

Multi-Winner Election Control via Social Influence

Mohammad Abouei Mehrizi and Gianlorenzo D'Angelo

Gran Sasso Science Institute, L'Aquila, Italy.

{mohammad.aboueimehrizi,gianlorenzo.dangelo}@gssi.it

Abstract. The extensive use of social media in political campaigns has motivated the recent study of *election control problem in social networks*. In an election, we are given a set of voters, each having a preference list over a set of candidates, that are distributed on a social network. The winners of the election are computed by aggregating the preference lists of voters according to a so-called voting rule. We consider a scenario where voters may change their preference lists as a consequence of the messages received by their neighbors in a social network. Specifically, we consider a political campaign that spreads messages in a social network in support or against a given candidate and the spreading follows a dynamic model for information diffusion. When a message reaches a voter, this latter changes its preference list according to an update rule. The election control problem asks to find a bounded set of nodes to be the starter of a political campaign in support (constructive problem) or against (destructive problem) a given target candidate c , in such a way that the margin of victory of c w.r.t. its most voted opponents is maximized. It has been shown that several variants of the problem can be solved within a constant factor approximation of the optimum, which shows that controlling elections by means of social networks is doable and constitutes a real problem for modern democracies. Most of the literature, however, focuses on the case of single-winner elections.

In this paper, we define the election control problem in social networks for *multi-winner elections* with the aim of modeling parliamentary elections. Differently from the single-winner case, we show that the multi-winner election control problem is NP -hard to approximate within any factor in both constructive and destructive cases. We then study a relaxation of the problem where votes are aggregated on the basis of parties (instead of single candidates), which is a variation of the so-called *straight-party voting* used in some real parliamentary elections. We show that the latter problem remains NP -hard but can be approximated within a constant factor.

1 Introduction

Nowadays, social media are extensively used and have become a crucial part of our life. Generating information and spreading in social media is one of the cheapest and most effective ways of advertising and sharing content and opinions. People feel free to share their opinion, information, news, or also gain something by learning or teaching

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

in social media; on the other hand, they also use social media to get the latest news and information. Many people even prefer to check social media rather than news websites.

Social media are also exploited during election campaigns to support some party or a specific candidate. Many political parties diffuse targeted messages in social media with the aim of convincing users to vote for their candidates. Usually, these messages are posted by influential users and diffused on the network following a cascade effect, also called social influence. There are shreds of evidence of control election using the effect of social influence by spreading some pieces of information, including fake news or misinformation [21]. The presidential election of the United States of America is a real example. It has been shown that on average, ninety-two percent of Americans remembered pro-Trump false news, and twenty-three percent of them remembered the pro-Clinton fake news [2]. There are more real-life examples that have been presented in the literature [4,16,19,24].

This motivated the study of election control problems in social networks by using dynamic models for influence diffusion. We are given a social network of voters, a set of candidates, and a dynamic model for diffusion of information that models the spread of messages produced by political campaigns. The problem asks to find a bounded set of voters/nodes to be the starter of a political campaign in support of a given target candidate c , in such a way that the margin of victory of c w.r.t. its opponents is maximized. Each voter has its own preference list over the candidates and the winner of an election is determined by aggregating all preference lists according to some specific voting rule. Voters are autonomous, however their opinions about the candidates, and hence their preference lists, may change as a consequence of messages received by neighbors. When a message generated by a political campaign reaches a node, this latter changes its preference list according to some specific update rule. When the campaign aims to make the target candidate win, we refer to the *constructive* problem, while when the aim is to make c lose, we refer to the *destructive* problem. This problem recently received some attention (see the next paragraph). Most of the works in the area, however, focus on single-winner voting systems, while several scenarios require voting systems with multiple winners, e.g., parliamentary elections.

In this paper, we consider the problem of *multi-winner election control via social influence*, where there are some parties, each with multiple candidates, and we want to find at most B nodes to spread a piece of news in the social network in such a way that a target party elects a large number of its candidates. In this model, more than one candidate will be elected as the winner, and parties try to maximize some function of the number of winners from their party. We considered this problem for some well-known objective functions in both constructive and destructive cases.

Related Work. There is an extensive literature about manipulation or control of elections, we refer to the survey in [14] for relevant work on election control without the use of social networks. In the following, we focus on election control problems where the voters are the nodes of a social network, which recently received some attention.

Finding strategies to maximize the spread of influence in a network is one of the main topics in network analysis. Given a network and a dynamic model for the diffusion of influence, find a bounded set of nodes to be the starters of a dynamic process of influence spread in such a way that the number of eventually influenced users is

maximized. The problem, known as Influence Maximization (IM), has been introduced by Domingos and Richardson [12,22] and formalized by Kempe et al., who gave a $(1 - 1/e)$ -approximation algorithm [18] for two of the most used dynamic models, namely Independent Cascade Model (ICM) and Linear Threshold Model (LTM). We point the reader to the book by Chen et al. [9] and to [18].

Wilder and Vorobeychik [25] started the study of election control by means of IM. They defined an optimization problem that combines IM and election control called *election control through influence maximization* that is defined as follows. We are given a set of candidates, a social network of voters, each having a preference list over the candidates, a budget B , and a specific target candidate c_* . The network allows the diffusing influence of individuals according to ICM. When a node/voter v is influenced, it changes its preference list in such a way that the rank of c_* in the preference list of v is promoted (constructive) or demoted (destructive) by one position. At the end of a diffusion process, the voters elect a candidate according to the plurality rule [26]. The problem asks to find a set of at most B nodes to start a diffusion process in such a way that the chances for c_* to win (constructive) or lose (destructive) at the end of the diffusion are maximized. Wilder and Vorobeychik used the Margin of Victory (MoV) as an objective function and showed that there exists a greedy algorithm that approximates an optimal solution by a factor $1/3(1 - 1/e)$ for constructive and $1/2(1 - 1/e)$ for the destructive case. The same problem has been extended to LTM and general scoring rules [26] by Corò et al. [10,11]. They have shown that the problem can be approximated within the same bound. A similar problem has been studied in [13]. The authors consider a network where each node is a set of voters with the same preference list, and edges connect nodes whose preference lists differ by the ordering of a single pair of adjacent candidates. They use a variant of LTM for influence diffusion and show that the problem of making a specific candidate win is *NP*-hard and fixed-parameter tractable w.r.t. the number of candidates. Bredereck and Elkind [6] considered the following election control problem. Given a network where the influence spread according to a variant of LTM in which each node has a fixed threshold, and all edges have the same weight, find an optimal way of bribing nodes or add/delete edges in order to make the majority of nodes to vote for a target candidate. A different line of research investigates a model in which each voter is associated with a preference list over the candidates, and it updates its list according to the majority of opinions of its neighbors in the network [3,5,7]. All the previous works on election control through IM consider single-winner voting systems. Multi-winner voting systems raised recent and challenging research trends, we refer to a recent book chapter [15] and references therein.

Our results. We introduce the multi-winner election control problem via social influence and show that it is *NP*-hard to approximate within any factor $\alpha > 0$, for two common objective functions known as *margin of victory* and *difference of winners* using a general scoring rule. This is in contrast with the previous work on single-winner election control through IM, in which it is possible to approximate the optimum within a constant factor. The hardness results hold for both constructive and destructive cases. Given the hardness result, we focus on a relaxed version of the problem, which is a variation of straight-party voting. We show that this latter remains *NP*-hard but admits a constant factor approximation algorithm for both constructive and destructive cases.

2 Multi-winner Election Control

In this section, we introduce the multi-winner election control problem. We consider elections with k winners and general scoring rule as a voting system, which includes many well-known scoring rules, such as plurality, approval, Borda, and veto [26]. We first introduce the models that we use for diffusion of influence and for updating the preference list of voters. Then we introduce the objective functions for the election control problem in both constructive and destructive cases.

