Exploiting Social Influence to Control Elections Based on Scoring Rules

Federico Corò[*], **Emilio Cruciani**, **Gianlorenzo D'Angelo** and **Stefano Ponziani**
Gran Sasso Science Institute, L'Aquila, Italy
{federico.coro, emilio.cruciani, gianlorenzo.dangelo, stefano.ponziani}@gssi.it

Abstract

We consider the election control problem in social networks which consists in exploiting social influence in a network of voters to change their opinion about a target candidate with the aim of increasing his chances to win (constructive control) or lose (destructive control) the election. Previous works on this problem focus on plurality voting systems and on a influence model in which the opinion of the voters about the target candidate can only change by shifting its ranking by one position, regardless of the amount of influence that a voter receives. We introduce *Linear Threshold Ranking*, a natural extension of *Linear Threshold Model*, which models the change of opinions taking into account the amount of exercised influence. In this general model, we are able to approximate the maximum score that a target candidate can achieve up to a factor of $1 - 1/e$ by showing submodularity of the objective function. We exploit this result to provide a $\frac{1}{3}(1 - 1/e)$-approximation algorithm for the *constructive* election control problem and a $\frac{1}{2}(1 - 1/e)$-approximation ratio in the *destructive* scenario. The algorithm can be used in *arbitrary scoring rule voting systems*, including *plurality rule* and *borda count*.

1 Introduction

All of us have specific personal opinions on certain topics, such as lifestyle or consumer products. These opinions, normally formed on personal life experience and information, can be conditioned by the interaction with our friends leading to a change in our original opinion on a particular topic if a large part of our friends holds a different opinion. Moreover, opinions can propagate through a social network, generating a diffusion process. This phenomenon of opinion diffusion has been intensely investigated from many different perspectives, from sociology to economics. In recent years, there has been a growing interest on the relationship between social networks and political campaigning. Political campaigns nowadays use online social networks to lead elections in their favor; for example, they can target specific voters with fake news [Allcott and Gentzkow, 2017]. A real-life example of political intervention in this context occurred in the US Congressional elections in 2010, where a set of users were encouraged to vote with a message on Facebook. These messages directly influenced the real-world voting behavior of millions of people [Bond *et al.*, 2012]. Another example is that of French elections in 2017, where automated accounts in Twitter spread a considerable portion of political content trying to influence the outcome [Ferrara, 2017].

There exist an extensive literature on manipulating elections without considering the underlying social network structure of the voters, e.g., swap bribery [Elkind *et al.*, 2009], shift bribery [Bredereck *et al.*, 2016]; we point the reader to a recent survey [Faliszewski *et al.*, 2016]. Nevertheless, there are only few studies that exploit opinion diffusion in social networks to change the outcome of elections. The Independent Cascade Model [Kempe *et al.*, 2015] has been considered as diffusion process to guarantee that a target candidate wins/loses [Bartholdi *et al.*, 1992; Hemaspaandra *et al.*, 2007]. The constructive (destructive) election control problem has been introduced in [Wilder and Vorobeychik, 2018a] and consists in changing voters' opinions with the aim of maximizing (minimizing) the margin and probability of victory of a specific target candidate. A variant of the Linear Threshold Model [Kempe *et al.*, 2015] with weights on the vertices has been considered on a graph in which each node is a cluster of voters with a specific list of candidates and there is an edge between two nodes if they differ by the ordering of a single pair of adjacent candidates [Faliszewski *et al.*, 2018]. Moreover, it has been studied how to manipulate the network in order to have control on the majority opinion, e.g., bribing or adding/deleting edges, on a simple Linear Threshold Model where each node holds a binary opinion, each edge has a fixed weight, and all vertices have a threshold fixed to $1/2$ [Bredereck and Elkind, 2017]. The study of opinion diffusion modeled as a majority dynamics has attracted much attention in recent literature [Auletta *et al.*, 2015; Brill *et al.*, 2016; Botan *et al.*, 2017]. In these models each agent has an initial preference list and at each time step a subset of agents updates their opinions according to some majority -based rule that depends on their neighbors in the network.

In this work we focus on the election control through social influence problem [Wilder and Vorobeychik, 2018a]: Given

[*]Contact Author

a social network of voters, we want to select a subset of voters such that their influence will change the opinion about a target candidate, maximizing its chances to win or lose (we remark that we are in the specific scenario in which only the opinions about a target candidate can be changed). In previous work, the only voting system studied is the plurality rule. Moreover, in the diffusion model considered in the literature, an influenced voter changes the position of the target candidate in its ranking by shifting it up or down by one position, regardless of the amount of influence received [Wilder and Vorobeychik, 2018a]. Here we study the election control problem in any scoring rule voting system and in a different diffusion model, that takes into account the degree of influence that voters exercise on the others and is able to describe the scenario in which a high influence on someone can radically change its opinion.

