Orienteering for Electioneering

Jonah Kallenbach

Department of Computer Science, Harvard University.

Robert Kleinberg

Department of Computer Science, Cornell University.

Scott Duke Kominers

Harvard Business School, Department of Economics, Center of Mathematical Sciences and Applications, and Center for Research on Computation and Society, Harvard University, and National Bureau of Economic Research.

Abstract

In this paper, we introduce a combinatorial optimization problem that models the investment decision a political candidate faces when treating the opponent's campaign plan as given. Our formulation accounts for both the time cost of traveling between districts and the time expended while campaigning within districts. We describe a polynomial-time algorithm that computes a $(2+\epsilon)$-approximation to the optimal solution of a discrete version of our problem by reducing the problem to another combinatorial optimization problem known as Orienteering.

1. Introduction

In the finite time between the declaration of candidacy for political office and the date of the election, a candidate must determine the optimal way to spend his or her time and resources campaigning. An optimized campaign strategy could mean the difference between victory and defeat, especially in close contests.

Time budgeting presents a difficult optimization problem: the candidate must split his or her time between campaigning in key districts and traveling among districts, all the while considering where the opponent is spending time (and attempting to counter him or her there). Time budget constraints on campaigns were particularly salient in the nineteenth and early twentieth centuries—before air travel, planning an intelligent sequence of campaign stops was essential for every candidate. We believe that such planning issues remain relevant today, as in many election contexts either campaign time is limited (as in typical congressional elections) or the area to be covered is so large that choices must be made regarding which areas to visit (as in most national elections).

For an illustrative example, we consider the flight distances between the cities where Hillary Clinton made campaign stops in the March, 2016 leg of her United States presidential campaign (fetched from `www.hillaryspeeches.com/speech-archive/2016-2/march-2016/`).

We estimate flight distances between each of the cities where Clinton stopped using `www.travelmath.com`, assuming that if a flight time is < 30 minutes then Clinton drove instead (in which case we use `travelmath.com`'s driving time estimation service). We find that on average Clinton spent at least 111.3 minutes per day traveling—on the same order of magnitude as many campaign stops, and accounting for about 7.7% of the total time in the month of March. (This is an extremely conservative underestimate—e.g., if Clinton had a campaign stop in one city on a given day and a stop in a distant city the next, we only considered the time it would take to fly in a straight line from one city to the other. This does not account for time spent traveling to and from airports, checking into hotels, traveling to campaign events themselves, security, and so forth.)

A large body of work has applied economic and game-theoretic models to the problem of resource allocation in campaigns. Perhaps most famous is the work of Snyder [18], which demonstrated existence, uniqueness, and characterization results for Nash equilibria of a campaign resource allocation game under relatively general conditions (see also [19] for empirical estimates of the Snyder [18] model). Related campaign models include those of Persson and Tabellini [14], Lizzeri and Persico [11], Strömberg [20], and Hummel and Holden [8]. General models of "all-pay contests," of which some political campaigns are examples, have recently been studied by, e.g., Konrad and Kovenock [9] and Siegel [17]. However, to our knowledge, no work thus far has taken into account candidates' time budgets, examining how candidates split their time between campaigning within areas and traveling among them.

When one accounts for travel time, the candidate's resource allocation problem becomes a mixture of discrete and continuous optimization. We introduce a precise formulation of this problem, which we term ELECTIONEERING, and show that it is NP-hard. Our formulation optimizes one candidate's allocation of campaign resources while treating the opponent's resource allocation as given (e.g., if the opposition candidate has published how much time he or she will be spending in each district, and the payoff function is known).

The discrete portion of the ELECTIONEERING problem involves optimizing over sequences of districts to visit; in this respect, it bears some resemblance to the traveling TRAVELING SALESMAN (TSP). We show that ELECTIONEERING does indeed have an extremely close connection to a variant of the TSPknown as ORIENTEERING. The ORIENTEERING problem, like our ELECTIONEERING problem, is NP-hard. Fortunately, there are polynomial-time algorithms to compute approximately optimal solutions to ORIENTEERING. The first such algorithm, due to Blum et al. [5], computes a solution that approximates the optimum within a factor 4; this approximation factor was improved to 3 by Bansal et al. [3], and then to $2+\epsilon$ (for any constant $\epsilon > 0$), by Chekuri et al. [7]. (The algorithm of Chekuri et al. [7] runs in polynomial time, but the exponent in the running time bound tends to ∞ as ϵ tends to 0, which is why the existence of a polynomial-time $(2+\epsilon)$-approximation algorithm for every $\epsilon > 0$ does not imply the existence of a polynomial time $(2+o(1))$-approximation algorithm.)

