On the Approximability of Dodgson and Young Elections*

Ioannis Caragiannis[†] Jason A. Covey[‡] Michal Feldman[§] Christopher M. Homan[¶]
Christos Kaklamanis[‖] Nikos Karanikolas[**] Ariel D. Procaccia[††]
Jeffrey S. Rosenschein[‡‡]

Abstract

The voting rules proposed by Dodgson and Young are both designed to find the alternative closest to being a Condorcet winner, according to two different notions of proximity; the score of a given alternative is known to be hard to compute under either rule.

In this paper, we put forward two algorithms for approximating the Dodgson score: a combinatorial, greedy algorithm and an LP-based algorithm, both of which yield an approximation ratio of H_{m-1}, where m is the number of alternatives and H is the harmonic number. We also prove that our algorithms are optimal within a factor of 2, unless problems in $\mathcal{NP}$ have quasi-polynomial time algorithms. Despite the intuitive appeal of the greedy algorithm, we argue that the LP-based algorithm has an advantage from a social choice point of view.

Further, we demonstrate that computing any reasonable approximation of the ranking produced by Dodgson's rule is $\mathcal{NP}$-hard. This result provides a complexity-theoretic explanation of sharp discrepancies that have been observed in the Social Choice Theory literature when comparing Dodgson elections with simpler voting rules.

Finally, we show that the problem of calculating the Young score is $\mathcal{NP}$-hard to approximate by any factor. This leads to an inapproximability result for the Young ranking.

Keywords: Computational social choice, Approximation algorithms

*A preliminary version of the results in this paper appeared in the *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '09)*.

[†]Research Academic Computer Technology Institute and Department of Computer Engineering and Informatics, University of Patras, 26500 Rio, Greece, email: caragian@ceid.upatras.gr

[‡]Department of Computer Science, Rochester Institute of Technology, 102 Lomb Memorial Drive, Rochester, NY 14623-5603, email: jac8687@cs.rit.edu

[§]School of Business Administration and Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, and Microsoft Israel R&D Center, email: mfeldman@huji.ac.il

[¶]Department of Computer Science, Rochester Institute of Technology, 102 Lomb Memorial Drive Rochester, NY 14623-5603, email: cmh@cs.rit.edu

[‖]Research Academic Computer Technology Institute and Department of Computer Engineering and Informatics, University of Patras, 26500 Rio, Greece, email: kakl@ceid.upatras.gr

[**]Research Academic Computer Technology Institute and Department of Computer Engineering and Informatics University of Patras, 26500 Rio, Greece, email: nkaranik@ceid.upatras.gr

[††]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA, email: arielpro@seas.harvard.edu

[‡‡]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, email: jeff@cs.huji.ac.il

1 Introduction

The discipline of voting theory deals with the following setting: a group of n agents each ranks a set of m alternatives; one alternative is to be elected. The big question is: which alternative best reflects the social good? The French philosopher and mathematician Marie Jean Antoine Nicolas de Caritat, marquis de Condorcet, suggested the following intuitive criterion: the winner should be an alternative that beats every other alternative in a *pairwise election*, i.e., an alternative that a majority of the agents prefers over any other alternative. Sadly, it is fairly easy to see that the preferences of the majority may be cyclic, hence a *Condorcet winner* does not necessarily exist. This unfortunate phenomenon is known as the *Condorcet paradox* (see Black [5]).

In order to circumvent this result, several researchers have proposed choosing an alternative that is "as close as possible" to a Condorcet winner. Different notions of proximity can be considered, and yield different voting rules. One such suggestion was advocated by Charles Dodgson, better known by his pen name Lewis Carroll, author of "Alice's Adventures in Wonderland". The *Dodgson score* [5] of an alternative, with respect to a given set of agents' preferences, is the minimum number of exchanges between adjacent alternatives in the agents' rankings one has to introduce in order to make the given alternative a Condorcet winner. A *Dodgson winner* is any alternative with a minimum Dodgson score.

Young [33] raised a second option: measuring the distance by agents. Specifically, the *Young score* of an alternative is the size of the largest subset of agents such that, if only these ballots are taken into account, the given alternative becomes a Condorcet winner. A *Young winner* is any alternative with the maximum Young score. Alternatively, one can perceive a Young winner as the alternative that becomes a Condorcet winner by removing the fewest agents.

Though these two voting rules sound appealing and straightforward, they are notoriously complicated to resolve. As early as 1989, Bartholdi, Tovey and Trick [2] showed that computing the Dodgson score is $\mathcal{NP}$-complete, and that pinpointing a Dodgson winner is $\mathcal{NP}$-hard. This important paper was one of the first to introduce complexity-theoretic considerations to social choice theory. Hemaspaandra et al. [14] refined the aforementioned result by showing that the Dodgson winner problem is complete for Θ_2^p, the class of problems that can be solved by $\mathcal{O}(\log n)$ queries to an $\mathcal{NP}$ set. Subsequently, Rothe et al. [29] proved that the Young winner problem is also complete for Θ_2^p.

The aforementioned complexity results give rise to the agenda of *approximately* calculating an alternative's score, under the Dodgson and Young schemes. This is clearly an interesting computational problem, as an application area of algorithmic techniques.

However, from the point of view of social choice theory, it is not immediately apparent that an approximation of a voting rule is satisfactory, since an "incorrect" alternative—in our case, one that is not closest to a Condorcet winner—might be elected. Nevertheless, we argue that the use of such an approximation is strongly motivated. Indeed, at least in the case of the Dodgson and Young rules, the winner is an "approximation" in the first place, in instances where no Condorcet winner exists. Moreover, the approximation algorithm is equivalent to a new voting rule, which is guaranteed to elect an alternative that is not far from being a Condorcet winner. In other words, a perfectly sensible definition of a "socially good" winner, given the circumstances, is simply the alternative chosen by the approximation algorithm. Note that the approximation algorithm can be designed to satisfy the Condorcet criterion, i.e., always elect a Condorcet winner if one exists. This is always true for an approximation of the Dodgson score, as the Dodgson score of a Condorcet winner is zero. Moreover, approximation algorithms can be designed to satisfy other, less trivial, social choice desiderata, and hence may ultimately be considered socially sensible voting rules.

Related work. The agenda of approximating voting rules was recently pursued by Ailon et al. [1], Coppersmith et al. [10], and Kenyon-Mathieu and Schudy [17]. These works deal, directly or indirectly, with the Kemeny rank aggregation rule, which chooses a ranking of the alternatives instead of a single winning alternative. The Kemeny rule picks the ranking that has the maximum number of agreements with the agents' individual rankings regarding the correct order of pairs of alternatives. Ailon et al. improve the trivial 2-approximation algorithm to an involved, randomized algorithm that gives an 11/7-approximation; Kenyon-Mathieu and Schudy further improve the approximation, and obtain a PTAS.

Two recent papers have directly put forward algorithms for the Dodgson winner problem [15, 22]. Both papers independently build upon the same basic idea: if the number of agents is significantly larger than the number of alternatives, and one looks at a uniform distribution over the preferences of the agents, with high probability one obtains an instance on which it is trivial to compute the Dodgson score of a given alternative. This directly gives rise to an algorithm that can usually compute the Dodgson score (under the assumption on the number of agents and alternatives). However, this is not an approximation algorithm in the usual sense, since the algorithm *a priori* gives up on certain instances, whereas an approximation algorithm is judged by its worst-case guarantees. In addition, this algorithm would be useless if the number of alternatives is not small compared with the number of agents.[1]

Betzler et al. [4] have investigated the parameterized computational complexity of the Dodgson and Young rules. The authors have devised a fixed parameter algorithm for exact computation of the Dodgson score, where the fixed parameter is the "edit distance," i.e., the number of exchanges. Specifically, if k is an upper bound on the Dodgson score of a given alternative, n is the number of agents, and m the number of alternatives, the algorithm runs in time $\mathcal{O}(2^k \cdot nk + nm)$. Notice that in general it may hold that $k = \Omega(nm)$. In contrast, computing the Young score is $W[2]$-complete; this implies that there is no algorithm that computes the Young score exactly, and whose running time is polynomial in nm and only exponential in k, where the parameter k is the number of remaining votes. These results complement ours nicely, as we shall also demonstrate that computing the Dodgson score is in a sense easier than computing the Young score, albeit in the context of approximation.

Putting computational complexity aside, some research by social choice theorists has considered comparing the ranking produced by Dodgson, i.e., the ordering of the alternatives by nondecreasing Dodgson score, with elections based on simpler voting rules. Such comparisons have always revealed sharp discrepancies. For example, the Dodgson winner can appear in any position in the Kemeny ranking [26] and in the ranking of any positional scoring rule [27] (e.g., Borda or Plurality), Dodgson rankings can be exactly the opposite of Borda [20] and Copeland rankings [18], while the winner of Kemeny or Slater elections can appear in any position of the Dodgson ranking [19].

More distantly related to our work is research that is concerned with exactly resolving hard-to-compute voting rules by heuristic methods. Typical examples include papers regarding the Kemeny rule [9] and the Slater rule [8]. Another more remotely related field of research is concerned with finding approximate, efficient representations of voting rules, by eliciting as little information as possible; this line of research employs techniques from learning theory [24].

Our results. In the context of approximating the Dodgson score, we devise a greedy algorithm for the Dodgson score which has an approximation ratio of H_{m-1}, where m is the number of alternatives and H is

[1] This would normally not happen in political elections, but can certainly be the case in many other settings. For instance, consider a group of agents trying to reach an agreement on a joint plan, when multiple alternative plans are available. Specifically, think of a group of investors deciding which company to invest in.

the harmonic number. We then propose a second algorithm that is based on solving a linear programming relaxation of the Dodgson score and has the same approximation ratio. Although the former algorithm gives us a better intuition into the combinatorial structure of the problem, we show that the latter has the advantage of being score monotonic, which is a desirable property from a social choice point of view. We further observe that it follows from the work of McCabe-Dansted [21] that the Dodgson score cannot be approximated within sublogarithmic factors by polynomial-time algorithms unless $\mathcal{P} = \mathcal{NP}$. We prove a more explicit inapproximability result of $(1/2 - \epsilon)\ln m$, under the assumption that problems in $\mathcal{NP}$ do not have algorithms running in quasi-polynomial time; this implies that the approximation ratio achieved by our algorithms is optimal up to a factor of 2.

Some of the results mentioned above [26, 27, 18, 19, 20] establish that there are sharp discrepancies between the Dodgson ranking and the rankings produced by other rank aggregation rules. Some of these rules (e.g., Borda and Copeland) are polynomial-time computable, so the corresponding results can be viewed as negative results regarding the approximability of the Dodgson ranking by polynomial-time algorithms. We show that the problem of distinguishing between whether a given alternative is the unique Dodgson winner or in the last $O(\sqrt{m})$ positions in any Dodgson ranking is $\mathcal{NP}$-hard. This theorem provides a complexity-theoretic explanation for some of the observed discrepancies, but in fact is much wider in scope as it applies to any efficiently computable rank aggregation rule.

The problem of calculating the Young score seems at first glance simple compared with the Dodgson score (we discuss in Section 6 why this seems so). Therefore, we found the following result quite surprising: it is $\mathcal{NP}$-hard to approximate the Young score within any factor. Specifically, we show that it is $\mathcal{NP}$-hard to distinguish between the case where the Young score of a given alternative is 0, and the case where the score is greater than 0. As a corollary we obtain an inapproximability result for the Young ranking. We also show that it is $\mathcal{NP}$-hard to approximate the dual Young score within $O(n^{1-\epsilon})$, for any constant $\epsilon > 0$.