Original Contribution

- We introduce the *Linear Threshold Ranking*, a natural and powerful extension of the *Linear Threshold Model* for the election scenario that takes into account the degree of influence of the voters on each other.
- We show that maximizing the score of a target candidate is monotone and submodular for *arbitrary* scoring function (including popular voting systems, e.g., *plurality rule* or *borda count*), with any number of candidates. This implies that a greedy hill-climbing algorithm achieves a $1 - 1/e$ factor approximation for the problem of maximizing the score.
- Exploiting the previous result, we achieve a $\frac{1}{3}(1 - 1/e)$ factor to the problem of maximizing the *Margin of Victory* of a target candidate in *arbitrary scoring rule voting systems* with any number of candidates.
- We give a simple reduction that maps *destructive control* problems to constructive control ones and allows us to achieve a $\frac{1}{2}(1 - 1/e)$-approximation.

Due to space constraints, some of the proofs are omitted or only sketched. Full proofs, along with an experimental study, can be found in the full version.[1]

2 Background

In this section we present some notions and concepts about *voting systems* and *influence maximization* on social networks that will be used in the design and analysis of the algorithm.

Voting Systems

Voting systems are sets of rules that regulate all aspects of elections and that determine their outcome. Herein we consider two *single-winner* voting systems: (i) *Plurality rule*: Each voter can only express a single preference among the candidates; the winner is the candidate with the highest number of voter, i.e., the *plurality*. (ii) *Scoring rule*: Each voter expresses his preference as a *ranking*; each candidate is then assigned a *score*, computed as a function of the positions he was ranked. The former is arguably the simplest voting rule and one of the most commonly used. The latter is a general definition that include several popular election methods by choosing an adequate *scoring function*: (i) *plurality rule*: 1 point to the first candidate and 0 to all the others; (ii) *t-approval*: 1 point to the first t candidates and 0 to the others (each voter approves t candidates); (iii) *t-veto* or *anti-plurality*: 1 point to the first $m - t$ candidates and 0 to the remaining t, with m the number of candidates; (iv) *borda count*: $m - l$ points to candidate in position l.

Influence Maximization

Influence maximization is the problem of finding the subset of nodes of a graph that maximizes the spread of information. *Linear Threshold Model* (LTM) is one of the most used models to study influence diffusion in social networks [Kempe *et al.*, 2015]. Given a graph $G = (V, E)$, in LTM each node $v \in V$ has a threshold $t_v \in [0, 1]$ sampled uniformly at random and independently and each edge $(u, v) \in E$ has a weight $b_{uv} \in [0, 1]$ such that, for each $v \in V$, the sum of the weights of the incoming edges is less than or equal to 1, i.e., $\sum_{(u,v)\in E} b_{uv} \leq 1$. Let $A_t \subseteq V$ be the set of *active* nodes at time t, where A_0 is the set of nodes that are active at the beginning of the process. In LTM a node v becomes active if the sum of the weights of the edges coming from active nodes is greater than or equal to its threshold t_v, i.e., $v \in A_t$ if and only if $v \in A_{t-1}$ or $\sum_{u\in A_{t-1}:(u,v)\in E} b_{uv} \geq t_v$. The process has *quiesced* at the first time $\tilde{t}$ in which the set of active nodes does not change anymore, i.e., time $\tilde{t}$ is such that $A_{\tilde{t}} = A_{\tilde{t}+1}$. We define the eventual set of active nodes as $A := A_{\tilde{t}}$.

The most central result in LTM is the following [Kempe *et al.*, 2015]: Starting from any set A_0, the distribution of A is equivalent to the distribution of reachable nodes in the set of random graphs called *live-edge graphs* [Kempe *et al.*, 2015]. In live-edge graphs, subgraphs where each node has at most one incoming edge, the problem of selecting the initial set of nodes to maximize the diffusion is monotone and *submodular*;[2] hence, it can be approximated to a factor of $1 - 1/e$ using a simple greedy hill-climbing algorithm [Nemhauser *et al.*, 1978]. While it is $\#P$-hard to compute the expected number of active nodes, there exists a simulation-based approach in which the spread of influence can be evaluated by sampling a polynomial number of live-edges [Kempe *et al.*, 2015, Proposition 4.1].

3 Linear Threshold Ranking

We consider the scenario in which a set of candidates are running for the elections and a social network of voters will decide the winner. Some attacker could be interested in changing the outcome of the elections by sending ads and/or (possibly fake) news about a specific candidate to a subset of voters, that could share the news and influence their friends. Is it possible for the attacker to select a subset of voters to control the election, i.e., to change voters' opinions about a target candidate, maximizing his chances to win/lose the elections?

More formally, let $G = (V, E)$ be a directed graph representing the underlying social network. For each node $v \in V$

[1] https://arxiv.org/abs/1902.07454

[2] For a set N, a function $z : 2^N \to \mathbb{R}$ is *submodular* if $\forall S, T \subseteq N$ s.t. $S \subseteq T$, $\forall e \in N \setminus T$ holds $z(S \cup \{e\}) - z(S) \geq z(T \cup \{e\}) - z(T)$.

we call N_v the set of incoming neighbors of v. Let $C = \{c_1, \dots, c_m\}$ be a set of m candidates nominated for the elections; we refer to our *target candidate*, i.e., the one that we want to make win/lose the elections, as $c_\star$. Each $v \in V$ has a permutation π_v of C, i.e., its list of preferences for the elections; we denote the position of candidate c_i in the preference list of node v as $\pi_v(c_i)$.

