence, to win the election it is necessary and sufficient to influence k nodes, which can be accomplished by any choice of k seeds from among those with c_* as their second choice. Fix a seed set B containing $k-1$ of these nodes and consider any $A \subset B$ (that is, A is strictly smaller). We have POV(B) = POV(A) = 0. By adding a node $v \in V_{c_*}^2 \setminus B$ to B, we have POV($B \cup \{v\}$) − POV(B) = 1. However, since $|A| < k - 1$, POV($A \cup \{v\}$) = 0 and hence POV($A \cup \{v\}$) − POV(A) = 0. This contradicts the definition of submodularity. Essentially, the POV objective displays a sharp threshold behavior, where additional seed nodes have no value until we are close to winning. This behavior in fact translates into the following strong hardness result:

THEOREM 4.2. *It is NP-hard to compute an α-approximation to the problem $\max_{|S| \leq k}$ POV(S) for any $\alpha > 0$, even for two candidates and even when the instance is deterministic.*

PROOF. We consider a deterministic objective: the ICM with all propagation probabilities either 0 or 1. Without loss of generality, we have only a single scenario and will drop the dependence on y in f. Suppose that we have an α-approximation for POV-maximization. We show how we can use this algorithm to optimally solve the influence maximization problem (i.e., maximizing $f(\cdot, V)$), which is known to NP-hard since it includes maximum coverage as a special case. Let OPT_{IM} be the optimal value of the influence maximization problem and $OPT_{POV}(\Delta)$ be the optimal value for Problem 1 with the given threshold. Fix any $\Delta > 0$. If $\Delta \leq OPT_{IM}$, then there is a set S with POV(S) = 1 and hence $OPT_{POV}(\Delta) = 1$. Otherwise, there is no set with value Δ and $OPT_{POV}(\Delta) = 0$. Since the objective to the POV problem is either 0 or 1, any α-approximation algorithm for it must return 1 whenever $OPT_{POV}(\Delta) = 1$. Now, we can just enumerate over $\Delta = 1 \dots n$, where n is the number of nodes in the graph. At each value of Δ, we ask the α-approximation algorithm to solve the POV maximization problem with that value of Δ. We return the solution corresponding to the highest value of Δ for which we can find a set with $f(S, V) \geq \Delta$. By the above, this set is an optimal solution to the influence maximization problem. □

We remark that since this hardness result has broader implications. Recall that POV(S) = $\Pr_y \left[f(S, y, V_{c_*}^2) \geq \Delta \right]$ where by Theorem 4.1, $f(S, y, V_{c_*}^2)$ is a submodular function. Therefore, the inapproximability result in Theorem 4.2 shows that it is in general hard to approximate the problem of of maximizing the probability that a submodular function exceeds a given threshold value. This is a natural objective in other domains, e.g. for a risk-averse decision maker who wants to control the probability of a bad outcome.

We pair this hardness result with a positive algorithmic result regarding bicriteria approximations. A bicriteria approximation algorithm gives up solution quality in more than one dimension, and is of interest when hardness results preclude the usual notion of approximation (as for our problem). We provide an algorithm which has a solution quality guarantee whenever the election is winnable by a "large margin". That is, there is a seed set with high probability of greatly exceeding Δ votes. Our algorithm will attempt to maximize the probability of exceeding exactly Δ votes, but has a guarantee relative to the optimal value for threshold $\frac{1}{\alpha}\Delta$ for some $\alpha < 1$. That is, it is only compared to the optimal value of a harder problem.

Our algorithm is a greedy strategy based on the surrogate function $h(S) = \mathbb{E}_y \left[\min\{\beta, f(S, y, V_{c_*}^2)\} \right]$, where β is a chosen threshold value. The intuition is to replace the sharp discontinuity of the original POV objective by a surrogate which interpolates smoothly up to the threshold β. However, we do not give any "credit" for nodes reached beyond β since (unlike in the MOV case) we only care about crossing the threshold. It is easy to see that the minimum of a submodular function and a constant is itself submodular [31]. Hence, h is submodular and amenable to greedy optimization. POV-CONSTRUCTIVE (Algorithm 2) iterates over a series of possible thresholds for h, optimizes each one greedily, and outputs the best of the resulting seed sets. Specifically, it tries every value of β from $\Delta \dots n$. For each β, it finds a seed set S_β by greedily optimizing $\mathbb{E}_y \left[\min\{\beta, f(S, y, V_{c_*}^2)\} \right]$ (Algorithm 2, Lines 3-4). Then, it outputs the S_β which maximizes $\Pr_y \left[f(S_\beta, y, V_{c_*}^2) \geq \Delta \right]$, i.e., the one which has the best probability of exceeding the true objective (Line 5).

