

The tree search game for two players

RAVI B. BOPPANA

*Department of Mathematics, Massachusetts Institute of Technology
Cambridge, Massachusetts
U.S.A.
rboppana@mit.edu*

JOEL BREWSTER LEWIS

*Department of Mathematics, The George Washington University
Washington, DC
U.S.A.
jblewis@gwu.edu*

Abstract

We consider a two-player search game on a tree T . One vertex (unknown to the players) is randomly selected as the target. The players alternately guess vertices. If a guess v is not the target, then both players are informed in which subtree of $T \setminus v$ the target lies. The winner is the player who guesses the target.

When both players play optimally, we show that each of them wins with probability approximately $1/2$. When one player plays optimally and the other plays randomly, we show that the player with the optimal strategy wins with probability between $9/16$ and $2/3$ (asymptotically). When both players play randomly, we show that each wins with probability between $13/30$ and $17/30$ (asymptotically).

1 Introduction

We consider the following competitive variant of traditional binary search: two players seek an (unknown, uniformly random) element of the set $\{1, \dots, n\}$. The players alternately guess elements of the set; if a guess is incorrect, then both players are informed whether the secret number is larger or smaller than the guess. The winner is the player who guesses the secret number.

We consider the following more general variant of the game, using a model of binary search on trees introduced by Onak and Parys [5]: the starting position is not

a set of n numbers, but rather a labeled tree T on n vertices, one vertex of which has been selected (uniformly at random) as the target. The two players alternately choose vertices; if a guess v is incorrect, then both players are informed in which subtree of $T \setminus v$ the target vertex lies. The winner is the player who guesses the target vertex. One immediately recovers the previous game upon choosing T to be a path on n vertices.

We consider the variant of the game in which both players play according to an optimal strategy, as well as variants where one or both players play uniformly at random. When both players play optimally, we completely analyze the game: if n (the number of vertices) is even, then the game is fair, while if n is odd, then the first player wins with probability $\frac{1}{2} + \frac{1}{2n}$, regardless of the structure of the tree. We also describe all optimal strategies in this case.

When one or both players play randomly (selecting a vertex uniformly at random from among the vertices that could be the target), the probabilities of winning are surprisingly more complicated to analyze. We complete the analysis in the case of paths and stars, and conjecture that the probability of a first-player win always lies between these two extremes (Conjectures 4.9 and 5.3); we are able to establish the conjectures in some cases. Specifically, when one player plays optimally and the other plays randomly, we show that the optimally playing competitor wins with probability between $9/16$ and $2/3$ (asymptotically). When both players play uniformly at random, we show that each wins with probability between $13/30$ and $17/30$ (for $n \geq 2$).

The structure of the paper is as follows: in Section 2, we establish the terminology and notation that is used throughout the paper as well as some basic lemmas. In Section 3, we completely analyze the game in the case of two optimally playing competitors. In Section 4, we study the game in the case that one player plays according to an optimal strategy while the other chooses vertices uniformly at random. In Section 5, we study the game in the case that *both* players play uniformly at random. Finally, in Section 6, we give a number of open problems including variants of the game that we believe might be of interest.

2 Background and notation

We begin by establishing some basic terminology and notation for the rest of the paper.

As usual, a *tree* is a connected acyclic graph. Our trees are undirected and unrooted, with a finite but positive number of vertices. We denote by $V(T)$ the vertex set of tree T . The *order* $|T|$ of tree T is $|V(T)|$, the number of vertices of T . We denote by $E(T)$ the edge set of tree T . We denote by $v \sim w$ the relation that vertices v and w are joined by an edge of the tree.

The *degree* $\deg(v)$ of a vertex v of a tree T is the number of edges of T incident to v . A *leaf* is a vertex of degree 1. Every tree T of order n has $n - 1$ edges, and the

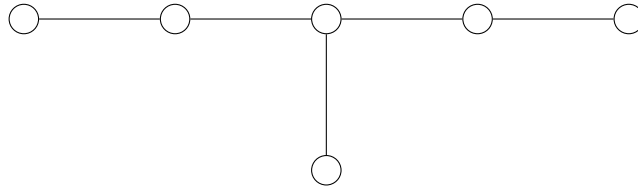


Figure 1: The unique tree for which the inequality in Lemma 2.1(g) fails

sum of the vertex degrees of T is $2(n - 1)$. Given a vertex v of T , let $T \setminus v$ be the graph that results from deleting v and its edges from T ; the graph $T \setminus v$ is a forest with $\deg(v)$ components, each of which is a tree. Given an edge e of T , let $T \setminus e$ be the graph that results from deleting e from T ; the graph $T \setminus e$ is a forest with two components, each of which is a tree.