Model for influence diffusion. We use the *Independent Cascade Model (ICM)* for influence diffusion [18]. In this model, we are given a directed graph $G = (V, E)$, where each edge $(u, v) \in E$ has a weight $b_{uv} \in [0, 1]$. The influence starts with a set of seed nodes S and keeps activating the nodes in at most $|V|$ discrete steps. In the first step, all the seed nodes S become active. In the next steps $i > 1$, all the nodes that were active in step $i - 1$ remain active, moreover, each node u that became active at step $i - 1$ tries to activate its outgoing neighbors at step i with probability b_{uv} , for each node $v \in N_u^o$. An active node will try to activate its outgoing neighbors independently and only once. The process stops when no new node becomes active. We denote by A_S the set of nodes that are eventually active by the diffusion started by the seeds S .

Model for multi-winner election control. We consider a multi-winner election in which k candidates will be elected. Let $G = (V, E)$ be a directed social graph, where the nodes are the voters in the election, and the edges represent social relationships among users. The voters influence each other the same as ICM. We consider t parties C_1, C_2, \dots, C_t , each having k candidates, $C_i = \{c_1^i, c_2^i, \dots, c_k^i\}$, $1 \leq i \leq t$. Without loss of generality, we assume that C_1 is the target party. The set of all candidates is denoted by C , i.e. $C = \bigcup_{i=1}^t C_i$. Each voter $v \in V$ has a preference list π_v over the candidates. For each $c \in C$, we denote by $\pi_v(c) \in \{1, 2, \dots, tk\}$ the rank (or position) of the candidate c in the preference list of node v .

Given a budget B , we want to select a set of B seed nodes that maximizes the number of candidates in C_1 who win the election after a political campaign that spread according to ICM starting from nodes S (see next section for a formal definition of the objective functions).¹

After S , nodes in A_S will change the positions of candidates in C_1 in their preferences list. In contrast, nodes not in A_S will maintain their original preference list. The update rule for active nodes depends on the position of the target candidates and the goal of the campaign, i.e., if it is a constructive or a destructive one. We denote the preference list of node v after the process by $\tilde{\pi}_v$. If $v \notin A_S$, then $\tilde{\pi}_v = \pi_v$. In the following, we focus on nodes $v \in A_S$.

In the constructive case, like in the model in [25], the position of the target candidates in the list of active nodes will be decreased by one, if there is at least one opponent candidate in a smaller rank. The candidates who are overtaken will be demoted by the number of target candidates that were just after them. Formally, in the constructive case,

¹ In the remainder of the paper, by *after S* , we mean *after the diffusion process started from the set of seed nodes S* .

the position of the candidates after the diffusion starting from seed S will change as follows. For each node $v \in A_S$ and for each target candidate $c \in C_1$, the new position of c in v is

$$\tilde{\pi}_v(c) = \begin{cases} \pi_v(c) - 1 & \text{if } \exists c' \in C \setminus C_1 \text{ s.t. } \pi_v(c') < \pi_v(c) \\ \pi_v(c) & \text{otherwise,} \end{cases}$$

while, for each opponent candidate $c \in C \setminus C_1$, if there exists a candidate $c' \in C_1$ s.t. $\pi_v(c') = \pi_v(c) + 1$ we have $\tilde{\pi}_v(c) = \pi_v(c) + |\{c'' \in C_1 \mid \pi_v(c'') > \pi_v(c) \wedge (\nexists \bar{c} \in C \setminus C_1 : \pi_v(c) < \pi_v(\bar{c}) < \pi_v(c'')\})|$, otherwise, we set $\tilde{\pi}_v(c) = \pi_v(c)$.

In the destructive case, we want to reduce the number of winners in C_1 and then each node $v \in A_S$ increases their position by one, if it is possible. Formally, after S the preferences list of the candidates will change as follows. For each node $v \in A_S$ and for each target candidate $c \in C_1$, the new position of c in v is

$$\tilde{\pi}_v(c) = \begin{cases} \pi_v(c) + 1 & \text{if } \exists c' \in C \setminus C_1 \text{ s.t. } \pi_v(c') > \pi_v(c) \\ \pi_v(c) & \text{otherwise,} \end{cases}$$

while for $c \in C \setminus C_1$, if there exists a candidate $c' \in C_1$ s.t. $\pi_v(c') = \pi_v(c) - 1$ we have $\tilde{\pi}_v(c) = \pi_v(c) - |\{c'' \in C_1 \mid \pi_v(c'') < \pi_v(c) \wedge (\nexists \bar{c} \in C \setminus C_1 : \pi_v(c'') < \pi_v(\bar{c}) < \pi_v(c)\})|$, otherwise, we set $\tilde{\pi}_v(c) = \pi_v(c)$.

As an example, if there are two parties with three candidates each, and the initial preferences list of a node is $(c_1^2, c_1^1, c_2^1, c_2^2, c_3^1, c_3^2)$, then if the node becomes active its preferences list in the constructive case will be $(c_1^1, c_2^1, c_1^2, c_3^1, c_2^2, c_3^2)$, i.e., all of the candidates c_i^1 will promote, and all the overtaken candidates will demote; while in the destructive case, it will be $(c_1^2, c_2^2, c_1^1, c_2^1, c_3^2, c_3^1)$, and all of the candidates in our target party demote, and all the overtaken candidates will promote.

The above rule for updating the preference lists is commonly used in the literature [10,25]. In this model, we consider just one message, which contains some positive/negative information about the target party that will affect all the target candidates.

We consider a non-increasing scoring function $f(i)$, $1 \leq i \leq |C|$, such that for all $j > i > 0$ we have $f(j) \leq f(i)$. A candidate $c \in C$ gets $f(\pi_v(c))$ and $f(\tilde{\pi}_v(c))$ points from voter v before and after a diffusion, respectively. In other words, each voter will reveal his preferences list, and each candidate will get some score according to his position in the list and the scoring function. Also, we assume w.l.o.g. that there exist $1 \leq i < j \leq |C|$ such that $f(i) > f(j)$, i.e., the function does not return a fixed number for all ranks. The score of a candidate c is the sum of the scores received by all voters. The k candidates with the highest score will be elected.

We denote by $\mathcal{F}(c, S)$ the expected overall score received by candidate c after S , formally, $\mathcal{F}(c, S) = \mathbb{E}_{A_S} \left[\sum_{v \in V} f(\tilde{\pi}_v(c)) \right]$ and $\mathcal{F}(c, \emptyset) = \sum_{v \in V} f(\pi_v(c))$.

Objective Functions. The objective function for the constructive election control problem in the single-winner case is maximizing the *margin of victory* (MoV) defined in [25]. Let us consider the difference between the votes for the target candidate and those for the most voted opponent candidate. MoV is the change of this value after S . Note that the most voted opponent before and after S might change. The notion of MoV captures the goal of a candidate to have the largest margin in terms of votes w.r.t. any other candidate. We extend the above definition of MoV in the case of *multi-winner*

election control. Since the main goal is to elect more candidates from the target party, then we define the objective function in terms of the number of winning candidates in our target party before and after S .