Structure of the paper. In Section 2, we introduce some notations and definitions. In Section 3, we present our upper bounds for approximating the Dodgson score. We study the monotonicity properties of our algorithms in Section 4. In Section 5, we present our lower bounds for approximating the Dodgson score and ranking. Finally, in Section 6, we prove that the Young score, dual Young score, and Young ranking are inapproximable.

2 Preliminaries

Let $N = \{1, \ldots, n\}$ be a set of agents, and let A be the set of *alternatives*. We denote $|A| = m$, and denote the alternatives themselves by letters, such as $a \in A$. Indices referring to agents appear in superscript. Each agent $i \in N$ holds a binary relation R^i over A that satisfies irreflexivity, asymmetry, transitivity and totality. Informally, R^i is a ranking of the alternatives. Let $L = L(A)$ be the set of all rankings over A; we have that each $R^i \in L$. We denote $R^N = \langle R^1, \ldots, R^n \rangle \in L^N$, and refer to this vector as a *preference profile*. We may also use Q^i to denote the preferences of agent i, in cases where we want to distinguish between two different rankings R^i and Q^i. For sets of alternatives $B_1, B_2 \subseteq A$, we write $B_1 R^i B_2$ if for all $a \in B_1$ and $b \in B_2$, aR^ib.

Let $a, b \in A$. Denote

$$P(a, b) = \{i \in N : \ aR^ib\} \ .$$

We say that *a beats b in a pairwise election* if $|P(a, b)| > n/2$, that is, a is preferred to b by the majority of agents. A *Condorcet winner* is an alternative that beats every other alternative in a pairwise election.

The *Dodgson score* of a given alternative a^*, with respect to a given preference profile R^N, is the least number of exchanges between adjacent alternatives in R^N needed to make a^* a Condorcet winner. For instance, let $N = \{1, 2, 3\}$, $A = \{a, b, c\}$, and let R^N be given by:

R^1	R^2	R^3
a	b	a
b	a	c
c	c	b

In this example, the Dodgson score of a is 0 (a is a Condorcet winner), the score of b is 1, and the score of c is 3. Bartholdi et al. [2] have shown that computing the Dodgson score is an $\mathcal{NP}$-complete problem.

The *Young score* of a^* with respect to R^N is the size of the largest subset of agents for whom a^* is a Condorcet winner. This is the definition given by Young himself [33], and used in subsequent articles [29]. If, for every nonempty subset of agents, a^* is not a Condorcet winner, its Young score is 0. In the above example, the Young score of a is 3, the score of b is 1, and the score of c is 0.

Notice that, equivalently, a Young winner is an alternative such that one has to remove the minimum number of agents (we call this number the *dual* Young score) in order to make it a Condorcet winner. However, these two definitions are not equivalent in the context of approximation; we employ the former (original, prevalent) definition, but touch on the latter as well.

As the Young winner problem is known to be intractable [29], the Young score problem must also be hard; otherwise, we would be able to calculate the scores of all the alternatives efficiently, and identify the alternatives with minimum score.

3 Approximability of Dodgson Scores

We begin by presenting our approximation algorithms for the Dodgson score. Let us first introduce some notation.

Let $a^* \in A$ be a distinguished alternative, whose Dodgson score we wish to compute. Define the *deficit* of a^* with respect to $a \in A$, simply denoted def(a) when the identity of a^* is clear, as the number of additional agents that must rank a^* above a in order for a^* to beat a in a pairwise election. For instance, if four agents prefer a to a^* and only one agent prefers a^* to a, then def$(a) = 2$. If a^* beats a in a pairwise election (namely a^* is preferred by the majority of agents) then def$(a) = 0$. We say that alternatives $a \in A$ with def$(a) > 0$ are *alive*. Alternatives that are not alive, i.e., those with def$(a) = 0$, are *dead*.

3.1 A Greedy Algorithm

In this section we present a combinatorial, greedy algorithm for approximating the Dodgson score of a given alternative. Consider, once again, a special alternative a^*, and recall that a live alternative is one with a positive deficit. In each step, the algorithm selects the most cost-effective push of alternative a^* in the preference of some agent. The *cost-effectiveness* of pushing a^* in the preference of an agent $i \in N$ is the ratio between the total number of positions a^* is moved upwards in the preference of i *compared with the original profile* R^N, and the number of *currently* live alternatives that a^* overtakes as a result of this push. For example, if for some agent the algorithm raises a^* by one position where the alternative over which a^* is raised is dead, and later by a second position that causes a^* to overtake a live alternative, then the cost-effectiveness of the push is two and not one, since a^* ends up being two positions higher than its original position and only overtakes one live alternative.

After selecting the most cost-effective push, the algorithm decreases def(a) by one for each live alternative a that a^* overtakes. Alternatives $a \in A$ with def$(a) = 0$ become dead. The algorithm terminates when no live alternatives remain. The input and output of the algorithm are as before.

Greedy Algorithm:

1. Let A' be the set of live alternatives, namely those alternatives $a \in A$ with def$(a) > 0$.
2. While $A' \neq \emptyset$:
   - Perform the most cost-effective push, namely push a^* in the preferences of agent $i \in N$ in a way that minimizes the ratio between the *total* number of positions moved upwards in the preferences of i and the number of currently live alternatives overtaken by a^*.
   - Recalculate A'.
3. Return the number of exchanges performed.

By the definition of the algorithm, it is clear that it produces a profile where a^* is a Condorcet winner. It is important to notice that if a^* is initially a Condorcet winner then the algorithm calculates a score of zero, so as a voting rule the algorithm satisfies the Condorcet criterion.

Theorem 3.1. *For any input a^* and R^N with m alternatives, the greedy algorithm returns an H_{m-1}-approximation of the Dodgson score of a^*, where H_k is the k-th harmonic number.*

The proof of Theorem 3.1 uses the dual fitting technique, and is based on the connection between our problem and the Constrained Set Multicover problem [25].

Proof. We may view the problem of approximating the Dodgson score as the following covering problem with different covering requirements and constraints. The ground set is the set of live alternatives. For each live alternative $a \in A \setminus \{a^*\}$, its deficit def$(a)$ is in fact its covering requirement, i.e., the number of different sets it has to belong to in the final cover. For each agent $i \in N$ that ranks a^* in place r^i, we have a subcollection $\mathcal{S}^i$ consisting of the sets S^i_k for $k = 1, \ldots, r^i - 1$, where the set S^i_k contains the (initially) live alternatives that appear in positions $r^i - k$ to $r^i - 1$ in the preference of agent i. The set S^i_k has cost k. Now, the covering problem to be solved is the following. We wish to select at most one set from each of the different subcollections so that each alternative $a \in A \setminus \{a^*\}$ appears in at least def(a) sets and the total cost of the selected sets is minimized. The optimal cost is the Dodgson score of a^* and, hence, the cost of any approximate cover that satisfies the covering requirements and the constraints is an upper bound on the Dodgson score.

In terms of this covering problem, the greedy algorithm mentioned above can be thought of as working as follows. In each step, it selects the most cost-effective set where the cost-effectiveness of a set is defined as the ratio between the cost of the set and the number of live alternatives it covers that have not been previously covered by sets belonging to the same subcollection. For these live alternatives, the algorithm decreases their covering requirements at the end of the step. The algorithm terminates when all alternatives have died (i.e., their covering requirement has become zero). The output of the algorithm consists of the maximum-cost sets that were picked from each subcollection.

We remark that the covering problem we use is closely related to the Constrained Set Multicover problem considered in Rajagopalan and Vazirani [25] (see also [32, pp. 112–116]), with the additional constraint that at most one set has to be selected from each subcollection.

We now turn to the formal part of the proof. We find it convenient to formulate the Dodgson score problem as the following integer linear program.

$$\begin{aligned} \text{minimize} \quad & \sum_{i\in N}\sum_{k=1}^{r^i-1} k\cdot x_{S_k^i} \\ \text{subject to} \quad & \forall a\in A\setminus\{a^*\},\ \sum_{i\in N}\sum_{S\in\mathcal{S}^i:a\in S} x_S \geq \text{def}(a) \\ & \forall i\in N,\ \sum_{S\in\mathcal{S}^i} x_S \leq 1 \\ & x\in\{0,1\} \end{aligned}$$

The variable x_S associated with a set S denotes whether S is included in the solution ($x_S = 1$) or not ($x_S = 0$). We relax the integrality constraint in order to obtain a linear programming relaxation and we compute its dual linear program.

$$\begin{aligned} \text{maximize} \quad & \sum_{a\in A\setminus\{a^*\}} \text{def}(a)\cdot y_a - \sum_{i\in N} z^i \\ \text{subject to} \quad & \forall i\in N, k=1,\ldots,r^i-1,\ \sum_{a\in S_k^i} y_a - z^i \leq k \\ & \forall i\in N,\ z^i\geq 0 \\ & \forall a\in A\setminus\{a^*\},\ y_a\geq 0 \end{aligned}$$

For a set S that is picked by the algorithm to cover alternative $a\in A\setminus\{a^*\}$ for the j-th time (the j-th copy of a), we set $p(a,j)$ to be equal to the cost-effectiveness of S when it is picked. Informally, p distributes equally the cost of S among the copies of the live alternatives it covers. For each $i\in N$ and $k=1,\ldots,r^i-1$, we define a set T_k^i that, if S_k^i was picked by the algorithm, contains the alternatives in S_k^i that were alive when S_k^i was picked (namely those whose covering requirement decreased), and is empty otherwise. In the case where $a\in T_k^i$, we use $j(a,S_k^i)$ to denote the index of the copy of a that S_k^i covers when it is picked.

Now, we shall show that by setting

$$y_a = \frac{p(a,\text{def}(a))}{H_{m-1}}$$

for each alternative $a\in A\setminus\{a^*\}$, where H_{m-1} is the $m-1$ harmonic number, and

$$z^i = \frac{1}{H_{m-1}}\sum_{k=1}^{r^i-1}\sum_{a\in T_k^i}\left(p(a,\text{def}(a)) - p(a,j(a,S_k^i))\right)$$

for each agent $i\in N$, the constraints of the dual linear program are satisfied. The variables y_a are clearly non-negative. Since the algorithm always selects the most cost-effective set, it holds that $p(a,\text{def}(a)) \geq p(a,j)$ for every alternative a with $\text{def}(a)>0$ and $j\leq \text{def}(a)$ and, hence, z^i is non-negative.