Let $B \in \mathbb{N}$ be the initial budget, i.e., the maximum size of the set of active nodes A_0 from which the LTM process starts. After the LTM process has quiesced, the position of $c_\star$ in the preference list of each node changes according to a function of its incoming active neighbors. The threshold t_v of each node $v \in V$ models its strength in retaining its original opinion about candidate $c_\star$: The higher is the threshold the lower is the probability that v is influenced by its neighbors. The weight on an edge b_{uv} measures the influence that node u has on node v. Taking into account the role of such parameters, we define the number of positions that $c_\star$ goes up in π_v as

$$\pi_v^{\uparrow}(c_\star) := \min\left(\pi_v(c_\star) - 1,\ \left\lfloor \frac{\alpha(\pi_v(c_\star))}{t_v} \sum_{u \in A,\, (u,v) \in E} b_{uv} \right\rfloor\right),$$

where $\alpha : \{1, \dots, m\} \to [0, 1]$ is a function that depends on the position of $c_\star$ in π_v and models the rate at which $c_\star$ shifts up. Note that α can be set arbitrarily to model different scenarios, e.g., shifting up of one position from the bottom of the list could be easier than moving from the second to the first position. Note that $\pi_v^{\uparrow}(c_\star)$ can be any integer value in $\{0, \dots, \pi_v(c_\star) - 1\}$: The floor function guarantees a positive integer value and the minimum between such value and $\pi_v(c_\star) - 1$ guarantees that the final position of $c_\star$ is at least 1. We call this process the *Linear Threshold Ranking* (LTR).

After LTR, the candidates might have a new position in the preference list of each node $v \in V$, that we denote as $\tilde{\pi}_v$. In particular, the new position of $c_\star$ will be $\tilde{\pi}_v(c_\star) := \pi_v(c_\star) - \pi_v^{\uparrow}(c_\star)$; the candidates that are overtaken by $c_\star$ will shift one position down.

In the problem of *election control* we want to maximize the chances of the target candidate to win the elections under LTR. To achieve that, we maximize its expected *Margin of Victory* (MoV) w.r.t. the most voted opponent, akin to that defined in [Wilder and Vorobeychik, 2018a].[3] Let us consider the general case of the *scoring rule*, where a *non-increasing scoring function* $f : \{1, \dots, m\} \to \mathbb{N}$ assigns a score to each position. Let c and $\tilde{c}$ be the candidates, different from $c_\star$, with the highest score before and after LTR, respectively. Let

$$\mu(\emptyset) := \textstyle\sum_{v \in V} (f(\pi_v(c)) - f(\pi_v(c_\star))) \quad (1)$$

$$\mu(A_0) := \textstyle\sum_{v \in V} (f(\tilde{\pi}_v(\tilde{c})) - f(\tilde{\pi}_v(c_\star))) \quad (2)$$

be the *margin* (i.e., difference in score) between the most voted opponent and $c_\star$ before and after LTR, respectively. Thus, the *election control* problem is formalized as that of finding a set of nodes A_0 such that

$$\begin{aligned} \max_{A_0} \quad & \mathbf{E}[\mathrm{MoV}(A_0)] := \mathbf{E}[\mu(\emptyset) - \mu(A_0)] \\ \text{s.t.} \quad & |A_0| \le B, \end{aligned}$$

namely to find an initial set of seed nodes of at most size B that maximizes the expected MoV, i.e., change in margin.[4]

To solve the problem we focus on the score of the target candidate. Let us define

$$F(\emptyset) := \textstyle\sum_{v \in V} f(\pi_v(c_\star)) \quad (3)$$

$$F(A_0) := \mathbf{E}\left[\textstyle\sum_{v \in V} f(\tilde{\pi}_v(c_\star))\right] \quad (4)$$

as the total expected score obtained by candidate $c_\star$ before and after LTR, respectively. In Sections 4 and 5 we prove that the score of the target candidate is a monotone submodular function w.r.t. the initial set of seed nodes A_0 in both the *plurality* and the *scoring* rule; this allows us to get a $(1 - 1/e)$-approximation of the maximum score through the use of a greedy algorithm that iteratively selects the node that maximizes the increment in score [Nemhauser *et al.*, 1978]; we denote this algorithm as GREEDY. Note that maximizing the score of the target candidate is a NP-hard problem: Consider the case in which there are only two candidates, $\alpha(1) = \alpha(2) = 1$, all nodes have $c_\star$ as second preference, and the scoring function is that of the plurality rule; maximizing the score is equal to maximizing the number of active nodes in LTM because when a node becomes active the target candidate shifts of at least one position up (in this case, in first position); thus the two problems are equivalent. Influence maximization in LTM is NP-hard [Kempe *et al.*, 2015] and maximizing the score in LTR is also NP-hard because it generalizes it. Moreover, in this instance, the maximum value of MoV is equal to twice the maximum score; then maximizing MoV is also NP-hard.

Although maximizing the score is not equivalent to maximizing MoV, in Section 6 we show that we can use GREEDY to obtain a constant factor approximation to MoV. Finally, in Section 7, we consider the problem of *destructive control*, in which we want the target candidate to lose the elections and prove a constant factor approximation also in this scenario.

4 Maximizing the Score: Plurality Rule

As a warm-up, in this section we focus on the *plurality rule*. We give an algorithm to select an initial set of seed nodes to maximize the expected number of nodes that will change their opinion and have $c_\star$ as first preference at the end of LTR.