Given a set A_S of nodes that are active at the end of a diffusion process started from S , we denote by $\mathcal{F}_{A_S}(c)$ the score that a candidate $c \in C$ receives if the activated nodes are A_S , and by $\mathcal{Y}_{A_S}(c)$ the number of candidates that have less score than the candidate c . As a tie-breaking rule, we assume that c_i^j has priority over $c_{i'}^{j'}$ if $j < j'$, or $j = j'$ and $i < i'$. In particular, the target candidates have priority over opponents when they have the same score. Then, for each $c_i^j \in C$, $i \in \{1, \dots, k\}$, $j \in \{1, \dots, t\}$, $\mathcal{Y}_{A_S}(c_i^j, S)$ is defined as $\mathcal{Y}_{A_S}(c_i^j) = \left| \{c_{i'}^{j'} \in C \mid \mathcal{F}_{A_S}(c_i^j) > \mathcal{F}_{A_S}(c_{i'}^{j'}) \vee (\mathcal{F}_{A_S}(c_i^j) = \mathcal{F}_{A_S}(c_{i'}^{j'}) \wedge (j < j' \vee (j = j' \wedge i < i')))\} \right|$.

For a party C_i , we define $\mathcal{F}(C_i, S)$ as the expected number of candidates in C_i that win the election after S ; formally,

$$\mathcal{F}(C_i, S) = \mathbb{E}_{A_S} \left[\sum_{c \in C_i} \mathbb{1}_{\mathcal{Y}_{A_S}(c) \geq (t-1)k} \right]. \quad (1)$$

We denote by C_b and C_A^S the opponent party with the highest number of winners *before* and *after* S , respectively. For the constructive case, the *margin of victory* (MoV_c) for party C_1 , w.r.t. seeds S , is defined as follows:

$$\text{MoV}_c(C_1, S) = \mathcal{F}(C_1, S) - \mathcal{F}(C_A^S, S) - (\mathcal{F}(C_1, \emptyset) - \mathcal{F}(C_b, \emptyset)),$$

while for the destructive case, it is defined as:

$$\text{MoV}_d(C_1, S) = \mathcal{F}(C_1, \emptyset) - \mathcal{F}(C_b, \emptyset) - (\mathcal{F}(C_1, S) - \mathcal{F}(C_A^S, S)).$$

The *Constructive (Destructive, resp.) Multi-winner Election Control problem* (CMEC (DMEC, resp.)) asks to find a set S of B seed nodes that maximizes $\text{MoV}_c(C_1, S)$ ($\text{MoV}_d(C_1, S)$, resp), where $B \in \mathbb{N}$ is a given budget.

In some scenarios, it is enough to maximize the difference between the number of our target candidates who win the election before and after S ; we call this objective function the *difference of winners* (DoW), and for the constructive case we define it as follows:

$$\text{DoW}_c(C_1, S) = \mathcal{F}(C_1, S) - \mathcal{F}(C_1, \emptyset).$$

While for the destructive model it is defined as:

$$\text{DoW}_d(C_1, S) = \mathcal{F}(C_1, \emptyset) - \mathcal{F}(C_1, S).$$

The problems of finding a set of at most B seed nodes that maximize DoW_c and DoW_d , for a given integer B , are called *Constructive Difference of Winners* (CDW) and *Destructive Difference of Winners* (DDW), respectively.

3 Hardness Results

In this section, we show the hardness of approximation results for the problems defined in the previous section. We first focus on the constructive case and prove that CMEC and CDW are *NP*-hard to approximate within any approximation factor $\alpha > 0$. Then, we show that the same results hold for DMEC and DDW. All the results hold even when the instance is deterministic (i.e. $b_{uv} = 1$, for each $(u, v) \in E$) and when $t = 3$ and $k = 2$. Note that for $t = 1$ the problem is trivial and for $k = 1$ the problem reduces to the single-winner case.

Constructive Election Control. We first give an intuition of the hardness of approximation proof, which is formally given in Theorem 1. Consider an instance of the constructive case in which $t = k = 2$, $C_1 = \{c_1^1, c_2^1\}$, $C_2 = \{c_1^2, c_2^2\}$, and $C = C_1 \cup C_2$. The weight of all edges are equal to 1, that is, the diffusion is a deterministic process. Also, assume the scoring rule is plurality, i.e., $f(1) = 1, f(2) = f(3) = f(4) = 0$. Moreover, the nodes are partitioned into two sets of equal size, V_1 and V_2 . In the preferences lists of all nodes in V_1 , candidate c_1^1 is in the first position and c_2^1 is in the second position, while in the preferences lists of nodes in V_2 , candidate c_2^2 is in first position and c_1^2 is in second position. In this instance, initially party C_1 does not have any elected candidate, that is, $\mathcal{F}(c_1^1, \emptyset) = \mathcal{F}(c_2^1, \emptyset) = 0$, $\mathcal{F}(c_1^2, \emptyset) = |V_1|$, $\mathcal{F}(c_2^2, \emptyset) = |V_2|$, $\mathcal{F}(C_1, \emptyset) = 0$, and $\mathcal{F}(C_2, \emptyset) = 2$.

Consider a diffusion process starting from seeds S that activate nodes A_S (note that, since the weights are all equal to 1, A_S is a deterministic set for any fixed S). The number of candidates that receives fewer votes than a target candidate c_i^1 after the diffusion process is $\mathcal{Y}_{A_S}(c_i^1) = \left| \{c_{i'}^{j'} \in C \mid \mathcal{F}_{A_S}(c_i^1) > \mathcal{F}_{A_S}(c_{i'}^{j'}) \vee (\mathcal{F}_{A_S}(c_i^1) = \mathcal{F}_{A_S}(c_{i'}^{j'}) \wedge (j' = 2 \vee i < i'))\} \right|$.

Let us consider the case $i = 1$ and analyze the conditions that a seed set S must satisfy in order to include a candidate in the above set, i.e., make c_1^1 win. We analyze the three other candidates $c_{i'}^{j'}$ separately.

- If $j' = 2$ and $i' = 1$, then we must have $\mathcal{F}_{A_S}(c_1^1) \geq \mathcal{F}_{A_S}(c_1^2)$. Since the preferences list of each active nodes in V_1 is updated in a way that c_1^1 moves to the first position and c_2^1 moves to the second position, and the active nodes in V_2 do not affect the rankings of c_1^1 and c_2^1 , we have that $\mathcal{F}_{A_S}(c_1^1) = |A_S \cap V_1|$ and $\mathcal{F}_{A_S}(c_1^2) = |V_1 \setminus A_S|$. Therefore, $\mathcal{F}_{A_S}(c_1^1) \geq \mathcal{F}_{A_S}(c_1^2)$ if and only if $|A_S \cap V_1| \geq |V_1 \setminus A_S| = |V_1| - |V_1 \cap A_S|$, which means that $|A_S \cap V_1| \geq |V_1|/2$.
- If $j' = 2$ and $i' = 2$, then we must have $\mathcal{F}_{A_S}(c_1^1) \geq \mathcal{F}_{A_S}(c_2^2)$. In this case, we still have $\mathcal{F}_{A_S}(c_1^1) = |A_S \cap V_1|$, and, since c_2^2 is moved down by one position for each active node in V_2 , then $\mathcal{F}_{A_S}(c_2^2) = |V_2 \setminus A_S|$. This implies that $\mathcal{F}_{A_S}(c_1^1) \geq \mathcal{F}_{A_S}(c_2^2)$ if and only if $|A_S \cap V_1| \geq |V_2 \setminus A_S| = |V_2| - |V_2 \cap A_S|$, which means $|A_S \cap V_1| + |A_S \cap V_2| \geq |V_2|$.
- If $j' = 1$ and $i' = 2$, then we must have $\mathcal{F}_{A_S}(c_1^1) \geq \mathcal{F}_{A_S}(c_2^1)$. We again have $\mathcal{F}_{A_S}(c_1^1) = |A_S \cap V_1|$, and, since c_2^1 is moved by one position up for each active node in V_2 , then $\mathcal{F}_{A_S}(c_2^1) = |A_S \cap V_2|$. Therefore, $\mathcal{F}_{A_S}(c_1^1) \geq \mathcal{F}_{A_S}(c_2^1)$ if and only if $|A_S \cap V_1| \geq |A_S \cap V_2|$.