In order to show that the first constraint of the dual linear program is also satisfied, consider an agent $i \in N$ and integer λ such that $1 \leq \lambda \leq r^i - 1$. We have

$$\begin{aligned}
\sum_{a \in S^i_\lambda} y_a - z^i &= \frac{1}{H_{m-1}} \left[\sum_{a \in S^i_\lambda} p(a, \mathrm{def}(a)) - \sum_{k=1}^{r^i-1} \sum_{a \in T^i_k} (p(a, \mathrm{def}(a)) - p(a, j(a, S^i_k))) \right] \\
&\leq \frac{1}{H_{m-1}} \left[\sum_{a \in S^i_\lambda} p(a, \mathrm{def}(a)) - \sum_{a \in T^i_\lambda} (p(a, \mathrm{def}(a)) - p(a, j(a, S^i_\lambda))) \right] \\
&= \frac{1}{H_{m-1}} \left[\sum_{a \in S^i_\lambda \setminus T^i_\lambda} p(a, \mathrm{def}(a)) + \sum_{a \in T^i_\lambda} p(a, j(a, S^i_\lambda)) \right] \qquad (1)
\end{aligned}$$

Let $s = |S^i_\lambda|$ and $s' = |S^i_\lambda \setminus T^i_\lambda|$. We number the alternatives of $S^i_\lambda \setminus T^i_\lambda$ in the order in which they die during the execution of the algorithm. Let this order be $a_1, a_2, \ldots, a_{s'}$. When a_t dies, we have that

$$p(a_t, \mathrm{def}(a_t)) \leq \frac{\lambda}{s - t + 1} \qquad (2)$$

since otherwise the set S^i_λ would have been used to cover a_t. We distinguish between the following two cases:

Case 1: If $T^i_\lambda = \emptyset$, using inequalities (1) and (2), we obtain

$$\sum_{a \in S^i_\lambda} y_a - z^i \leq \frac{1}{H_{m-1}} \sum_{t=1}^{s} p(a_t, \mathrm{def}(a_t)) \leq \frac{1}{H_{m-1}} \sum_{t=1}^{s} \frac{\lambda}{s-t+1} \leq \lambda \ .$$

Case 2: If $T^i_\lambda \neq \emptyset$ then $s' \leq s - 1$. We have

$$\sum_{a \in T^i_\lambda} p(a, j(a, S^i_\lambda)) = \lambda \ ,$$

since the cost of the set S^i_λ is equally distributed among the copies it covers. Also, using (2),

$$\sum_{a \in S^i_\lambda \setminus T^i_\lambda} p(a, \mathrm{def}(a)) = \sum_{t=1}^{s'} p(a_t, \mathrm{def}(a_t)) \leq \sum_{t=1}^{s-1} \frac{\lambda}{s-t+1} = \lambda(H_{s-1} - 1) \ .$$

So, inequality (1) again yields

$$\sum_{a \in S^i_\lambda} y_a - z^i \leq \lambda \ ,$$

implying that the constraints of the dual linear program are always satisfied.

Now, denote by OPT the optimal objective value of the integer linear program. By duality, we have that any feasible solution to the dual of its linear programming relaxation has objective value at most OPT.

Hence,

$$
\begin{aligned}
H_{m-1} \cdot \text{OPT} &\geq H_{m-1} \left(\sum_{a \in A \setminus \{a^*\}} \text{def}(a) \cdot y_a - \sum_{i \in N} z^i \right) \\
&= \sum_{a \in A \setminus \{a^*\}} \text{def}(a) \cdot p(a, \text{def}(a)) - \sum_{i \in N} \sum_{k=1}^{r^i - 1} \sum_{a \in T_k^i} \big(p(a, \text{def}(a)) - p(a, j(a, S_k^i)) \big) \\
&= \sum_{i \in N} \sum_{k=1}^{r^i - 1} \sum_{a \in T_k^i} p(a, j(a, S_k^i)) \\
&= \sum_{i \in N} \sum_{k \in \{1, \ldots, r^i - 1\} : T_k^i \neq \emptyset} k \ .
\end{aligned}
$$

The theorem follows since the last expression clearly upper-bounds the cost of the algorithm. □

3.2 An LP-based Algorithm

The analysis of the greedy algorithm suggests an LP-based algorithm for approximating the Dodgson score of an alternative a^* without explicitly providing a way to push a^* upwards in the preference of some agents so that a^* becomes the Condorcet winner. This algorithm uses the same LP relaxation of the Dodgson score that was used in the analysis of the greedy algorithm. The algorithm computes the optimal objective value, and returns this value multiplied by H_{m-1} as a score of the alternative a^*.

For completeness, we reformulate the LP in a more detailed form that takes the preference profile as a parameter as well; this shall be useful in the following section, where we discuss the monotonicity properties of the algorithm. Given a profile $R = R^N$ with a set of agents N and a set of m alternatives A, we denote by $r^i(R)$ the rank of alternative a^* in the preference of agent i. We use the notation def(a, R) for the deficit of a^* against an alternative a in the profile R. Recall that alternatives $a \in A \setminus \{a^*\}$ such that def$(a, R) > 0$ are said to be alive. For every agent $i \in N$ that ranks a^* in place $r^i(R)$, we denote by $\mathcal{S}^i(R)$ the subcollection that consists of the sets $S_k^i(R)$ for $k = 1, ..., r^i(R) - 1$, where the set $S_k^i(R)$ contains the live alternatives that appear in positions $r^i(R) - k$ to $r^i(R) - 1$ in the preference of agent i. We denote by $\mathcal{S}(R)$ the union of the subcollections $\mathcal{S}^i(R)$ for $i \in N$.

The LP-based algorithm uses the following LP relaxation of the Dodgson score of alternative a^* in the profile R:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i \in N} \sum_{k=1}^{r^i(R)-1} k \cdot x_{S_k^i(R)} \\
\text{subject to} \quad & \forall a \in A \setminus \{a^*\}, \sum_{i \in N} \sum_{S \in \mathcal{S}^i(R) : a \in S} x_S \geq \text{def}(a, R) \\
& \forall i \in N, \sum_{S \in \mathcal{S}^i(R)} x_S \leq 1 \\
& \forall S \in \mathcal{S}(R),\ 0 \leq x_S \leq 1
\end{aligned}
$$

We remark that, as it is the case with the greedy algorithm, if a^* is initially a Condorcet winner then the algorithm calculates a score of zero. The approximation ratio of the LP-based algorithm now follows easily. Indeed, by the analysis in the proof of Theorem 3.1, the cost of the solution obtained by the greedy algorithm is never smaller than the score returned by the LP-based algorithm. Hence, the Dodgson score of alternative a^* is never larger than the score returned by the LP-based algorithm. Furthermore, the Dodgson score of a^* induces an integral feasible solution for the LP relaxation and, hence, the score returned by the LP-based algorithm is at most the Dodgson score of alternative a^* multiplied by H_{m-1}. This is formalized in the following theorem.

Theorem 3.2. *For any input a^* and R^N with m alternatives, the LP-based algorithm returns an H_{m-1}-approximation of the Dodgson score of a^*, where H_k is the k-th harmonic number.*

4 Interlude: On the Desirability of Approximation Algorithms as Voting Rules

In Section 1 we suggested that an approximation algorithm for the Dodgson score should be considered as a new voting rule. This implies that our approximation algorithms should be compared according to two conceptually different, but not orthogonal, dimensions: their algorithmic properties and their social choice properties. From an algorithmic point of view, the greedy algorithm gives us a better sense of the combinatorial structure of the problem. In the sequel we suggest, however, that the LP-based algorithm has some desirable properties from a social choice point of view.

In most algorithmic mechanism design settings [23], such as combinatorial auctions or scheduling, one usually seeks approximation algorithms that are truthful, i.e., the agents cannot benefit by lying. However, the well-known Gibbard-Satterthwaite Theorem [13, 30] precludes voting rules that are both truthful and reasonable, in a sense. Therefore, other desiderata are looked for in voting rules.

We have been careful to emphasize that both the greedy algorithm and the LP-based algorithm satisfy the Condorcet property. Let us now consider the *monotonicity* property, one of the major desiderata on the basis of which voting rules are compared. Many different notions of monotonicity can be found in the literature; for our purposes, a (score-based) voting rule is *score monotonic* if and only if pushing an alternative in the preferences of the agents cannot worsen the score of the alternative, that is, increase it when a lower score is desirable (as in Dodgson), or decrease it when a higher score is desirable. All prominent score-based voting rules (e.g., positional scoring rules, Copeland, Maximin) are score monotonic; it is straightforward to see that the Dodgson and Young rules are score monotonic as well.

We first claim that our LP-based algorithm is score monotonic.

Theorem 4.1. *The LP-based algorithm is score monotonic.*

Proof. We will consider the two linear programs used by the LP-based algorithm for computing the score of an alternative a^* in a profile $R = R^N$ and another profile $\bar{R}$ which is obtained from R by pushing alternative a^* upwards in the preferences of some of the agents. Given an optimal solution x for the first LP, we will construct a feasible solution $\bar{x}$ for the second LP, which does not exceed the value of the first LP. This is a sufficient condition for the assertion of the theorem.

By the definition of profile $\bar{R}$, it holds that $r^i(R) \geq r^i(\bar{R})$ for every $i \in N$. We partition the subcollection $\mathcal{S}^i(R)$ into the following two disjoint subcollections:

$$\mathcal{S}^{i,1}(R) = \{S^i_k(R) : k = r^i(R) - r^i(\bar{R}) + 1, ..., r^i(R) - 1\}$$

and

$$\mathcal{S}^{i,2}(R) = \{S^i_k(R) : k = 1, ..., r^i(R) - r^i(\bar{R})\} \ .$$

For every $i \in N$, there is a one-to-one and onto correspondence between the sets in $\mathcal{S}^i(\bar{R})$ and the sets in $\mathcal{S}^{i,1}(R)$, where for $k = 1, ..., r^i(\bar{R})$ the set $S^i_k(\bar{R})$ of $\mathcal{S}^i(\bar{R})$ corresponds to the set $S^i_{k+r^i(R)-r^i(\bar{R})}(R)$ of $\mathcal{S}^{i,1}(R)$ and vice versa. The solution $\bar{x}$ for the second LP is constructed by simply setting

$$\bar{x}_{S^i_k(\bar{R})} = x_{S^i_{k+r^i(R)-r^i(\bar{R})}(R)}$$

for $i \in N$ and $k = 1, ..., r^i(\bar{R}) - 1$.

We will first prove that the solution $\bar{x}$ is a feasible solution for the second LP. The definition of the $\bar{x}$-variables clearly imply that the second and third sets of constraints are satisfied (since the solution x is feasible). Also, the first set of constraints is trivially satisfied for each alternative a with $\text{def}(a, \bar{R}) = 0$. Assume now that alternative a has $\text{def}(a, \bar{R}) > 0$. Let e^i_a be 1 if agent i ranks alternative a above a^* in R and below it in $\bar{R}$. Then, it can be easily seen that $\text{def}(a, R) = \text{def}(a, \bar{R}) + \sum_{i\in N} e^i_a > 0$. Hence, by the correspondence between the sets in $\mathcal{S}^i(\bar{R})$ and the sets in $\mathcal{S}^{i,1}(R)$, it follows that for every set $S \in \mathcal{S}^i(\bar{R})$ that contains alternative a, its corresponding set in $\mathcal{S}^{i,1}(R)$ also contains a. Using this observation and the definition of the solution $\bar{x}$, we obtain that

$$\begin{aligned}\sum_{i\in N} \sum_{S\in\mathcal{S}^i(\bar{R}):a\in S} \bar{x}_S &= \sum_{i\in N} \sum_{S\in\mathcal{S}^{i,1}(R):a\in S} x_S \\ &= \sum_{i\in N} \left(\sum_{S\in\mathcal{S}^i(R):a\in S} x_S - \sum_{S\in\mathcal{S}^{i,2}(R):a\in S} x_S \right) .\end{aligned}$$

Let $\alpha = \sum_{S\in\mathcal{S}^{i,2}(R):a\in S} x_S$. Observe that if $e^i_a = 0$, then no set $S \in \mathcal{S}^{i,2}(R)$ contains a, thus $\alpha = 0$. Otherwise, if $e^i_a = 1$, then the second constraint in the LP implies that $\alpha \le 1$. In other words, in any case α is upper-bounded by e^i_a. Using this observation and, additionally, the fact that $\text{def}(a, R) = \text{def}(a, \bar{R}) + \sum_{i\in N} e^i_a$, we conclude that

$$\sum_{i\in N} \sum_{S\in\mathcal{S}^i(\bar{R}):a\in S} \bar{x}_S \ge \sum_{i\in N} \sum_{S\in\mathcal{S}^i(R):a\in S} x_S - \sum_{i\in N} e^i_a \ge \text{def}(a, R) - \sum_{i\in N} e^i_a = \text{def}(a, \bar{R}) \ ,$$

as desired.