In the next section, we formalize the ELECTIONEERING problem, compare it with the Snyder [18] model of campaign resource allocation, and show that it is NP-hard. The remainder of the paper presents an algorithm that approximates the optimal solution to ELECTIONEERING by reducing it to a problem we call ORIENTEERING-WITH-MINING: In Sections 3 and 4, we present an approximate reduction from ORIENTEERING-WITH-MINING to ORIENTEERING, which implies that the known approximation algorithms for the latter problem can be used to compute approximately optimal solutions of the former. In Section 5, we then present an algorithm that solves Electioneering by reducing it to Orienteering-with-Mining. In Section 6, we consider an extension of our model that incorporates time dependencies between campaign strategies in different districts. We provide concluding discussion in Section 7.

2. Electioneering

In this section we formalize the problem of choosing an optimal political campaign given (a) the support for each candidate prior to the campaign in each district and (b) the opponents' chosen campaign strategies. We assume there is a set N of districts; each district $n \in N$ has e_n electoral votes. We suppose that there are I candidates, $1, \ldots, I$, who campaign across districts to try to maximize their expected electoral vote totals. By introducing the electoral votes e_n, our model allows the possibility of including an "electoral college" like the United States has in its current presidential election process. To use a majority vote rule instead, we just set each district's electoral vote count equal to that district's population.

We assume that there is time $T \geq 0$ until the election, and that traveling between districts n and n' takes travel time $d(n, n')$; these travel times are assumed to be non-negative and symmetric (i.e., $d(n, n') = d(n', n)$), and to satisfy the triangle inequality (i.e., $d(n, n'') \leq d(n, n'') + d(n', n'')$). We imagine that at the start of time, the candidates $i \in \{1, \ldots, I\}$ respectively already have support levels $x'_{i,n}$ in each district n, equivalent to their having spent time $x'_{i,n}$ campaigning in that district. The candidates respectively allocate time $x_{i,n} \geq 0$ to campaigning in each district n.

We let N_i denote the set of districts that candidate i campaigns in, i.e.,

$$N_i := \{n \in N : x_{i,n} > 0\}.$$

We require that

$$\sum_{n \in N} x_{i,n} \leq T - D_i \tag{1}$$

for all i, where D_i is the minimal travel time required to visit each district in N_i. (We can also impose a budget constraint on the opponent's strategies analogous to (1). However, such a constraint is not necessary in our solution or analysis.)

We denote by h the payoff function for campaign time spent in a district, that is, the relation between time spent in a district and the expected electoral vote gain for that district; we assume that h is increasing and weakly concave, and normalize so that prior support levels enter additively in the argument of h. (A nice h to start with is $h(x) = x^b$ for $b \leq 1$.) We have that candidate i's probability of winning a particular district (dependent on candidate i's allocation $x_{i,n}$ of time to that district and opponents' allocations $x_{-i,n}$) is:

$$\tilde{p}_{i,n}(x_{1,n}, \ldots, x_{I,n}) := \frac{h(x'_{i,n} + x_{i,n})}{h(x'_{i,n} + x_{i,n}) + \sum_{j \neq i} h(x'_{j,n} + x_{j,n})}.$$

The Electioneering problem seeks a campaign time allocation x_i for candidate i that maximizes the resulting (expected) electoral vote total,

$$\sum_{n \in N} \tilde{p}_{i,n}(x_{1,n}, \ldots, x_{I,n}) \cdot e_n,$$

subject to the constraint (1), given prior allocations x'_i, and the time allocations of the other candidates.

2.1. Relation to the Snyder [18] model

Snyder [18] introduced a model in which candidates allocate campaign resources across districts to maximize the expected number of electoral votes they receive. In Snyder's [18] model, there are two candidates (called 1 and 2 herein) and some prior probability a_n that candidate 1 wins, so that the probability candidate 1 wins after $x_{1,n}$ and $x_{2,n}$ resources have been allocated in district n is

$$p_{1,n}(x_{1,n}, x_{2,n}) = \frac{a_n h(x_{1,n})}{a_n h(x_{1,n}) + (1 - a_n) h(x_{2,n})}.$$

Unlike in our work, Snyder [18] did not explicitly account for travel time as part of the candidate's budget constraint. Our functional form for within-district campaign impact is not directly comparable to Snyder's, but is arguably richer—it is capable of representing prior advantage in a deeper way than Snyder's [18] model, as it can encode both a prior probability and the tightness of the bounds on that probability. For example, it may be the case that in two districts the advantage of candidate 1 is 60%-40%, but one of those districts may be competitive while the other is not. For example, if one district is particularly large or has voters with particularly deeply entrenched views, this effect can be captured by the magnitude of prior support levels in our model, but cannot be modeled in Snyder's [18] framework. Note that if we set $I = 2$, $x'_{1,n} = x'_{2,n} = 0$, and $d \equiv 0$, we get back the Snyder [18] model, albeit without the priors a_n.