Similar conditions hold for $i = 2$.

In order to elect candidate c_1^1 we should have $\mathcal{Y}_{A_S}(c_1^1) \geq (t-1)k = 2$, which means, we should find a seed set that satisfies at least two of the above conditions (or the corresponding conditions to elect c_2^1). Note that finding a seed set S of size at most B that satisfies any pair of the above conditions is a *NP*-hard problem since it requires to solve the IM problem, which is *NP*-hard even when the weight of all edges is 1 [18]. Let us assume that an optimal solution is able to elect both candidates in C_1 (e.g. by influencing $|V_1|/2$ nodes from V_1 , and $|V_2|/2$ nodes from V_2), then the optimal MoV_c and DoW_c are equal to 4 and 2, respectively. Moreover, in this case $C_A^S = C_B = C_2$, then $\text{MoV}_c(C_1, S) = \mathcal{F}(C_1, S) - \mathcal{F}(C_1, \emptyset) + \mathcal{F}(C_2, \emptyset) - \mathcal{F}(C_2, S)$. Since $\mathcal{F}(C_2, \emptyset) - \mathcal{F}(C_2, S) = \mathcal{F}(C_1, S) - \mathcal{F}(C_1, \emptyset)$, i.e., for each candidate lost by C_2 there is a candidate gained by C_1 , then $\text{MoV}_c(C_1, S) = 2(\mathcal{F}(C_1, S) - \mathcal{F}(C_1, \emptyset)) = 2\text{DoW}_c(C_1, S)$. Since $\mathcal{F}(C_1, \emptyset) = 0$, any approximation algorithm for CDW or CMEC must find a seed set S s.t. $\mathcal{F}(C_1, S) > 0$ and this requires to elect at least one candidate in C_1 (see Equation (1)), which is *NP*-hard. It follows that it is *NP*-hard to approximate CMEC and CDW within any factor, as formally shown in the next theorem.

Theorem 1 *It is NP-hard to approximate CMEC and CDW within any factor $\alpha > 0$.*

Proof. We reduce the decision version of the deterministic IM problem, to CMEC and CDW, where *deterministic* refers to the weight of the edges in the graph, i.e., the weight of all edges is equal to 1. Let us define the decision version of the IM problem as follows: *Given a directed graph $G = (V, E)$ and budget $B \leq |V|$. Is there a set of seed nodes $S \subseteq V$ such that $|S| \leq B$ and $A_S = V$?*

Let $\mathcal{I}(G, B)$ be a deterministic instance of the decision IM problem (then, using a given seed set S , we can find the exact number of activated nodes in polynomial time). We create an instance $\mathcal{I}'(G', B)$ of CMEC and CDW, where $G' = (V \cup V', E \cup E')$. We use the same budget B for both problems. We first investigate the case where $t = 3, k = 2$, and consider two different cases as follows.

C1. If $f(1) = f(2) = f(3) = a, f(4) = f(5) = f(6) = b$ for $a, b \in \mathbb{R} \wedge a > b \geq 0$, we call this case *exceptional*, and do as follows.

For each $v \in V$ we add one more node in V' and it has just one incoming edge from v , i.e., $\forall v \in V : v_1 \in V', (v, v_1) \in E'$.

We set the preferences of all nodes $v \in V$ and its new outgoing neighbor as follows: $v = (c_1^2 \succ c_2^2 \succ c_1^3 \succ c_1^1 \succ c_2^1 \succ c_2^3), v_1 = (c_1^3 \succ c_2^3 \succ c_1^2 \succ c_2^1 \succ c_1^1 \succ c_2^2)$ where $a \succ b$ means a is preferred to b .

C2. For any non-increasing scoring function except the exceptional ones, we call it general and do as follows.

For each $v \in V$ we add three more nodes in V' and each of them has just one incoming edge from v , i.e., $\forall v \in V : v_1, v_2, v_3 \in V', (v, v_1), (v, v_2), (v, v_3) \in E'$.

We set the preferences of all nodes $v \in V$ and its new outgoing neighbors as follows: $v = (c_1^2 \succ c_1^1 \succ c_1^3 \succ c_2^2 \succ c_2^1 \succ c_2^3), v_1 = (c_2^2 \succ c_2^1 \succ c_2^3 \succ c_1^2 \succ c_1^1 \succ c_1^3), v_2 = (c_1^2 \succ c_1^3 \succ c_1^1 \succ c_2^2 \succ c_2^3 \succ c_2^1), v_3 = (c_2^2 \succ c_2^3 \succ c_2^1 \succ c_1^2 \succ c_1^3 \succ c_1^1)$.

In both cases, the weight of all edges is 1, i.e., $b_{uv} = 1$ for all $(u, v) \in E \cup E'$.

The score of candidates before any diffusion is as follows.

- C1.** $\mathcal{F}(c_1^1, \emptyset) = \mathcal{F}(c_2^1, \emptyset) = |V|(f(4) + f(5)) = 2b|V|$, $\mathcal{F}(c_1^2, \emptyset) = \mathcal{F}(c_3^1, \emptyset) = |V|(f(1) + f(3)) = 2a|V|$, $\mathcal{F}(c_2^2, \emptyset) = \mathcal{F}(c_3^2, \emptyset) = |V|(f(2) + f(6)) = (a+b)|V|$. Since $a > b \geq 0$, it yields that $\mathcal{F}(C_2, \emptyset) = \mathcal{F}(C_3, \emptyset) = 1$, $\mathcal{F}(C_1, \emptyset) = 0$, and none of our target candidates win the election.
- C2.** $\mathcal{F}(c_1^1, \emptyset) = \mathcal{F}(c_2^1, \emptyset) = |V|(f(2) + f(3) + f(5) + f(6))$, $\mathcal{F}(c_1^2, \emptyset) = \mathcal{F}(c_2^2, \emptyset) = |V|(2f(1) + 2f(4))$, $\mathcal{F}(c_1^3, \emptyset) = \mathcal{F}(c_2^3, \emptyset) = |V|(f(2) + f(3) + f(5) + f(6))$. Since $f(\cdot)$ is a non-increasing function, it yields $\mathcal{F}(C_2, \emptyset) = 2$ and $\mathcal{F}(C_1, \emptyset) = \mathcal{F}(C_3, \emptyset) = 0$ and none of our target candidates win the election.

In $\mathcal{I}'(G', B)$, in both cases, all of the nodes $v \in V \cup V'$ become active if and only if all of the nodes $v \in V$ become active. Indeed, by definition, if $V \subseteq A_S$, then for each node $u \in V'$ there exists an incoming neighbor $v \in V$ s.t. $(v, u) \in E'$ and $b_{vu} = 1$, then if v is active, also u becomes active.

Suppose there exists an α -approximation algorithm called α -*appAlg* for CDW (resp. CMEC) and it returns $S \subseteq V \cup V'$ as a solution. We show that, by using the seed nodes S returned by the algorithm α -*appAlg*, we can find the answer for the decision IM problem. We will show that $\text{DoW}_c(C_1, S) > 0$ (resp. $\text{MoV}_c(C_1, S) > 0$) if and only if S activates all of the nodes, i.e., $A_S = V \cup V'$. That is $\text{DoW}_c(C_1, S) > 0$ (resp. $\text{MoV}_c(C_1, S) > 0$) if and only if the answer to the decision IM problem is YES.

W.l.o.g., we assume $S \subseteq V$, because if there exists a node $u \in S \cap V'$, we can replace it with the node $v \in V$ s.t. $(v, u) \in E'$. Since $b_{uv} = 1$, this does not decrease the value of $\text{DoW}_c(C_1, S)$ or $\text{MoV}_c(C_1, S)$.