It is not hard to see that the objective of the second LP is upper bounded by the objective of the first one. Indeed, the coefficient of each $\bar{x}$-variable in the objective of the second LP is at most equal to the coefficient of the x-variable of the corresponding set in $\mathcal{S}^{i,1}(R)$ in the first LP, i.e., the variable $\bar{x}_{S^i_k(\bar{R})}$ is multiplied by k in the objective of the second LP while the variable $x_{S^i_{k+r^i(R)-r^i(\bar{R})}(R)}$ is multiplied by $k+r^i(R)-r^i(\bar{R}) \ge k$ in the first LP. □

In contrast, let us now consider the greedy algorithm. We design a preference profile and a push of a^* that demonstrate that the algorithm is not score monotonic. Agents 1 through 6 vote according to the profile R^N given in Figure 1(a). The positions marked by "." are placeholders for the rest of the alternatives, in some arbitrary order. Let $A' = \{a_1, \ldots, a_4\}$, $A'' = \{b_1, \ldots, b_{17}\}$. Notice that $\text{def}(a) = 1$ for all $a \in A'$, $\text{def}(b) = 0$ for all $b \in A''$. The optimal sequence of exchanges moves a^* all the way to the top of the preferences of agent 2, with a cost of seven. The greedy algorithm, given this preference profile, indeed chooses this sequence.

On the other hand, consider the profile $(R^1, R^2, Q^3, Q^4, Q^5, Q^6)$ given in Figure 1(b) (where the position of a^* was improved by two positions in the preferences of agents 3 through 6). First notice that the deficits

R^1	R^2	R^3	R^4	R^5	R^6
a_4	a_4				
a_3	a_3	a_4			
a_2	a_2	b_4	a_3		
a_1	a_1	b_5	b_9	a_2	
.	b_1	b_6	b_{10}	b_{13}	a_1
.	b_2	b_7	b_{11}	b_{14}	b_{16}
.	b_3	b_8	b_{12}	b_{15}	b_{17}
.	a^*	a^*	a^*	a^*	a^*
.	.	.	.	.	.
a^*	.	.	.	.	.

(a) Original Profile.

R^1	R^2	Q^3	Q^4	Q^5	Q^6
a_4	a_4				
a_3	a_3	a_4			
a_2	a_2	b_4	a_3		
a_1	a_1	b_5	b_9	a_2	
.	b_1	b_6	b_{10}	b_{13}	a_1
.	b_2	a^*	a^*	a^*	a^*
.	b_3	b_7	b_{11}	b_{14}	b_{16}
.	a^*	b_8	b_{12}	b_{15}	b_{17}
.	.	.	.	.	.
a^*	.	.	.	.	.

(b) Improvement of a^*.

Figure 1: The greedy algorithm is not score monotonic: an example.

have not changed compared to the profile R^N. The greedy algorithm would in fact push a^* to the top of the preferences of agents 6, 5, 4, and 3 (in this order), with a total cost of ten. Note that the optimal solution still has a cost of seven. In conclusion, the greedy algorithm is not score monotonic while the LP-based algorithm is score monotonic.

It should be mentioned that the following stronger notion of monotonicity is often considered in the literature: pushing a winning alternative in the preferences of the agents cannot harm it, that is, cannot make it lose the election. We say that a voting rule that satisfies this property is *monotonic*. Interestingly, Dodgson itself is not monotonic [6, 31], a fact that is considered by many to be a serious flaw. However, this does not preclude the existence of an approximation algorithm for the Dodgson score that is monotonic as a voting rule. Additionally, there are other prominent social choice properties that are often considered, e.g., homogeneity: a voting rule is said to be homogeneous if duplicating the electorate does not change the outcome of the election. The existence of algorithms that approximate the Dodgson score well and also satisfy additional social choice properties is addressed in the very recent paper of Caragiannis et al. [7], which builds upon the results of the current paper.

5 Lower Bounds for the Dodgson Rule

McCabe-Dansted [21] gives a polynomial-time reduction from the Minimum Dominating Set problem to the Dodgson score problem with the following property: given a graph G with k vertices, the reduction creates a preference profile with $n = \Theta(k)$ agents and $m = \Theta(k^4)$ alternatives, such that the size of the minimum dominating set of G is $\lfloor k^{-2}\mathrm{sc}_D(a^*)\rfloor$, where $\mathrm{sc}_D(a^*)$ is the Dodgson score of a distinguished alternative $a^* \in A$. Since the Minimum Dominating Set problem is known to be $\mathcal{NP}$-hard to approximate to within logarithmic factors [28], this implies that the Dodgson score problem is also hard to approximate to a factor of $\Omega(\log m)$. Due to the relation of Minimum Dominating Set to Minimum Set Cover, using an inapproximability result due to Feige [11], the explicit inapproximability bound can become $\left(\frac{1}{4} - \epsilon\right) \ln m$ under the assumption that problems in $\mathcal{NP}$ do not have quasi-polynomial-time algorithms.[2] This means that our algorithms are asymptotically optimal.

[2]Both inapproximability bounds have not been explicitly observed by McCabe-Dansted.

5.1 Inapproximability of the Dodgson Score

In the following, we present an alternative and more natural reduction directly from Minimum Set Cover that allows us to obtain a better explicit inapproximability bound. This bound implies that our greedy algorithm is optimal up to a factor of 2. The proof of the following theorem is given below.

Theorem 5.1. *There exists $\beta > 0$ such that it is $\mathcal{NP}$-hard to approximate the Dodgson score of a given alternative in an election with m alternatives to within a factor of $\beta \ln m$. Furthermore, for any $\epsilon > 0$, there is no polynomial-time $\left(\frac{1}{2} - \epsilon\right) \ln m$-approximation for the Dodgson score of a given alternative unless problems in $\mathcal{NP}$ of input size k have algorithms running in time $k^{\mathcal{O}(\log \log k)}$.*

Our inapproximability result for Dodgson score uses a reduction from Minimum Set Cover and the following well-known statements of its inapproximability.

Theorem 5.2 (Raz and Safra [28]). *There exists a constant $\alpha > 0$ such that, given an instance $(U, \mathcal{S})$ of Minimum Set Cover with $|U| = n$ and an integer $K \leq n$, it is $\mathcal{NP}$-hard to distinguish between the following two cases:*

- *$(U, \mathcal{S})$ has a cover of size at most K.*
- *Any cover of $(U, \mathcal{S})$ has size at least $\alpha K \ln n$.*

Theorem 5.3 (Feige [11]). *For any constant $\epsilon > 0$, given an instance $(U, \mathcal{S})$ of Minimum Set Cover with $|U| = n$ and an integer $K \leq n$, there is no polynomial-time algorithm that distinguishes between the following two cases:*

- *$(U, \mathcal{S})$ has a cover of size at most K, and*
- *Any cover of $(U, \mathcal{S})$ has size at least $(1 - \epsilon) K \ln n$,*

unless $\mathcal{NP} \subseteq DTIME(n^{O(\log \log n)})$.

The known inapproximability results for the problems that we use in our proofs are better understood by considering their relation to a generic NP-hard problem such as Satisfiability (see [25], Chapter 29). For example, what Theorem 5.2 essentially states is that there exists a polynomial-time reduction which, on input an instance ϕ of Satisfiability, constructs an instance $(U, \mathcal{S})$ of Minimum Set Cover with the following properties: if ϕ is satisfiable, $(U, \mathcal{S})$ has a cover of size at most K, and if ϕ is not satisfiable, any cover of $(U, \mathcal{S})$ has size at least $\alpha K \ln n$. This reduction implies that it is NP-hard to approximate Minimum Set Cover within a factor of $\alpha \ln n$. The interpretation of the remaining inapproximability results that are used or proved in the paper is similar.

Given an instance of Minimum Set Cover consisting of a set of n elements, a collection of sets over these elements and an integer $K \leq n$, we construct a preference profile with $m = (1 + \zeta)n + \lceil \alpha \zeta K n \ln n \rceil + 1$ alternatives and a specific alternative a^* in which we show that if we could distinguish in polynomial time between the following two cases:

- a^* has Dodgson score at most $(1 + \zeta) K n$, and
- a^* has Dodgson score at least $\alpha \zeta K n \ln n$,

then we could have distinguished between the two cases of Theorems 5.2 and 5.3 for the original Minimum Set Cover instance, contradicting the above inapproximability statements. Here, α is the inapproximability constant in Theorem 5.2 or 5.3 (in the latter $\alpha = 1 - \epsilon$), and ζ is an arbitrarily large positive constant. In this way, we obtain an inapproximability bound of $\frac{\alpha\zeta}{1+\zeta}\ln n$. Since $m = (1+\zeta)n + \lceil \alpha\zeta K n \ln n \rceil + 1$, it holds that $\ln n \geq \frac{1}{2}\ln m - \mathcal{O}(\ln\ln m)$, and hence the inapproximability bound for Dodgson score can be expressed in terms of the number of alternatives m as stated in Theorem 5.1.

We now present our reduction. Given an instance $(U, \mathcal{S})$ of Minimum Set Cover consisting of a set U of n elements, a collection $\mathcal{S}$ of sets $S_1, S_2, \ldots, S_{|\mathcal{S}|}$ and an integer $K \leq n$, we construct the following preference profile. There are the following alternatives:

- A set of n basic alternatives each corresponding to an element of U.
- A set Z of ζn alternatives where ζ is a positive constant.
- A set F of $\lceil \alpha\zeta K n \ln n \rceil$ alternatives, where α is the constant from Theorem 5.2.
- A specific alternative a^*.

There are the following $2|\mathcal{S}| + 1$ agents:

- A critical agent ℓ^i for each set $S_i \in \mathcal{S}$.
- An indifferent agent r^i for each set $S_i \in \mathcal{S}$.
- A special agent v^*.

The preferences of the agents are defined as follows:

- The special agent v^* ranks a^* in the first position of its preferences and the rest of the alternatives occupy the remaining positions in arbitrary order (i.e., $a^* \, R^{v^*} U \cup Z \cup F$).
- The critical agent ℓ^i ranks the basic alternatives corresponding to the elements of S_i in the first positions of its preference (in arbitrary order), next the alternatives of Z, next a^*, next the alternatives of F, and, in the last positions of its preference, the basic alternatives corresponding to the elements in $U \backslash S_i$ (i.e., $S_i \, R^{\ell^i} Z \, R^{\ell^i} a^* \, R^{\ell^i} F \, R^{\ell^i} U \backslash S_i$).
- We construct the ranking of the indifferent agents as follows. For each element u of U, we remove u from one of the sets of $\mathcal{S}$ (selecting arbitrarily among the sets of $\mathcal{S}$ containing u). Let $S'_1, S'_2, \ldots, S'_{|\mathcal{S}|}$ be the resulting sets and denote by $\mathcal{S}'$ their collection. The indifferent agent r^i ranks the basic alternatives corresponding to the elements in $U \setminus S'_i$ in the first positions of its preference, next the alternatives of F, next a^*, next the alternatives of Z and, in the last positions of its preference, the basic alternatives corresponding to elements in S'_i (if any)—i.e., $U \setminus S'_i \, R^{r^i} F \, R^{r^i} a^* \, R^{r^i} Z \, R^{r^i} S'_i$.