2.2. NP-hardness of Electioneering

Like many other problems that involve optimizing over sequences of vertices in a graph, the ELECTIONEERING problem is NP-hard. To see this formally, we present a reduction from the HAMILTONIAN PATH problem to ELECTIONEERING. For the proof, we look at the simple case in which there are only two candidates, 1 and 2; this suffices to show that the problem is NP-hard.

Recall that HAMILTONIAN PATH is the problem of deciding, for a given undirected graph, whether there exists a path that visits every vertex exactly once. Letting G denote an instance of HAMILTONIAN PATH with vertex set V, we construct an instance of ELECTIONEERING with $N = V \cup \{s\}$, where the number of electoral votes in district n is given by

$$e_n = \begin{cases} 1 & n \neq s \\ 0 & n = s. \end{cases}$$

The travel times $d(n, n')$ are defined as follows: $d(n, n') = 0$ if $n = n'$, $d(n, n') = 1$ if $n \neq n'$ and one of n, n' is equal to s, and otherwise $d(n, n')$ is the length of the shortest path from n to n' in G. For the payoff function h, we simply use $h(x) = x$, and for candidate 2's allocations we use $x_{2,n} = |V|^{-2}$ and $x'_{2,n} = 0$. The time budget for candidate 1 is set to $T = |V| + 1$, and the prior allocations $x'_{1,n}$ are set to 0 for all n.

If G has a Hamiltonian path, then candidate 1 has an election strategy that involves visiting all the districts in the order dictated by the Hamiltonian path (entailing a travel cost of $|V|$) and investing time $x_{1,n} = |V|^{-1}$ campaigning in each district. The expected number of electoral votes earned by this strategy is

$$|V| \cdot \frac{h(|V|^{-1})}{h(|V|^{-1}) + h(|V|^{-2})} = |V| \cdot \frac{|V|^{-1}}{|V|^{-1} + |V|^{-2}} = |V| \cdot \frac{1}{1 + |V|^{-1}} > |V| \cdot \left(1 - |V|^{-1}\right) = |V| - 1.$$

If G does not have a Hamiltonian path, then any election strategy for candidate 1 must either omit one of the districts, or it must entail a travel cost of $|V| + 1$. In the former case, the expected number of electoral votes is less than $|V| - 1$ (each district visited contributes strictly less than one electoral vote in expectation) while in the latter case, the expected number of electoral votes is 0, as candidate 1's remaining budget leaves no time at all for campaigning in any of the districts.

Thus, the question of whether G has a Hamiltonian path is equivalent to the question of whether candidate 1 can win more than $|V| - 1$ electoral votes, in expectation, in the Electioneering instance constructed by the reduction. Consequently, maximizing the expected electoral vote total (as in Electioneering) would allow us to solve the Hamiltonian Path problem.

3. Orienteering-with-Mining

We now introduce the Orienteering problem. We are given a graph with edge lengths and with prizes at each vertex; we seek to construct a path of length at most T, starting from a designated vertex s, that maximizes the total prize value of the vertices visited. Intuitively, Orienteering represents an agent traveling from vertex to vertex, collecting prizes, with a limited budget of travel time.

First, we present a simple generalization of the Orienteering problem: Orienteering-with-Mining, in which collecting prizes from vertices requires time investment, and time spent at each individual vertex has diminishing marginal returns. We consider an undirected graph $G = (V, E)$ with vertex set V and edge set E. We assume that there is a *distance* function $d : E \to \mathbb{R}_{\geq 0}$ on edges that satisfies the metric space axioms, and a *prize* function $\pi_v(t) : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ for each vertex $v \in V$. (Here, we use the notation $\mathbb{R}_{\geq 0}$ to denote the set of nonnegative real numbers.) We assume moreover that each prize function exhibits diminishing marginal returns, i.e., that each π_v is non-decreasing and weakly concave.

A *path (of length ℓ)* is a function $P : \{1, \ldots, \ell\} \to V$ encoding an ordered sequence of vertices $P(1), \ldots, P(\ell) \in V$. An *investment path (of length ℓ)* $\tilde{P} : \{1, \ldots, \ell\} \to V \times \mathbb{R}_{\geq 0}$ encodes both a sequence of vertices and an amount of time investment at each step in the sequence. We denote by $\tilde{P}_V$ the path underlying $\tilde{P}$, that is, the projection of $\tilde{P}$ to V; analogously, we denote by $\tilde{P}_{\mathbb{R}}$ the projection of $\tilde{P}$ to $\mathbb{R}_{\geq 0}$.