We now show that if $\text{DoW}_c(C_1, S) > 0$ (resp. $\text{MoV}_c(C_1, S) > 0$), then $A_S = V \cup V'$. By contradiction, assume that S will not activate all of the nodes, i.e., there exists a node v in $V \setminus A_S$. Then, the score of the candidates will be as follows.

- C1.** $\mathcal{F}(c_1^1, S) = \mathcal{F}(c_2^1, S) \leq (a+b)(|V|-1) + 2b$, $\mathcal{F}(c_1^2, S) = \mathcal{F}(c_3^1, S) \geq (a+b)(|V|-1) + 2a$, $\mathcal{F}(c_2^2, S) = \mathcal{F}(c_3^2, S) = |V|(f(2) + f(6)) = (a+b)|V|$. Since $a > b \geq 0$, then none of the target candidates will be among the winners, i.e., $\mathcal{F}(C_2, S) = \mathcal{F}(C_3, S) = 1$ and $\mathcal{F}(C_1, S) = 0$ and $\text{DoW}_c(C_1, S) = \text{MoV}_c(C_1, S) = 0$.
- C2.** $\mathcal{F}(c_1^1, S) = \mathcal{F}(c_2^1, S) \leq (|V|-1)(f(1) + f(2) + f(4) + f(5)) + (f(2) + f(3) + f(5) + f(6))$, $\mathcal{F}(c_1^2, S) = \mathcal{F}(c_2^2, S) \geq (|V|-1)(f(1) + f(2) + f(4) + f(5)) + (2f(1) + 2f(4))$, $\mathcal{F}(c_1^3, S) = \mathcal{F}(c_2^3, S) \geq (|V|-1)(f(3) + f(4) + 2f(6)) + (f(2) + f(3) + f(5) + f(6))$. Since $f(\cdot)$ is a non-increasing function, then $\mathcal{F}(C_1, S) = \mathcal{F}(C_3, S) = 0$ and $\mathcal{F}(C_2, S) = 2$. Therefore $\text{DoW}_c(C_1, S) = \text{MoV}_c(C_1, S) = 0$.

In both cases we have a contradiction. To show the other direction, if all of the nodes become active, then the score of candidates will be as follows.

- C1.** For each $1 \leq i \leq 2$ and $1 \leq j \leq 3$, $\mathcal{F}(c_i^j, S) = (a+b)|V|$. Due to the tie-breaking rule it follows that both of our target candidates will be among the winners, i.e., $\mathcal{F}(C_1, S) = 2$ and $\mathcal{F}(C_2, S) = \mathcal{F}(C_3, S) = 0$.
- C2.** $\mathcal{F}(c_1^1, S) = \mathcal{F}(c_2^1, S) = \mathcal{F}(c_1^2, S) = \mathcal{F}(c_2^2, S) = |V|(f(1) + f(2) + f(4) + f(5))$, $\mathcal{F}(c_1^3, S) = \mathcal{F}(c_2^3, S) = |V|(f(3) + f(4) + 2f(6))$. Then, $\mathcal{F}(C_1, S) = 2$ and $\mathcal{F}(C_2, S) = \mathcal{F}(C_3, S) = 0$.

Therefore we have $\text{DoW}_c(C_1, S) > 0$ (resp. $\text{MoV}_c(C_1, S) > 0$), and it concludes the proof. The proof can be generalized for any $t, k > 2$, see [1]. \square

Destructive Election Control. The following theorem shows the hardness of approximation of the destructive case. The proof is similar to that of Theorem 1, and hence it is omitted. Note that if we consider maximizing DoW_d , the destructive case can be reduced to the constructive model. We cannot apply the same reduction to the problem of maximizing MoV_d as the opponent party with the highest number of winners (i.e., C_B, C_A^S) may be different from that of the constructive case.

Theorem 2 *It is NP-hard to approximate DMEC and DDW within any factor $\alpha > 0$.*

4 Straight-Party Voting

Since all the variants of the multi-winner election control problems considered so far are NP-hard to be approximated within any factor, we now consider a relaxation of the problem in which, instead of focusing on the number of elected target candidates, we focus on the overall number of votes obtained by the target party. The rationale is that, even if a party is not able to (approximately) maximize the number of its winning candidates because it is computationally unattainable, it may want to maximize the overall number of votes, in the hope that these are not too spread among the candidates and still leads to a large number of seats in the parliament.

Moreover, the voting system that we obtain by the relaxation is used in some real parliamentary elections [23], and is called of *Straight-party voting* (SPV) or *straight-ticket voting* [8,20]. SPV was used very much until around the 1960s and 1970s in the United States. After that, the United States has declined SPV among the general voting; nevertheless, strong partisans are still voting according to SPV. Interestingly, the first time that every state voting for a Democrat for Senate also voted Democratic for president (and the same stability for Republicans) was the 2016 elections of the United states [17].

Note that in this model, if we consider that the controller targets a single candidate instead of a party and the preference lists are over candidates, then we can easily reduce the problem to the single-winner case. The same holds if the controller targets a party and the preference lists are over parties. Therefore, we assume that voters have preference lists over the candidates, but since the voting system is SPV and voters have to vote for a party, then they will cast a vote for each party based on the position of the candidates of the party in their preferences list, e.g., if the preferences list of a node $v \in V$ is $c_1^1 \succ c_1^2 \succ c_2^2 \succ c_2^1$, then the scores of v for party C_1 will be $f(1) + f(4)$, and $f(2) + f(3)$ for party C_2 .

Let us define $\mathcal{F}_{spv}(C_i, \emptyset)$ and $\mathcal{F}_{spv}(C_i, S)$ as sum of the scores obtained by party C_i in SPV before and after S , respectively, as follows.

$$\mathcal{F}_{spv}(C_i, \emptyset) = \sum_{v \in V} \sum_{c \in C_i} f(\pi_v(c)), \quad \mathcal{F}_{spv}(C_i, S) = \mathbb{E}_{A_S} \left[\sum_{v \in V} \sum_{c \in C_i} f(\tilde{\pi}_v(c)) \right].$$

As in the previous case, we denote by C_B and C_A^S the most voted opponents of C_1 before and after S , respectively. We define MoV_c and MoV_d for SPV as

$$\text{MoV}_c^{spv}(C_1, S) = \mathcal{F}_{spv}(C_1, S) - \mathcal{F}_{spv}(C_A^S, S) - (\mathcal{F}_{spv}(C_1, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset)),$$

$$\text{MoV}_d^{spv}(C_1, S) = \mathcal{F}_{spv}(C_1, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) - (\mathcal{F}_{spv}(C_1, S) - \mathcal{F}_{spv}(C_A^S, S)).$$

Also, we define *difference of votes* for constructive (DoV_c) and destructive (DoV_d) as

$$\begin{aligned} \text{DoV}_c^{spv}(C_1, S) &= \mathcal{F}_{spv}(C_1, S) - \mathcal{F}_{spv}(C_1, \emptyset), \\ \text{DoV}_d^{spv}(C_1, S) &= \mathcal{F}_{spv}(C_1, \emptyset) - \mathcal{F}_{spv}(C_1, S). \end{aligned}$$

Theorem 3 *Maximizing MoV and DoV in the constructive and destructive SPV problems is NP-hard.*

Proof. As in Theorem 1, we use the decision version of the deterministic IM problem. Let $\mathcal{I}(G, B)$ be a deterministic instance of the decision IM problem. We create an instance $\mathcal{I}'(G', B)$ of SPV, where $G' = G$ and B is the same budget for both problems. Assume $t = k = 2$, and we are using a non-increasing scoring function. Consider the minimum j such that $1 \leq i < j \leq |C|$ and $f(i) > f(j)$, i.e., j is the minimum rank that has less score than rank $i = j - 1$. Note that $2 \leq j \leq 4$. We set the preferences list of each voter $v \in V$ in graph G' as follows.