Let A_0 be the initial set of seed nodes and A the set of active nodes at the end of the process. An active node v with $\pi_v(c_\star) > 1$ will have $c_\star$ as first preference if $\pi_v^{\uparrow}(c_\star) = \pi_v(c_\star) - 1$, that is if and only if $\frac{\alpha(\pi_v(c_\star))}{t_v} \sum_{u \in A \cap N_v} b_{uv} \ge \pi_v(c_\star) - 1$ or, equivalently, $t_v \le \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star) - 1} \sum_{u \in A \cap N_v} b_{uv}$.

As in influence maximization problems, we define an alternative random process based on live-edge graphs, since in live-edge graph process we don't know the value of t_v and we cannot compute which nodes satisfy the above formula.

Definition 1. *Live-edge Coin Flip process* (LCF):

1. Each node $v \in V$ selects at most one of its incoming edges with probability proportional to the weight of that edge, i.e., edge (u, v) is selected with probability b_{uv}, and no edge is selected with probability $1 - \sum_{u \in N_v} b_{uv}$.

[3] We study the *change* in the margin, and not just the margin, to have well defined approximation ratios when the margin is negative.

[4] MoV is positive since the scoring function f is nonincreasing.

2. Each node v with $\pi_v(c_\star) > 1$ that is reachable from A_0 in the live-edge graph flips a biased coin and changes its list according to the outcome. In detail, v picks a random real number $s_v \in [0, 1]$ and sets the position of $c_\star$ according to s_v: If $s_v \le \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-1}$, node v sets $\tilde{\pi}_v(c_\star) = 1$ and shifts all the other candidates down by one position; otherwise, v maintains its original ranking.

We show that the two processes are equivalent, i.e., starting from any initial set A_0 each node has the same probability to end up with $c_\star$ in first position in both processes. This allows us to compute the function $F(A_0)$, for a given A_0, by simply solving a reachability problem in graphs.

We denote by $\mathcal{G}$ the set of all possible live-edge graphs sampled from G. For every $G' = (V, E') \in \mathcal{G}$ we denote by $\mathbf{P}(G')$ the probability that G' is sampled, namely

$$\mathbf{P}(G') = \prod_{v:(u,v)\in E'} b_{uv} \prod_{v:\nexists(u,v)\in E'} \left(1 - \sum_{w:(w,v)\in E} b_{wv}\right).$$

We denote by $R(A_0)$ the set of nodes reachable from A_0 at the end of the LCF process and by $R_{G'}(A_0)$ the set of nodes reachable from A_0 in G' and by $\mathbf{1}_{(G',v)}$ the indicator function that is 1 if $v \in R_{G'}(A_0)$ and 0 otherwise.

Lemma 1. *For any seed set A_0 and any node v it holds that*

$$\mathbf{P}(v \in R(A_0)) = \sum_{U\subseteq N_v} \sum_{u\in U} b_{uv} \cdot \mathbf{P}((R(A_0) \cap N_v) = U).$$

Lemma 1 tells us how to compute the probability that a node v is reachable from A_0 at the end of the LCF process by using live-edge graphs or by using the probability of the incoming neighbors of v to be reachable from A_0. The next theorem shows the equivalence between LTR and LCF.

Theorem 1. *Given a set of initially active nodes A_0, let A'_{LTR} and A'_{LCF} be the set of nodes such that $\tilde{\pi}_v(c_\star) = 1$ at the end of LTR and LCF, respectively, both starting from A_0. Then, for each $v \in V$, $\mathbf{P}(v \in A'_{LTR}) = \mathbf{P}(v \in A'_{LCF})$.*

Proof. We exclude from the analysis the nodes v with $\pi_v(c_\star) = 1$ since they keep their original ranking in both models. Let us start by analyzing the LTR process. Let A be the set of nodes activated in LTR from starting from A_0. Given a set U of in-neighbors of v, we can write the probability that $v \in A'_{LTR}$ given that U are the only active in-neighbors of v (i.e. $A \cap N_v = U$) as

$$\mathbf{P}(v \in A'_{LTR} \mid (A \cap N_v) = U)$$
$$= \mathbf{P}\left(t_v \le \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-1} \sum_{u\in U} b_{uv}\right) = \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-1} \sum_{u\in U} b_{uv}.$$

The overall probability $\mathbf{P}(v \in A'_{LTR})$ is equal to

$$\sum_{U\subseteq N_v} \mathbf{P}(v \in A'_{LTR} \mid (A \cap N_v) = U)\, \mathbf{P}(U = (A \cap N_v))$$
$$= \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-1} \sum_{U\subseteq N_v} \sum_{u\in U} b_{uv} \cdot \mathbf{P}((A \cap N_v) = U).$$

Let us now analyze the LCF process. To have $v \in A'_{LCF}$ we need that $v \in R(A_0)$ and that the coin toss has a positive outcome. Thus, $\mathbf{P}(v \in A'_{LCF}) = \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-1} \mathbf{P}(v \in R(A_0))$. Finally, by using Lemma 1 and the equivalence between the live-edge process and LTM [Kempe *et al.*, 2015, Proposition 4.1]) the theorem follows. □