Clearly, a^* is preferred to any alternative in Z by the special agent and by the $|\mathcal{S}|$ indifferent agents, i.e., by a majority of agents. Similarly, a^* is preferred to any alternative in F by the special agent and by the $|\mathcal{S}|$ critical agents. Now, for each element of U, denote by f_u the number of sets in $\mathcal{S}$ that contain u. Then, a^* is preferred to u by the special agent, by the $|\mathcal{S}| - f_u$ critical agents corresponding to sets in $\mathcal{S}$ that do not contain u, and by the $f_u - 1$ indifferent agents corresponding to sets in $\mathcal{S}'$ that contain u (i.e., by $|\mathcal{S}|$ agents in total). Hence, a^* has a deficit of exactly 1 with respect to each of the alternatives in U.

Theorem 5.1 follows by the next two lemmata that give bounds on the Dodgson score of alternative a^* in the two cases of interest: when $(U, \mathcal{S})$ has a cover of size at most K (Lemma 5.4) and when any cover of $(U, \mathcal{S})$ has size at least $\alpha K \ln n$ (Lemma 5.5).

Lemma 5.4. *If $(U, \mathcal{S})$ has a cover of size K, then a^* has Dodgson score at most $(1+\zeta)Kn$.*

Proof. Let $H \subseteq \mathcal{S}$ be a cover for $(U, \mathcal{S})$ with $|H| = K$. By the definition of a cover, H covers all elements of U. Hence, by pushing a^* to the first position in the preference of the critical agent ℓ^i such that $S_i \in H$, a^* will decrease its deficit with respect to each of the basic alternatives by 1, and hence it will become a Condorcet winner. The total number of positions a^* rises is at most $|H| \cdot (|Z| + n) = (1+\zeta)nK$. $\square$

Lemma 5.5. *If any cover of $(U, \mathcal{S})$ has size at least $\alpha K \ln n$, then a^* has Dodgson score at least $\alpha\zeta K n \ln n$.*

Proof. We first assume that the minimum number of positions a^* has to rise in order to beat the basic alternatives and become a Condorcet winner includes raising a^* by at least $|F|$ positions in the ranking of some indifferent agent r^i. Hence, a^* rises $|F|$ positions in the preference of r^i in order to reach position $|U \backslash S_i'| + 1$ and at least n additional positions in order to beat the basic alternatives. Hence, its Dodgson score is at least $|F| + n \geq \alpha\zeta K n \ln n$.

Now, assume that the minimum number of positions a^* has to rise in order to beat the basic alternatives does not include raising a^* by at least $|F|$ positions in the ranking of some indifferent agent. We will show that if the Dodgson score of a^* is less than $\alpha\zeta K n \ln n$, then there exists a cover of $(U, \mathcal{S})$ of size less than $\alpha K \ln n$, contradicting the assumption of the lemma.

Let H be the set of critical agents such that a^* is pushed $|Z|$ positions in their preferences. In total, a^* rises $|H| \cdot |Z|$ positions in order to reach position $|S_i| + 1$ in the preferences of each critical agent ℓ^i belonging to H, plus at least n additional positions in order to decrease its deficit with respect to each of the alternatives in U by 1, that is, at least $\zeta|H|n + n$ positions in total. Hence, by denoting the Dodgson score of a^* by $sc_D(a^*)$, we have $|H| \leq \frac{1}{\zeta n} sc_D(a^*) - \frac{1}{\zeta} < \alpha K \ln n$. The proof is completed by observing that the union of the sets S_i for each critical agent ℓ^i belonging to H contains all the basic alternatives, i.e., H corresponds to a cover for $(U, \mathcal{S})$ of size less than $\alpha K \ln n$. $\square$

5.2 Inapproximability of Dodgson Rankings

A related question is the approximability of the Dodgson ranking, that is, the ranking of alternatives given by ordering them by nondecreasing Dodgson score. To the best of our knowledge, no rank aggregation function, which maps preference profiles to rankings of the alternatives, is known to provably produce rankings that are close to the Dodgson ranking [26, 27, 18, 19, 20] (see the survey of related work in Section 1).

Our next result establishes that efficient approximation algorithms are unlikely to exist unless $\mathcal{P} = \mathcal{NP}$, by proving that the problem of distinguishing between whether a given alternative is the unique Dodgson winner or in the last $O(\sqrt{m})$ positions is $\mathcal{NP}$-hard.

Theorem 5.6. *Given a preference profile with m alternatives and an alternative a^*, it is $\mathcal{NP}$-hard to decide whether a^* is a Dodgson winner or has rank at least $m - 6\sqrt{m}$ in any Dodgson ranking.*

Our proof uses a reduction from Minimum Vertex Cover in 3-regular graphs, and exploits a very weak statement concerning its inapproximability (marginally stronger than its $\mathcal{NP}$-hardness) that follows from the work of Berman and Karpinski [3]. The approach is similar to the proof of Theorem 5.1, albeit considerably more involved. This result provides a complexity-theoretic explanation for the sharp discrepancies observed in the Social Choice Theory literature when comparing Dodgson elections with simpler, efficiently computable, voting rules.

Our proof relies on the inapproximability of Minimum Vertex Cover in 3-regular graphs [3]. In particular, our reduction uses the following very weak variant of an inapproximability result presented in [3].

Theorem 5.7 (Berman and Karpinski [3], see also [16]). *Given a 3-regular graph G with $n = 22t$ nodes for some integer $t > 0$ and an integer K in $[n/2, n-6]$, it is $\mathcal{NP}$-hard to distinguish between the following two cases:*

- *G has a vertex cover of size at most K.*
- *Any vertex cover of G has size at least $K+6$.*

Given an instance of Minimum Vertex Cover consisting of a 3-regular graph G with $n = 22t$ nodes $v_0, v_1, \ldots, v_{n-1}$ and an integer $K \in [n/2, n-6]$, we construct in polynomial time a preference profile in which if we could distinguish whether a particular alternative is a Dodgson winner or not very far from the last position in any Dodgson ranking, then we could also distinguish between the two cases mentioned in Theorem 5.7 for the original Minimum Vertex Cover instance. The Dodgson election has the following sets of alternatives:

- A special alternative a^*.
- A set F of $4Kn/11 + 3n/2$ alternatives. These alternatives are partitioned into n disjoint blocks F_0, F_1, ..., F_{n-1} so that each block contains either $\lceil 4K/11 + 3/2 \rceil$ or $\lfloor 4K/11 + 3/2 \rfloor$ alternatives.
- A set A of n alternatives $a_0, a_1, \ldots, a_{n-1}$.
- An alternative u_j for each edge e_j of G. Let U be the set of these $3n/2$ alternatives. We denote by S_i the set of the three alternatives of U that correspond to the edges of G which are incident to node v_i.

For each node v_i of G, there are two agents: one left agent ℓ^i and one right agent r^i. The preferences of the left agent ℓ^i are as follows:

- The three alternatives of S_i are ranked by agent ℓ^i in the first three positions of its preference (in arbitrary order).
- From position 4 to position $4n/11 + 3$, agent ℓ^i ranks the alternatives $a_i, a_{(i+1) \bmod n}, \ldots, a_{(i+4n/11-1) \bmod n}$ in this order.
- In position $4n/11 + 4$, agent ℓ^i ranks a^*.
- From position $4n/11 + 5$ to position $4Kn/11 + 41n/22 + 4$, agent ℓ^i ranks the alternatives of F in the following order. The alternatives of F_i are ranked in positions from $4n/11 + 5$ to $4n/11 + 4 + |F_i|$ (in arbitrary order). Next, agent ℓ^i ranks the alternatives of sets $F_0, \ldots, F_{i-1}, F_{i+1}, \ldots, F_{n-1}$ in this order (the relative order of the alternatives of the same block is arbitrary).
- From position $4Kn/11 + 41n/22 + 5$ to position $4Kn/11 + 5n/2 + 4$, agent ℓ^i ranks the alternatives $a_{(i+4n/11) \bmod n}, a_{(i+4n/11+1) \bmod n}, \ldots, a_{(i-1) \bmod n}$ in this order.
- In the last $3n/2 - 3$ positions, agent ℓ^i ranks the alternatives of $U \setminus S_i$ (in arbitrary order).

The preferences of the right agent r^i are as follows:

- In the first $3n/2 - 3$ positions, agent r^i ranks the alternatives of $U \setminus S_i$ in reverse relative order to the order ℓ^i ranks them.

- From position $3n/2 - 2$ to position $4Kn/11 + 3n - |F_i| - 3$, agent r^i ranks the alternatives of the blocks $F_{n-1}, F_{n-2}, \dots, F_{i+1}, F_{i-1}, \dots, F_0$ in this order so that the alternatives of block F_j are ranked in reverse relative order to the order ℓ^i ranks them.

- From position $4Kn/11 + 3n - |F_i| - 2$ to position $4Kn/11 + 40n/11 - |F_i| - 5$, agent r^i ranks the alternatives $a_{(n-i-1) \bmod n}, a_{(n-i-2) \bmod n}, \dots, a_{(4n/11-i+2) \bmod n}$ in this order.

- In position $4Kn/11 + 40n/11 - |F_i| - 4$, agent r^i ranks a^*.

- From position $4Kn/11 + 40n/11 - |F_i| - 3$ to position $4Kn/11 + 40n/11 - 4$, agent r^i ranks the alternatives of F_i in reverse relative order to the order ℓ^i ranks them.

- From position $4Kn/11 + 40n/11 - 3$ to position $4Kn/11 + 4n - 2$, agent r^i ranks the alternatives $a_{(4n/11-i+1) \bmod n}$, $a_{(4n/11-i) \bmod n}$, $\dots$, $a_{(n-i) \bmod n}$ in this order.

- The three alternatives of S_i are ranked in the last three positions in the preference of agent r^i, in reverse relative order to the order ℓ^i ranks them.

We observe that a^* beats all alternatives but the alternatives of U. In particular, a^* is preferred to each alternative of F by $n+1$ agents. Specifically, a^* is ranked above an alternative belonging to the block F_i by the n left agents and by the right agent r^i. Also, the alternative a_i is ranked below a^* by the $7n/11$ left agents $\ell^{(i+1) \bmod n}, \ell^{(i+2) \bmod n}, \dots, \ell^{(i+7n/11-1) \bmod n}$ and by the $4n/11+2$ right agents $r^{(i+7n/11-1) \bmod n}, r^{(i+7n/11) \bmod n}, \dots, r^i$. Hence, a^* beats all alternatives in set A as well since it is ranked above each of them by $n+2$ agents. Also, a^* is ranked above the alternative u_j corresponding to the edge e_j of G by the left agents ℓ^i and $\ell^{i'}$ and by all right agents besides r^i and $r^{i'}$ so that nodes v_i and $v_{i'}$ are the endpoints of edge e_j in G. Hence, a^* has a deficit of 1 with respect to each of the alternatives in U.

We also observe that the alternatives in F beat each alternative in A. Note that in each agent where a^* is preferred to an alternative in A besides by the right agent r^i, an alternative of block F_i is also preferred to the alternative in A. Hence, each alternative of F beats each alternative of A since it is ranked above it by $n+1$ agents. Furthermore, similarly to a^*, each alternative in F is preferred to each alternative of U by n agents. Also, when an alternative f of F is ranked above another alternative f' of F by agent ℓ^i, f' is ranked above f by agent r^i. Hence, an alternative of F has a deficit of 1 with respect to U and each other alternative in F, and a deficit of 2 with respect to a^*.

Furthermore, observe that each alternative in A is ranked above the alternative u_j corresponding to the edge e_j of G by the left agents ℓ^i and $\ell^{i'}$ and by all right agents besides r^i and $r^{i'}$ so that nodes v_i and $v_{i'}$ are the endpoints of edge e_j in G, i.e., by n agents. Also, when an alternative a of A is preferred to another alternative a' of A by agent ℓ^i, a' is preferred to a by agent r^i. Hence, an alternative in A has a deficit of 1 with respect to each of the alternatives in U and F and each other alternative in A, and a deficit of 3 with respect to a^*. This immediately yields that the Dodgson score of each alternative in U is at least $4Kn/11 + 4n + 2$.