- C1.** If $j = 2$. $v : c_1^2 \succ c_1^1 \succ c_2^1 \succ c_2^2$. In this case, $\mathcal{F}_{spv}(C_1, \emptyset) = (f(2) + f(3))|V|$.
- C2.** If $j = 3$. $v : c_1^2 \succ c_2^2 \succ c_1^1 \succ c_2^1$. In this case, $\mathcal{F}_{spv}(C_1, \emptyset) = (f(3) + f(4))|V|$.
- C3.** If $j = 4$. $v : c_1^1 \succ c_1^2 \succ c_2^2 \succ c_2^1$. In this case, $\mathcal{F}_{spv}(C_1, \emptyset) = (f(1) + f(4))|V|$.

By this preferences assignment, if all of the nodes become active after S , then the score and DoV_c^{spv} for party C_1 will be as following.

$$\text{C1. } \mathcal{F}_{spv}(C_1, S) = (f(1) + f(2))|V|,$$

$$\text{DoV}_c^{spv}(C_1, S) = (f(1) + f(2))|V| - (f(2) + f(3))|V| = (f(1) - f(3))|V|;$$

and, if at least one node is not active $\text{DoV}_c^{spv}(C_1, S) < (f(1) - f(3))|V|$.

$$\text{C2. } \mathcal{F}_{spv}(C_1, S) = (f(2) + f(3))|V|,$$

$$\text{DoV}_c^{spv}(C_1, S) = (f(2) + f(3))|V| - (f(3) + f(4))|V| = (f(2) - f(4))|V|;$$

also, if at least one node is not active $\text{DoV}_c^{spv}(C_1, S) < (f(2) - f(4))|V|$.

$$\text{C3. } \mathcal{F}_{spv}(C_1, S) = (f(1) + f(3))|V|,$$

$$\text{DoV}_c^{spv}(C_1, S) = (f(1) + f(3))|V| - (f(1) + f(4))|V| = (f(3) - f(4))|V|;$$

moreover, if at least one node is not active $\text{DoV}_c^{spv}(C_1, S) < (f(3) - f(4))|V|$.

Then by this reduction, we can distinguish between the case that all of the nodes become active or not, which is the answer of IM problem. In this case, since there are just two parties, whatever C_2 loses will go for C_1 . Then, $\text{MoV}_c^{spv}(C_1, S) = 2\text{DoV}_c^{spv}(C_1, S)$, which means we also can answer to IM problem by maximizing MoV_c^{spv} . The generalized version of this proof, $t, k > 0$, is available in the extended version [1].

Regarding the destructive case, the reduction is similar to the constructive one, except that we set the preferences of the voters s.t. at least one of the candidates $c \in C_1$ can decrease the score of C_1 . The same approach gets the NP-hardness result. \square

We now give an approximation algorithm for the problems of maximizing DoV_c^{spv} and DoV_d^{spv} that is based on a reduction to the node-weighted version of the IM problem. We construct an instance of this problem where the weight to each node $v \in V$, which is equal to the increase in the score of C_1 when v becomes active. The node-weighted IM problem can be approximated by a factor of $1 - \frac{1}{e} - \epsilon$, for any $\epsilon > 0$, by using the standard greedy algorithm [18].

Theorem 4 *There exists an algorithm that approximates DoV_c^{spv} and DoV_d^{spv} within a factor $(1 - \frac{1}{e}) - \epsilon$ from the optimum, for any $\epsilon > 0$.*

Proof. We first consider the constructive case, i.e., DoV_c^{spv} . Let us define $\bar{C}_1^v \subseteq C_1$ as a set of candidates in our target party whose rank is decreased if v become active; in other words, $\bar{C}_1^v = \{c \in C_1 : \exists c' \in C \setminus C_1, \pi_v(c') < \pi_v(c)\}$. In this case, a node $v \in V$ can increase the score of C_1 by $\sum_{c \in \bar{C}_1^v} f(\pi_v(c) - 1) - f(\pi_v(c))$.² Given an instance of $\mathcal{I}(G, B)$ of the DoV_c^{spv} maximization problem, we define an instance $\mathcal{I}'(G, B, w)$ of the node-weighted IM problem, where w is a node-weight function defined as $w(v) = \sum_{c \in \bar{C}_1^v} (f(\pi_v(c) - 1) - f(\pi_v(c)))$, for all $v \in V$. Given a set S of nodes, we denote by $\sigma(S)$ the expected weight of active nodes in G , when the diffusion starts from S . We will show that $\text{DoV}_c^{spv}(C_1, S) = \sigma(S)$ for any set $S \subseteq V$, since the standard greedy algorithm guarantees an approximation factor of $1 - \frac{1}{e} - \epsilon$, for the node-weight IM problem, for any $\epsilon > 0$, this shows the statement.

Given a set S , $\sigma(S)$ can be computed as follows:

$$\sigma(S) = \mathbb{E}_{A_S} \left[\sum_{v \in A_S} w(v) \right] = \sum_{A_S \subseteq V} \sum_{v \in A_S} w(v) \mathbb{P}(A_S),$$

where $\mathbb{P}(A_S)$ is the probability that $A_S \subseteq V$ is the set of nodes activated by S .

By definition $\text{DoV}_c^{spv}(C_1, S) = \mathcal{F}_{spv}(C_1, S) - \mathcal{F}_{spv}(C_1, \emptyset)$, where $\mathcal{F}_{spv}(C_1, \emptyset) = \sum_{v \in V} \sum_{c \in C_1} f(\pi_v(c))$ and

$$\begin{aligned} \mathcal{F}_{spv}(C_1, S) &= \mathbb{E}_{A_S} \left[\sum_{v \in V} \sum_{c \in C_1} f(\tilde{\pi}_v(c)) \right] = \sum_{v \in V} \sum_{c \in C_1} \mathbb{E}_{A_S} [f(\tilde{\pi}_v(c))] \\ &= \sum_{v \in V} \left(\sum_{c \in \bar{C}_1^v} \mathbb{E}_{A_S} [f(\tilde{\pi}_v(c))] + \sum_{c \in C_1 \setminus \bar{C}_1^v} f(\pi_v(c)) \right), \end{aligned}$$

where the last equality is due to the fact that, a node v doesn't change the positions of candidates in $C_1 \setminus \bar{C}_1^v$. Let us focus on the first term of the above formula,

$$\sum_{v \in V} \sum_{c \in \bar{C}_1^v} \mathbb{E}_{A_S} [f(\tilde{\pi}_v(c))]$$

² We assume function $f(\cdot)$ is defined in such away that $f(i-1) - f(i)$, for $i = 2, \dots, m$, does not depend exponentially on the graph size (e.g. it is a constant). The influence maximization problem with arbitrary node-weights is still an open problem [18].

$$\begin{aligned}
&= \sum_{v \in V} \sum_{c \in \bar{C}_1^v} \sum_{A_S \subseteq V} (f(\pi_v(c)) - 1) \mathbf{1}_{v \in A_S} + f(\pi_v(c)) \mathbf{1}_{v \notin A_S} \mathbb{P}(A_S) \\
&= \sum_{A_S \subseteq V} \left(\sum_{v \in A_S} \sum_{c \in \bar{C}_1^v} f(\pi_v(c)) - 1 + \sum_{v \notin A_S} \sum_{c \in \bar{C}_1^v} f(\pi_v(c)) \right) \mathbb{P}(A_S).
\end{aligned}$$

It follows that

$$\begin{aligned}
\text{DoV}_c^{spv}(C_1, S) &= \sum_{v \in V} \sum_{c \in \bar{C}_1^v} \mathbb{E}_{A_S} [f(\tilde{\pi}_v(c)) - f(\pi_v(c))] \\
&\quad \sum_{A_S \subseteq V} \sum_{v \in A_S} \sum_{c \in \bar{C}_1^v} (f(\pi_v(c)) - 1 - f(\pi_v(c))) \mathbb{P}(A_S) = \sigma(S),
\end{aligned}$$

since the term related to candidates in $C_1 \setminus \bar{C}_1^v$ and to nodes not in A_S are canceled out.