Denote by P_n the path graph on n vertices and by S_n the star graph on n vertices. In particular, $P_1 = S_1$, $P_2 = S_2$, and $P_3 = S_3$ are the unique trees on 1, 2, and 3 vertices, respectively.

At several points, we will be concerned with counting leaves and other small subtrees near “the boundary” of a given tree. To that end, define for each tree T and each integer k the *limb set*

$$L_k(T) = \{(v, T') : v \in V(T) \text{ and } T' \text{ is a component of } T \setminus v \text{ of order } k\}$$

and the *limb number*

$$\ell_k(T) = |L_k(T)|.$$

If $k \leq 0$ or $k \geq |T|$, then $\ell_k(T) = 0$. We record below some basic information about the limb numbers.

Lemma 2.1. *Let T be a tree on n vertices.*

- (a) $\ell_1(T)$ is the number of leaves of T .
- (b) If $n > 1$, then $\ell_1(T) \geq 2$.
- (c) For every integer k , we have $\ell_k(T) = \ell_{n-k}(T)$.
- (d) The sum $\sum_k \ell_k(T)$ is $2(n - 1)$.
- (e) The sum $\sum_k k\ell_k(T)$ is $n(n - 1)$.
- (f) For every integer k , we have $\ell_1(T) \geq \ell_k(T)$.
- (g) If T is not the tree in Figure 1, then $\ell_1(T) + \ell_3(T) \geq \ell_2(T) + \ell_4(T)$.

Proof.

- (a) The map that sends the leaf w to the pair (v, T') where v is the unique neighbor of w and T' is the tree with vertex set $\{w\}$ is a bijection between the set of leaves of T and $L_1(T)$. Hence $\ell_1(T)$ is the number of leaves of T .
- (b) It is a standard result that may be found in many textbooks on graph theory that every finite tree with more than one vertex has at least two leaves; for example, see [8, Lemma 2.1.3]. A short proof: since the n vertex degrees are positive integers and their sum is $2(n - 1)$, at least two of them must be equal to 1.
- (c) We index the (disjoint) union $\bigcup_k L_k(T)$ as follows: if v is a vertex of T and w is a neighbor of v , then let $T_{v,w}$ be the component of $T \setminus v$ that contains w . Thus

$$L_k(T) = \{(v, T_{v,w}) : v \sim w \text{ and } |T_{v,w}| = k\}.$$

Now fix an edge $e = \{v, w\}$ of T . Since $T_{v,w} \cup T_{w,v} = T \setminus e$, one has $|T_{v,w}| + |T_{w,v}| = n$. It follows that $(v, T_{v,w}) \in L_k(T)$ if and only if $(w, T_{w,v}) \in L_{n-k}(T)$, and hence $\ell_k(T) = \ell_{n-k}(T)$.

- (d) The map that sends the pair $(v, T_{v,w})$ to the edge $\{v, w\}$ is a two-to-one function from the union $\bigcup_k L_k(T)$ to the edge set $E(T)$. Hence $\sum_k \ell_k(T) = 2|E(T)| = 2(n - 1)$.
- (e) By part (c), we have

$$\begin{aligned} 2 \sum_k k \ell_k(T) &= \sum_k k \ell_k(T) + \sum_k k \ell_{n-k}(T) \\ &= \sum_k k \ell_k(T) + \sum_k (n - k) \ell_k(T) \\ &= n \sum_k \ell_k(T). \end{aligned}$$

Dividing by 2 and using part (d) gives

$$\sum_k k \ell_k(T) = \frac{n}{2} \sum_k \ell_k(T) = \frac{n}{2} \cdot 2(n - 1) = n(n - 1).$$

- (f) If $k \leq 0$, then $\ell_k(T) = 0 \leq \ell_1(T)$. If $k = 1$, then $\ell_k(T) = \ell_1(T)$. Hence we may assume that $k > 1$. By part (c), we may further assume that $k \leq \frac{n}{2}$.