The reason that we need to enumerate over values for β, instead of just using the true threshold Δ, is that optimizing the surrogate h might result in a solution which has value below β in every scenario. However, we can show that if $OPT(\beta)$ is high, there must be many scenarios where S_β has value close to β (a notion formalized in our proof). Hence, if we try a sufficiently large $\beta > \Delta$ and $OPT(\beta)$ is still high, there must be many scenarios with value at least Δ. This is formalized in the following guarantee:

THEOREM 4.3. *In an election with two candidates,* POVConstructive *produces a solution S such that*

$$\Pr_y[f(S, y, V_{c_*}^2) \geq \Delta] \geq \max_{0 < \alpha < 1} \frac{\left(1 - \frac{1}{e}\right) OPT_{POV}\left(\frac{1}{\alpha}\Delta\right) - \alpha}{1 - \alpha}$$

PROOF. POVCONSTRUCTIVE enumerates over values of α by trying thresholds $\beta = \Delta...n$. Fix a specific β and set $\alpha = \frac{\Delta}{\beta}$. We will prove that $\Pr_y[f(S_\beta, y, V_{c_*}^2) \geq \Delta] \geq \frac{(1-\frac{1}{e})OPT_{POV}(\frac{1}{\alpha}\Delta)-\alpha}{1-\alpha}$. This suffices the prove the theorem because we output the best of the S_β. A minor point is that the theorem takes the max over $0 < \alpha < 1$, while we try only the discrete points $\alpha = \frac{\Delta}{\Delta}, \frac{\Delta}{\Delta+1}, ..., \frac{\Delta}{n}$. However, these are equivalent because $f(S, y, V_{c_*}^2)$ is always integral.

We divide the set of scenarios into those where the S_β has value at least $\alpha\beta$ and those where it has less value. Let $A = \{y : f(S_\beta, y, V_{c_*}^2) \geq \alpha\beta\}$ and $B = \{y : f(S_\beta, y, V_{c_*}^2) < \alpha\beta\}$. We have

$$\frac{1}{m}\sum_{y \in A} \min\{\beta, f(S_\beta, y, V_{c_*}^2)\} + \frac{1}{m}\sum_{y \in B} \min\{\beta, f(S_\beta, y, V_{c_*}^2)\}$$
$$\geq \left(1 - \frac{1}{e}\right)\max_{|S| \leq k} h(S) \geq \left(1 - \frac{1}{e}\right)\beta m \cdot OPT_{POV}(\beta)$$

where the first inequality follows from submodularity and the second follows since the solution attaining value $OPT_{POV}(\beta)$ for the POV maximization problem is a feasible solution to the problem $\max_{|S| \leq k} h(S)$ which has value at least $\beta m \cdot OPT_{POV}(\beta)$. We are interested in the minimum possible size of A given that the total value is lower bounded as above. By inspection, $|A|$ is minimized when $\min\{\beta, f(S, y)\} = \beta$ for each $y \in A$ and $f(S, y) = \alpha\beta$ for each $y \notin A$. In this case, we have

$$\frac{|A|}{m} + \alpha\left(1 - \frac{|A|}{m}\right) \geq \left(1 - \frac{1}{e}\right)OPT_{POV}(\beta)$$

and hence

$$\frac{|A|}{m} = \frac{1}{m}\sum_{i=1}^{m} 1[f(S_\beta, y, V_{c_*}^2) \geq \alpha\beta] = \Pr_y\left[f(S_\beta, y, V_{c_*}^2) \geq \Delta\right]$$
$$\geq \frac{\left(1 - \frac{1}{e}\right)OPT_{POV}\left(\frac{1}{\alpha}\Delta\right) - \alpha}{1 - \alpha}$$

which completes the proof. □

Theorem 4.3 in fact applies to the general problem of maximizing the probability that a submodular function exceeds a threshold value (complementing our hardness result in Theorem 4.2). As discussed above, this may be of interest independently of election control.

5 MULTIPLE CANDIDATES

We now consider election control with more than two candidates. There is a target candidate c_* and other candidates $c_1...c_\ell$. Note that constructive and destructive control are distinct in this setting. We will give algorithms for both cases for both the MOV and POV objectives.

The problem becomes significantly harder in the multicandidate setting because we must now reason simultaneously about several objectives – whether each alternate candidate c_i will accumulate more votes than c_*. We demonstrate that, up to the loss of a constant in the approximation ratio, it suffices to concentrate only on the number of votes gained or lost by c_* (not the margin against each c_i individually). This concept yields (bicriteria) approximation algorithms for each setting along the lines of the two-candidate case.

