Robust Bounds on Choosing from Large Tournaments

Christian Saile
Technical University of Munich
Germany

Warut Suksompong
University of Oxford
United Kingdom

Tournament solutions provide methods for selecting the "best" alternatives from a tournament and have found applications in a wide range of areas. Previous work has shown that several well-known tournament solutions almost never rule out any alternative in large random tournaments. Nevertheless, all analytical results thus far have assumed a rigid probabilistic model, in which either a tournament is chosen uniformly at random, or there is a linear order of alternatives and the orientation of all edges in the tournament is chosen with the same probabilities according to the linear order. In this work, we consider a significantly more general model where the orientation of different edges can be chosen with different probabilities. We show that a number of common tournament solutions, including the top cycle and the uncovered set, are still unlikely to rule out any alternative under this model. This corresponds to natural graph-theoretic conditions such as irreducibility of the tournament. In addition, we provide tight asymptotic bounds on the boundary of the probability range for which the tournament solutions select all alternatives with high probability.

1. Introduction

Tournaments play an important role in numerous situations as a means of representing entities and a dominance relationship between them. For instance, both the outcome of a round-robin sports competition and the majority relation of voters in an election can be represented by a tournament. A question that occurs frequently is therefore the following: Given a tournament, how can we choose the "best" alternatives in a consistent manner? This question has been addressed by a rich and beautiful literature on tournament solutions, which have found applications in areas ranging from sports competitions (Ushakov, 1976)

1

to multi-criteria decision analysis (Arrow and Raynaud, 1986; Bouyssou, 2004) to biology (Schjelderup-Ebbe, 1922; Landau, 1953; Slater, 1961; Allesina and Levine, 2011). Over the past half century several tournament solutions have been proposed, two of the oldest and best-known of which are the *top cycle* (Good, 1971; Schwartz, 1972; Miller, 1977) and the *uncovered set* (Miller, 1980).[1]

Given that the purpose of tournament solutions is to discriminate the "best" alternatives from the remaining ones, it perhaps comes as a surprise that many common tournament solutions—including the top cycle, the uncovered set, the Banks set, and the minimal covering set—select all alternatives with high probability in a large random tournament (Fey, 2008; Scott and Fey, 2012). Put differently, the aforementioned tournament solutions almost never exclude any alternative in a tournament chosen at random. Nevertheless, these results are based on the *uniform random model*, in which all tournaments are drawn with equal probability, or equivalently each edge is oriented in one direction or the other with equal probability independently of other edges. For a large majority of applications of tournaments, one would not expect that this assumption holds. Indeed, stronger teams are likely to beat weaker teams in a sports competition, and candidates with a large base of support have a higher chance of winning an election. Moreover, real-world tournaments often exhibit a certain degree of transitivity: If alternatives a, b, and c are such that a dominates b and b dominates c, then it is more likely that a dominates c than the other way around.

A more general model of random tournaments is the *Condorcet random model*, previously considered by Frank (1968), Łuczak et al. (1996), Vassilevska Williams (2010) and Kim et al. (2017). In this model, there is a linear order of alternatives, which can be interpreted as an ordering of the alternatives from strongest to weakest. For each pair of alternatives, the probability that the edge is oriented from the alternative that occurs later in the linear order to the alternative that occurs earlier in the linear order is p, independently of other pairs of alternatives.[2] Crucially, the value of p is the same for all pairs of alternatives. The Condorcet random model generalizes the uniform random model, since the latter can be obtained from the former by taking $p = 1/2$. Łuczak et al. (1996) showed that under the Condorcet random model, the top cycle selects all alternatives as long as $p \in \omega(1/n)$. The same authors show furthermore that this bound is tight, that is, the statement no longer holds if $p \in O(1/n)$.[3]

Although the Condorcet random model addresses the issues raised above with regard to the uniform random model, it is still rather unrealistic for two important reasons. Firstly, in tournaments in the real world, the orientation of different edges are typically determined by different probabilities. For instance, in a sports tournament the probability that a very strong team beats a very weak team is usually higher than the probability that a

---

[1] For a thorough treatment of tournament solutions, we refer the reader to excellent surveys by Laslier (1997) and Brandt et al. (2016).

[2] By symmetry, we may assume without loss of generality that $p \leq 1/2$.

[3] See, e.g., Cormen et al. (2009) for the definitions of asymptotic notations.

moderately strong team beats a moderately weak team; a similar phenomenon can be observed in elections. Secondly, even though one can roughly order the alternatives in a tournament according to their strength, it is often the case that not all probabilities of the orientation of the edges respect the ordering. Indeed, this precisely corresponds to the notion of "bogey teams"—weak teams that nevertheless frequently beat certain supposedly stronger teams. Given the limitations of the uniform random model and the Condorcet random model, it is natural to ask whether previous results continue to hold under more general and realistic models of random tournaments, or whether they break down as soon as we move beyond these restricted models.

   In this paper, we show that a number of tournament solutions, including the top cycle and the uncovered set, still choose all alternatives with high probability under a significantly more general model of random tournaments. Unlike the Condorcet random model, our model does not rely on an ordering of the alternatives. Instead, the orientation of each edge is determined by probabilities within the range $[p, 1-p]$ for some parameter p, and these probabilities are allowed to vary across edges. The only substantive assumption that we make is that the orientations of different edges are chosen independently from one another. Under this model, which is more general than both the uniform random model and the Condorcet random model, we establish in Section 3 that the top cycle almost never rules out any alternative as long as $p \in \omega(1/n)$, thus generalizing the result by Łuczak et al. (1996). We also show that our bound is asymptotically tight, and that analogous results hold for two other tournament solutions based on the set of Condorcet winners and losers as well. Moreover, we prove in Section 4 that the uncovered set is likely to include the whole set of alternatives when $p \in \omega(\sqrt{\log n/n})$. This bound is again asymptotically tight, and the same holds for another tournament solution based on the uncovered set. Since the condition that the top cycle or the uncovered set chooses all alternatives have meaningful graph-theoretic interpretations—the top cycle is the whole set of alternatives if and only if the tournament is strongly connected,[4] and the uncovered set fails to exclude any alternative exactly when all alternatives are *kings*[5]—we believe that our results are of independent interest in graph theory and discrete mathematics. Furthermore, the generality of our model allows us to derive consequences in Section 5 for a different model in which tournaments are generated from random voter preferences, and we complement our theoretical results with experimental data in Section 6.

---

[4] A strongly connected tournament is also said to be *strong*. Strong connectedness is equivalent to *irreducibility* and to the property of having a Hamiltonian cycle (Moon, 1968).

[5] A *king* is an alternative that can reach any other alternative via a directed path of length at most two (Maurer, 1980). Therefore, all alternatives of a tournament are kings if and only if every pair of alternatives can reach each other via a directed path of length at most two. Such a tournament has been studied in graph theory and called an *all-kings tournament* (Reid, 1982).

1.1. Related Work

The study of the behavior of tournament solutions in large random tournaments goes back to Moon and Moser (1962), who showed that the top cycle almost never rules out any alternative in a large tournament chosen uniformly at random. In fact, they proved a stronger statement that the probability that the top cycle excludes at least one alternative is inverse exponential in the number of alternatives; the estimate was later made more precise by Moon (1968) in his seminal book on tournaments. Bell (1981) also considered the top cycle but assumed that tournaments are generated from the preferences of a large number of voters, each with a uniform random ranking over the alternatives; he likewise found that the top cycle selects all alternatives with high probability under this assumption. Fey (2008) and later Scott and Fey (2012) established results on several tournament solutions including the uncovered set, the Banks set, the Copeland set, the minimal covering set, and the bipartisan set using the uniform random model. While the uncovered set, the Banks set, and the minimal covering set are likely to include all alternatives in a large random tournament, the same event is unlikely to occur for the Copeland set. On the other hand, the bipartisan set chooses on average half of the alternatives in a random tournament of any fixed size (Fisher and Ryan, 1995); it is the unique most discriminating tournament solution satisfying standard properties proposed in the literature (Brandt et al., 2018).

The discriminative power of tournament solutions has also been investigated empirically by Brandt and Seedig (2016). Building on the observation that the distributions of real-world tournaments are typically far from uniform, these authors examined the behavior of eleven common tournament solutions on tournaments generated according to stochastic preference models and empirical data. The stochastic models that they used include the impartial culture model, the Mallows mixtures model, and the Pólya-Eggenberger urn model. They reported that under these more realistic models, most tournament solutions are in fact much more discriminating than the analytical results for uniform random tournaments suggest.

2. Preliminaries

A tournament T consists of a set $A = \{a_1, a_2, \ldots, a_n\}$ of alternatives and a dominance relation. The dominance relation is an asymmetric and connex binary relation on A represented by a directed edge between each unordered pair of distinct alternatives in A. We say that alternative a_i *dominates* another alternative a_j if there is an edge from a_i to a_j. An alternative is said to be a *Condorcet winner* if it dominates all of the remaining alternatives, and a *Condorcet loser* if it is dominated by all of the remaining alternatives. We extend the dominance relation to sets and say that a set $A' \subseteq A$ of alternatives dominates another set $A'' \subseteq A$ of alternatives disjoint from A' if for all $a' \in A'$ and $a'' \in A''$, a' dominates a''. A tournament is commonly interpreted as the outcome of a round-robin sports competition and as the majority relation of an odd number of voters with linear

preferences. In the former interpretation, alternative a_i dominating alternative a_j means that the player or team represented by a_i beats the player or team represented by a_j in the competition. In the latter interpretation, the same dominance relation signifies that more than half of the voters prefer a_i to a_j.