The *travel time* for a path P of length ℓ,

$$\tau(P) := \sum_{k=1}^{\ell-1} d(P(k), P(k+1)),$$

records the time required to traverse the vertices of P in sequence. For an investment path $\tilde{P}$ of length ℓ, the *time spent at vertex $v \in V$* is

$$\tau_v(\tilde{P}) := \sum_{k=1}^{\ell} \tilde{P}_{\mathbb{R}}(k) \cdot \mathbb{1}_{\tilde{P}_V(k)=v}.$$

For an investment path $\tilde{P}$, the *total time*

$$\tau(\tilde{P}) := \tau(\tilde{P}_V) + \sum_{v \in V} \tau_v(\tilde{P}) \tag{2}$$

records the total time investment required in following $\tilde{P}$, combining travel time and investment time at each vertex. The *prize associated to investment path* $\tilde{P}$ is

$$\Pi(\tilde{P}) := \sum_{v \in V} \pi_v(\tau_v(\tilde{P})). \tag{3}$$

The ORIENTEERING-WITH-MINING problem seeks to find the investment path $\tilde{P}$ of total time at most T, starting from a fixed root node s, that maximizes the associated prize:

$$\begin{aligned} \text{max:}\ & \Pi(\tilde{P}) \\ \text{subject to:}\ & \tau(\tilde{P}) \leq T \\ & \tilde{P}_V(1) = s. \end{aligned}$$

The ORIENTEERING problem is the special case of ORIENTEERING-WITH-MINING in which π_v is a constant function for each $v \in V$. Indeed, if π_v is constant for each $v \in V$, then any optimal investment path will spend 0 time at each vertex, as that eliminates the second term in (2) without changing (3).

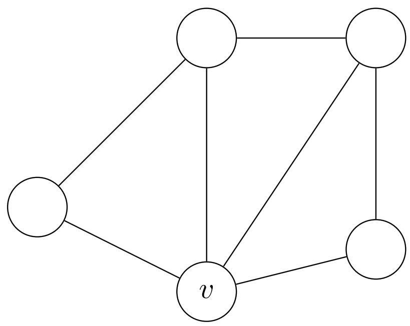
4. Reducing Orienteering-with-Mining to Orienteering

We now construct a graph $G' = (V', E')$ and ORIENTEERING problem $\langle G', d', \pi', s \rangle$ that yields an approximation to the ORIENTEERING-WITH-MINING problem $\langle G, d, \pi, s \rangle$. We discretize time spent at nodes into units of length m; we leave travel times between nodes unchanged. Formally, $\langle G', d', \pi', s \rangle$ is defined as follows:

- We construct $V' \supset V$ by supplementing V with T/m dummy nodes $\nu_1^v, \ldots \nu_{T/m}^v$ for each node $v \in V$.
- We construct $E' \supset E$ by supplementing E with an edge connecting each dummy node ν_k^v to its parent node v.
- We take $d' \equiv d$ on V, and set $d'(v, \nu_k^v) = m/2$ for each dummy node ν_k^v associated to v.
- We take $\pi'_v \equiv \pi_v(0)$ for $v \in V$, and then set

  $$\pi'_{\nu_k^v} = \pi_v(k \cdot m) - \pi_v((k-1) \cdot m),$$

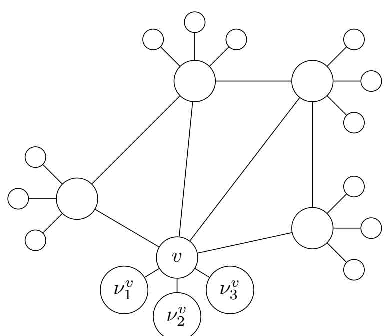
  that is, we set the prize for visiting ν_k^v to be equal to the marginal reward of spending m additional units of time at v under π (after having already spent $(k-1) \cdot m$ units of time at v).



$$T = 3$$

$$\pi_v(t) = \frac{h(x'_{1,v}+t)}{h(x'_{1,v}+t)+h(x_{2,v}+x'_{2,v})}$$

(a) An Orienteering-with-Mining problem, with the reward for visiting node v given by $\pi_v(t)$.



$$T = 3,\ m = 1$$

$$\pi_{\nu_k^v} = \frac{h(x'_{1,v}+km)}{h(x'_{1,v}+km)+h(x_{2,v}+x'_{2,v})} - \frac{h(x'_{1,v}+(k-1)m)}{h(x'_{1,v}+(k-1)m)+h(x_{2,v}+x'_{2,v})}$$

(b) The graph in (a) augmented with dummy nodes ν_k^v, to be solved with Orienteering.