We start out with the MOV_C objective (constructive control for the margin of victory), since the idea is simpler to illustrate in this case. The basic intuition is that the change in margin between candidate c_i and c_* can be re-expressed as follows:

$$g_C(S, y, c_i) = \sum_{v \in V_{c_*}^2 \setminus V_{c_i}^1} \chi(v, S, y) + 2\sum_{v \in V_{c_*}^2 \cap V_{c_i}^1} \chi(v, S, y)$$
$$= f\left(S, y, V_{c_*}^2\right) + f\left(S, y, V_{c_*}^2 \cap V_{c_i}^1\right)$$

Now, we can express the final margin in scenario y as

$$m_C(S, y) = f\left(S, y, V_{c_*}^2\right) + \min_{c_j}\left(f\left(S, y, V_{c_*}^2 \cap V_{c_j}^1\right) + \max_{c_i}|V_{c_i}^1| - \left|V_{c_j}^1\right|\right)$$

where the first term is common to all candidates and reflects the total number of voters who switch to c_* and the min term selects the c_j who has the most remaining votes. In general, this second term can be very difficult to approximate because it is the minimum of submodular functions, which is not in general submodular (or even approximable [31]). We might hope that there is some special structure to the election control problem, but this is not the case:

THEOREM 5.1. *For any $\epsilon > 0$, it is NP-hard to compute any $\Omega\left(\frac{1}{n^{1-\epsilon}}\right)$-approximation to the problem*

$$\max_{|S| \leq k} \min_{c_j}\left(f\left(S, y, V_{c_*}^2 \cap V_{c_i}^1\right) + \max_{c_i}|V_{c_i}^1| - |V_{c_j}^1|\right)$$

PROOF. We will consider instances where all of the c_j start with an equal number of votes and so $\max_{c_i}|V_{c_i}^1| - |V_{c_j}^1| = 0$ for all c_j. Thus, the problem is just $\max_{|S| \leq k} \min_{c_j} f\left(S, y, V_{c_*}^2 \cap V_{c_i}^1\right)$. We reduce from the robust influence maximization (RIM) problem [9, 27]. In RIM, we are given a set of objectives $f_1...f_r$, each of which represent expected influence spread in an instance of the independent cascade model on a common underlying graph G. He and Kempe [27] show that it is NP-hard to compute $\Omega\left(\frac{1}{n^{1-\epsilon}}\right)$-approximation to the problem $\max_{|S| \leq k} \min_{i=1...r} f_i(S)$. Their proof holds when each f_i is deterministic (assigns probability 0 or 1 to each edge), so we will assume that the instance is in this form. Let G_i be a graph in which each edge of G assigned probability 0 by f_i has been removed. We create a graph G' as follows. G' contains each G_i as a disconnected subgraph. For every $v \in G$, we add a vertex v' to

G'. v' has an outgoing edge to the copy of node v in each of the G_i subgraphs. Each such edge has propagation probability 1. There is a target candidate c_* and r additional candidates $c_1...c_r$. Each of the v' nodes that were added has c_* as their first choice. Each node in subgraph G_i has c_i as their first choice and c_* as their second choice.

Suppose that we have an α-approximation algorithm for our problem for some $\alpha = \Omega\left(\frac{1}{n^{1-\epsilon}}\right)$. Without loss of generality, we will assume that this algorithm only selects nodes from the v' (since if a seed set contains the copy of v in any subgraph, we can only obtain greater influence spread by exchanging it for v'). Note that, for any such set of seed nodes, $f_i(S) = f(S, V_{c_*}^2 \cap V_{c_i}^1)$ Thus, if S is an α-approximate solution for our problem, it is also an $\Omega\left(\frac{1}{n^{1-\epsilon}}\right)$-approximate solution to the RIM problem. $\qquad\square$

Therefore, we should not hope for any algorithm which can closely approximate the entirety of the objective; the min component is too difficult to handle. However, we can leverage the fact that the first term, $f(S, y, V_{c_*}^2)$, is easy to optimize because it is just a submodular function. Hence, the objective is the sum of an easy term and a hard term. Importantly, we can show that optimizing just the easy term (which is what MOVCONSTRUCTIVE does) is sufficient to obtain a constant factor approximation.

THEOREM 5.2. *MOVConstructive obtains a* $\frac{1}{3}\left(1 - \frac{1}{e}\right)$- *approximation to the* MOV_C *problem with any number of candidates.*

PROOF. Let $c(S, y) = \arg\min_{c_i} f(V_{c_*}^2 \cap V_{c_i}^1) - |V_{c_i}^1|$ be the candidate achieving the minimum in the definition of m_C. Let S^* be an optimal seed set. Note that for all scenarios y, seed sets S, and candidates c_i, $f(S, y, V_{c_*}^2) \geq f(S, y, V_{c_*}^2 \cap V_{c_i}^1)$. Hence, we have

$$\mathbb{E}_y\left[f\left(S^*, y, V_{c_*}^2\right)\right] \geq \frac{1}{3}\mathbb{E}_y\left[f\left(S^*, y, V_{c_*}^1\right) + f\left(S^*, y, V_{c_*}^1 \cap V_{c(S^*,y)}^1\right)\right.$$
$$\left. + f\left(S^*, y, V_{c_*}^1 \cap V_{c(S,y)}^1\right)\right]$$