We now exploit Theorem 1 to show how to compute the value of $F(A_0)$. In the case of plurality rule we have that

$$F(A_0) = \mathbf{E}[|A'_{LTR}|] = \sum_{v\in V} \mathbf{P}(v \in A'_{LCF})$$
$$= F(\emptyset) + \sum_{v\in V, \pi_v(c_\star)>1} \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-1} \mathbf{P}(v \in R(A_0)).$$

Thanks to Lemma 1, we can rewrite the above formula as

$$F(A_0) - F(\emptyset) = \sum_{r=2}^{m} \frac{\alpha(r)}{r-1} \sum_{G'\in\mathcal{G}} \mathbf{P}(G')\, |R_{G'}(A_0, V^r_{c_\star})|,$$

where, for a graph $G' \in \mathcal{G}$ and an integer $r \le m$, we denoted by $V^r_{c_i}$ the set of nodes that have candidate c_i in position r and $R_{G'}(A_0, V^r_{c_\star}) = \{v : v \in R_{G'}(A_0) \wedge \pi_v(c_\star) = r\}$.

It follows that the function $F(A_0)$ is a non-negative linear combination of functions $|R_{G'}(A_0, V^r_{c_\star})|$. In the next lemma, we show that these functions are monotone and submodular w.r.t. A_0 and this implies that also $F(A_0) - F(\emptyset)$ is monotone and submodular w.r.t. A_0. Therefore, we can use GREEDY to find a set A_0 such that $F(A_0) - F(\emptyset)$ is at least $1 - 1/e$ times the optimum [Nemhauser *et al.*, 1978].

Lemma 2. *Given a graph $G' \in \mathcal{G}$ and a positive integer $r \le m$, the size of $R_{G'}(A_0, V^r_{c_\star})$ in G' is a monotone submodular function of A_0.*

5 Maximizing the Score: Scoring Rule

In this section we extend the results of Section 4 to the general case of the *scoring rule*, in which a *scoring function* f assigns a score to each candidate according to the positions he was ranked in the voters' lists. The overall approach is similar, but more general: We first define an alternative random process, called *Live-edge Dice Roll* (LDR), and show its equivalence to LTR; then we use LDR to compute $F(A_0)$ and show that it is a monotone submodular function of the initial set of active nodes A_0. This latter result allows us to compute a set A_0 that has an approximation guarantee of $1 - 1/e$ on the maximization of the score of the target candidate with GREEDY. Process LDR is defined as follows.

Definition 2. *Live-edge Dice Roll process* (LDR):

1. Each node $v \in V$ selects at most one of its incoming edges with probability proportional to the weight of that edge, i.e., edge (u, v) is selected with probability b_{uv}, and no edge is selected with probability $1 - \sum_{u\in N_v} b_{uv}$.
2. Each node v with $\pi_v(c_\star) > 1$ that is reachable from A_0 in the live-edge graph rolls a biased $\pi_v(c_\star)$-sided dice and changes its list according to the outcome. This is

equivalent to picking a random real number s_v in $[0,1]$ and setting the position of $c_\star$ according to s_v as follows:

$$\tilde{\pi}_v(c_\star) = \begin{cases} 1 & \text{if } s_v \le \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-1}, \\ \ell & \text{if } \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-\ell+1} < s_v \le \frac{\alpha(\pi_v(c_\star))}{\pi_v(c_\star)-\ell}, \\ & \quad \text{for } \ell = 2, \dots, \pi_v(c_\star)-1, \\ \pi_v(c_\star) & \text{if } s_v > \alpha(\pi_v(c_\star)). \end{cases}$$

If $\tilde{\pi}_v(c_\star) \neq \pi_v(c_\star)$, all candidates between $\tilde{\pi}_v(c_\star)$ and $\pi_v(c_\star)-1$ are shifted down by one position.

We show that LTR and LDR have the same distribution.

Theorem 2. *Given a set of initially active nodes A_0 and a node $v \in V$, let $\tilde{\pi}_v^{LTR}(c_\star)$ and $\tilde{\pi}_v^{LDR}(c_\star)$ be the position of node v at the end of LTR and LDR, respectively, both starting from A_0. Then, $\mathbf{P}\left(\tilde{\pi}_v^{LTR}(c_\star)=\ell\right) = \mathbf{P}\left(\tilde{\pi}_v^{LDR}(c_\star)=\ell\right)$, for each $\ell = 1,\dots,\pi_v(c_\star)$.*

Proof. Let A be the set of active nodes at the end of the LTR process that starts from A_0. The probability that an active node moves candidate $c_\star$ from position r to position ℓ is:

$$\mathbf{P}(r,\ell) := \begin{cases} \frac{\alpha(r)}{r-1} & \text{if } \ell = 1, \\ \frac{\alpha(r)}{r-\ell} - \frac{\alpha(r)}{r-\ell+1} & \text{if } \ell = 2,\dots,r-1, \\ 1-\alpha(r) & \text{if } \ell = r, \end{cases}$$

for each $r, \ell \in \{1,\dots,m\}$, $\ell \le r$. In particular, for a node v, the probability that the second step of LDR yields $\tilde{\pi}_v(c_\star) = \ell$, for $\ell = 1,\dots,\pi_v(c_\star)$, is $\mathbf{P}(\pi_v(c_\star),\ell)$.