Suppose that (v', T') and (v'', T'') are distinct elements of $L_k(T)$. We claim that T' and T'' are vertex-disjoint. This clearly holds when $v' = v''$, so we may assume that $v' \neq v''$. We have $T' = T_{v',w'}$ for some neighbor w' of v' and $T'' = T_{v'',w''}$ for some neighbor w'' of v'' . Are w' and/or w'' on the (unique) path between v' and v'' in T ? The four cases are illustrated in Figure 2. If either w' or w'' (but not both) were on the path between v' and v'' (Case 1 or 2), then one of the

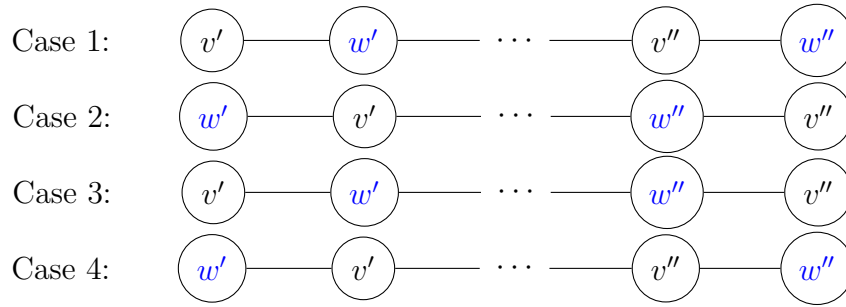


Figure 2: The four cases in the proof of Lemma 2.1(f)

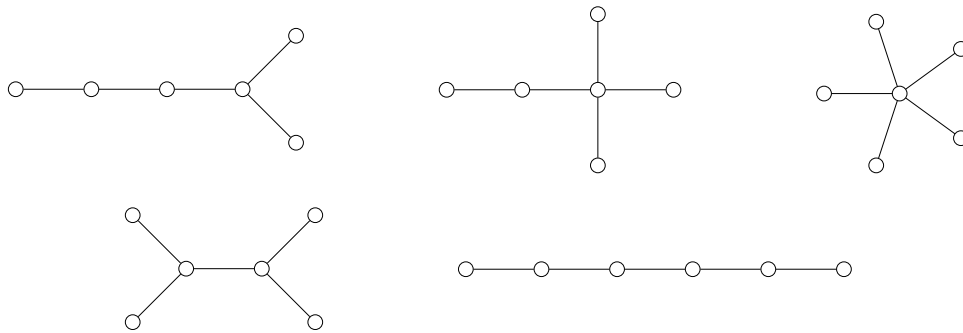


Figure 3: The trees with six vertices other than the tree in Figure 1, with associated limb numbers $(\ell_1(T), \ell_2(T), \ell_3(T), \ell_4(T)) = (3, 1, 2, 1), (4, 1, 0, 1), (5, 0, 0, 0), (4, 0, 2, 0),$ and $(2, 2, 2, 2)$

trees T' and T'' would strictly contain the other, which is impossible since they have equal orders. If both w' and w'' are on the path (Case 3), which includes as special cases the situation $w' = w''$ and the situation $w' = v''$ and $w'' = v'$, then $V(T') \cup V(T'') = V(T)$, while $|T'| + |T''| = 2k \leq n$; hence $V(T')$ and $V(T'')$ partition $V(T)$ and in particular are disjoint. If neither w' nor w'' is on the path (Case 4), then T' and T'' are disjoint. In every case, T' and T'' are disjoint.

By part (b), since $k > 1$, for every pair $(v, T_{v,w})$ in $L_k(T)$, there is a vertex $\phi(v, T_{v,w})$ different from w that is a leaf of subtree $T_{v,w}$ and hence a leaf of T . By vertex disjointness, the map ϕ is an injective function from $L_k(T)$ to the set of leaves of T . Consequently, by part (a), we have $\ell_k(T) \leq \ell_1(T)$.

- (g) If $n \leq 4$, then $\ell_4(T) = 0$ and part (f) implies that $\ell_1(T) \geq \ell_2(T)$. If $n = 5$, then part (c) implies that $\ell_1(T) = \ell_4(T)$ and $\ell_2(T) = \ell_3(T)$. If $n = 6$, since T is not the tree in Figure 1, T is one of the trees in Figure 3, and one checks by hand that $\ell_1(T) + \ell_3(T) \geq \ell_2(T) + \ell_4(T)$ in all cases. If $n = 7$, then parts (c) and (f) imply that $\ell_3(T) = \ell_4(T)$ and $\ell_1(T) \geq \ell_2(T)$. Now suppose that $n \geq 8$. We construct an injective map $\phi : L_2(T) \cup L_4(T) \hookrightarrow L_1(T) \cup L_3(T)$. Given a pair $(v, T') \in L_2(T) \cup L_4(T)$, let w be the neighbor of v in T' . The graph $T' \setminus w$ is a forest on an odd number of vertices, so it has a connected component T'' of