Figure 1: An illustration of our reduction applied to Electioneering, with two candidates 1 and 2.

The graph G' is constructed by adding a series of dummy nodes attached to each $v \in V$; these nodes take time $m/2$ to travel to (or from), and encode the returns to spending an additional m units of time at v in $\langle G, d, \pi, s\rangle$, as pictured in Figure 1. Because π_v has diminishing marginal returns, the prize-maximizing path through G' will always visit dummy nodes in sequence $(\nu_1^v, \nu_2^v, \ldots)$. The discrete approximation can get arbitrarily close to the correct solution as we adjust m.

Now, we can run the polynomial-time approximation algorithm of Chekuri et al. [7] on the ORIENTEERING problem $\langle G', d', \pi', s\rangle$, to obtain an α_B-approximation to the reward conferred by the best investment path $\tilde{P}$ on the $\langle G', d', \pi', s\rangle$, where $\alpha_B = 2 + \epsilon$ for an arbitrarily small constant $\epsilon > 0$. (The algorithm's running time is $|V'|^{O(1/\epsilon^2)}$, so in order to obtain a polynomial running time we require α_B to be bounded away from 2.) We let ALG be the utility produced by running the approximation algorithm of Chekuri et al. [7] on $\langle G', d', \pi', s\rangle$; we seek to compare ALG to the utility OPT $:= \pi(\tilde{P}_{\mathrm{OPT}})$ obtained in $\langle G, d, \pi, s\rangle$ under the optimal (investment) path $\tilde{P}_{\mathrm{OPT}}$.

We do a worst case analysis to bound the additional error that we introduce by using a discrete approximation to ORIENTEERING-WITH-MINING, relative to the true continuous version. We let DIS $= \pi(\tilde{P}_{\mathrm{DIS}})$ be the utility achieved by the best investment path $\tilde{P}_{\mathrm{DIS}}$ under discrete approximation. Reasoning conservatively, we can ignore travel costs, as any travel route among nodes in V that is used under OPT can be replicated under DIS. When "mining" units of prize from each node $v \in V$, DIS can lose to OPT by at most $\pi'_{\nu_1^v}$, as at a given node, DIS can never be behind OPT by more than a single discrete unit (and the biggest discrete unit for v is $\pi'_{\nu_k^v}$). Additionally, DIS may slightly overestimate OPT, as follows: at time $T - m/2$, DIS may spend the last $m/2$ time units to travel to a dummy node associated to v without returning, collecting at most $\max_{v\in V}[\pi'_{\nu_1^v} - \pi_v(m/2)]$ extra units of reward relative to OPT. That is, at the end of its path, DIS can collect a reward as if it spent m extra units of time at a node in V, despite only having spent $m/2$ units of time. So we have that:

$$\mathrm{DIS} \geq \mathrm{OPT} - \left(\sum_{v\in V} \pi'_{\nu_1^v}\right) + \left(\max_{v\in V}[\pi'_{\nu_1^v} - \pi_v(m/2)]\right).$$

Combining the preceding observations, we see that:

$$\mathrm{ALG} \geq \frac{1}{\alpha_B}\mathrm{DIS} \geq \frac{1}{\alpha_B}\left(\mathrm{OPT} - \left(\sum_{v\in V} \pi'_{\nu_1^v}\right) + \left(\max_{v\in V}[\pi'_{\nu_1^v} - \pi_v(m/2)]\right)\right). \tag{4}$$

The expression (4) completely describes the approximation that our algorithm provides. As $m \to 0$, we have $\sum_{v\in V} \pi'_{\nu_1^v} \to 0$ and $(\max_{v\in V}[\pi'_{\nu_1^v} - \pi_v(m/2)]) \to 0$, so we can make the error in our approximation to ORIENTEERING-WITH-MINING arbitrarily close to the error bound that Chekuri et al. [7] found for approximating ORIENTEERING. (The Chekuri et al. [7] approximation factor does not depend on the number of nodes in the input graph, and hence does not depend on the level of discretization—although the running time of the Chekuri et al. [7] algorithm does.)

5. Algorithm for Electioneering

Now, as an application of the results from the preceding section, we present an approximation algorithm for the Electioneering problem introduced in Section 2. Our algorithm proceeds by reducing Electioneering to Orienteering-with-Mining.