Note that $\mathbb{E}_y[f(\cdot, y, V_{c_*}^2)]$ is a monotone submodular function, which MOVCONSTRUCTIVE greedily maximizes. Let S be the resulting seed set. We have

$$\mathbb{E}_y\left[f\left(S, y, V_{c_*}^2\right) + f\left(S, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)\right] \geq \mathbb{E}_y\left[f\left(S, y, V_{c_*}^2\right)\right]$$
$$\geq \frac{1}{3}\left(1 - \frac{1}{e}\right)\mathbb{E}_y\left[f\left(S^*, y, V_{c_*}^2\right) + f\left(S^*, y, V_{c_*}^2 \cap V_{c(S^*,y)}^1\right)\right.$$
$$\left. + f\left(S^*, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)\right]$$

which allows us to bound the margin of victory relative to S^* as

$MOV_C(S)$
$$= \mathbb{E}_y\left[f\left(S, y, V_{c_*}^2\right) + \min_{c_j} f\left(S, y, V_{c_*}^2 \cap V_{c_j}^1\right) + \max_{c_i}|V_{c_i}^1| - |V_{c_j}^1|\right]$$
$$= \mathbb{E}_y\left[f\left(S, y, V_{c_*}^2\right) + f\left(S, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)\right] + \max_{c_i}|V_{c_i}^1| - \mathbb{E}_y\left[|V_{c(S,y)}^1|\right]$$
$$\geq \frac{1}{3}\left(1 - \frac{1}{e}\right)\mathbb{E}_y\left[f\left(S^*, y, V_{c_*}^1\right) + f\left(S^*, y, V_{c_*}^2 \cap V_{c(S^*,y)}^1\right)\right.$$

$$\left. + f\left(S^*, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)\right)\right] + \max_{c_i}|V_{c_i}^1| - \mathbb{E}_y\left[|V_{c(S,y)}^1|\right]$$

and some additional algebra (deferred to the supplement) yields

$$MOV_C(S) \geq \frac{1}{3}\left(1 - \frac{1}{e}\right)\left(MOV_C(S^*) + \mathbb{E}_y\left[f\left(S^*, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)\right.\right.$$
$$\left.\left. + |V_{c(S^*,y)}^1| - |V_{c(S,y)}^1|\right]\right).$$

Now by definition of $c(S^*, y)$, $f(S^*, y, V_{c_*}^2 \cap V_{c(S^*,y)}^1) - |V_{c(S^*,y)}^1| \leq f(S^*, y, V_{c_*}^2 \cap V_{c(S,y)}^1) - |V_{c(S,y)}^1|$ and so

$$|V_{c(S^*,y)}^1| - |V_{c(S,y)}^1| \geq f\left(S^*, y, V_{c_*}^2 \cap V_{c(S^*,y)}^1\right) - f\left(S^*, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)$$

This yields

$$MOV_C(S) \geq \frac{1}{3}\left(1 - \frac{1}{e}\right)\left(MOV_C(S^*) + \mathbb{E}_y\left[f\left(S^*, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)\right.\right.$$
$$\left. + f\left(S^*, y, V_{c_*}^2 \cap V_{c(S^*,y)}^1\right) - f\left(S^*, y, V_{c_*}^2 \cap V_{c(S,y)}^1\right)\right]\right)$$
$$= \frac{1}{3}\left(1 - \frac{1}{e}\right)\left(MOV_C(S^*) + \mathbb{E}_y\left[f\left(S^*, y, V_{c_*}^2 \cap V_{c(S^*,y)}^1\right)\right]\right)$$
$$\geq \frac{1}{3}\left(1 - \frac{1}{e}\right)MOV_C(S^*).$$

$\qquad\square$

We also have a corresponding result for the destructive control case. Here, we can rewrite the change in margin as $m_D(S, y) = f(S, y, V_{c_*}^1) + \max_{c_i}\left(f(S, y, V_{c_*}^1 \cap V_{c_i}^2) + |V_{c_i}^1| - \max_{c_j}|V_{c_j}^1|\right)$. MOVDE-STRUCTIVE greedily optimizes the submodular function $\mathbb{E}_y\left[f(S, y, V_{c_*}^1)\right]$, which we show is a good surrogate for $\mathbb{E}_y[m_D(S, y)]$.

THEOREM 5.3. *MOVDestructive obtains a* $\frac{1}{2}\left(1 - \frac{1}{e}\right)$-*approximation to the multicandidate* MOV_D *problem.*

The proof, which is similar to that of Theorem 5.2, can be found in the supplement.