We are interested in tournament solutions, which are functions that map each tournament to a nonempty subset of its alternatives, usually referred to as the *choice set*. Two simple tournament solutions are *COND*, which chooses a Condorcet winner if one exists and chooses all alternatives otherwise,[6] and the set of *Condorcet non-losers* (*CNL*), which consists of all alternatives that are not Condorcet losers. Other tournament solutions considered in this paper are the following:

- The *top cycle* (*TC*) is the (unique) smallest set of alternatives such that all alternatives in the set dominate all alternatives not in the set;

- The *uncovered set* (*UC*) consists of all alternatives that can reach all other alternatives via a domination path of length at most two;[7]

- The *iterated uncovered set* (*UC*$^\infty$) is the result of iteratively computing the uncovered set until there is no further reduction.

The inclusions $UC^\infty(T) \subseteq UC(T) \subseteq TC(T) \subseteq CNL(T)$ and $TC(T) \subseteq COND(T)$ hold for any tournament T.

Next, we describe the random models for generating tournaments that we consider in this paper. We will work with the first model in Sections 3 and 4 and the second model in Section 5.

- Model 1: For each pair of distinct alternatives a_i, a_j, there is an edge from a_i to a_j with probability $p_{i,j}$ and an edge from a_j to a_i with probability $p_{j,i} = 1 - p_{i,j}$, independently of other pairs of alternatives.

- Model 2: There is a constant number k of voters, where k is odd. For each voter v and each pair of distinct alternatives a_i, a_j, the voter prefers a_i to a_j with probability $q_{v,i,j}$ and prefers a_j to a_i with probability $q_{v,j,i} = 1 - q_{v,i,j}$, independently of other voters and other pairs of alternatives.[8] The majority relation, in which alternative a_i dominates another alternative a_j if and only if more than half of the voters prefer a_i to a_j, forms a tournament with A as its set of alternatives.

---

[6]Note that the set of Condorcet winners is not a tournament solution because it can be empty.

[7]This is known in graph theory as the set of *kings* (cf. Footnote 5). An alternative definition, which is also the origin of the name "uncovered set", is based on the *covering relation*. An alternative a_i is said to *cover* another alternative a_j if (i) a_i dominates a_j, and (ii) any alternative that dominates a_i also dominates a_j. The uncovered set corresponds to the set of alternatives that are not covered by any other alternative.

[8]One way to interpret the possible intransitivity of the preferences is as a result of noise in the voters' true preferences. Laslier (2010) introduced the term *Rousseauist cultures* for this kind of models.

Several models for generating random tournaments considered in previous work are special cases of our models. For example, the *uniform random model* (Fey, 2008; Scott and Fey, 2012) corresponds to taking $p_{i,j} = 1/2$ for all i, j in Model 1 or taking $q_{v,i,j} = 1/2$ for all v, i, j in Model 2 with any k. The *Condorcet random model* (Frank, 1968; Łuczak et al., 1996; Vassilevska Williams, 2010; Kim et al., 2017) corresponds to taking $p_{i,j} = p$ for all $i < j$ in Model 1, for some fixed value of p. The *Condorcet random model for voters* (Brandt and Seedig, 2016) corresponds to taking $q_{v,i,j} = p$ for all v and all $i < j$ in Model 2, for some fixed value of p. Following standard terminology, we say that an event occurs "with high probability" or "almost surely" if the probability that the event occurs converges to 1 as n, the number of alternatives, goes to infinity.

We end this section by listing some standard tools for deriving probabilistic bounds. Our first lemma is the Chernoff bound, which gives us an upper bound on the probability that a sum of independent random variables is far away from its expected value.

Lemma 1 (Chernoff bound)**.** *Let $X_1, X_2, \ldots, X_r$ be independent random variables that take on values in the interval $[0, 1]$, and let $S = X_1 + X_2 + \cdots + X_r$. For every $\delta \geq 0$, we have*

$$\Pr[S \geq (1 + \delta)\, \mathbb{E}[S]] \leq \exp\left(\frac{-\delta^2\, \mathbb{E}[S]}{3} \right).$$

The next two lemmas allow us to estimate the expression $1 + x$ from above and below.

Lemma 2 (Bernoulli's inequality)**.** *For all real numbers $r \geq 1$ and $x \geq -1$, we have*

$$(1 + x)^r \geq 1 + rx.$$

Lemma 3. *For all real numbers x, we have $1 + x \leq e^x$.*

3. Top Cycle

In this section, we consider the top cycle. We show that when each probability $p_{i,j}$ is between $f(n)$ and $1 - f(n)$ for some function $f(n) \in \omega\, (1/n)$, TC chooses all alternatives with high probability (Theorem 1). By using the inclusion relationships between TC, $COND$, and CNL, we obtain analogous statements for $COND$ and CNL. We also show that our results are asymptotically tight—for all three tournament solutions, the statement ceases to hold if $f(n) \in O\, (1/n)$ (Theorem 2).

We begin with our main result of the section.

Theorem 1. *Let $f : \mathbb{Z}^+ \to \mathbb{R}_{\geq 0}$ be a function such that $f(n) \leq 1/2$ for all n and $f(n) \in \omega\, (1/n)$. Assume that a tournament T is generated according to Model 1, and that*

$$p_{i,j} \in [f(n), 1 - f(n)]$$

for all $i \neq j$. Then with high probability, $TC(T) = A$.

Theorem 1 generalizes a result by Łuczak et al. (1996) that establishes the claim for the case where $p_{i,j} = f(n)$ for all $i < j$ (or, by symmetry, the case where $p_{i,j} = 1 - f(n)$ for all $i < j$). We remark that their proof relies crucially on the assumption that there is a linear order of alternatives and all edges are more likely to be oriented in one direction than in the other direction according to the order. Indeed, this assumption allows the authors to show that with high probability, any alternative can be reached by the strongest alternative and can reach the weakest alternative via a domination path of length at most two each. Moreover, with the assumption $f(n) \in \omega(1/n)$ one can show that the weakest alternative can almost surely reach the strongest alternative via a domination path of length four, thus establishing the strong connectivity of the tournament. In contrast, we do not assume that the edges in the tournament are likely to be oriented in one direction or the other. As such, we will need a completely different approach for our proof.

Before we go into the proof of Theorem 1, we first give a high-level overview. We observe that $TC(T) \neq A$ exactly when there exists a proper, nontrivial subset of alternatives B that dominates the complement set of alternatives $A \backslash B$. Using the union bound, we then upper bound the probability that $TC(T) \neq A$ by the sum over all sets B of the probabilities that B dominates $A \backslash B$; this sum can be written entirely in terms of the variables $p_{i,j}$ for $i < j$. From here, a relatively straightforward calculation yields the desired result for $f(n) \in \Omega(\log n / n)$. However, substantially more work is required to shave off the logarithmic factor and tighten the bound. To this end, we note that the aforementioned sum is linear in all of the variables $p_{i,j}$, implying that its maximum is attained when all variables take on a value at one of the two boundaries of their domain. Using a number of helper lemmas (Lemmas 4, 5, and 6), which include an interesting combinatorial extension of Karamata's inequality (Lemma 5), we show that the sum is in fact maximized when all variables take on a value at the *same* boundary (i.e., when the tournament is generated by the Condorcet random model). This allows us to bound the sum directly by plugging in the value at a boundary and complete the proof.

In what follows, we assume that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ are vectors of nonnegative integers with n components. We start by defining majorization, a preorder on vectors that we will use frequently in our proof.

Definition 1 (Majorization). For a vector $\mathbf{x}$, let $\mathbf{x}^{\downarrow} = (x_1^{\downarrow}, x_2^{\downarrow}, \ldots, x_n^{\downarrow})$ be the vector with the same components, but sorted in descending order. Given two vectors $\mathbf{x}, \mathbf{y}$, we say that $\mathbf{x}$ *majorizes* $\mathbf{y}$, and write $\mathbf{x} \succ \mathbf{y}$, if the following two conditions are satisfied:

(i) $\sum_{i=1}^{j} x_i^{\downarrow} \geq \sum_{i=1}^{j} y_i^{\downarrow}$ for $j = 1, 2, \ldots, n-1$;

(ii) $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.

When one vector majorizes another vector, Karamata's inequality allows us to compare the sum of an arbitrary convex function at the components of one vector to the corresponding sum of the other vector.

Lemma 4 (Karamata's inequality)**.** *Let $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be a convex function, and let $\boldsymbol{x}, \boldsymbol{y}$ be vectors with n components such that $\boldsymbol{x} \succ \boldsymbol{y}$. Then*

$$\sum_{i=1}^{n} f(x_i) \geq \sum_{i=1}^{n} f(y_i).$$

We next show that if one vector majorizes another vector, then an analogous statement holds for the two vectors that arise from taking the sum of all subsets with any fixed number of components of the original vectors.

Definition 2. Let n be a positive integer and $k \in \{1, 2, \ldots, n\}$. For a vector $\mathbf{x}$ with n components, define $\mathbf{x}^{(k)}$ to be the vector with $\binom{n}{k}$ components consisting of all sums of k distinct components of $\mathbf{x}$ in nonincreasing order.

For example, if $n = 4$ and $\mathbf{x} = (2, 4, 5, 7)$, then $\mathbf{x}^{(2)} = (12, 11, 9, 9, 7, 6)$.

Lemma 5. *If two vectors $\boldsymbol{x}, \boldsymbol{y}$ with n components are such that $\boldsymbol{x} \succ \boldsymbol{y}$, then we also have $\boldsymbol{x}^{(k)} \succ \boldsymbol{y}^{(k)}$ for all $k = 1, 2, \ldots, n$.*

For the sake of continuity, we leave the proof of Lemma 5 along with that of the next lemma to the appendix.