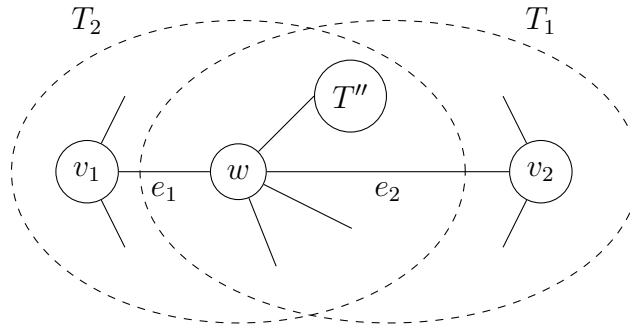


Figure 4: The form of tree T in the proof of Lemma 2.1(g)

odd order; define $\phi(v, T') = (w, T'')$. By construction, this map has the correct domain and range; it remains to show that it is injective. Suppose otherwise, so that $\phi(v_1, T_1) = \phi(v_2, T_2) = (w, T'')$ and $(v_1, T_1) \neq (v_2, T_2)$. Then by the definition of ϕ , the tree T has the form illustrated in Figure 4, namely, v_1 is adjacent to w by an edge e_1 and T_1 is the component of containing w when e_1 is removed from T , and v_2 is adjacent to w by an edge e_2 and T_2 is the component containing w when e_2 is removed from T . In this case, for any vertex v of T , at most one of e_1 and e_2 lies on the path connecting v to w , and therefore $V(T_1) \cup V(T_2) = V(T)$ and $w \in V(T_1) \cap V(T_2)$. However, this contradicts the hypotheses $|T| \geq 8$, $|T_1| \leq 4$, and $|T_2| \leq 4$. Thus ϕ is injective, as claimed. \square

In upcoming parts of the paper, the following type of tree will be useful. A *spider* is a tree with exactly one vertex of degree bigger than 2. (In the literature, spiders are also sometimes called *star-like trees* or *subdivisions of stars*.) The vertex of degree bigger than 2 is called the *head* of the spider. Let T be a spider with head h . Each component of $T \setminus h$ is called a *leg* of the spider. Each leg is a path, and the number of legs of the spider is the degree of its head. The *length* of a leg is its order, the number of vertices on it.

Given an integer d bigger than 2 and a list $\lambda_1, \dots, \lambda_d$ of positive integers, let $S_{\lambda_1, \dots, \lambda_d}$ denote the spider with a head of degree d and legs of length $\lambda_1, \dots, \lambda_d$. Thus the order of $S_{\lambda_1, \dots, \lambda_d}$ is $1 + \sum_i \lambda_i$. For example, the tree in Figure 1 is the spider $S_{2,2,1}$, and the three trees in the first row of Figure 3 are the spiders $S_{3,1,1}$, $S_{2,1,1,1}$, and $S_{1,1,1,1,1}$.

The next lemma will use paths and spiders to characterize trees with two or three leaves. We will use the following indicator notation: $\mathbf{1}_{\text{true}} = 1$ and $\mathbf{1}_{\text{false}} = 0$.

Lemma 2.2.

- (a) If T is a tree such that $\ell_1(T) = 2$, then T is a path.
- (b) If T is the path P_n and $1 \leq k \leq n - 1$, then $\ell_k(P_n) = 2$.
- (c) If T is a tree such that $\ell_1(T) = 3$, then T is a spider with three legs.

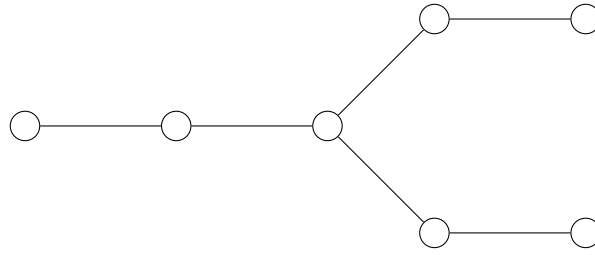


Figure 5: The unique tree for which $\ell_1 = 3$, $\ell_2 = 3$, and $\ell_3 = 0$

(d) If T is the spider $S_{\lambda_1, \dots, \lambda_d}$ and $1 \leq k \leq \sum_{i=1}^d \lambda_i$, then

$$\ell_k(T) = \sum_{i=1}^d \mathbf{1}_{k \leq \lambda_i} + \sum_{i=1}^d \mathbf{1}_{k \geq n - \lambda_i}.$$

(e) If T is a tree such that $\ell_1(T) = 3$, $\ell_2(T) = 3$, and $\ell_3(T) = 0$, then T is the spider $S_{2,2,2}$, shown in Figure 5.