Now, we extend these ideas to obtain similar guarantees for the POV_C and POV_D objectives. Starting with POV_C, recall that our objective is to maximize $\Pr_y[m_C(S, y) \geq \Delta_C]$, the probability that the change in margin exceeds the number of votes needed to win. We will prove a guarantee for the same algorithm POVCONSTRUCTIVE as from the two-candidate case. Recall that POVCONSTRUCTIVE optimizes the surrogate $\mathbb{E}_y\left[\min\left(\beta, f\left(S, y, V_{c_*}^2\right)\right)\right]$, enumerating over possible values of the threshold β. We have the following bicriteria approximation guarantee:

THEOREM 5.4. *Let* $OPT(\Delta)$ *denote the optimal value of the problem* $\max_{|S| \leq k} Pr_y[m_C(S, y) \geq \Delta]$. *Let S be the set produced by POV-Constructive. We have*

$$POV_C(S) \geq \max_{0 < \alpha < 1} \frac{\frac{e-1}{3e-1}OPT\left(\frac{1}{\alpha}\Delta_C\right) - \alpha}{1 - \alpha}$$

The proof can be found in the supplement. The main difference from the two candidate case is that we do not directly optimize $\frac{1}{m} \sum_y \min(\beta, m_C(S, y))$ since it may no longer be submodular. Instead, we greedily optimize the submodular surrogate function $\frac{1}{m} \sum_y \min\left(\beta, f(S, y, V_{c_*}^2)\right)$ and show that this surrogate approximates $\frac{1}{m} \sum_y \min(\beta, m_C(S, y))$. From there, the same argument as in Theorem 4.3 extends to the multicandidate case. Analogous reasoning also yields a bicriteria guarantee for destructive control:

THEOREM 5.5. *Let OPT(Δ) denote the optimal value of the problem* $\max_{|S| \leq k} Pr_y[m_D(S, y) \geq \Delta]$. *Let S be the set produced by* POVDestructive. *We have*

$$POV_D(S) \geq \max_{0 < \alpha < 1} \frac{\frac{e-1}{3e-1}OPT(\frac{1}{\alpha}\Delta) - \alpha}{1 - \alpha}$$

6 MIXED-INTEGER PROGRAMS

Thus far, we have only considered approximation algorithms for election control, motivated by computational hardness results for exact optimization. Now we give mixed-integer linear programming (MILP) formulations to find exact solutions. This serves two purposes. First, it allows us to study the effectiveness of election control for problem instances with are within the range of state of the art MILP solvers. Second, we can determine the empirical effectiveness of the approximation algorithms proposed in earlier sections.

There are two principal difficulties in obtaining MILP formulations. First, the objective is stochastic, ranging over an exponential number of scenarios. Second, even for a single fixed scenario, the number of nodes reached by a seed set is a nonlinear function.

We first show how to linearize the problem when have only a single scenario y. Recall that y corresponds to a sampled graph G_y, where every edge e is removed independently with probability $1 - p_e$. Our MILP will have a binary variable $s_v \in \{0, 1\}$ for each node $v \in V$, where $s_v = 1$ indicates that v is a seed node. We will maximize an objective over all $s_v \in \{0, 1\}^{|V|}$ which satisfy $\sum_{v \in V} s_v \leq k$ (at most k nodes are seeded). The challenge is to embed the nonlinear objective into the constraints of the MILP. Let $x_v^y, v \in V$ be a binary variable indicating whether v is influenced in scenario y. We must constrain x_v^y to be 1 only if v truly is reachable in G_y from some node with $s_v = 1$. To accomplish this, let $R(v, y)$ be the set of nodes which have a directed path to v in scenario y. $R(v, y)$ does not depend on the decision variables s and can easily precomputed. Using this set, we constrain the x variables as:

$$x_v^y \leq \sum_{u \in R(v, y)} s_u \quad \forall v \in V.$$

Now ee deal with stochasticity using sample average approximation. We first sample scenarios $G_{y_i}, i = 1...r$, maintaining a separate copy $x_v^{y_i}$ for each sampled scenario. Finally, we average over the variables in each scenario to obtain the final objective.

Using these components, we now give concrete formulations for each of the problem instances that we consider. We will assume that scenarios $y_1...y_r$ have been sampled, where r is a tunable parameter trading off computational cost and sampling error.

6.0.1 Constructive control. We create a variable $g_C(y_i, c_j)$ for each scenario y_i and candidate c_j which represents the change in the margin between c_j and c_* in scenario y_i. Using these variables, we set a variable $m_C(y_i)$ for each scenario y_i which represents the overall change in margin. These variables are set using the constraints

$$g_C(y_i, c_j) \leq \sum_{v \in V_{c_*}^2} x_v^{y_i} + \sum_{v \in V_{c_*}^2 \cap V_{c_j}^2} x_v^{y_i}$$

$$m_C(y_i) \leq g_C(y_i, c_j) + \max_{c_i} |V_{c_i}^1| - |V_{c_j}^1| \forall i, j$$

which gives the following MILP to maximize the MOV:

$$\max_{s, x, g_C, m_C} \frac{1}{r} \sum_{i=1}^{r} m_C(y_i)$$

$$\sum_{v \in V} s_v \leq k$$

The next formulation maximizes the POV:

$$\max_{s, x, g_C, m_C, u} \frac{1}{r} \sum_{i=1}^{r} u_i$$

$$- M(1 - u_i) + \max_{c_i} |V_{c_i}^1| - |V_{c_*}^1| + 1 - m_C(y_i) \leq 0$$

$$\sum_{v \in V} s_v \leq k, \quad u_i \in \{0, 1\} \quad i = 1...r$$

Here, u_i is a binary variable representing whether c_* wins the election in scenario y_i, while M is a large number.