Our final lemma shows that the outdegree vector of a transitive tournament majorizes the corresponding vector of any tournament. Given a tournament T and alternative a in the tournament, denote by $\deg_T(a)$ the outdegree of a in T.

Lemma 6. *Let W be a transitive tournament on alternatives $d_1, d_2, \ldots, d_n$, where d_i dominates d_j for all $i < j$. For any tournament U with alternatives $b_1, b_2, \ldots, b_n$, we have*

$$(\deg_W(d_1), \deg_W(d_2), \ldots, \deg_W(d_n)) \succ (\deg_U(b_1), \deg_U(b_2), \ldots, \deg_U(b_n)).$$

With Lemmas 4, 5, and 6 in hand, we are now ready to prove Theorem 1.

Proof of Theorem 1. Let $c \geq 10$ be a constant. Since $f(n) = \omega(1/n)$, there exists N' such that $f(n) \geq c/n$ for all $n \geq N'$. Let $N = \max(N', 4c)$. We will show that for $n \geq N$, the probability that TC does not choose the whole set of alternatives is at most $16ce^{-\frac{c}{2}}$. Since the expression converges to 0 as c approaches infinity, this will establish the desired result.

Assume that $n \geq N$. Observe that $TC(T) \neq A$ exactly when there is a proper, nontrivial set of alternatives that dominate the complement set of alternatives. Hence

$$\Pr[TC(T) \neq A] = \Pr[B \text{ dominates } A \backslash B \text{ for some } \emptyset \neq B \subset A]$$

$$\leq \sum_{k=1}^{n-1} \sum_{\substack{B \subset A \\ |B|=k}} \Pr[B \text{ dominates } A \backslash B]$$

$$= \sum_{k=1}^{n-1} \sum_{\substack{B \subset A \\ |B|=k}} \prod_{\substack{1 \le i \ne j \le n \\ a_i \in B \\ a_j \in A \setminus B}} p_{i,j}$$

$$= \sum_{k=1}^{n-1} \sum_{\substack{B \subset A \\ |B|=k}} \left(\prod_{\substack{1 \le i < j \le n \\ a_i \in B \\ a_j \in A \setminus B}} p_{i,j} \prod_{\substack{1 \le i < j \le n \\ a_i \in A \setminus B \\ a_j \in B}} (1 - p_{i,j}) \right), \qquad (1)$$

where we use the union bound for the inequality. We will derive an upper bound for expression (1). Note that if we view the terms $p_{i,j}$ with $i < j$ as variables, then the expression in linear in each variable. This implies that the maximum of expression (1) over the range $p_{i,j} \in [c/n, 1 - c/n]$ is attained when each $p_{i,j}$ is either c/n or $1 - c/n$ (but not necessarily when all $p_{i,j}$ are identical). We henceforth assume that for each $i < j$, either $p_{i,j} = c/n$ or $p_{i,j} = 1 - c/n$. We will show that expression (1) is maximized when $p_{i,j} = 1 - c/n$ for all $i < j$ (or alternatively, when $p_{i,j} = c/n$ for all $i < j$). In fact, we will show the stronger statement that for each particular value of k in the outermost summation, the expression inside the outermost summation is also maximized when $p_{i,j} = 1 - c/n$ for all $i < j$.

Fix $k \in \{1, 2, \ldots, n-1\}$. Define a tournament U on n alternatives $b_1, b_2, \ldots, b_n$ as follows: For any $i, j \in \{1, 2, \ldots, n\}$ with $i < j$, there is an edge from b_i to b_j if $p_{i,j} = 1 - c/n$ and an edge from b_j to b_i if $p_{i,j} = c/n$. We have

$$\sum_{\substack{B \subset A \\ |B|=k}} \left(\prod_{\substack{1 \le i < j \le n \\ a_i \in B \\ a_j \in A \setminus B}} p_{i,j} \prod_{\substack{1 \le i < j \le n \\ a_i \in A \setminus B \\ a_j \in B}} (1 - p_{i,j}) \right)$$

$$= \sum_{\substack{B \subset A \\ |B|=k}} \prod_{\substack{1 \le i \ne j \le n \\ a_i \in B \\ a_j \in A \setminus B}} p_{i,j}$$

$$= \left(1 - \frac{c}{n}\right)^{k(n-k)} \sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{c}{n-c}\right)^{|\{(a_i,a_j) \in B \times A \setminus B \text{ such that } (b_i,b_j) \text{ is not an edge in } U\}|}$$

$$= \left(1 - \frac{c}{n}\right)^{k(n-k)} \sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{c}{n-c}\right)^{(n-1)k - \binom{k}{2} - \sum_{i:a_i \in B} \deg_U(b_i)}$$

9

$$= \left(1 - \frac{c}{n}\right)^{k(n-k)} \left(\frac{c}{n-c}\right)^{(n-1)k - \binom{k}{2}} \sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{n-c}{c}\right)^{\sum_{i:a_i \in B} \deg_U(b_i)}. \tag{2}$$

Let W be a transitive tournament on n alternatives $d_1, d_2, \ldots, d_n$, where d_i dominates d_j for all $i < j$. In particular, $\deg_W(d_i) = n - i$ for all $i = 1, 2, \ldots, n$. To show that expression (1) is maximized when $p_{i,j} = 1 - c/n$ for all $i < j$, it suffices to show that expression (2) is maximized when $U = W$. The terms outside the summation do not depend on the tournament U that we choose, so for the purpose of maximizing expression (2) we may ignore them.

From Lemma 6, we know that

$$(\deg_W(d_1), \deg_W(d_2), \ldots, \deg_W(d_n)) \succ (\deg_U(b_1), \deg_U(b_2), \ldots, \deg_U(b_n)).$$

Lemma 5 then implies that

$$(\deg_W(d_1), \deg_W(d_2), \ldots, \deg_W(d_n))^k \succ (\deg_U(b_1), \deg_U(b_2), \ldots, \deg_U(b_n))^k.$$

Using Lemma 4 with the convex function $f(x) = \left(\frac{n-c}{c}\right)^x$, we find that

$$\sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{n-c}{c}\right)^{\sum_{i:a_i \in B} \deg_U(b_i)} \leq \sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{n-c}{c}\right)^{\sum_{i:a_i \in B} \deg_W(d_i)}.$$

It follows that expression (2) is maximized when $U = W$, as claimed.

We return to expression (1), which we now know is maximized when $p_{i,j} = 1 - c/n$ for all $i < j$. Substituting $p_{i,j} = 1 - c/n$ for all $i < j$, expression (1) becomes

$$\sum_{k=1}^{n-1} \left(\left(1 - \frac{c}{n}\right)^{k(n-k)} \sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{c}{n-c}\right)^{(n-1)k - \binom{k}{2} - \sum_{i:a_i \in B} \deg_W(d_i)} \right)$$

$$\leq 2 \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{-\frac{ck(n-k)}{n}} \sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{2c}{n}\right)^{(n-1)k - \binom{k}{2} - \sum_{i:a_i \in B}(n-i)} \right)$$

$$\leq 2 \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{-\frac{ck}{2}} \sum_{\substack{B \subset A \\ |B|=k}} \left(\frac{2c}{n}\right)^{\sum_{i:a_i \in B} i - \binom{k+1}{2}} \right),$$

where we use Lemma 3, the assumption $n \geq 4c$, and the symmetry between the terms with $k = i$ and $k = n - i$ for the first inequality. Observe that $\sum_{i:a_i \in B} i - \binom{k+1}{2}$ is

10

always nonnegative, and is zero exactly when $B = \{1, 2, \ldots, k\}$. Moreover, for any $j = \{1, 2, \ldots, k\}$, the number of subsets $B \subset A$ with $|B| = k$ such that $\sum_{i:a_i \in B} i - \binom{k+1}{2} \le j$ is at most n^j. Indeed, if a subset B satisfies this inequality, the $n - j$ smallest elements of B must be $1, 2, \ldots, n - j$, which leaves at most n^j choices for the remaining elements. Note also that $|\{B \subset A \mid |B| = k\}| = \binom{n}{k} \le n^k$. We have

$$2 \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{-\frac{ck}{2}} \sum_{\substack{B \subset A \\ |B| = k}} \left(\frac{2c}{n} \right)^{\sum_{i:a_i \in B} i - \binom{k+1}{2}} \right)$$

$$\le 2 \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{-\frac{ck}{2}} \sum_{\substack{B \subset A \\ |B| = k}} \left(\frac{2c}{n} \right)^{\min\left(k, \sum_{i:a_i \in B} i - \binom{k+1}{2}\right)} \right)$$

$$\le 2 \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{-\frac{ck}{2}} \sum_{j=0}^{k} \left(n^j \cdot \left(\frac{2c}{n} \right)^j \right) \right)$$

$$= 2 \sum_{k=1}^{\lfloor n/2 \rfloor} \left(e^{-\frac{ck}{2}} \sum_{j=0}^{k} (2c)^j \right)$$

$$\le 2 \sum_{k=1}^{\lfloor n/2 \rfloor} e^{-\frac{ck}{2}} (4c)^k$$

$$= 2 \sum_{k=1}^{\lfloor n/2 \rfloor} \left(4c e^{-\frac{c}{2}} \right)^k$$

$$\le 2 \sum_{k=1}^{\infty} \left(4c e^{-\frac{c}{2}} \right)^k$$

$$= \frac{8c e^{-\frac{c}{2}}}{1 - 4c e^{-\frac{c}{2}}}$$

$$\le 16c e^{-\frac{c}{2}},$$

where we use the assumption $c \ge 10$ for the last inequality.