The destructive case is similar to the constructive one except that a node $v \in V$ can decrease the score of C_1 by $\sum_{c \in C_1: \exists c' \in C \setminus C_1, \pi_v(c') > \pi_v(c)} f(\pi_v(c)) - f(\pi_v(c) + 1)$. Therefore the same approach, where the weights are set to the above value, yields the same approximation factor for DoV_d^{spv} . \square

In the following theorems, we show that using Theorem 4, we get a constant approximation factor for the problem of maximizing MoV. Specifically, we show that by maximizing DoV_c^{spv} we get an extra $1/3$ approximation factor for the problem of maximizing MoV_c^{spv} . For the destructive case, the extra approximation factor is $1/2$. It follows that, by using the greedy algorithm for maximizing DoV_c^{spv} and DoV_d^{spv} , we obtain approximation factors of $\frac{1}{3}(1 - \frac{1}{e}) - \epsilon$ and $\frac{1}{2}(1 - \frac{1}{e}) - \epsilon$, for any $\epsilon > 0$, of the maximum MoV_c^{spv} and MoV_d^{spv} , respectively.

Theorem 5 *There exists an algorithm that approximates MoV_c^{spv} within a factor $\frac{1}{3}(1 - \frac{1}{e}) - \epsilon$ from the optimum, for any $\epsilon > 0$.*

Proof. Let S and S^* be the solution returned by the greedy algorithm for DoV_c^{spv} maximization and a solution that maximizes MoV_c^{spv} , respectively. For each party $C_i \neq C_1$, we denote by $\text{DoV}_c^-(C_i, S)$ the score lost by C_i after S , that is $\text{DoV}_c^-(C_i, S) = \mathcal{F}(C_i, \emptyset) - \mathcal{F}(C_i, S) \geq 0$. Let $\alpha_\epsilon := (1 - \frac{1}{e}) - \epsilon$. Since S is a factor α_ϵ from the optimum DoV_c^{spv} , the following holds.

$$\begin{aligned}
\text{MoV}_c^{spv}(C_1, S) &= \mathcal{F}_{spv}(C_1, S) - \mathcal{F}_{spv}(C_A^S, S) - \left(\mathcal{F}_{spv}(C_1, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) \right) \\
&= \text{DoV}_c^{spv}(C_1, S) + \text{DoV}_c^-(C_A^S, S) - \mathcal{F}_{spv}(C_A^S, \emptyset) + \mathcal{F}_{spv}(C_B, \emptyset) \\
&\geq \alpha_\epsilon \text{DoV}_c^{spv}(C_1, S^*) - \mathcal{F}_{spv}(C_A^S, \emptyset) + \mathcal{F}_{spv}(C_B, \emptyset) \\
&\stackrel{(a)}{\geq} \frac{1}{3} \alpha_\epsilon \left[\text{DoV}_c^{spv}(C_1, S^*) + \text{DoV}_c^-(C_A^{S^*}, S^*) + \text{DoV}_c^-(C_A^S, S^*) \right] \\
&\quad - \mathcal{F}_{spv}(C_A^S, \emptyset) + \mathcal{F}_{spv}(C_B, \emptyset) \\
&\stackrel{(b)}{\geq} \frac{1}{3} \alpha_\epsilon \left[\text{DoV}_c^{spv}(C_1, S^*) + \text{DoV}_c^-(C_A^{S^*}, S^*) + \text{DoV}_c^-(C_A^S, S^*) \right]
\end{aligned}$$

$$\begin{aligned}
& -\mathcal{F}_{spv}(C_A^S, \emptyset) + \mathcal{F}_{spv}(C_B, \emptyset) + \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) - \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) \\
&= \frac{1}{3}\alpha_\epsilon \left[\text{MoV}_c^{spv}(C_1, S^*) + \text{DoV}_c^-(C_A^S, S^*) + \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) - \mathcal{F}_{spv}(C_A^S, \emptyset) \right] \\
&\stackrel{(c)}{\geq} \frac{1}{3}\alpha_\epsilon \text{MoV}_c^{spv}(C_1, S^*) \geq \left(\frac{1}{3} \left(1 - \frac{1}{e} \right) - \epsilon \right) \text{MoV}_c^{spv}(C_1, S^*),
\end{aligned}$$

for any $\epsilon > 0$. Inequality (a) holds because, by definition, the score lost by C_A^S and $C_A^{S^*}$ will be added to the score of C_1 . Inequality (b) holds since, by definition of C_B , $\mathcal{F}_{spv}(C_B, \emptyset) \geq \mathcal{F}_{spv}(C_A^S, \emptyset)$ and then $-\mathcal{F}_{spv}(C_A^S, \emptyset) + \mathcal{F}_{spv}(C_B, \emptyset) \geq 0$. Inequality (c) holds because

$$\begin{aligned}
& \text{DoV}_c^-(C_A^{S^*}, S^*) - \text{DoV}_c^-(C_A^S, S^*) \\
&= \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) - \mathcal{F}_{spv}(C_A^{S^*}, S^*) - \mathcal{F}_{spv}(C_A^S, \emptyset) + \mathcal{F}_{spv}(C_A^S, S^*) \\
&\stackrel{(d)}{\leq} \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) - \mathcal{F}_{spv}(C_A^S, \emptyset),
\end{aligned}$$

which implies that

$$\text{DoV}_c^-(C_A^S, S^*) + \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) - \mathcal{F}_{spv}(C_A^S, \emptyset) \geq \text{DoV}_c^-(C_A^{S^*}, S^*) \geq 0.$$

Inequality (d) holds since, by definition of $C_A^{S^*}$, $\mathcal{F}(C_A^S, S^*) \leq \mathcal{F}(C_A^{S^*}, S^*)$. \square

Theorem 6 *There exists an algorithm that approximates MoV_d^{spv} within a factor $\frac{1}{2}(1 - \frac{1}{e}) - \epsilon$ from the optimum, for any $\epsilon > 0$.*

Proof. Let S and S^* be the solution returned by the greedy algorithm for DoV_d^{spv} maximization and a solution that maximizes MoV_d^{spv} , respectively. For each party $C_i \neq C_1$, we denote by $\text{DoV}_c^+(C_i, S)$ the score gained by C_i after S , that is $\text{DoV}_c^+(C_i, S) = \mathcal{F}(C_i, S) - \mathcal{F}(C_i, \emptyset) \geq 0$. Let $\alpha_\epsilon := (1 - \frac{1}{e}) - \epsilon$. Since S is a factor α_ϵ from the optimum DoV_c^{spv} , the following holds.