We have that $\mathbf{P}\left(\tilde{\pi}_v^{LTR}(c_\star)=\ell\right)$ is equal to

$$\sum_{U \subseteq N_v} \mathbf{P}\left(\tilde{\pi}_v^{LTR}(c_\star)=\ell \,\middle|\, (A\cap N_v)=U\right) \mathbf{P}\left((A\cap N_v)=U\right).$$

If U is the maximal subset of active neighbors of v (i.e., $U = A \cap N_v$), then we can write the probability that $\tilde{\pi}_v^{LTR}(c_\star)=\ell$ given U as follows:

$$\mathbf{P}\left(\tilde{\pi}_v^{LTR}(c_\star)=\ell \,\middle|\, (A\cap N_v)=U\right) = \mathbf{P}(\pi_v(c_\star),\ell)\sum_{u\in U} b_{uv}.$$

Therefore, $\mathbf{P}\left(\tilde{\pi}_v^{LTR}(c_\star)=\ell\right)$ is equal to

$$\mathbf{P}(\pi_v(c_\star),\ell)\sum_{U\subseteq N_v}\sum_{u\in U} b_{uv}\mathbf{P}\left((A\cap N_v)=U\right).$$

Recall that, in LDR, $\mathbf{P}\left(\tilde{\pi}_v^{LDR}(c_\star)=\ell\right)$ is equal to $\mathbf{P}(v \in R(A_0)) \cdot \mathbf{P}(\pi_v(c_\star),\ell)$. By Lemma 1, it follows that

$$\mathbf{P}(v\in R) = \sum_{U\subseteq N_v}\sum_{u\in U} b_{uv}\mathbf{P}\left((R\cap N_v)=U\right)$$

and hence $\mathbf{P}\left(\tilde{\pi}_v^{LDR}(c_\star)=\ell\right)$ is equal to

$$\mathbf{P}(\pi_v(c_\star),\ell)\sum_{U\subseteq N_v}\sum_{u\in U} b_{uv}\mathbf{P}\left((R(A_0)\cap N_v)=U\right).$$

Finally, using [Kempe *et al.*, 2015, Proposition 4.1], we get that $\mathbf{P}\left((R(A_0)\cap N_v)=U\right) = \mathbf{P}\left((A\cap N_v)=U\right)$. □

With some algebra, and by applying Lemma 1 and Theorem 2, we get the following formulation of $F(A_0)$:

$$F(A_0) = \sum_{r=1}^{m}\sum_{\ell=1}^{r} f(\ell)\mathbf{P}(\pi_v(c_\star),\ell)\sum_{G'\in\mathcal{G}}\mathbf{P}(G')\,|R_{G'}(A_0,V_{c_\star}^r)|.$$

Therefore, $F(A_0)$ is a non-negative linear combination of the monotone submodular function $|R_{G'}(A_0,V_{c_\star}^r)|$ (see Lemma 2), and hence $F(A_0)-F(\emptyset)$ is also monotone and submodular. Thus, we can use GREEDY to find a $(1-1/e)$-approximation to the problem of maximizing the score of the target candidate [Nemhauser *et al.*, 1978].

6 Maximizing the Margin of Victory

In previous sections we have shown that the problem of maximizing the score of the target candidate can be approximated within a factor $1-1/e$ by using GREEDY. In the following we show how to achieve a constant factor approximation to the original problem of maximizing the MOV by only maximizing the score of the target candidate. Given the equivalence of LCF and LDR with LTR, we can formulate our objective function as the average $\text{MOV}_{G'}$ computed on a sampled live-edge graph G', namely $\mathbf{E}[\text{MOV}(A_0)] = \mathbf{E}[\text{MOV}_{G'}(A_0)]$, where $\text{MOV}_{G'}(A_0) = \mu_{G'}(\emptyset) - \mu_{G'}(A_0)$, and $\mu_{G'}$ is the change in margin on a fixed G'.

We formulate the margin on the live-edge graphs in a way that is akin to that of [Wilder and Vorobeychik, 2018a]: We can exploit such formulation to prove our constant factor approximation with the same proof structure since also in our case the objective function is monotone and submodular (Lemma 2). For the *plurality rule* we have that

$$\mathbf{E}[\text{MOV}_{G'}(A_0)] := \sum_{r=2}^{m}\frac{\alpha(r)}{r-1}|R_{G'}(A_0,V_{c_\star}^r)| + \min_{c_z}\max_{c_i}|V_{c_i}^1| - |V_{c_z}^1| + \sum_{r=2}^{m}\frac{\alpha(r)}{r-1}|R_{G'}(A_0,V_{c_\star}^r\cap V_{c_z}^1)|,$$

where: the first term is the number of points gained by $c_\star$ after LTR; the second term (the first inside the minimum) is the number of points of the most voted opponent before LTR; the third one is the total number of points that the most voted opponent after LTR had before the process; the fourth term is the number of points that the most voted opponent after LTR lost because of the shifting of $c_\star$. Similarly, for the general case of arbitrary *scoring rule* we have