In conclusion, when $n \ge N$, the probability that $TC(T) \ne A$ is at most $16c e^{-\frac{c}{2}}$, completing our proof. $\qquad\square$

Since $TC(T) \subseteq COND(T)$ and $TC(T) \subseteq CNL(T)$, we immediately obtain the following corollary.

Corollary 1. *Let $f : \mathbb{Z}^+ \to \mathbb{R}_{\geq 0}$ be a function such that $f(n) \leq 1/2$ for all n and $f(n) \in \omega(1/n)$. Assume that a tournament T is generated according to Model 1, and that*

$$p_{i,j} \in [f(n), 1 - f(n)]$$

for all $i \neq j$. Then with high probability, $COND(T) = CNL(T) = A$.

Next, we show that Theorem 1 and Corollary 1 are tight in the sense that if $f(n) \in O(1/n)$, the results no longer hold.

Theorem 2. *Let $c \geq 0$ be a constant. Assume that a tournament T is generated according to Model 1, and that*

$$p_{i,j} \leq \frac{c}{n}$$

for all $i > j$. Then for large enough n, with at least constant probability both $TC(T)$ and $COND(T)$ contain a single alternative. Moreover, for large enough n, with at least constant probability $CNL(T)$ does not contain all alternatives.

Proof. The probability that a_1 dominates all of the remaining alternatives is at least

$$\left(1 - \frac{c}{n}\right)^{n-1} \to e^{-c}$$

as $n \to \infty$. When this occurs, both TC and $COND$ only choose a_1.

An analogous argument shows that a_n is dominated by all of the remaining alternatives with at least constant probability for large enough n. When this occurs, CNL chooses all alternatives except a_n. $\qquad\square$

Theorems 1 and 2 and Corollary 1 allow us to obtain the following corollary on the Condorcet random model.

Corollary 2. *Let $f : \mathbb{Z}^+ \to \mathbb{R}_{\geq 0}$ be a function such that $f(n) \leq 1/2$ for all n. Assume that a tournament T is generated according to Model 1, and that $p_{i,j} = f(n)$ for all $i > j$.*

- *If $f(n) \in \omega(1/n)$, then with high probability, $TC(T) = COND(T) = CNL(T) = A$.*

- *If $f(n) \in o(1/n)$, then with high probability, $TC(T)$ and $COND(T)$ contain a single alternative, and $CNL(T)$ does not contain all alternatives.*

- *If $f(n) \leq c/n$ for some constant $c \geq 0$, then for large enough n, with at least constant probability $TC(T)$ and $COND(T)$ contain a single alternative. Moreover, for large enough n, with at least constant probability $CNL(T)$ does not contain all alternatives.*

Łuczak et al. (1996) also considered the case where $p_{i,j} = c/n$ for all $i > j$ and showed that the probability that TC selects all alternatives converges to $(1 - e^{-c})^2$ in this special case. Our next theorem establishes an analogous result for $COND$ and CNL.

Theorem 3. *Let $c \geq 0$ be a constant. Assume that a tournament T is generated according to Model 1, and that*

$$p_{i,j} = \frac{c}{n}$$

for all $i > j$. Then the probability that $COND(T) = A$ converges to $1 - e^{-c}$ as $n \to \infty$. The same statement holds for CNL.

Proof. We show the result for $COND$; a similar argument holds for CNL. We have

$$\Pr[COND(T) \neq A] = \sum_{i=1}^{n} \Pr[a_i \text{ is a Condorcet winner}]$$

$$= \sum_{i=1}^{n} \left(1 - \frac{c}{n}\right)^{n-i} \left(\frac{c}{n}\right)^{i-1}$$

$$= \left(1 - \frac{c}{n}\right)^{n-1} \cdot \sum_{i=0}^{n-1} \left(\frac{c}{n-c}\right)^{i}.$$

The first term converges to e^{-c} as $n \to \infty$. For the second term, notice that it is always at least 1. Moreover, when $n \geq (k+1)c$ for some positive $k > 1$, the term is at most

$$1 + \frac{1}{k} + \frac{1}{k^2} + \cdots = \frac{k}{k-1},$$

which approaches 1 for large n. Hence the second term converges to 1, and therefore the probability that $COND(T) \neq A$ converges to e^{-c}, yielding the desired result. $\quad\square$

4. Uncovered Set

In this section, we turn our focus to the uncovered set. We show that when each probability $p_{i,j}$ is between $f(n)$ and $1 - f(n)$ for some function $f(n) \geq c\sqrt{\log n/n}$ with $c > \sqrt{2}$ a constant, UC chooses all alternatives with high probability (Theorem 4). As with TC, we also show that our result is asymptotically tight—if $f(n) \leq 0.6\sqrt{\log n/n}$, the statement no longer holds (Theorem 5). It follows that similar results hold for UC^{∞}, implying that $\Theta(\sqrt{\log n/n})$ is the threshold where the two tournament solutions go from almost always choosing all alternatives to excluding at least one alternative with high probability.

Our first result of the section shows that UC chooses the whole set of alternatives for a wide range of distributions over tournaments.

Theorem 4. *Let $c > \sqrt{2}$ be a constant. Assume that a tournament T is generated according to Model 1, and that*

$$p_{i,j} \in \left[c\sqrt{\frac{\log n}{n}}, 1 - c\sqrt{\frac{\log n}{n}}\right]$$

for all $i \neq j$. Then with high probability, $UC(T) = A$.

13

Proof. Choose N such that $\frac{c^2(N-2)}{N} > 2$, and let $n \geq N$. Fix a pair of distinct alternatives a_i, a_j. We first bound the probability that a_i cannot reach a_j via a domination path of length at most two. For each $l \notin \{i, j\}$, the probability that there is an edge from a_i to a_l and an edge from a_l to a_j is at least $\left(c\sqrt{\log n / n}\right)^2 = c^2 \log n / n$. The probability that a_i cannot reach a_j via a domination path of length at most two is therefore bounded above by

$$\left(1 - \frac{c^2 \log n}{n}\right)^{n-2} \leq e^{-\frac{c^2(n-2)\log n}{n}}$$
$$= n^{-\frac{c^2(n-2)}{n}},$$

where we use Lemma 3 for the inequality.

Observe that $UC(T) = A$ exactly when any alternative can reach any other alternative via a domination path of length at most two. Using the union bound over all (ordered) pairs of distinct alternatives i, j, we find that the probability that some alternative cannot reach some other alternative via a domination path of length at most two is no more than

$$n(n-1)n^{-\frac{c^2(n-2)}{n}} \leq n^{2-\frac{c^2(n-2)}{n}},$$

which vanishes for large n. $\qquad\square$

Since the uncovered set is the finest tournament solution satisfying the axioms of Condorcet consistency, neutrality, and expansion (Moulin, 1986), Theorem 4 implies that any tournament solution that satisfies these three axioms also selects all alternatives with high probability when the tournament is generated according to the assumptions of the theorem.

Next, we show that the statement of Theorem 4 breaks down if $f(n) \leq 0.6\sqrt{\log n / n}$, thus confirming that the assumption of the theorem cannot be relaxed asymptotically.

Theorem 5. *Assume that a tournament T is generated according to Model 1, and that*

$$p_{i,j} \leq 0.6\sqrt{\frac{\log n}{n}}$$

for all $i > j$. Then with high probability, $UC(T) \neq A$.

Proof. Let $A_1 = \{a_1, a_2, \ldots, a_{\lfloor n^{0.49} \rfloor}\}$, and let A_2 be the set of alternatives that a_n dominates. We first prove the following claim.

Claim: With high probability, the following two events occur simultaneously: (i) a_n does not dominate any of the alternatives in A_1, and (ii) $|A_2| \leq 0.61\sqrt{n \log n}$.

Proof of Claim: First, using Lemma 2, the probability that a_n dominates at least one of the alternatives in A_1 is at most

$$1 - \left(1 - 0.6\sqrt{\frac{\log n}{n}}\right)^{\lfloor n^{0.49} \rfloor} \leq 1 - \left(1 - 0.6\lfloor n^{0.49} \rfloor \sqrt{\frac{\log n}{n}}\right)$$

14

$$\leq 0.6 \cdot \frac{\sqrt{\log n}}{n^{0.01}},$$

which converges to 0 as $n \to \infty$.

Next, for each $a_i \in A$ with $i = 1, 2, \ldots, n - 1$, let X_i be an indicator random variable that indicates whether a_n dominates a_i or not: X_i takes on the value 1 if a_n dominates a_i and 0 otherwise. We have

$$\mathbb{E}[X_i] \leq 0.6\sqrt{\frac{\log n}{n}} \leq \frac{0.6\sqrt{n \log n}}{n - 1}.$$

Define $X_i' = X_i + \frac{0.6\sqrt{n \log n}}{n-1} - \mathbb{E}[X_i]$, and $X' = \sum_{i=1}^{n-1} X_i'$. We now have

$$\mathbb{E}[X_i'] = \frac{0.6\sqrt{n \log n}}{n - 1} \text{ and } \mathbb{E}[X'] = 0.6\sqrt{n \log n}.$$

Moreover, observe that $|A_2| = \sum_{i=1}^{n-1} X_i$. By Lemma 1, it follows that

$$\begin{aligned}
\Pr\left[|A_2| > 0.61\sqrt{n \log n}\right] &\leq \Pr\left[X' > 0.61\sqrt{n \log n}\right] \\
&= \Pr\left[X' > \frac{0.61}{0.6} \cdot \mathbb{E}[X']\right] \\
&\leq \exp\left(-\left(\frac{0.61}{0.6} - 1\right)^2 \cdot \frac{\mathbb{E}[X']}{3}\right) \\
&= \exp\left(-\frac{1}{3600} \cdot \frac{0.6\sqrt{n \log n}}{3}\right),
\end{aligned}$$

which again vanishes for large n.