We assume that candidates $j \neq i$ have made their allocations across districts, x_j. The problem i faces is then an Orienteering-with-Mining problem in which

- the graph G is the graph of districts, with associated distances given by travel times, and
- each district $n \in N$ has reward function

$$\pi_n(t) = \frac{h(x'_{i,n} + t)}{h(x'_{i,n} + t) + \sum_{j \neq i} h(x'_{j,n} + x_{j,n})}.$$

With the Electioneering→Orienteering-with-Mining reduction just described, the weights on the dummy nodes in our reduction would be

$$\pi'_{\nu_k^n} = \frac{h(x'_{i,n} + km)}{h(x'_{i,n} + km) + \sum_{j \neq i} h(x_{j,n} + x'_{j,n})} - \frac{h(x'_{i,n} + (k-1)m)}{h(x'_{i,n} + (k-1)m) + \sum_{j \neq i} h(x_{j,n} + x'_{j,n})}$$

for $k \in \{1, \ldots, T/m\}$. Then simply applying the algorithm presented in Section 4 yields an approximation to Electioneering that achieves the approximation bound (4), that is, within a factor of α_B of

$$\mathrm{OPT} - \left(\sum_{n \in N} \pi'_{\nu_1^n} \right) + \left(\max_{n \in N} [\pi'_{\nu_1^n} - \pi_n(m/2)] \right).$$

A graphical illustration of our reduction in action on the Electioneering problem is given in Figure 1.

6. Extension: Incorporating time dependencies

As an extension of our model, we consider the possibility of side constraints that introduce time dependencies between campaign strategy decisions in different districts. Such constraints are natural, for example, if a candidate must pass through one district *en route* to others. Unfortunately—at least when considered in generality—they appear to introduce significant computational difficulties.

The side constraints that we consider take the following form: For each district $n \in N$ and $\tau > 0$, let $\theta_{n,\tau}$ denote a predicate that is satisfied by investment paths that devote at least τ units of time to visiting n. (Here, abusing terminology only slightly, we extend our concept of "investment paths" from the setting of Orienteering to Electioneering, using this structure for the travel sequences implicitly computed in (1).) Let Λ denote a finite set of *clauses*, each of which is a disjunction of one or more predicates of the form $\theta_{n,\tau}$. For each district $n \in N$ there is a constraint ϕ_n, which is a conjunction of (0 or more) clauses in Λ; an empty conjunction $\phi_n = \top$ is interpreted as a trivial constraint that is always satisfied.

An investment path *satisfies the constraint set* $\Phi = \{\phi_n\}_{n \in N}$ if it is the case that, for every district n occurring along the path, the sub-path that precedes the first visit to n satisfies the Boolean combination of predicates represented by ϕ_n. Let Electioneering-with-Dependencies denote the problem of finding an investment path starting in district s that maximizes the expected electoral vote total, subject to a specified set of time dependency constraints, Φ, and a constraint that the total time investment is at most T.

The Electioneering-with-Dependencies problem is complex enough that it is NP-hard to decide whether the optimum value of an instance of Electioneering-with-Dependencies is non-zero, and thus also NP-hard to approximate the optimum to within any finite factor. This can be seen using the following reduction from the Hitting Set problem, in which one is given a finite set U, a collection of subsets $S_1, \ldots, S_r \subseteq U$, and a desired number of elements k, and one must decide whether there exists a set of at most k elements that has non-empty intersection with each of $S_1, \ldots, S_r$.

Indeed, given an instance of Hitting Set, $\langle U, S_1, \ldots, S_r \rangle$, we construct an Electioneering-with-Dependencies instance with districts $N = \{s\} \cup U \cup \{v_1, \ldots, v_r\} \cup \{w\}$. The are two candidates, 1 and 2, with $x'_{1,n} = x'_{2,n} = 0$ and $x_{2,n} = 1$ for each district n. The payoff function of each district is $h(x) = x$, and the electoral vote counts satisfy

$$e_n = \begin{cases} 1 & n = w \\ 0 & \text{otherwise,} \end{cases}$$

so that a candidate's expected electoral vote total equals his or her probability of winning district w. We set $\phi_{v_j} = \bigvee_{u \in S_j} \theta_{u,1}$ for each $j \in \{1, \ldots, r\}$ and we set $\phi_w = \bigwedge_{j=1}^{r} \theta_{v_j,1}$: to campaign in v_j, a candidate must first spend at least 1 unit of time campaigning in at least one of the districts in S_j, and to campaign in w, a candidate must first campaign in each of the districts $v_1, \ldots, v_r$. For every other district n, the constraint ϕ_n is the trivial constraint $\top$. The reduction is completed by defining the distance between any two distinct districts to be 1, and defining the time budget to be $T = 2k + 2r + 2$.