$$\mathbf{E}[\text{MOV}_{G'}(A_0)] := \sum_{r=2}^{m}\sum_{\ell=1}^{r-1}\mathbf{P}(r,\ell)|R_{G'}(A_0,V_{c_\star}^r)|(f(\ell)-f(r)) + \min_{c_z}\left(\max_{c_i}\sum_{r=1}^{m}f(r)|V_{c_i}^r| - \sum_{r=1}^{m}f(r)|V_{c_z}^r| + \sum_{r=2}^{m}\sum_{\ell=1}^{r-1}\sum_{h=\ell}^{r-1}\mathbf{P}(r,\ell)\,|R_{G'}(A_0,V_{c_\star}^r\cap V_{c_z}^h)|(f(h)-f(h+1))\right)$$

where the meaning of the terms is similar to above. This latter formulation is just a generalization of the plurality case whenever we choose f such that $f(1)=1$ and $f(r)=0$, for each $r \in \{2, \ldots, m\}$. In this way we would have that the gain in score would be just 1 and that $\frac{\alpha(r)}{r-1} = \mathbf{P}(r, 1)$.

In the following Theorem we prove that, up to the loss of a constant-factor in the approximation ratio, it suffices to concentrate only on the score of the target candidate $c_\star$ and not on the margin w.r.t. the most voted opponent.

Theorem 3. GREEDY *is a* $\frac{1}{3}(1 - 1/e)$*-approximation algorithm for the problem of election control in arbitrary scoring rule voting systems.*

Roughly speaking, the factor $\frac{1}{3}$ appears because we lower bound three terms in the MOV formulation to reconstruct the optimum in the approximation.

7 Destructive Election Control

In this section we focus on the *destructive election control* problem. Here we define, for each node $v \in V$, the number of positions of which $c_\star$ shifts down after LTR process as

$$\pi_v^{\downarrow}(c_\star) := \min\left(m - \pi_v(c_\star),\ \left\lfloor \frac{\alpha(\pi_v(c_\star))}{t_v} \sum_{u \in A,\, (u,v) \in E} b_{uv} \right\rfloor\right).$$

The final position of $c_\star$ in v will be $\tilde{\pi}_v(c_\star) := \pi_v(c_\star) + \pi_v^{\downarrow}(c_\star)$ and the overall score that $c_\star$ gets is

$$F_D(A_0) := \mathbf{E}\left[\sum_{v \in V} f(\pi_v(c_\star) + \pi_v^{\downarrow}(c_\star))\right].$$

The problem can be defined as that of finding an initial set of seed nodes A_0 that maximizes the expected MOV_D:

$$\begin{array}{ll} \max_{A_0} & \mathbf{E}[\text{MOV}_D(A_0)] := \mathbf{E}[\mu(A_0) - \mu(\emptyset)] \\ \text{s.t.} & |A_0| \le B, \end{array}$$

Similarly to the constructive case we are able to achieve a constant factor approximation, to do that we provide a reduction from the destructive to the constructive case. Given an instance of destructive control, we build an instance of constructive control in which we simply reverse the rankings of each node and complement the scoring function to its maximum value. Roughly speaking, this reduction maintains invariant the absolute value of the change in margin of the score of any candidate between the two cases. Formally, for each $v \in V$, the new instance has a preference list defined as $\pi'_v(c) := m - \pi_v(c) + 1$ for each candidate $c \in C$, and, for each position $r \in \{1, \ldots, m\}$, has a scoring function defined as $f'(r) := f_{\max} - f(m - r + 1)$, where $f_{\max} := \max_{r \in \{1,\ldots,m\}} f(r)$. For each $v \in V$, the ranking of $c_\star$ in the new instance is $\pi'_v(c_\star) := m - \pi_v(c_\star) + 1$.

For each solution A_0 found in the new instance, i.e., a constructive one, the overall score of $c_\star$ after the process is $F'(A_0) := \mathbf{E}\left[\sum_{v \in V} f'(\pi'_v(c_\star) - \pi'^{\uparrow}_v(c_\star))\right]$.

The reduction allows us to maximize the score of the target candidate in the constructive case and then to map it back to destructive case. Differently from the *constructive* scenario, we get a factor $\frac{1}{2}$ because we can reconstruct the optimum in the approximation by only lower bounding two terms.

Theorem 4. GREEDY *is a* $\frac{1}{2}(1 - 1/e)$*-approximation algorithm for the problem of* destructive *election control in arbitrary scoring rule voting systems.*

8 Conclusions and Future Work

We introduced *Linear Threshold Ranking*, which describes the change of opinions taking into account the amount of exercised influence. We provided a constant factor approximation algorithm to the problems of *constructive* and *destructive* election control in arbitrary scoring rule voting systems. We simulated our model on real-world networks using synthetic election data, i.e., random degrees of influences and random preference lists. We used several combinations of parameters (B, $|C|$, α, π_v) on 4 networks exhibiting heterogeneous topologies. We observed that GREEDY can find a solution that makes the target candidate win the elections between 50% and 88% of the times, depending on the scenario.

Nowadays social media are are significant sources of information for voters and the massive usage of these channels for political campaigning is a turning point. Potential attackers can manipulate the outcome of elections through the spread of targeted ads and/or fake news. Being able to control the information spread can have a great impact, but it is not easy to achieve since traditional media sources are relatively transparent: It is essential to protect the integrity of electoral processes to ensure the proper operation of democratic institutions. Our results indicate that social influence is a salient threat to election integrity: We provide an approximation algorithm to maximize the MOV of a target candidate, that could be used to control election results and is of fundamental importance to protect their fairness.