6.0.2 Destructive control. Now, we use an analogous set of constraints to set variables $g_D(y_i, c_j)$ and $m_D(y_i)$:

$$g_D(y_i, c_j) \leq \sum_{v \in V_{c_*}^1} x_v^{y_i} + \sum_{v \in V_{c_*}^1 \cap V_{c_j}^2} x_v^{y_i}$$

$$- M(1 - z_i^j) + m_D(y_i) - \left(g_D(y_i, c_j) + |V_{c_j}^1| - \max_k |V_{c_k}^1|\right) \leq 0 \quad \forall i, j$$

$$\sum_j z_i^j \geq 1 \ \forall i, \quad z_i^j \in \{0, 1\} \ \forall i, j$$

The second and third constraints use a new set of binary variables z_i^j, where $z_i^j = 1$ indicates that in scenario y_i, $m_D(y_i)$ is at most the change in margin between c_j and c_*. The constraint $\sum_j z_i^j \geq 1$ requires that $m_D(y_i)$ must be bounded by one such value, and so can be at most the maximum margin. Using these variables, the MOV and POV MILPs are analogous to those for constructive control.

7 EXPERIMENTS

We now present experimental results comparing the performance of our approximation algorithms to the solutions found via mixed integer programming. We show results on four datasets. First, *netscience*, a collaboration network of researchers in network science, with 1461 nodes [38]. Second, *facebook*, the subgraph centered on 10 Facebook users, with 2888 nodes [37]. Third, *polblogs*, a network of links between political blogs, with 1224 nodes [38]. Fourth, *irvine*, a graph representing instant messages exchanged between students at U.C. Irvine, with 1889 nodes [37]. We select these datasets because they represent the kinds of networks on which political messages (such as

Table 1: Percent of MILP value obtained by approximation algorithm.

|	netscience			facebook			polblogs			irvine		
k =	25	50	100	25	50	100	25	50	100	25	50	100
Constructive												
$|C|$= 2	99.5	99.3	100.	100	100	100	100	100	99.4	100	100	99.4
$|C|$= 5	82.8	90.1	91.5	90.8	90.9	90.9	97.4	97.8	99.5	97.7	95.4	96.8
$|C|$= 10	80.9	89.1	98.3	80.9	83.7	88.6	98.7	99.2	99.6	96.9	97.4	98.9
Destructive												
$|C|$= 2	99.9	99.6	99.9	100	100	100	100	100	99.4	100	99.8	98.8
$|C|$= 5	73.8	73.2	83.3	87.7	79.7	81.6	100	97.8	99.3	100	99.0	100
$|C|$= 10	75.9	87.2	97.2	81.8	85.0	89.0	98.9	99.3	99.6	100	97.3	99.0
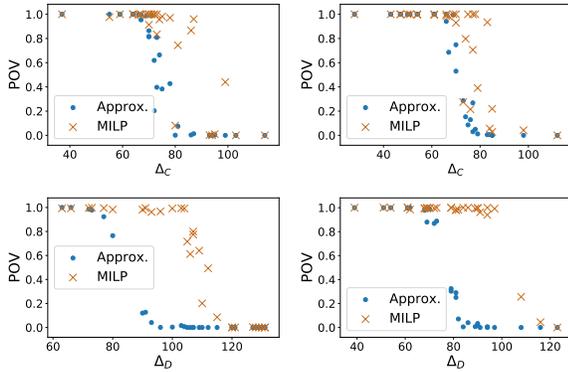
Figure 1: Probability of victory. Top: constructive. Bottom: destructive. Left: netscience. Right: polblogs.

fake news) spread. We also note that our approximation algorithms can easily be scaled to much larger networks since we can apply the same techniques developed in the influence maximization literature [14, 42]. However, our focus here is to characterize the performance of our algorithms in comparison to the optimal solution, so we select datasets which are feasible for mixed integer programming. For each network, we randomly generated 30 sets of voter preferences.

We start out with the MOV objective. Table 1 shows the percentage of the MILP's value which is obtained by our approximation algorithms (MOVCONSTRUCTIVE and MOVDESTRUCTIVE respectively), averaging over the 30 instances on each network with propagation probability $p = 0.1$. We vary the number of seed nodes k and the number of candidates $|C|$. We see that the approximation algorithms perform well across all settings, obtaining expected change in margin at least 73% of that of the MILP. Variance over the 30 sampled instances was uniformly low (with standard deviation less than 0.05 of the reported mean in almost all cases). For 2-candidate elections the approximation algorithms obtain nearly 100% of the optimal value on all networks. The empirical approximation ratio degrades with the number of candidates, particularly when the budget k is small. Overall, our approximation algorithms are highly effective

for election control via the MOV in both constructive and destructive control, particularly with a moderate number of candidates.