Using the union bound over the two events, we have our claim. $\qquad\square$

From now on, we assume that a_n does not dominate any of the alternatives in A_1 and that $|A_2| \leq 0.61\sqrt{n \log n}$. Under this assumption, a_n can reach all of the alternatives in A_1 via a domination path of length at most two if and only if each alternative in A_1 is dominated by some alternative in A_2. Note that the event that this holds for a particular alternative in A_1 is independent of the corresponding events for other alternatives in A_1. It follows that

$$\begin{aligned}
&\Pr\left[a_n \text{ can reach all } a_i \in A_1 \text{ via a domination path of length at most two}\right] \\
&= \Pr\left[a_n \text{ can reach a fixed } a_i \in A_1 \text{ via a domination path of length two}\right]^{\lfloor n^{0.49} \rfloor} \\
&= \left(1 - \Pr\left[\text{a fixed } a_i \in A_1 \text{ dominates } a_j \text{ for all } a_j \in A_2\right]\right)^{\lfloor n^{0.49} \rfloor} \\
&\leq \left(1 - \left(1 - 0.6\sqrt{\frac{\log n}{n}}\right)^{0.61\sqrt{n \log n}}\right)^{\lfloor n^{0.49} \rfloor}
\end{aligned}$$

$$= \left(1 - \left(\left(1 - 0.6\sqrt{\frac{\log n}{n}} \right)^{0.61\sqrt{\frac{n}{\log n}}} \right)^{\log n} \right)^{\lfloor n^{0.49} \rfloor}$$

$$\leq \left(1 - \left(1 - 0.6\sqrt{\frac{\log n}{n}} \cdot 0.61\sqrt{\frac{n}{\log n}} \right)^{\log n} \right)^{\lfloor n^{0.49} \rfloor}$$

$$= \left(1 - 0.634^{\log n} \right)^{\lfloor n^{0.49} \rfloor}$$

$$\leq \left(1 - n^{-0.46} \right)^{\lfloor n^{0.49} \rfloor}$$

$$\leq e^{-n^{-0.46} \lfloor n^{0.49} \rfloor},$$

where we use Lemma 2, the estimate $0.634 > e^{-0.46}$, and Lemma 3 for the second, third, and fourth inequalities, respectively.

Finally, since

$$\lim_{n \to \infty} e^{-n^{-0.46} \lfloor n^{0.49} \rfloor} = 0,$$

the probability that $a_n \notin UC(T)$ converges to 1 as n goes to infinity. This implies that with high probability, $UC(T)$ is not the whole set of alternatives, as desired. $\qquad \square$

Since $UC(T) = A$ exactly when $UC^\infty(T) = A$, we immediately have the following corollary.

Corollary 3. *Assume that a tournament T is generated according to Model 1.*

- *Let $c > \sqrt{2}$ be a constant. If $p_{i,j} \in \left[c\sqrt{\frac{\log n}{n}}, 1 - c\sqrt{\frac{\log n}{n}} \right]$ for all $i \neq j$, then with high probability, $UC^\infty(T) = A$.*

- *If $p_{i,j} \leq 0.6\sqrt{\frac{\log n}{n}}$ for all $i > j$, then with high probability, $UC^\infty(T) \neq A$.*

Theorems 4 and 5 and Corollary 3 allow us to obtain the following corollary on the Condorcet random model.

Corollary 4. *Let $f : \mathbb{Z}^+ \to \mathbb{R}_{\geq 0}$ be a function such that $f(n) \leq 1/2$ for all n. Assume that a tournament T is generated according to Model 1, and that $p_{i,j} = f(n)$ for all $i > j$.*

- *If $f(n) \in \omega\left(\sqrt{\log n/n}\right)$ or $f(n) \geq c\sqrt{\log n/n}$ for some constant $c > \sqrt{2}$, then with high probability, $UC(T) = UC^\infty(T) = A$.*

- *If $f(n) \in o\left(\sqrt{\log n/n}\right)$ or $f(n) \leq 0.6\sqrt{\log n/n}$, then with high probability, $UC(T) \neq A$ and $UC^\infty(T) \neq A$.*

16

5. Majority Tournaments

Thus far, we have established probabilistic results for a general model in which the distribution over tournaments is defined by the probabilities that an alternative dominates another alternative in the tournament (Model 1). As we mentioned in Section 2, a common interpretation of tournaments is as the majority relation of an odd number of voters who are endowed with linear preferences over a set of alternatives. In this section, we investigate a more specific model in which the distribution over tournaments is determined by the probability that a voter prefers an alternative to another alternative (Model 2). It turns out that the generality of our results for Model 1 will allow us to derive similar results for Model 2 as consequences.

We first consider the coarser tournament solutions TC, $COND$, and CNL.

Theorem 6. *Let $f : \mathbb{Z}^+ \to \mathbb{R}_{\geq 0}$ be a function such that $f(n) \leq 1/2$ for all n, and $f(n) \in \omega\left(1/n^{2/(k+1)}\right)$. Assume that a tournament T is generated according to Model 2, and that*

$$q_{v,i,j} \in [f(n), 1 - f(n)]$$

for all voters v and all $i \neq j$. Then with high probability, $TC(T) = COND(T) = CNL(T) = A$.

Proof. Since $TC(T) \subseteq COND(T)$ and $TC(T) \subseteq CNL(T)$, it suffices to prove the statement for TC. Let $c > 0$ be a constant. Since $f(n) \in \omega\left(1/n^{2/(k+1)}\right)$, there exists N' such that $f(n) \geq c/n^{2/(k+1)}$ for all $n \geq N'$. Let $N = \max(N', (2c)^{(k+1)/2})$ and $n \geq N$, and fix a pair of distinct alternatives a_i, a_j. Let $p_{i,j}$ denote the probability that a_i dominates a_j in T. Observe that $p_{i,j}$ is minimized when $q_{v,i,j} = c/n^{2/(k+1)}$ for all voters v. When $q_{v,i,j}$ takes on this value for all v, the probability that a_i dominates a_j is at least the probability that exactly $(k+1)/2$ voters prefer a_i to a_j. The latter probability is

$$
\binom{k}{\frac{k+1}{2}} \left(\frac{c}{n^{\frac{2}{k+1}}}\right)^{\frac{k+1}{2}} \left(1 - \frac{c}{n^{\frac{2}{k+1}}}\right)^{\frac{k-1}{2}}
$$
$$
\geq \binom{k}{\frac{k-1}{2}} \cdot \frac{c^{\frac{k+1}{2}}}{n} \cdot \left(\frac{1}{2}\right)^{\frac{k-1}{2}}
$$
$$
\geq \left(\frac{k}{\frac{k-1}{2}}\right)^{\frac{k-1}{2}} \cdot \frac{c}{n} \cdot \left(\frac{1}{2}\right)^{\frac{k-1}{2}}
$$
$$
\geq 2^{\frac{k-1}{2}} \cdot \frac{c}{n} \cdot \left(\frac{1}{2}\right)^{\frac{k-1}{2}}
$$
$$
\geq \frac{c}{n},
$$

where we use the assumption $n \geq (2c)^{(k+1)/2}$ for the first inequality and the approximation $\binom{n}{k} \geq (n/k)^k$ for the second inequality. Hence $p_{i,j} \geq c/n$ for all $n \geq N$. This implies that there exists a function $f(n) \in \omega(1/n)$ such that $p_{i,j} \in [f(n), 1 - f(n)]$. Using Theorem 1, we have that $TC(T) = A$ with high probability, as desired. $\qquad\square$

We now consider the finer tournament solutions UC and UC^∞.

Theorem 7. *Let $c > \sqrt{2}$ be a constant. Assume that a tournament T is generated according to Model 2, and that*

$$q_{v,i,j} \in \left[c \left(\frac{\log n}{n} \right)^{\frac{1}{k+1}}, 1 - c \left(\frac{\log n}{n} \right)^{\frac{1}{k+1}} \right]$$

for all voters v and all $i \neq j$. Then with high probability, $UC(T) = UC^\infty(T) = A$.