$$\begin{aligned}
& \text{MoV}_d^{spv}(C_1, S) = \mathcal{F}_{spv}(C_1, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) - (\mathcal{F}_{spv}(C_1, S) - \mathcal{F}_{spv}(C_A^S, S)) \\
&= \text{DoV}_d^{spv}(C_1, S) + \text{DoV}_c^+(C_A^S, S) + \mathcal{F}_{spv}(C_A^S, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) \\
&\geq \alpha_\epsilon \text{DoV}_d^{spv}(C_1, S^*) + \text{DoV}_c^+(C_A^S, S) + \mathcal{F}_{spv}(C_A^S, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) \\
&\stackrel{(a)}{\geq} \frac{1}{2}\alpha_\epsilon \left[\text{DoV}_d^{spv}(C_1, S^*) + \text{DoV}_c^+(C_A^{S^*}, S^*) \right] + \text{DoV}_c^+(C_A^S, S) + \\
&\quad \mathcal{F}_{spv}(C_A^S, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) \\
&\stackrel{(b)}{\geq} \frac{1}{2}\alpha_\epsilon \left[\text{DoV}_d^{spv}(C_1, S^*) + \text{DoV}_c^+(C_A^{S^*}, S^*) + \text{DoV}_c^+(C_A^S, S) + \right. \\
&\quad \left. \mathcal{F}_{spv}(C_A^S, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) + \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) - \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) \right] \\
&= \frac{1}{2}\alpha_\epsilon \left[\text{MoV}_d(C_1, S^*) + \text{DoV}_c^+(C_A^S, S) - \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) + \mathcal{F}_{spv}(C_A^S, \emptyset) \right] \\
&\stackrel{(c)}{\geq} \frac{1}{2}\alpha_\epsilon \text{MoV}_d(C_1, S^*) \geq \left(\frac{1}{2} \left(1 - \frac{1}{e} \right) - \epsilon \right) \text{MoV}_d(C_1, S^*),
\end{aligned}$$

for any $\epsilon > 0$. Inequality (a) holds because $C_A^{S^*}$ can gain at most all of the scores lost by C_1 . Inequality (b) holds since we have

$$\begin{aligned} \text{DoV}_c^+(C_A^S, S) + \mathcal{F}_{spv}(C_A^S, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) \\ &= \mathcal{F}_{spv}(C_A^S, S) - \mathcal{F}_{spv}(C_A^S, \emptyset) + \mathcal{F}_{spv}(C_A^S, \emptyset) - \mathcal{F}_{spv}(C_B, \emptyset) \\ &= \mathcal{F}_{spv}(C_A^S, S) - \mathcal{F}_{spv}(C_B, \emptyset), \end{aligned}$$

and, by definition of C_A^S , $\mathcal{F}_{spv}(C_A^S, S) \geq \mathcal{F}_{spv}(C_B, S) \geq \mathcal{F}_{spv}(C_B, \emptyset)$. Inequality (c) holds because

$$\begin{aligned} \text{DoV}_c^+(C_A^S, S) - \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) + \mathcal{F}_{spv}(C_A^S, \emptyset) \\ &= \mathcal{F}_{spv}(C_A^S, S) - \mathcal{F}_{spv}(C_A^S, \emptyset) - \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) + \mathcal{F}_{spv}(C_A^S, \emptyset) \\ &= \mathcal{F}_{spv}(C_A^S, S) - \mathcal{F}_{spv}(C_A^{S^*}, \emptyset) \end{aligned}$$

and, by definition of C_A^S , $\mathcal{F}_{spv}(C_A^S, S) \geq \mathcal{F}_{spv}(C_A^{S^*}, S) \geq \mathcal{F}_{spv}(C_A^{S^*}, \emptyset)$. \square

5 Conclusions and Future Work

Controlling elections through social networks is a significant issue in modern society. Political campaigns are using social networks as effective tools to influence voters in real-life elections. In this paper, we formalized the multi-winner election control problem through social influence. We proved that finding an approximation to the maximum margin of victory or difference of winners, for both constructive and destructive cases, is *NP*-hard for any approximation factor. We relaxed the problem to a variation of straight-party voting and showed that this case is approximable within a constant factor in both constructive and destructive cases. To our knowledge, these are the first results on multi-winner election control via social influence.

The results in this paper open several research directions. We plan to study the problem in which the adversary can spread a different (constructive/destructive) message for each candidate, using different seed nodes. In these cases, a good strategy could be that of sending a message regarding a third party (different from the target one and the most voted opponent), and our results cannot be easily extended.

References

1. Mohammad Abouei Mehrizi and Gianlorenzo D'Angelo. Multi-Winner Election Control via Social Influence. *arXiv e-prints*, page arXiv:2005.04037, May 2020.
2. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *J. Economic Perspectives*, 31(2):211–36, 2017.
3. Vincenzo Auletta, Ioannis Caragiannis, Diodato Ferraioli, Clemente Galdi, and Giuseppe Persiano. Minority becomes majority in social networks. In *WINE2015*, pages 74–88, 2015.
4. 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.

5. Sirin Botan, Umberto Grandi, and Laurent Perrussel. Propositionwise opinion diffusion with constraints. In *4th AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE)*, 2017.
6. Robert Brederick and Edith Elkind. Manipulating opinion diffusion in social networks. In *IJCAI2017*, pages 894–900, 2017.
7. Markus Brill, Edith Elkind, Ulle Endriss, and Umberto Grandi. Pairwise diffusion of preference rankings in social networks. In *IJCAI2016*, pages 130–136, 2016.
8. Bryan A Campbell and Michael D Byrne. Straight-party voting: what do voters think? *IEEE Transactions on Information Forensics and Security*, 4(4):718–728, 2009.
9. W. Chen, C. Castillo, and L. V. S. Lakshmanan. *Information and Influence Propagation in Social Networks*. Morgan & Claypool, 2013.
10. Federico Corò, Emilio Cruciani, Gianlorenzo D'Angelo, and Stefano Ponziani. Exploiting social influence to control elections based on scoring rules. In *IJCAI2019*, pages 201–207, 2019.
11. Federico Corò, Emilio Cruciani, Gianlorenzo D'Angelo, and Stefano Ponziani. Vote for me! election control via social influence in arbitrary scoring rule voting systems. In *AAMAS2019*, pages 1895–1897, 2019.
12. Pedro Domingos and Matt Richardson. Mining the network value of customers. In *KDD2001*, pages 57–66. ACM, 2001.
13. Piotr Faliszewski, Rica Gonen, Martin Koutecký, and Nimrod Talmon. Opinion diffusion and campaigning on society graphs. In *IJCAI2018*, pages 219–225, 2018.
14. Piotr Faliszewski and Jörg Rothe. Control and bribery in voting. In *Handbook of Computational Social Choice*, pages 146–168. Cambridge University Press, 2016.
15. Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner voting: A new challenge for social choice theory. In Ulle Endriss, editor, *Trends in Computational Social Choice*, chapter 2, pages 27–48. AI Access, 2017.
16. Emilio Ferrara. Disinformation and social bot operations in the run up to the 2017 french presidential election. *First Monday*, 22(8), 2017.
17. M.R. Hershey. *Party Politics in America*. Taylor & Francis, 2017.
18. David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. *Theory of Computing*, 11:105–147, 2015.
19. Daniel Kreiss. Seizing the moment: The presidential campaigns' use of twitter during the 2012 electoral cycle. *New Media & Society*, 18(8):1473–1490, 2016.
20. Herbert M Kritzer. Roll-off in state court elections: Change over time and the impact of the straight-ticket voting option. *Journal of Law and Courts*, 4(2):409–435, 2016.
21. Gordon Pennycook, Tyrone Cannon, and David G Rand. Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 2018.
22. Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD2002*, pages 61–70. ACM, 2002.
23. J.M. Ruhl and R.H. McDonald. *Party Politics And Elections In Latin America*. Taylor & Francis, 2019.
24. Sebastian Stier, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter. *Political Communication*, 35(1):50–74, 2018.
25. Bryan Wilder and Yevgeniy Vorobeychik. Controlling elections through social influence. In *AAMAS2018*, pages 265–273, 2018.
26. William S. Zwicker. Introduction to the theory of voting. In *Handbook of Computational Social Choice*, pages 23–56. Cambridge University Press, 2016.