We now turn to the POV objectives. We show results for $k = 50$, $|C|= 5$, comparing our bicriteria approximation algorithms to the corresponding MILP formulations. To keep the experiments timely, we ran each approximation algorithm for 150 random values of the threshold β instead of enumerating over all (empirically, this resulted in very similar solution quality). Figure 1 shows the results on netscience and polblogs for constructive and destructive control. Results for facebook and irvine can be found in the supplement. The x axis shows the starting margin (Δ_C or Δ_D) in each randomly generated instance while the y axis shows the POV. The instances fall into three groups. First, when the margin is small, both the approximation algorithms and the MILP have a high POV. Second, when the margin is large, both have a small POV. Third are intermediate points where the approximation algorithm and MILP strongly diverge. Averaged over all random instances, the approximation algorithm obtains 40-60% of the MILP's value (depending on the network). However, there are instances among the intermediate cases where, e.g., the approximation algorithm obtains a POV of 0.1%, but the MILP finds a solution with POV 99%. We conclude that election control for the POV objective can be very computationally difficult in narrowly winnable elections, dovetailing with our theoretical results.

8 CONCLUSION

Fake news and other targeted misinformation are an increasingly prevalent way of interfering with democratic elections. We introduce and study the problem of election control through social influence, providing algorithms and hardness results for maximizing both the margin and probability of victory for an attacker in both constructive and destructive control. Our results indicate that social influence is a salient threat to election integrity, particularly in the MOV case where we provide high-quality approximation algorithms. Maximizing the probability of victory is manageable in easier instances, but difficult both theoretically and empirically in narrow races.

REFERENCES

[1] John S Ahlquist, Kenneth R Mayer, and Simon Jackman. 2014. Alien abduction and voter impersonation in the 2012 US general election: Evidence from a survey list experiment. *Election Law Journal* 13, 4 (2014), 460–475.

[2] Hunt Allcott, Matthew Gentzkow, et al. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 2 (2017), 211–236.

[3] John J Bartholdi, Craig A Tovey, and Michael A Trick. 1992. How hard is it to control an election? *Mathematical and Computer Modelling* 16, 8-9 (1992), 27–40.

[4] Dorothea Baumeister, Gábor Erdélyi, Olivia J Erdélyi, and Jörg Rothe. 2015. Complexity of manipulation and bribery in judgment aggregation for uniform premise-based quota rules. *Mathematical Social Sciences* 76 (2015), 19–30.

[5] Ted Brader. 2005. Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions. *American Journal of Political Science* 49, 2 (2005), 388–405.

[6] Robert Bredereck and Edith Elkind. 2017. Manipulating opinion diffusion in social networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*.

[7] Alexandra Carpentier and Michal Valko. 2016. Revealing graph bandits for maximizing local influence. In *Artificial Intelligence and Statistics*. 10–18.

[8] Jiehua Chen, Piotr Faliszewski, Rolf Niedermeier, and Nimrod Talmon. 2015. Elections with Few Voters: Candidate Control Can Be Easy.. In *AAAI*, Vol. 15. 2045–2051.

[9] Wei Chen, Tian Lin, Zihan Tan, Mingfei Zhao, and Xuren Zhou. 2016. Robust influence maximization. *arXiv preprint arXiv:1601.06551* (2016).

[10] Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1029–1038.

[11] Wei Chen, Yajun Wang, and Yang Yuan. 2013. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*. 151–159.

[12] Feng Chi and Nathan Yang. 2011. Twitter adoption in Congress. *Review of Network Economics* 10, 1 (2011).

[13] Ray Christensen and Thomas J Schultz. 2014. Identifying election fraud using orphan and low propensity voters. *American Politics Research* 42, 2 (2014), 311–337.

[14] Edith Cohen, Daniel Delling, Thomas Pajor, and Renato F Werneck. 2014. Sketch-based influence maximization and computation: Scaling up with guarantees. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 629–638.

[15] Vincent Conitzer. 2013. The maximum likelihood approach to voting on social networks. In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*. IEEE, 1482–1487.

[16] Vincent Conitzer, Tuomas Sandholm, and Jérôme Lang. 2007. When are elections with few candidates hard to manipulate? *Journal of the ACM (JACM)* 54, 3 (2007), 14.

[17] Stefano DellaVigna and Ethan Kaplan. 2007. The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics* 122, 3 (2007), 1187–1234.

[18] Nan Du, Le Song, Manuel Gomez Rodriguez, and Hongyuan Zha. 2013. Scalable influence estimation in continuous-time diffusion networks. In *Advances in neural information processing systems*. 3147–3155.