Proof. Since $UC(T) = A$ implies that $UC^\infty(T) = A$, it suffices to prove the statement for UC. Let $n \geq (2c)^{2k+2}$, and fix a pair of distinct alternatives a_i, a_j. Let $p_{i,j}$ denote the probability that a_i dominates a_j in T. Observe that $p_{i,j}$ is minimized when $q_{v,i,j} = c \left(\log n / n \right)^{1/(k+1)}$ for all voters v. When $q_{v,i,j}$ takes on this value for all v, the probability that a_i dominates a_j is at least the probability that exactly $(k+1)/2$ voters prefer a_i to a_j. The latter probability is

$$\binom{k}{\frac{k+1}{2}} \left(c \left(\frac{\log n}{n} \right)^{\frac{1}{k+1}} \right)^{\frac{k+1}{2}} \left(1 - c \left(\frac{\log n}{n} \right)^{\frac{1}{k+1}} \right)^{\frac{k-1}{2}}$$

$$\geq \binom{k}{\frac{k-1}{2}} \left(c \left(\frac{\log n}{n} \right)^{\frac{1}{k+1}} \right)^{\frac{k+1}{2}} \left(\frac{1}{2} \right)^{\frac{k-1}{2}}$$

$$\geq \left(\frac{k}{\frac{k-1}{2}} \right)^{\frac{k-1}{2}} \left(c \left(\frac{\log n}{n} \right)^{\frac{1}{k+1}} \right)^{\frac{k+1}{2}} \left(\frac{1}{2} \right)^{\frac{k-1}{2}}$$

$$\geq 2^{\frac{k-1}{2}} c^{\frac{k+1}{2}} \sqrt{\frac{\log n}{n}} \left(\frac{1}{2} \right)^{\frac{k-1}{2}}$$

$$\geq c \sqrt{\frac{\log n}{n}},$$

where we use the assumption $n \geq (2c)^{2k+2}$ for the first inequality and the approximation $\binom{n}{k} \geq (n/k)^k$ for the second inequality. Using Theorem 4, we have that $UC(T) = A$ with high probability, as desired. $\qquad\square$

6. Experiments

To complement our theoretical results, in this section we investigate the asymptotic behavior of random tournaments according to the Condorcet random model as well as another more realistic model that we call the *gap model*.

6.1. Condorcet Random Model

Starting from a set of alternatives $\{a_1, a_2, \ldots, a_n\}$, we generate random tournaments according to the Condorcet random model by inserting for each pair of alternatives a_i, a_j with $i > j$ an edge from a_i to a_j with probability p and an edge in the reverse direction with probability $1 - p$. The tournament solutions that we consider can all be computed efficiently: A simple counting algorithm suffices to compute $COND$, a depth-first search algorithm computes TC in linear time, and the asymptotic running time for computing UC equals that of matrix multiplication (Hudry, 2009). In our experimental setup, we draw 10000 random tournaments of each size $n \in \{5, 10, 20, 30, \ldots, 100\}$ for each $p \in \{0.5, 0.3, 1/n, 1/n^2, \sqrt{2 \log n / n}, 0.6\sqrt{\log n / n}\}$ and check for each tournament solution $S \in \{COND, UC, TC\}$ whether it selects all alternatives.[9,10] Out of that, we compute the percentage of tournaments in which all alternatives are selected. The resulting graphs are displayed in Figure 1.

For $p = 0.5$, which corresponds to the uniform random model, our experimental results in Figure 1(a) coincide with the main theorem of Fey (2008). The results moreover reveal that UC chooses all alternatives with high probability in tournaments with at least 50 alternatives while $COND$ and TC already do so in much smaller tournaments. As p decreases from 0.5 toward 0, the curves of $COND$, TC, and UC are shifted to the right; this is to be expected since for smaller p the tournament is more skewed, making it more likely for weaker alternatives to be excluded. Nevertheless, for any fixed p the fraction of tournaments in which all alternatives are chosen approaches 1. In particular, when $p = 0.3$, UC almost never rules out any alternative in tournaments of size 100 or more (Figure 1(b)).

Next, we look at the regimes where the probability p goes to 0 as n approaches infinity. For the case of $p = 1/n$ we find that, in line with Theorem 3, the probability that $COND$ selects all alternatives converges to $1 - e^{-1} \approx 0.6321$ (Figure 1(c)). Similarly, the probability that TC selects all alternatives converges to $(1 - e^{-1})^2 \approx 0.3996$ for the same value of p, confirming a result by Łuczak et al. (1996). Letting p approach 0 even faster, we find that for $p = 1/n^2$, both TC and $COND$ are discriminative with high probability (Figure 1(d)). As $1/n^2 \in o(1/n)$, this is consistent with Corollary 2. Note that UC is discriminative for almost all tournaments for both $p = 1/n$ and $p = 1/n^2$; indeed, this is implied by

---

[9]Our setting is slightly different for the last two values of p, as we explain later in this section.

[10]Since the probability that CNL selects all alternatives is equal to the corresponding probability for $COND$ for any fixed n by symmetry, and UC^∞ selects all alternatives exactly when UC does, the results for CNL and UC^∞ are captured by those for $COND$ and UC, respectively.
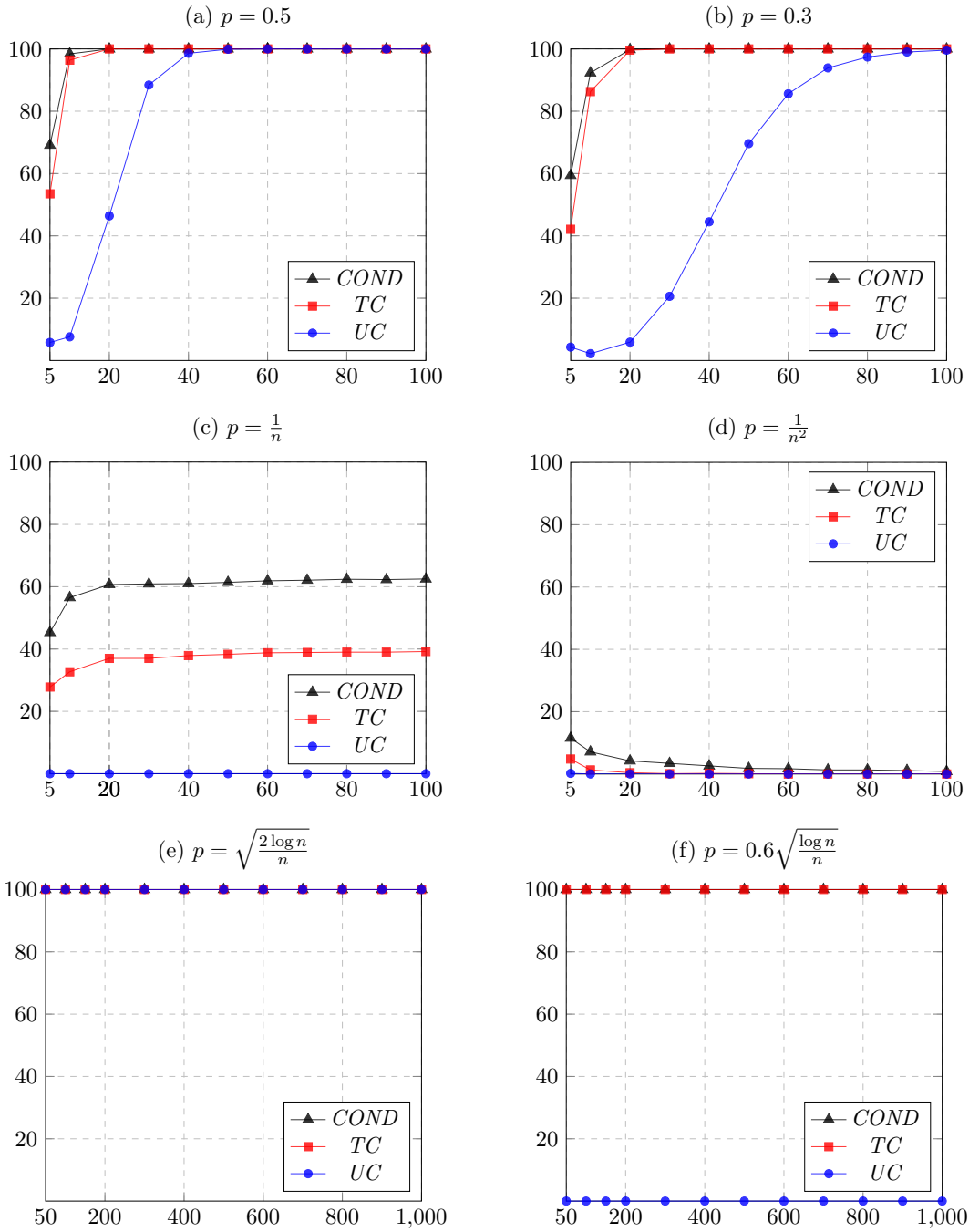
Figure 1: Percentage of tournaments for which the tournament solution chooses the entire set of alternatives in the Condorcet random model, for different values of the probability p. The horizontal and vertical axes correspond to the number of alternatives in the tournament and the percentage, respectively. Averages are taken over 10000 runs.

Corollary 4 since already $1/n \in o\left(\sqrt{\log n/n}\right)$.

Finally, we consider the regime $p = \Theta\left(\sqrt{\log n/n}\right)$, which according to Corollary 4 is the boundary between UC almost never ruling out any alternative and almost always ruling out at least one alternative. The experimental setting for $p = c\sqrt{\log n/n}$ with $c \in \{0.6, \sqrt{2}\}$ differs from the previous settings in that we only examined tournaments of size $n \geq 50$, since for small n the expression $\sqrt{2\log n/n}$ is larger than 0.5, making it unsuitable for our experiments. On the other hand, as p decreases rather slowly, we examined random tournaments up to size 1000 in order to increase the expressive power of our experiments. We find that $COND$ and TC select all alternatives with high probability for both values of c; this is in line with Corollary 2 and the observation that $c\sqrt{\log n/n} \in \omega\left(1/n\right)$. On the other hand, our experiments indicate that UC returns all alternatives in almost all tournaments in the case of $p = \sqrt{2\log n/n}$ (Figure 1(e)) but is discriminative in almost all tournaments when $p = 0.6\sqrt{\log n/n}$ (Figure 1(f)). These findings coincide with Corollary 4 and demonstrate the interesting fact that a small gap in the constant factor constitutes the threshold with regard to the discriminative power of UC.