Under our construction, if there is a hitting set $H = \{u_1, \ldots, u_k\}$, then there is a campaign strategy in which the candidate visits

$$s, u_1, \ldots, u_k, v_1, \ldots, v_r, w$$

in sequence, spending one unit of time in each of those districts except s, resulting in a positive probability of winning district w. Conversely, if there is a campaign strategy that results in a positive probability of winning district w, then the candidate must spend at least 1 unit of time in each of $v_1, \ldots, v_r$, plus 1 unit of time traveling to each of those districts and to w—a total time investment of $2r+1$. If H denotes the set of districts in U in which the candidate spends at least 1 unit of time, then H must be a hitting set for $S_1, \ldots, S_r$ because each of the clauses ϕ_{v_j} is satisfied, and $|H| \leq k$ because the candidate must invest at least 2 units of time per element of H (one time unit is spend traveling to each $u \in H$; the other is spent campaigning there), and the remaining time budget (when we subtract the amount of time invested in districts $v_1, \ldots, v_r, w$) is less than $T - (2r + 1) = 2k + 1$.

Even if we constrain each of the clauses in Λ to be a single predicate $\theta_{n,\tau}$, rather than a disjunction of such predicates, the Electioneering-with-Dependencies problem remains at least as hard to approximate as the Densest k-Subgraph problem, via the following reduction. Given graph $G = (V, E)$ and parameter k, we construct an instance of ELECTIONEERING-WITH-DEPENDENCIES in which $N = \{s, s'\} \cup V \cup E$ and the electoral vote counts are specified by

$$e_n = \begin{cases} 1 & n \in E \\ 0 & \text{otherwise.} \end{cases}$$

Distances are given by

$$\begin{aligned} & d(s, s') = 2 \\ \forall v \in V: \quad & d(s, v) = d(s', v) = \frac{3}{2}|E| \\ \forall \xi \in E: \quad & d(s, \xi) = d(s', \xi) = 1 \\ \forall x, y \in N: \quad & d(x, y) = d(x, s) + d(s, y) \ \text{ if } \{x, y\} \cap \{s, s'\} = \emptyset, \end{aligned}$$

and the payoff function for each district is $h(x) = x$. There are two candidates, 1 and 2, with $x'_{1,n} = x'_{2,n} = 0$ and $x_{2,n} = 1$ for each district n. The time budget is $T = (k+1)(3|E|+1)$. The time dependency constraint ϕ_n is defined to be trivial unless $n = \xi$ for some edge ξ with endpoints u and v; in that case $\phi_n = \theta_{s',1} \wedge \theta_{u,1} \wedge \theta_{v,1}$. In other words, before campaigning in a district corresponding to an edge of G, a candidate must first campaign in s' as well as the districts corresponding to both endpoints of E.

Note that if G contains a subgraph L with k vertices and k' edges, then there is a campaign strategy that consists of visiting the vertices of L, followed by s', followed by the edges of L, while spending 1 time unit in each district visited. This strategy obeys the budget and dependency constraints, and earns an expected electoral vote total of $\frac{1}{2}k'$. As a partial converse, any strategy that earns at least q electoral votes in expectation must visit at least q of the districts corresponding to edges of G, in expectation. To do so, the candidate must first visit—and spend at least 1 unit of time in—each of the districts corresponding to endpoints of those edges, as well as the district s'. Visiting r vertices followed by s', while spending one unit of time in each, incurs a time cost of $r(3|E|+1)+1$, which exceeds the budget T unless $r \le k$. Thus, an electoral strategy resulting in expected electoral vote total q implies the existence of a subgraph with $r \le k$ vertices and at least q edges. Thus, an α-approximation algorithm for ELECTIONEERING-WITH-DEPENDENCIES would imply the existence of a (2α)-approximation algorithm for DENSEST k-SUBGRAPH. The latter problem is widely conjectured to be hard to approximate to within an n^{ϵ} factor for some constant $\epsilon > 0$ (see, e.g., [4, 6]), although the only known proofs of super-constant hardness of approximation for DENSEST k-SUBGRAPH depend either on the Exponential Time Hypothesis [12] or on average-case hardness assumptions [1].

7. Discussion

In an actual election campaign with a limited number of districts to visit (e.g., fifty states in the case of American presidential campaigns), the worst-case computational hardness of the resource allocation problem may not pose a severe impediment to computing optimal solutions using intelligently chosen heuristics to prune the search space. (For example, Applegate et al. [2] report successfully solving an instance of the traveling salesman problem with 85,900 cities.)