Similarly, when an alternative u of U is preferred to another alternative u' of U by agent ℓ^i, u' is preferred to u by agent r^i. Hence, an alternative in U has a deficit of 1 with respect to each of the alternatives in A and F, each other alternative in U, and a^*. This immediately yields that the Dodgson score of each alternative in U is at least $4Kn/11 + 4n$.

The next lemma gives upper and lower bounds on the Dodgson score of the alternatives in F.

Lemma 5.8. *Each alternative in F has Dodgson score between $4Kn/11+3n+1$ and $4Kn/11+37n/11+2K/11+3/4$.*

Proof. Since each alternative in F has a deficit of 1 with respect to each alternative in U and each other alternative in F, and a deficit of 2 with respect to a^*, its Dodgson score is at least $|U| + |F| - 1 + 2 = 4Kn/11 + 3n + 1$.

Now, consider an alternative f belonging to block F_i. f is at distance at most

$$\lfloor \frac{|F_i| - 1}{2} + 1 \rfloor \leq \frac{|F_i| + 1}{2} \leq \frac{\lceil 4K/11 + 3/2 \rceil + 1}{2} \leq 2K/11 + 7/4$$

from a^* in the preferences of either the left agent ℓ^i or the right agent r^i. Hence, by raising f at most $2K/11 + 7/4$ positions in the preferences of either ℓ^i or r^i, its deficit with respect to a^* decreases by 1. Consider a left agent $\ell^{i'}$ with $i' \neq i$ and let F' be the subset of alternatives in F that are higher than f in the preferences of $\ell^{i'}$. By pushing f to the first position in the preferences of agent $\ell^{i'}$ (i.e., $4n/11 + 3 + |F'|$ additional positions), f decreases its deficit by 1 with respect to each alternative of F' and a^*, as well as with respect to the three alternatives of $S_{i'}$ in the first three positions in the preferences of $\ell^{i'}$. Now, consider the right agent $r^{i'}$. In the preferences of $r^{i'}$, f is ranked higher than the alternatives in F' and lower than the alternatives in $F \backslash F' - \{f\}$. Hence, by pushing f to the first position in the preferences of agent $r^{i'}$ (i.e., $|F \backslash F' - \{f\}| + 3n/2 - 3 = 4Kn/11 - |F'| + 3n - 4$ additional positions), f decreases its deficit by 1 with respect to each alternative of $F \backslash F' - \{f\}$ as well the alternatives of $U \backslash S_{i'}$ in the first $3n/2 - 3$ positions in the preferences of r^i. Hence, by pushing $4Kn/11 + 37n/11 + 2K/11 + 3/4$ positions, f becomes a Condorcet winner. □

The next two lemmata give bounds on the Dodgson score of alternative a^* in the two cases of interest: when G has a vertex cover of size at most K (Lemma 5.9), and when any vertex cover of G has size at least $K + 6$ (Lemma 5.10).

Lemma 5.9. *If G has a vertex cover of size at most K, then the Dodgson score of a^* is less than $4Kn/11 + 3n$.*

Proof. Let $H \subseteq V$ be a vertex cover of G with $|H| = K$. By the definition of the vertex cover, H covers all edges of G and this implies that $\cup_{i:v_i \in H} S_i = U$. Hence, by pushing a^* to the first position in the preferences of each of the K left agents ℓ^i such that $v_i \in H$, a^* decreases it deficit with respect to each of the alternatives in U by 1, and becomes a Condorcet winner. The total number of positions a^* rises is $K(4n/11 + 3) < 4Kn/11 + 3n$. The last inequality is true since $K < n$. □

Lemma 5.10. *If any vertex cover of G has size at least $K + 6$, then the Dodgson score of a^* is larger than $4Kn/11 + 37n/11 + 2K/11 + 3/4$.*

Proof. First assume that the minimum sequence of exchanges that makes a^* beat the alternatives of U and become a Condorcet winner includes pushing a^* to one of the first $3n/2 - 3$ positions in the preferences of some right agent r^i. Certainly, not all alternatives of U are beaten in this way since the three alternatives of S_i are ranked below a^* by agent r^i. So, in order to beat the remaining 3 alternatives of S_i, a^* has to either be pushed to one of the first three positions of a left agent or to one of the first $3n/2 - 3$ positions of another right agent $r^{i'}$ with $i' \neq i$. Hence, a^* must be first pushed to position $3n/2 - 2$ of agent r^i (i.e., $|F \backslash F_i| + 7n/11 - 2$ positions), to position 4 of a left agent (i.e., $4n/11$ additional positions) or to position $3n/2 - 2$ of agent $r^{i'}$ (i.e., $|F \backslash F_{i'}| + 7n/11 - 2$ additional positions), and then rise at least $3n/2$ additional positions in order to beat all alternatives of U. In total, a^* rises at least

$$\begin{aligned}
& |F\backslash F_i| + 7n/11 - 2 + \min\{4n/11, |F\backslash F_{i'}| + 7n/11 - 2\} + 3n/2 \\
\geq\ & |F| - |F_i| + 5n/2 - 2 \\
\geq\ & 4Kn/11 + 4n - \lceil 4K/11 + 3/2 \rceil - 2 \\
\geq\ & 4Kn/11 + 4n - 4K/11 - 9/2 \\
=\ & 4Kn/11 + 37n/11 + n/11 + 6n/11 - 4K/11 - 9/2 \\
\geq\ & 4Kn/11 + 37n/11 + 22/11 + 6(K+6)/11 - 4K/11 - 9/2 \\
>\ & 4Kn/11 + 37n/11 + 2K/11 + 3/4
\end{aligned}$$

positions. The fourth inequality holds since $n \geq 22$ and $n \geq K + 6$.

Now, assume that the minimum sequence of exchanges for making a^* a Condorcet winner does not include raising a^* to any of the first $3n/2 - 3$ positions of any right agent. We will show that if a^* has Dodgson score at most $4Kn/11 + 37n/11 + 2K/11 + 3/4$, then G has a vertex cover of size less than $K + 6$, contradicting the assumption of the lemma.

Let H be the set of left agents where a^* rises to one of the first three positions in order to beat all the alternatives of U. In total, a^* rises $4|H|n/11$ positions in order to reach position 4 in the preferences of each of the agents in H plus at least $3n/2$ additional positions in order to decrease its deficit with respect to the alternatives in U by at least 1, i.e., at least $4|H|n/11 + 3n/2$ positions in total. Hence, by denoting the Dodgson score of a^* by $sc_D(a^*)$, we have $|H| \leq \frac{11}{4n}(sc_D(a^*) - 3n/2)$.

Since $\cup_{i:\ell^i \in H} S_i = U$, the set of nodes of G consisting of nodes v_i such that agent ℓ^i belongs to H is a vertex cover of G of size $|H|$. Assuming that the Dodgson score of a^* is at most $4Kn/11 + 37n/11 + 2K/11 + 3/4$, we have

$$\begin{aligned}
|H| &\leq \frac{11}{4n}(sc_D(a^*) - 3n/2) \\
&\leq \frac{11}{4n}\left(4Kn/11 + 41n/22 + 2K/11 + 3/4\right) \\
&< K + 6\ ,
\end{aligned}$$

where the last inequality follows since $K \leq n - 6$. □

By Lemmata 5.8, 5.9, and 5.10, we obtain that if G has a vertex cover of size at most K, then a^* is the unique Dodgson winner, while if every vertex cover of G has size at least $K + 6$, then a^* is below all alternatives in F in any Dodgson ranking. Denote by $m = |F| + |A| + |U| + 1 = 4Kn/11 + 4n + 1$ the total number of alternatives. Then, the rank of a^* in the second case is at least

$$\begin{aligned}
|F| + 1 &= 4Kn/11 + 3n/2 + 1 = m - 5n/2 = m - \sqrt{25n^2/4} \\
&\geq m - \sqrt{25nK/2} \geq m - 6\sqrt{4Kn/11 + 4n + 1} = m - 6\sqrt{m}\ ,
\end{aligned}$$

where the first inequality follows since $K \geq n/2$. By Theorem 5.7, we obtain the desired result. □

An example. We present an example of the construction in the proof of Theorem 5.6. Consider an instance of Minimum Vertex Cover with the 22-node 3-regular graph of Figure 2, and $K = 12$.

The corresponding preference profile has 185 alternatives and 44 agents. In particular, the set F has 129 alternatives $f_0, f_1, \ldots, f_{128}$, which are partitioned into 22 blocks as follows. Block F_0 contains the

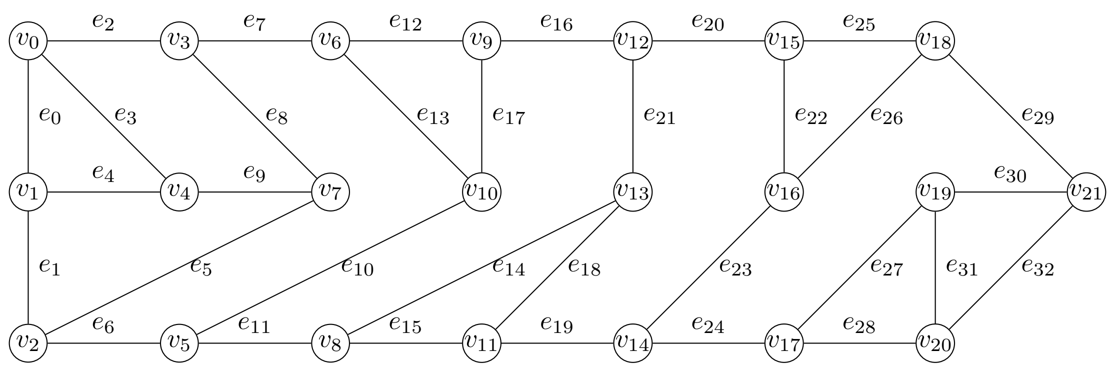

Figure 2: A 3-regular graph with 22 nodes.

six alternatives $f_0, f_1, \dots, f_5$, block F_1 contains the six alternatives $f_6, \dots, f_{11}$, $\dots$, block F_{18} contains the six alternatives $f_{108}, \dots, f_{113}$, block F_{19} contains the five alternatives $f_{114}, \dots, f_{118}$, $\dots$, and block F_{21} contains the alternatives $f_{124}, \dots, f_{128}$. The set A has 22 alternatives $a_0, \dots, a_{21}$. The set U has 33 alternatives $u_0, \dots, u_{32}$, one alternative for each edge of the graph. The agents are partitioned into 22 left agents and 22 right agents. In order to compute the preferences of an agent, say agent ℓ^{17}, we first compute the set S_{17}, which contains the alternatives corresponding to the edges incident to node v_{17} of the graph, i.e., $S_{17} = \{u_{24}, u_{27}, u_{28}\}$. Now, the preferences of agent ℓ^{17} are:

$$S_{17}\, R^{\ell^{17}} a_{17}\, R^{\ell^{17}} a_{18}\, R^{\ell^{17}} \dots R^{\ell^{17}} a_1\, R^{\ell^{17}} a_2\, R^{\ell^{17}} a^*\, R^{\ell^{17}} F_{17}\, R^{\ell^{17}} F_0\, R^{\ell^{17}} \dots F_{16}$$
$$R^{\ell^{17}} F_{18}\, R^{\ell^{17}} \dots R^{\ell^{17}} F_{21}\, R^{\ell^{17}} a_3\, R^{\ell^{17}} \dots R^{\ell^{17}} a_{16}\, R^{\ell^{17}} U \setminus S_{17}\ .$$

Similarly, the preferences of agent r^{17} are:

$$\overleftarrow{U \setminus S_{17}}\ R^{\ell^{17}}\ \overleftarrow{F_{21}}\ R^{\ell^{17}}\ \overleftarrow{F_{20}}\ R^{\ell^{17}} \dots R^{\ell^{17}}\ \overleftarrow{F_{18}}\ R^{\ell^{17}}\ \overleftarrow{F_{16}}\ R^{\ell^{17}} \dots R^{\ell^{17}}\ \overleftarrow{F_0}\ R^{\ell^{17}}$$
$$a_4\, R^{\ell^{17}} a_3\, R^{\ell^{17}} \dots R^{\ell^{17}} a_{15}\, R^{\ell^{17}} a^*\, R^{\ell^{17}}\ \overleftarrow{F_{17}}\ a_{16}\, R^{\ell^{17}} a_5\, R^{\ell^{17}}\ \overleftarrow{S_{17}}\ ,$$

where the symbol $\leftarrow$ on top of a set of alternatives is used to denote that their order in the preferences of r^{17} is the reverse of the order ℓ^{17} ranks them.