6.2. Gap Model

As we explained in the introduction, while the Condorcet random model is commonly used in theoretical analyses, it does not properly capture tournaments in the real world since it assigns the same probability to all edges regardless of the difference in strength between the two alternatives adjacent to that edge. We next consider a different model, which we call the *gap model*, that takes this issue into account.

Like in the Condorcet random model, in the gap model there is a linear order of alternatives from strongest to weakest as well as a parameter $p \leq 0.5$. However, the probability that a stronger alternative dominates a weaker alternative depends linearly on the size of the gap between the two alternatives in the linear order: For two alternatives a_i, a_j with $i < j$, there is an edge from a_i to a_j with probability $0.5 + \frac{(0.5-p)(j-i)}{n-1}$, and an edge in the reverse direction with the remaining probability. In particular, there is an edge from a_1 to a_n with probability $1-p$. We perform experiments on the gap model using the same values of p as we did for the Condorcet random model. The only exception is $p = 0.5$ which we replace by $p = 0$, the reason being that both models coincide when $p = 0.5$. Moreover, we double the sizes of the tournaments considered for the first four values of p in order to better illustrate the convergence behavior. The resulting graphs are displayed in Figure 2.

We now make some observations. Firstly, in Figure 2(b), we find that for $p = 0.3$ all three tournament solutions are unlikely to exclude any alternative in large tournaments; this is consistent with Theorems 1 and 4. The same phenomenon occurs for $p = \sqrt{2\log n/n}$ (Figure 2(e)). We remark that these phenomena cannot be explained by any theoretical result prior to our work. In the remaining figures, we likewise find that in the gap model, all tournament solutions cease to be discriminative as the size of the tournament grows. This
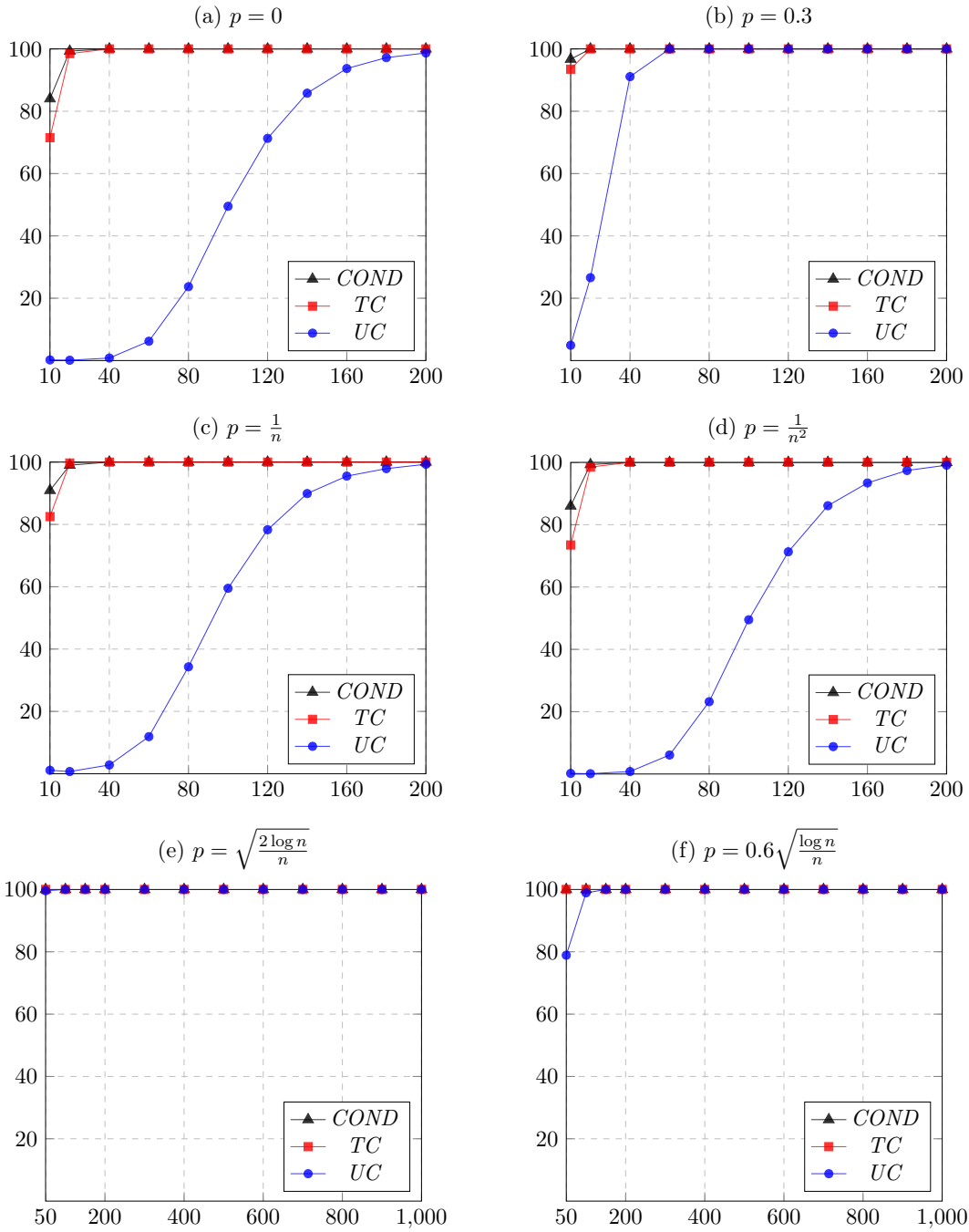
Figure 2: Percentage of tournaments for which the tournament solution chooses the entire set of alternatives in the gap model, for different values of the probability p. The horizontal and vertical axes correspond to the number of alternatives in the tournament and the percentage, respectively. Averages are taken over 10000 runs.

is the case even for the extreme value $p = 0$ (Figure 2(a)); the Condorcet random model for this value of p clearly always produces a Condorcet winner. Intuitively, the reason that all alternatives are likely to be chosen is that in the gap model, the overall difference in strength between alternatives is significantly less than in the Condorcet random model. Note that the observations for the latter values of p are not captured by our results, since our theorems require all edge probabilities to be in the range $[p, 1-p]$ for appropriate values of p. Indeed, an intriguing direction for future work would be to generalize our results so that some edge probabilities are allowed to be outside of this range.[11]

7. Conclusion

In this paper, we investigate the behavior of a number of tournament solutions in large random tournaments under a general probabilistic model. We establish tight asymptotic bounds on the boundary of the probability range for which each tournament solution is unlikely to exclude any alternative. In particular, we illustrate a difference between the discriminative power of the top cycle and the uncovered set; this difference is not evident in previous studies that focused on more restricted models. Indeed, while both tournament solutions include all alternatives with high probability in the uniform random model, our results suggest that the uncovered set is in fact considerably more discriminative than the top cycle.

Our work leaves many interesting open questions for future study. A natural next step would be to investigate the asymptotic behavior of other tournament solutions that have been previously studied in the uniform random model—including the Banks set (Fey, 2008), the minimal covering set (Scott and Fey, 2012), and the bipartisan set (Fisher and Ryan, 1995)—using our general probabilistic model. For instance, it is conceivable that the approach used by Fey (2008) to show that the Banks set almost never rules out any alternative in the uniform random model can be extended to establish an analogous statement when each edge probability is drawn from some constant range. It is not clear, however, whether the approach would still work if we allow the range to depend on the number of alternatives in the tournament like we do in the current work.

From a broader point of view, we believe that an important direction is to apply our model to other tournament problems beyond those concerning tournament solutions, for example the problem of finding a dominating set of minimum size. It is well-known that a dominating set of size at most $\log_2(n + 1)$ always exists and can be found using a simple greedy algorithm. While a dominating set can be as small as a singleton in tournaments that admit a Condorcet winner, Scott and Fey (2012) showed that for uniform random tournaments, a dominating set of logarithmic size is the best that one can hope for. More precisely, these authors showed that given any constant $0 < c < 1$, the smallest dominating set of a tournament chosen uniformly at random contains at least $c \log_2 n$ alternatives with

---

[11]A result of this flavor has been shown by Kim et al. (2017) in the context of single-elimination winners.

high probability. Establishing a similar result in our general probabilistic model is an intriguing technical challenge that would allow us to better understand the behavior of such structures in the real world.

Acknowledgments

References

S. Allesina and J. M. Levine. A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences (PNAS)*, 108(14):5638–5642, 2011.

K. J. Arrow and H. Raynaud. *Social Choice and Multicriterion Decision-Making*. MIT Press, 1986.

C. E. Bell. A random voting graph almost surely has a Hamiltonian cycle when the number of alternatives is large. *Econometrica*, 49(6):1597–1603, 1981.

D. Bouyssou. Monotonicity of 'ranking by choosing': A progress report. *Social Choice and Welfare*, 23(2):249–273, 2004.

F. Brandt and H. G. Seedig. On the discriminative power of tournament solutions. In *Selected Papers of the International Conference on Operations Research, OR2014*, Operations Research Proceedings, pages 53–58. Springer-Verlag, 2016.

F. Brandt, M. Brill, and P. Harrenstein. Tournament solutions. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 3. Cambridge University Press, 2016.

F. Brandt, M. Brill, H. G. Seedig, and W. Suksompong. On the structure of stable tournament solutions. *Economic Theory*, 65(2):483–507, 2018.

T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 3rd edition, 2009.