For using our approach in practice, however, two additional issues are paramount: structural estimation and equilibrium computation. We imagine that an algorithm like the one we provide here would be used as a subroutine for iteratively computing best responses in a heuristic equilibrium solver. As input for such a system, we would need an estimate for h, the cumulative return for campaign effort in a particular district. In principle, any campaign planner would already need such an estimate, although in practice the function would assuredly be multivariate, unlike the univariate function we have assumed here. For example, it is likely that spending a large, continuous block of time in a district is less effective than that same amount of time spread out over the campaign; this is not accounted for by our model. Our model also does not account for any externalities between campaign time spent in one district and electoral returns in another. While we have not found academic estimates of the returns to spending time at individual campaign sites, academic estimates of the value of campaign spending are available (these estimates are based on public records of campaign spending, vote totals, and exogenous instruments; see, e.g., Levitt [10], Palda and Palda [13], Samuels [16], Rekkas [15]). Meanwhile, equilibrium computation in our setting remains a subtle problem, which we regard as an enticing open question for future investigations of optimal election campaign strategy.

Acknowledgements

The authors thank Mark Hansen, the associate editor, and an anonymous referee for helpful comments. Kleinberg gratefully acknowledges the support of National Science Foundation grant CCF-1512964. Most of this research was completed while Kleinberg was a Researcher at Microsoft Research New England. Kominers gratefully acknowledges the support of National Science Foundation grants CCF-1216095 and SES-1459912, the Harvard Milton Fund, and the Ng Fund of the Harvard Center of Mathematical Sciences and Applications. Most of this research was completed while Kominers was a Junior Fellow at the Harvard Society of Fellows.

References

[1] Alon, N., Arora, S., Manokaran, R., Moshkovitz, D., Weinstein, O., 2011. Inapproximability of densest κ-subgraph from average case hardness, preprint.

[2] Applegate, D. L., Bixby, R. E., Chvátal, V., Cook, W., Espinoza, D. G., Goycooleaf, M., Helsgaun, K., 2009. Certification of an optimal TSP tour through 85,900 cities. Operations Research Letters 37 (1), 11–15.

[3] Bansal, N., Blum, A., Chawla, S., Meyerson, A., 2004. Approximation algorithms for deadline-TSP and vehicle routing with time-windows. In: Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing. ACM, pp. 166–174.

[4] Bhaskara, A., Charikar, M., Vijayaraghavan, A., Guruswami, V., Zhou, Y., 2012. Polynomial integrality gaps for strong sdp relaxations of densest k-subgraph. In: Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM, pp. 388–405.

[5] Blum, A., Chawla, S., Karger, D. R., Lane, T., Meyerson, A., Minkoff, M., 2007. Approximation algorithms for orienteering and discounted-reward TSP. SIAM Journal on Computing 37 (2), 653–670.

[6] Braverman, M., Ko, Y. K., Rubinstein, A., Weinstein, O., 2017. ETH-hardness for densest-k-subgraph with perfect completeness. In: Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM, pp. 1326–1341.

[7] Chekuri, C., Korula, N., Pál, M., 2012. Improved algorithms for orienteering and related problems. ACM Transactions on Algorithms 8 (3), Art. 23.

[8] Hummel, P., Holden, R., 2014. Optimal primaries. Journal of Public Economics 109, 64–75.

[9] Konrad, K. A., Kovenock, D., 2009. Multi-battle contests. Games and Economic Behavior 66 (1), 256–274.

[10] Levitt, S. D., 1994. Using repeat challengers to estimate the effect of campaign spending on election outcomes in the US House. Journal of Political Economy 102 (4), 777–798.

[11] Lizzeri, A., Persico, N., 2001. The provision of public goods under alternative electoral incentives. American Economic Review 91 (1), 225–239.

[12] Manurangsi, P., 2017. Almost-polynomial ratio ETH-hardness of approximating densest k-subgraph. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing. ACM, pp. 954–961.

[13] Palda, F., Palda, K., 1998. The impact of campaign expenditures on political competition in the French legislative elections of 1993. Public Choice 94 (1), 157–174.

[14] Persson, T., Tabellini, G., 1999. The size and scope of government: Comparative politics with rational politicians. European Economic Review 43 (4), 699–735.

[15] Rekkas, M., 2007. The impact of campaign spending on votes in multiparty elections. Review of Economics and Statistics 89 (3), 573–585.

[16] Samuels, D., 2001. Incumbents and challengers on a level playing field: Assessing the impact of campaign finance in Brazil. Journal of Politics 63 (2), 569–584.

[17] Siegel, R., 2009. All-pay contests. Econometrica 77 (1), 71–92.

[18] Snyder, J. M., 1989. Election goals and the allocation of campaign resources. Econometrica 57 (3), 637–660.

[19] Snyder, J. M., 1993. The market for campaign contributions: Evidence for the US Senate 1980–1986. Economics & Politics 5 (3), 219–240.

[20] Strömberg, D., 2002. Optimal campaigning in presidential elections: The probability of being Florida, CEPR Discussion Paper No. 3372.