6 Inapproximability of Young Scores and Rankings

Recall that the Young score of a given alternative $a^* \in A$ is the size of the largest subset of agents for which a^* is a Condorcet winner.

It is straightforward to obtain a simple ILP for the Young score problem. As before, let $a^* \in A$ be the alternative whose Young score we wish to compute. Let the variables of the program be $x^i \in \{0,1\}$ for all $i \in N$; $x^i = 1$ iff agent i is included in the subset of agents for a^*. Define constants $e^i_a \in \{-1, 1\}$ for all $i \in N$ and $a \in A \setminus \{a^*\}$, which depend on the given preference profile; $e^i_a = 1$ iff agent i ranks a^* higher than a. The ILP that computes the Young score of a^* is given by:

$$\begin{aligned}\text{maximize} \quad & \sum_{i \in N} x^i \\ \text{subject to} \quad & \forall a \in A \setminus \{a^*\},\ \sum_{i \in N} x^i e_a^i \geq 1 \\ & \forall i \in N,\ x^i \in \{0,1\}\end{aligned} \tag{3}$$

The ILP (3) for the Young score is seemingly simpler than the one for the Dodgson score. This might seem to indicate that the problem can be easily approximated by similar techniques. Therefore, the following result is quite surprising.

Theorem 6.1. *It is $\mathcal{NP}$-hard to approximate the Young score by any factor.*

This result becomes more self-evident when we notice that the Young score has the rare property of being nonmonotonic as an optimization problem, in the following sense: given a subset of agents that make a^* a Condorcet winner, it is not necessarily the case that a smaller subset of the agents would satisfy the same property. This stands in contrast to many approximable optimization problems, in which a solution which is worse than a valid solution is also a valid solution. Consider the Set Cover problem, for instance: if one adds more subsets to a valid cover, one obtains a valid cover. The same goes for the Dodgson score problem: if a sequence of exchanges makes a^* a Condorcet winner, introducing more exchanges on top of the existing ones would not undo this fact.

In order to prove the inapproximability of the Young score, we define the following problem.

$\mathrm{NonEmptySubset}$
Instance: An alternative a^*, and a preference profile $R^N \in L^N$.
Question: Is there a nonempty subset of agents $C \subseteq N$, $C \neq \emptyset$, for which a^* is a Condorcet winner?

To prove Theorem 6.1, it is sufficient to prove that $\mathrm{NonEmptySubset}$ is $\mathcal{NP}$-hard. Indeed, this implies that it is $\mathcal{NP}$-hard to distinguish whether the Young score of a given alternative is zero or greater than zero, which directly entails that the score cannot be approximated.

Lemma 6.2. $\mathrm{NonEmptySubset}$ *is $\mathcal{NP}$-complete.*

Proof. The problem is clearly in $\mathcal{NP}$; a witness is given by a nonempty set of agents for which a^* is a Condorcet winner.

In order to show $\mathcal{NP}$-hardness, we present a polynomial-time reduction from the $\mathcal{NP}$-hard Exact Cover by 3-Sets (X3C) problem [12] to our problem. An instance of the X3C problem includes a finite set of elements U, $|U| = n$ (where n is divisible by 3), and a collection $\mathcal{S}$ of 3-element subsets of U, $\mathcal{S} = \{S_1, \ldots, S_k\}$, such that for every $1 \leq i \leq k$, $S_i \subseteq U$ and $|S_i| = 3$. The question is whether the collection $\mathcal{S}$ contains an *exact cover* for U, i.e., a subcollection $\mathcal{S}^* \subseteq \mathcal{S}$ of size $n/3$ such that every element of U occurs in exactly one subset in $\mathcal{S}$.

We next give the details of the reduction from X3C to $\mathrm{NonEmptySubset}$. Given an instance of X3C, defined by the set U and a collection of 3-element sets $\mathcal{S}$, we construct the following instance of $\mathrm{NonEmptySubset}$.

Define the set of alternatives as $A = U \cup \{a\} \cup \{a^*\}$. Let the set of agents be $N = N' \cup N''$, where N' and N'' are defined as follows. The set N' is composed of k agents, corresponding to the k subsets in $\mathcal{S}$, such that for all $i \in N'$, agent i prefers the alternatives in $U \setminus S_i$ to a^*, and prefers a^* to all the alternatives in $S_i \cup \{a\}$ (i.e., $U \setminus S_i \; R^i \; a^* R^i \; S_i \cup \{a\}$).

Subset N'' is composed of $\frac{n}{3}-1$ agents who prefer a to a^* and a^* to U (i.e., for all $i \in N''$, $a \; R^i \; a^* \; R^i \; U$).

We next show that there is an exact cover in the given instance iff there is nonempty subset of agents for which a^* is a Condorcet winner in the constructed instance.

Sufficiency: Let $\mathcal{S}^*$ be an exact cover by 3-sets of U, and let $N^* \subseteq N'$ be the subset of agents corresponding to the $\frac{n}{3}$ subsets $S_i \in \mathcal{S}^*$. We show that a^* is a Condorcet winner for $C = N^* \cup N''$. Since $\mathcal{S}^*$ is an exact cover, for all $b \in U$ there exists exactly one agent in N^* that prefers a^* to b and $\frac{n}{3} - 1$ agents in N^* that prefer b to a^*. In addition, all $\frac{n}{3} - 1$ agents in N'' prefer a^* to b. Therefore, a^* beats b in a pairwise election.

It remains to show that a^* beats a in a pairwise election. This is true since all $\frac{n}{3}$ agents in N^* prefer a^* to a, and there are only $\frac{n}{3} - 1$ agents in N'' who prefer a to a^*. It follows that a^* is a Condorcet winner for $N^* \cup N''$.

Necessity: Assume the given instance of X3C has no exact cover. We have to show that there is no subset of agents for which a^* is a Condorcet winner. Let $C \subseteq N$, $C \neq \emptyset$, and let $N^* = C \cap N'$. We distinguish among three cases.

Case 1: $|N^*| = 0$. It must hold that $C \cap N'' \neq \emptyset$. In this case, a^* loses to a in a pairwise election, since all the agents in N'' prefer a to a^*.

Case 2: $0 < |N^*| \leq \frac{n}{3}$. Since there is no exact cover, the corresponding sets S_i cannot cover U. Thus there exists $b \in U$ that is ranked higher than a^* by all agents in N^*. In order for a^* to beat b in a pairwise election, C must include at least $|N^*|+1$ agents from N''. However, this means that a beats a^* in a pairwise election (since a is ranked lower than a^* by $|N^*|$ agents, and higher than a^* by at least $|N^*| + 1$ agents). It follows that a^* is not a Condorcet winner for C.

Case 3: $|N^*| > \frac{n}{3}$. Let us award each alternative $b \in A \setminus \{a^*\}$ a point for each agent that ranks it above a^*, and subtract a point for each agent that ranks it below a^*. a^* is a Condorcet winner iff the score of every other alternative, counted this way, is negative. This implies that a^* is a Condorcet winner only if for every subset $B \subseteq A$ of alternatives, the total score of the alternatives in B is at most $-|B|$.

We shall calculate the total score of the alternatives in U from the agents in N^*. Every agent in N^* prefers a^* to 3 alternatives in U and prefers $n-3$ alternatives in U to a^*. Thus, every agent in N^* contributes $(n-3) - 3 = n - 6$ points to the total score of U. Summing over all the agents in N^*, we have that the total score of U from N^* is $|N^*|(n-6)$. By $|N^*| > \frac{n}{3}$, we have that

$$|N^*|(n-6) \geq \left((\frac{n}{3} - 1) + 2 \right) (n-6) = (\frac{n}{3} - 1)n - 6 \; .$$

Recall that every agent in N'' prefers a^* to all alternatives in U. However, since $|N''| = \frac{n}{3} - 1$, agents from N'' can only subtract $(\frac{n}{3} - 1)n$ from the total score of U. We conclude that the total score of U is at least -6. Since we can assume that $|U| = n > 6$,[3] a^* cannot beat all the alternatives in U in pairwise elections. This concludes the proof. □

A short discussion is in order. Theorem 6.1 states that the Young score cannot be efficiently approximated to any factor. The proof shows that, in fact, it is impossible to efficiently distinguish between a zero and a nonzero score. However, the proof actually shows more: it constructs a family of instances, where it is hard to distinguish between a score of zero and almost $2m/3$. Now, if one looks at an alternative formulation of the Young score problem where all the scores are scaled by an additive constant, it is no longer true that

[3] X3C is obviously tractable for a constant n, as one can examine all the families $\mathcal{S}' \subseteq \mathcal{S}$ of constant size in polynomial time.

it is hard to approximate the score to *any* factor; however, the proof still shows that it is hard to approximate the Young score, even under this alternative formulation, to a factor of $\mathcal{O}(m)$.

The strong inapproximability result for the Young score intuitively implies that the Young ranking cannot be approximated. The following corollary, whose proof is a straightforward variation on the proof of Lemma 6.2, shows that this is indeed the case. It can be viewed as an analog of Theorem 5.6 for Young.