M. Fey. Choosing from a large tournament. *Social Choice and Welfare*, 31(2):301–309, 2008.

D. C. Fisher and J. Ryan. Tournament games and positive tournaments. *Journal of Graph Theory*, 19(2):217–236, 1995.

O. Frank. Stochastic competition graphs. *Review of the International Statistical Institute*, 36(3):319–326, 1968.

I. J. Good. A note on Condorcet sets. *Public Choice*, 10(1):97–101, 1971.

O. Hudry. A survey on the complexity of tournament solutions. *Mathematical Social Sciences*, 57(3):292–303, 2009.

M. P. Kim, W. Suksompong, and V. Vassilevska Williams. Who can win a single-elimination tournament? *SIAM Journal on Discrete Mathematics*, 31(3):1751–1764, 2017.

H. G. Landau. On dominance relations and the structure of animal societies: III. The condition for a score structure. *Bulletin of Mathematical Biophysics*, 15(2):143–148, 1953.

J.-F. Laslier. *Tournament Solutions and Majority Voting*. Springer-Verlag, 1997.

J.-F. Laslier. In silico voting experiments. In J.-F. Laslier and M. R. Sanver, editors, *Handbook on Approval Voting*, chapter 13, pages 311–335. Springer-Verlag, 2010.

T. Łuczak, A. Ruciński, and J. Gruszka. On the evolution of a random tournament. *Discrete Mathematics*, 148(1–3):311–316, 1996.

S. B. Maurer. The king chicken theorems. *Mathematics Magazine*, 53:67–80, 1980.

N. R. Miller. Graph-theoretic approaches to the theory of voting. *American Journal of Political Science*, 21(4):769–803, 1977.

N. R. Miller. A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, 24(1):68–96, 1980.

J. W. Moon. *Topics on Tournaments*. Holt, Reinhard and Winston, 1968.

J. W. Moon and L. Moser. Almost all tournaments are irreducible. *Canadian Mathematical Bulletin*, 5:61–65, 1962.

H. Moulin. Choosing from a tournament. *Social Choice and Welfare*, 3(4):271–291, 1986.

K. B. Reid. Every vertex a king. *Discrete Mathematics*, 38(1):93–98, 1982.

T. Schjelderup-Ebbe. Beiträge zur Sozialpsychologie des Haushuhns. *Zeitschrift für Psychologie*, 88:225–252, 1922.

T. Schwartz. Rationality and the myth of the maximum. *Noûs*, 6(2):97–117, 1972.

A. Scott and M. Fey. The minimal covering set in large tournaments. *Social Choice and Welfare*, 38(1):1–9, 2012.

P. Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48(3–4):303–312, 1961.

I. A. Ushakov. The problem of choosing the preferred element: An application to sport games. In R. E. Machol, S. P. Ladany, and D. G. Morrison, editors, *Management Science in Sports*, pages 153–161. North-Holland, 1976.

V. Vassilevska Williams. Fixing a tournament. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, pages 895–900. AAAI Press, 2010.

A. Appendix

A.1. Proof of Lemma 5

Before we prove the lemma, we first give a characterization of when one vector majorizes another. To this end, we define the notion of an equalizing move, which involves taking two components of a vector and bringing them "closer together".

Definition 3. Given a vector $\mathbf{x}$, an *equalizing move* on $\mathbf{x}$ takes two components $x_i > x_j$ and replaces them by $x_i - 1$ and $x_j + 1$, respectively.

Lemma 7. *For any vector $\mathbf{x}$, a vector $\mathbf{y}$ can be obtained from $\mathbf{x}$ by a finite number of equalizing moves if and only if $\mathbf{x} \succ \mathbf{y}$.*

Proof. The direction from left to right follows from the observation that an equalizing move never decreases the sum of the j highest components of the vector for any $j = 1, 2, \ldots, n-1$, and leaves the sum of all components invariant.

For the converse direction, we proceed by induction on n, the number of components of the vectors. The base case $n = 1$ holds trivially. Assume that the statement holds when there are at most $n-1$ components. To prove the statement when there are n components, assume for contradiction that it does not hold for some pairs $\mathbf{x}, \mathbf{y}$, i.e., $\mathbf{x} \succ \mathbf{y}$ but $\mathbf{y}$ cannot be obtained from $\mathbf{x}$ by a finite number of equalizing moves. For each pair assume without loss of generality that $x_1 \geq \cdots \geq x_n$ and $y_1 \geq \cdots \geq y_n$, and consider the pairs that minimize the difference $x_1 - y_1$ among all such pairs; this difference is guaranteed to be nonnegative by condition (i) of Definition 1. Among all of the pairs under consideration,

take one that minimizes the largest index k such that $x_k = x_1$. Let l be the smallest index such that $\sum_{i=1}^{l} x_i = \sum_{i=1}^{l} y_i$; the existence of l is guaranteed by condition (ii) of Definition 1. If $l < n$, we can apply the induction hypothesis on the first l components and the last $n-l$ components separately to obtain $\mathbf{y}$ from $\mathbf{x}$ using a finite number of equalizing moves, which would be a contradiction. Hence we may assume that $l = n$. In particular, $\sum_{i=1}^{j} x_i \geq \sum_{i=1}^{j} y_i + 1$ for all $j = 1, 2, \ldots, n-1$.

Let m be the smallest index such that $x_m = x_n$. If $x_k = x_m$ or $x_k = x_m + 1$, then the only vector with nonincreasing components that is majorized by $\mathbf{x}$ is $\mathbf{x}$ itself, a contradiction. So $x_k \geq x_m + 2$, and we may replace (x_k, x_m) by $(x_k - 1, x_m + 1)$ in an equalizing move. Let $\mathbf{x}' = (x_1', x_2', \ldots, x_n')$ be the vector resulting from this move, i.e., $x_k' = x_k - 1$, $x_m' = x_m + 1$, and $x_i' = x_i$ for all $i \notin \{k, m\}$. By definition of k and m, we have $x_1' \geq \cdots \geq x_n'$. Moreover, $\sum_{i=1}^{n} x_i' = \sum_{i=1}^{n} y_i$, and for any $j = 1, 2, \ldots, n-1$ we have $\sum_{i=1}^{j} x_i' \geq \sum_{i=1}^{j} x_i - 1 \geq \sum_{i=1}^{j} y_i$. This means that $\mathbf{x}' \succ \mathbf{y}$. If $k = 1$, we have $x_1' - y_1 < x_1 - y_1$, which means that we can make a sequence of equalizing moves on $\mathbf{x}'$ to obtain $\mathbf{y}$, a contradiction. Else, if $k > 1$, then we have $x_1' - y_1 = x_1 - y_1$ and $x_k' < x_1'$, so we can again make a sequence of equalizing moves on $\mathbf{x}'$ to obtain $\mathbf{y}$ and arrive at a contradiction, completing our proof. $\qquad\square$

We now proceed to the proof of Lemma 5. Suppose that $\mathbf{x} \succ \mathbf{y}$, and fix $k \in \{1, 2, \ldots, n\}$. By Lemma 7, there exists a sequence of equalizing moves that takes $\mathbf{x}$ to $\mathbf{y}$. It suffices to show that if an equalizing move takes $\mathbf{x}$ to $\mathbf{x}'$, then there is a corresponding sequence of equalizing moves that takes $\mathbf{x}^{(k)}$ to $\mathbf{x}'^{(k)}$. Indeed, if this is true, then the sequence of equalizing moves that takes $\mathbf{x}$ to $\mathbf{y}$ gives rise to a corresponding sequence of equalizing moves that takes $\mathbf{x}^{(k)}$ to $\mathbf{y}^{(k)}$. By Lemma 7 again, this will imply that $\mathbf{x}^{(k)} \succ \mathbf{y}^{(k)}$.

Consider an equalizing move that takes $\mathbf{x}$ to $\mathbf{x}'$; assume that the move replaces the components $x_i > x_j$ by $x_i - 1$ and $x_j + 1$, respectively. Note that the only components that change between $\mathbf{x}^{(k)}$ and $\mathbf{x}'^{(k)}$ are the ones that contain exactly one of x_i and x_j in their sum. These components can be paired up in such a way that for each pair, one component contains x_i, the other component contains x_j, and both components contain exactly the same subset of the remaining x_l's with $l \notin \{i, j\}$. For each pair, replacing x_i and x_j by $x_i - 1$ and $x_j + 1$ corresponds to an equalizing move. It follows that there exists a sequence of equalizing moves that takes $\mathbf{x}^{(k)}$ to $\mathbf{x}'^{(k)}$, as claimed.

A.2. Proof of Lemma 6

Note that $(\deg_W(d_1), \deg_W(d_2), \ldots, \deg_W(d_n)) = (n-1, n-2, \ldots, 0)$. We verify that both conditions in Definition 1 are satisfied.

Fix $k \in \{1, 2, \ldots, n-1\}$, and assume without loss of generality that $\deg_U(b_1) \geq \cdots \geq \deg_U(b_n)$. Let $B = \{b_1, b_2, \ldots, b_n\}$ and $B' = \{b_1, b_2, \ldots, b_k\}$. The number of edges from an alternative in B' to another alternative in B' is exactly $\binom{k}{2}$. On the other hand, the number of edges from an alternative in B' to an alternative in $B \backslash B'$ is at most $k(n-k)$.

27

It follows that

$$\deg_U(b_1) + \deg_U(b_2) + \cdots + \deg_U(b_k) \leq \binom{k}{2} + k(n-k)$$
$$= k\left(n - \frac{k+1}{2}\right)$$
$$= (n-1) + (n-2) + \cdots + (n-k),$$

so condition (i) is satisfied.

Finally, observe that

$$\deg_U(b_1) + \deg_U(b_2) + \cdots + \deg_U(b_n) = \binom{n}{2} = (n-1) + (n-2) + \cdots + 0,$$

so condition (ii) is also satisfied.