[19] Gábor Erdélyi, Edith Hemaspaandra, and Lane A Hemaspaandra. 2015. More natural models of electoral control by partition. In *International Conference on Algorithmic DecisionTheory*. Springer, 396–413.

[20] Gábor Erdélyi, Markus Nowak, and Jörg Rothe. 2009. Sincere-Strategy Preference-Based Approval Voting Fully Resists Constructive Control and Broadly Resists Destructive Control. *Mathematical Logic Quarterly* 55, 4 (2009), 425–443.

[21] Gábor Erdélyi, Christian Reger, and Yongjie Yang. 2017. The Complexity of Bribery and Control in Group Identification. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1142–1150.

[22] Piotr Faliszewski, Edith Hemaspaandra, and Lane A Hemaspaandra. 2009. How hard is bribery in elections? *Journal of Artificial Intelligence Research* 35 (2009), 485–532.

[23] Piotr Faliszewski, Edith Hemaspaandra, and Lane A Hemaspaandra. 2011. Multi-mode control attacks on elections. *Journal of Artificial Intelligence Research* 40, 1 (2011), 305–351.

[24] Piotr Faliszewski, Edith Hemaspaandra, Lane A Hemaspaandra, and Jörg Rothe. 2009. Llull and Copeland voting computationally resist bribery and constructive control. *Journal of Artificial Intelligence Research* 35 (2009), 275–341.

[25] Alan S Gerber, Dean Karlan, and Daniel Bergan. 2009. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics* 1, 2 (2009), 35–52.

[26] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. 2011. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*. ACM, 47–48.

[27] Xinran He and David Kempe. 2016. Robust Influence Maximization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. 885–894. https://doi.org/10.1145/2939672.2939760

[28] Edith Hemaspaandra, Lane A Hemaspaandra, and Jörg Rothe. 2007. Anyone but him: The complexity of precluding an alternative. *Artificial Intelligence* 171, 5-6 (2007), 255–285.

[29] Jesse Holcomb, Jeffrey Gottfried, and Amy Mitchell. 2013. News use across social media platforms. *Pew Research Journalism Project* (2013).

[30] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.

[31] Andreas Krause, H Brendan McMahan, Carlos Guestrin, and Anupam Gupta. 2008. Robust submodular observation selection. *Journal of Machine Learning Research* 9, Dec (2008), 2761–2801.

[32] Siyu Lei, Silviu Maniu, Luyi Mo, Reynold Cheng, and Pierre Senellart. 2015. Online influence maximization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 645–654.

[33] Hong Liu, Haodi Feng, Daming Zhu, and Junfeng Luan. 2009. Parameterized computational complexity of control problems in voting systems. *Theoretical Computer Science* 410, 27-29 (2009), 2746–2753.

[34] Andrea Loreggia, Nina Narodytska, Francesca Rossi, K Brent Venable, and Toby Walsh. 2015. Controlling elections by replacing candidates or votes. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1737–1738.

[35] Meghna Lowalekar, Pradeep Varakantham, and Akshat Kumar. 2016. Robust influence maximization. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1395–1396.

[36] Elchanan Mossel, Joe Neeman, and Omer Tamuz. 2014. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems* 28, 3 (2014), 408–429.

[37] KONECT network dataset collection. 2017. http://konect.uni-koblenz.de/networks/ego-facebook. (2017).

[38] Mark Newman. 2017. Network data. http://www-personal.umich.edu/~mejn/netdata/. (2017).

[39] Gordon Pennycook, Tyrone D Cannon, and David G Rand. 2017. Prior exposure increases perceived accuracy of fake news. (2017).

[40] Ariel D. Procaccia, Nisarg Shah, and Eric Sodomka. 2015. Ranked Voting on Social Networks. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. 2040–2046. http://ijcai.org/Abstract/15/289

[41] Manuel Gomez Rodriguez and Bernhard Schölkopf. 2012. Influence maximization in continuous time diffusion networks. *arXiv preprint arXiv:1205.1682* (2012).

[42] Youze Tang, Xiaokui Xiao, and Yanchen Shi. 2014. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. ACM, 75–86.

[43] Alan Tsang and Kate Larson. 2016. The echo chamber: Strategic voting and homophily in social networks. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 368–375.

[44] Sunil Wattal, David Schuff, Munir Mandviwalla, and Christine B Williams. 2010. Web 2.0 and politics: the 2008 US presidential election and an e-politics research agenda. *MIS quarterly* (2010), 669–688.

[45] Bryan Wilder, Amulya Yadav, Nicole Immorlica, Eric Rice, and Milind Tambe. 2017. Uncharted but not Uninfluenced: Influence Maximization with an uncertain network. In *Proceedings of the 16th Conference on Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1305–1313.

[46] Yongjie Yang, Yash Raj Shrestha, and Jiong Guo. 2016. How Hard Is Bribery with Distance Restrictions?. In *ECAI*. 363–371.