Compared to the only other work on election control via social influence [Wilder and Vorobeychik, 2018a], we consider general scoring functions and a more realistic model (LTM instead of ICM) that takes into account the amount of influence exercised on voters. We believe that our algorithm could be used in real-life scenarios to predict election results and to understand what degree of control has been exercised. Our results assume the knowledge of information that is not available, but can be estimated. Recent studies analyze the robustness of greedy w.r.t. inaccurate estimations of the degrees of influence. Nevertheless, experiments on greedy algorithm for Influence Maximization show that the worst case hardness theoretical results do not necessarily translate into bad performance on real-world datasets [He and Kempe, 2018].

As future research directions we would like to study our model in a scenarios which are not currently captured, including *multi-winner* and *proportional representation* systems. It is also worth to analyze approaches that mix constructive and destructive control. Moreover, we would like to extend our model in order to consider a more uncertain scenario, in which the preferences of voters are not known. Finally, it would be interesting to study how to prevent election control for the integrity of voting processes, e.g., through the placement of monitors in the network [Zhang *et al.*, 2015; Amoruso *et al.*, 2017] or by considering strategic settings [Yin *et al.*, 2018; Wilder and Vorobeychik, 2018b].

Acknowledgments

This work has been partially supported by the Italian MIUR PRIN 2017 Project ALGADIMAR "Algorithms, Games, and Digital Markets".

References

[Allcott and Gentzkow, 2017] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Working Paper 23089, National Bureau of Economic Research, January 2017.

[Amoruso *et al.*, 2017] Marco Amoruso, Daniele Anello, Vincenzo Auletta, and Diodato Ferraioli. Contrasting the spread of misinformation in online social networks. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 1323–1331, 2017.

[Auletta *et al.*, 2015] Vincenzo Auletta, Ioannis Caragiannis, Diodato Ferraioli, Clemente Galdi, and Giuseppe Persiano. Minority becomes majority in social networks. In *International Conference on Web and Internet Economics*, pages 74–88. Springer, 2015.

[Bartholdi *et al.*, 1992] John J. Bartholdi, III, Craig A. Tovey, and Michael A. Trick. How hard is it to control an election? *Math. Comput. Model.*, 16(8-9):27–40, August 1992.

[Bond *et al.*, 2012] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295, Sep 2012.

[Botan *et al.*, 2017] Sirin Botan, Umberto Grandi, and Laurent Perrussel. Propositionwise opinion diffusion with constraints. In *Proceedings of the 4th AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE)*, 2017.

[Bredereck and Elkind, 2017] Robert Bredereck and Edith Elkind. Manipulating opinion diffusion in social networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 894–900, 2017.

[Bredereck *et al.*, 2016] Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, and Nimrod Talmon. Complexity of shift bribery in committee elections. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, pages 2452–2458, 2016.

[Brill *et al.*, 2016] Markus Brill, Edith Elkind, Ulle Endriss, and Umberto Grandi. Pairwise diffusion of preference rankings in social networks. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 130–136, 2016.

[Elkind *et al.*, 2009] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. Swap bribery. In *Proceedings of the 2nd International Symposium on Algorithmic Game Theory*, pages 299–310, 2009.

[Faliszewski *et al.*, 2016] Piotr Faliszewski, Jörg Rothe, and Hervé Moulin. *Control and Bribery in Voting*, pages 146–168. Cambridge University Press, 2016.

[Faliszewski *et al.*, 2018] Piotr Faliszewski, Rica Gonen, Martin Koutecký, and Nimrod Talmon. Opinion diffusion and campaigning on society graphs. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 219–225, 7 2018.

[Ferrara, 2017] Emilio Ferrara. Disinformation and social bot operations in the run up to the 2017 french presidential election. *First Monday*, 22(8), 2017.

[He and Kempe, 2018] Xinran He and David Kempe. Stability and robustness in influence maximization. *ACM Trans. Knowl. Discov. Data*, 12(6):66:1–66:34, August 2018.

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

[Kempe *et al.*, 2015] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. *Theory of Computing*, 11(4):105–147, 2015.

[Nemhauser *et al.*, 1978] George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14(1):265–294, Dec 1978.

[Wilder and Vorobeychik, 2018a] Bryan Wilder and Yevgeniy Vorobeychik. Controlling elections through social influence. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, pages 265–273, Richland, SC, 2018.

[Wilder and Vorobeychik, 2018b] Bryan Wilder and Yevgeniy Vorobeychik. Defending Elections Against Malicious Spread of Misinformation. *ArXiv e-prints*, September 2018.

[Yin *et al.*, 2018] Yue Yin, Yevgeniy Vorobeychik, Bo An, and Noam Hazon. Optimal defense against election control by deleting voter groups. *Artificial Intelligence*, 259:32–51, 2018.

[Zhang *et al.*, 2015] Huiling Zhang, Md Abdul Alim, My T. Thai, and Hien Nguyen. Monitor placement to timely detect misinformation in online social networks. In *2015 IEEE International Conference on Communications*, pages 1152–1157, 2015.