Corollary 6.3. *For any constant $\epsilon > 0$, given a preference profile with m alternatives and an alternative a^*, it is $\mathcal{NP}$-hard to decide whether a^* has rank $\mathcal{O}(m^\epsilon)$ or is ranked in place m (that is, ranked last) in any Young ranking.*

Proof. Let $\epsilon > 0$ be a constant. We perform the same reduction as before, with the following differences. Let A' be the set of alternatives constructed in the reduction of Lemma 6.2, and $m' = |A'|$; we add a set B of $(m')^{1/\epsilon}$ additional alternatives, i.e., $A = A' \cup B$, $m = |A| = m' + (m')^{1/\epsilon}$. The set of agents is $N' \cup N'' \cup N^*$, the preferences of N' and N'' restricted to A' are as before, and all these agents rank B at the bottom. All the agents in N^* rank a^* last; for each $b \in A' \setminus \{a^*\}$, there is $i \in N^*$ that ranks b first and B just above a^*, i.e.,

$$b \; R^i \; A' \setminus \{a^*, b\} \; R^i \; B \; R^i \; a^* \quad .$$

For each $c \in B$, there is $i \in N^*$ that ranks c first and the rest of B just above a^*, namely

$$c \; R^i \; A' \setminus \{a^*\} \; R^i \; B \setminus \{c\} \; R^i \; a^* \quad .$$

Notice that the Young score of the alternatives in $A' \setminus \{a^*\}$ is at least one. The Young score of any alternative $c \in B$ is exactly one, since exactly one agent (in N^*) does not rank $A' \setminus \{a^*\}$ above c. Now, if there is an exact 3-cover, then the Young score of a^* is at least $2n/3 - 1$ (according to the proof of Lemma 6.2), so a^* is ranked above all the alternatives in B, that is, in the top $m' + 1 = \mathcal{O}(m^\epsilon)$ places. On the other hand, if there is no exact 3-cover, then the Young score of a^* is zero by the same arguments as in Lemma 6.2, since the agents in N^* all rank a^* last. Hence a^* is placed last in any Young ranking. □

As noted in Section 2, one can imagine another alternative formulation of the Young score. Indeed, one might ask: given a preference profile, what is the *smallest* number of agents that must be *removed* in order to make a^* a Condorcet winner? This minimization problem, where the score is the number of agents that are removed, is referred to as the *Dual Young score* by Betzler et al. [4]. Of course, a Young winner according to the primal formulation is always a winner according to the dual formulation, and vice versa. Notice that it is easy to obtain an ϵn-approximation under the dual formulation for any constant $\epsilon > 0$ by enumerating all subsets of agents of size at least $n - 1/\epsilon$ and checking whether a^* is the Condorcet winner in the preferences of these agents. Our next result states that the dual Young score is hard to approximate significantly better.

Theorem 6.4. *For any constant $\epsilon > 0$, the dual Young score is $\mathcal{NP}$-hard to approximate within $O(n^{1-\epsilon})$.*

Our proof relies on a statement regarding the inapproximability of Vertex Cover that is even weaker than the one we used in the proof for the inapproximability of the Dodgson ranking.

Theorem 6.5 (Berman and Karpinski [3], see also [16])**.** *Given a 3-regular graph G and an integer $K \geq 1$, it is $\mathcal{NP}$-hard to distinguish between the following two cases:*

- *G has a vertex cover of size at most K.*
- *Any vertex cover of G has size at least $K + 2$.*

Our reduction extends the one in the proof of Lemma 6.2. Consider a 3-regular graph $G = (V_1, E)$ with p nodes and an integer $K \geq 1$. Also, let $\epsilon \in (0, 1)$ be a constant and let $n = \lceil p^{1/\epsilon} \rceil$. Denote by $H = (V_2, F)$ the complete graph with $n - p$ nodes.

Define the set of alternatives as $A = E \cup F \cup \{a\} \cup \{a^*\}$. Let the set of agents be $N = N' \cup N'' \cup N'''$, where N', N'', and N''' are defined as follows. The set N' consists of p agents corresponding to the p nodes of G, such that for all $i \in N'$, agent i prefers the alternatives in $F \cup E \setminus E_i$ to a^* (where the set E_i consists of the edges of E which are incident to node i), and prefers a^* to all the alternatives in $E_i \cup \{a\}$ (i.e., $F \cup E \setminus E_i \; R^i \; a^* R^i \; E_i \cup \{a\}$). The set N'' contains $n - p$ agents corresponding to the $n - p$ nodes of H, such that for all $i \in N''$, agent i prefers the alternatives in $E \cup F \setminus F_i$ to a^* (where the set F_i consists of the edges of F which are incident to node i), and prefers a^* to all the alternatives in $F_i \cup \{a\}$ (i.e., $E \cup F \setminus F_i \; R^i \; a^* R^i \; F_i \cup \{a\}$). Subset N''' consists of $n - p + K - 2$ agents who prefer a to a^* and a^* to $E \cup F$ (i.e., $a \; R^i \; a^* \; R^i \; E \cup F$).

Theorem 6.4 now follows by Theorem 6.5 and the next two lemmata.

Lemma 6.6. *If G has a vertex cover of size at most K, then the dual Young score of alternative a^* is at most n^ϵ.*

Proof. Let $C \subseteq V_1$ be a vertex cover of G of size at most K. Consider the following sets of agents: a set $N^* \subseteq N'$ that contains the agents that correspond to nodes in the vertex cover C, a set N^+ of all the agents of N'' besides one, and the set N'''.

Recall that $N^* \cup N^+$ has size at most $n - p + K - 1$ while N''' has size $n - p + K - 2$. Since C is a vertex cover in G, each alternative in E is ranked lower than a^* by at least one agent of N^*. Also, the nodes corresponding to the agents in N^+ form a vertex cover of H. So, each alternative in F is ranked lower than a^* by at least one agent of N^+. Hence, by considering the agents of $N^* \cup N^+ \cup N'''$, a^* beats any other alternative in their pairwise comparison and its dual Young score is at most $p - |C| + 1 \leq p \leq n^\epsilon$. □

Lemma 6.7. *If G has no vertex cover of size less than $K + 2$, then the dual Young score of alternative a^* is n.*

Proof. We will show that there is no nonempty subset of agents that make a^* a Condorcet winner. Indeed, assume for contradiction that there exists such a subset that contains the sets of agents $N^* \subseteq N'$, $N^+ \subseteq N''$, and $N^- \subseteq N'''$.

If $|N^+| < n - p - 1$ or $|N^*| < K + 2$ then there exists an alternative of E or F which is not ranked lower than a^* by any agent of $N^* \cup N^+$. In both cases, N^- must have size at least $|N^*| + |N^+|$ in order for a^* to beat every alternative in $E \cup F$ in their pairwise comparison. However, a^* does not beat a and cannot be a Condorcet winner.

Therefore, it holds that $|N^+| \in \{n - p - 1, n - p\}$ and $|N^*| \geq K + 2$. If $|N^+| = n - p - 1$, then some alternative of F is ranked below a^* by at most one agent of N^+. It is also ranked above a^* by the agents of N^* and below it by the agents of N^-. In total, it is ranked above a^* by at least $n - p + K$ agents while it is ranked below a^* by at most $n - p + K - 1$ agents. Hence, a^* cannot be a Condorcet winner in this case.

If $|N^+| = n - p$ then each alternative in F is ranked below a^* by two agents of N^+. In total, it is ranked above a^* by at least $n - p + K$ agents while it is ranked below a^* by at most $n - p + K$ agents. Again, a^* cannot be a Condorcet winner. □

The proof of Theorem 6.4 provides an alternative proof of Theorem 6.1. In terms of the Young score, it implies that, for every constant $\epsilon > 0$, there are instances for which it is hard to distinguish between a score of zero and a score of at least $n - n^\epsilon$. So, for the formulation of the Young score where all the scores are scaled by an additive constant, it provides additional information which is complementary to the one provided by the proof of Theorem 6.1: it implies that it is hard to approximate the Young score, even under this alternative formulation, within a factor of $\mathcal{O}(n)$.

Acknowledgments: The authors wish to thank Felix Fischer and Nati Linial for helpful discussions. The work was done while Procaccia was at the Hebrew University of Jerusalem, and was supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities. The work of Procaccia and Rosenschein was partially supported by Israel Science Foundation grant #898/05. The work of Feldman was partially supported by the Israel Science Foundation (grant #1219/09) and by the Leon Recanati Fund of the Jerusalem school of business administration. The work of Caragiannis, Kaklamanis, and Karanikolas was partially supported by the European Union under IST FET Integrated Project AEOLUS "Algorithmic Principles for Building Efficient Overlay Computers" and COST Action IC0602 "Algorithmic Decision Theory", and by a "Caratheodory" basic research grant from the University of Patras.

References

[1] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM*, 55(5), 2008.

[2] J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6:157–165, 1989.

[3] P. Berman and M. Karpinski. On some tighter inapproximability results. In *Proceedings of the 26th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 200–209, 1999.

[4] N. Betzler, J. Guo, and R. Niedermeier. Parameterized computational complexity of Dodgson and Young elections. *Information and Computation*, 208(2):165–177, 2010.

[5] D. Black. *Theory of Committees and Elections*. Cambridge University Press, 1958.

[6] F. Brandt. Some remarks on Dodgson's voting rule. *Mathematical Logic Quarterly*, 55(4):460–463, 2009.

[7] I. Caragiannis, C. Kaklamanis, N. Karanikolas, and A. D. Procaccia. Socially desirable approximations for Dodgson's voting rule. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*, 2010 (to appear).

[8] V. Conitzer. Computing Slater rankings using similarities among candidates. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, pages 613–619, 2006.

[9] V. Conitzer, A. Davenport, and J. Kalagnanam. Improved bounds for computing Kemeny rankings. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, pages 620–626, 2006.

[10] D. Coppersmith, L. Fleischer, and A. Rudra. Ordering by weighted number of wins gives a good ranking for weighted tournaments. In *Proceedings of the 17th Annual ACM-SIAM symposium on Discrete algorithm (SODA)*, pages 776–782, 2006.

[11] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4):643–652, 1998.

[12] M. R. Garey and D. S. Johnson. *Computers and Intractability: a Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[13] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–602, 1973.

[14] E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM*, 44(6):806–825, 1997.

[15] C. Homan and L. A. Hemaspaandra. Guarantees for the success frequency of an algorithm for finding Dodgson-election winners. *Journal of Heuristics*, 15(4): 403-423, 2009.

[16] M. Karpinski. Approximating bounded degree instances of NP-hard problems. In *Proceedings of the 13th International Symposium on Fundamentals of Computation Theory (FCT)*, volume 2138 of *Lecture Notes in Computer Science*, pages 24–34. Springer, 2001.

[17] C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 95–103, 2007.

[18] C. Klamer. A comparison of the Dodgson method and the Copeland rule. *Economics Bulletin*, 4(8):1–7, 2003.

[19] C. Klamer. The Dodgson ranking and its relation to Kemeny's method and Slater's rule. *Social Choice and Welfare*, 23(1):91–102, 2004.

[20] C. Klamler. The Dodgson ranking and the Borda count: a binary comparison. *Mathematical Social Sciences*, 48(1):103–108, 2004.

[21] J. C. McCabe-Dansted. Approximability and computational feasibility of Dodgson's rule. Master's thesis, University of Auckland, 2006.

[22] J. C. McCabe-Dansted, G. Pritchard, and A. M. Slinko. Approximability of Dodgson's rule. *Social Choice and Welfare*, 31(2): 311-330, 2008

[23] N. Nisan. Introduction to mechanism design (for computer scientists). In N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, chapter 9. Cambridge University Press, 2007.

[24] A. D. Procaccia, A. Zohar, Y. Peleg, and J. S. Rosenschein. The learnability of voting rules. *Artificial Intelligence*, 173(12–13):1133–1149, 2009.

[25] S. Rajagopalan and V. V. Vazirani. Primal-dual RNC approximation algorithms for set cover and covering integer programs. *SIAM Journal on Computing*, 28:526–541, 1999.

[26] T. C. Ratliff. A comparison of Dodgson's method and Kemeny's rule. *Social Choice and Welfare*, 18(1):79–89, 2001.

[27] T. C. Ratliff. A comparison of Dodgson's method and the Borda count. *Economic Theory*, 20(2):357–372, 2002.

[28] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, pages 475–484, 1997.

[29] J. Rothe, H. Spakowski, and J. Vogel. Exact complexity of the winner problem for young elections. *Theory of Computing Systems*, 36(4):375–386, 2003.

[30] M. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.

[31] N. Tideman. *Collective Decisions and Voting: The Potential for Public Choice*. Ashgate, 2006.

[32] V. V. Vazirani. *Approximation Algorithms*. Springer, 2001.

[33] H. P. Young. Extending Condorcet's rule. *Journal of Economic Theory*, 16:335–353, 1977.