Proof.

- (a) From Lemma 2.1(a), $\ell_1(T)$ is the number of leaves of T . It is a standard exercise [8, Ex. 2.1.18] that the number of leaves of a tree is at least its maximum degree Δ , since $2(n - 1) = \sum_v d(v) \geq \ell_1(T) \cdot 1 + 1 \cdot \Delta + 2 \cdot (n - \ell_1(T) - 1)$. Thus if $\ell_1(T) = 2$, the tree T has maximum degree at most 2, i.e., it is a path.
- (b) Labelling the path $v_1 \sim v_2 \sim \dots \sim v_n$ and using the indexing from the proof of Lemma 2.1(c), we have for $k = 1, \dots, n - 1$ that $L_k(T) = \{(v_{k+1}, T_{v_{k+1}, v_k}), (v_{n-k}, T_{v_{n-k}, v_{n-k+1}})\}$.
- (c) If T has three leaves, then it is not a path, and so some vertex has degree at least 3. By the argument in part (a), in fact the maximum degree must be exactly 3. Refining the argument in part (a), if we have equality $\Delta = \ell_1(T)$ then it must be the case that all vertices other than the leaves and a single vertex of degree Δ have degree 2, and so the graph is a spider with Δ legs [8, Ex. 2.1.59(a)].
- (d) Deleting the vertices of the leg of length λ_i contributes $\mathbf{1}_{k < \lambda_i} + \mathbf{1}_{k \geq n - \lambda_i}$, with the first term counting the contribution of the paths that do not contain the head and the second term counting the subtrees containing the head; deleting the head contributes $\sum_i \mathbf{1}_{k = \lambda_i}$. Summing up over all legs gives the result.
- (e) Let T be a tree such that $\ell_1(T) = \ell_2(T) = 3$ and $\ell_3(T) = 0$. By part (c), since $\ell_1(T) = 3$, $T = S_{\lambda_1, \lambda_2, \lambda_3}$ is a spider with three legs. Since $\ell_3(T) = 0$, we have by part (d) that $\lambda_i < 3$ (because the first sum must not have any positive terms). Finally one easily checks by hand that among the remaining possibilities $S_{1,1,1}$, $S_{2,1,1}$, $S_{2,2,1}$, and $S_{2,2,2}$, only the last has $\ell_2(T) = 3$. □

3 Two players with optimal strategy

In this section, we consider the tree search game when both players are trying to maximize their probability of winning. The tree search game is a finite, two-player, zero-sum, sequential game, and so both players have an optimal deterministic strategy. For a tree T , let $\mathcal{P}(T)$ be the probability that the first player wins on T when both players play with an optimal strategy. Our first theorem exactly computes this win probability for every tree.

Theorem 3.1. *If T is a tree on n vertices, then*

$$\mathcal{P}(T) = \frac{1}{2} + \frac{1}{2n} \cdot \mathbf{1}_{n \text{ is odd}},$$

where $\mathbf{1}_{n \text{ is odd}}$ denotes the indicator function for the event that n is odd.

Proof. The result is easy to verify for $n = 1$. Let T be a tree on $n > 1$ vertices, and suppose the result is true for all trees on fewer than n vertices. For a vertex v of T , let $\mathcal{P}(T, v)$ be the probability that the first player wins on T if their first guess is v , so $\mathcal{P}(T) = \max_{v \in V(T)} \mathcal{P}(T, v)$.

With probability $\frac{1}{n}$, the first guess is correct. Otherwise, the target vertex lies in one of the $k = \deg(v)$ components T_1, \dots, T_k of the forest $T \setminus v$, and the probability that it lies in component T_i is $\frac{|T_i|}{n}$. Moreover, in this case, the second player wins with probability $\mathcal{P}(T_i)$, and so the first player wins with probability $1 - \mathcal{P}(T_i)$. Thus

$$\mathcal{P}(T, v) = \frac{1}{n} + \sum_{i=1}^k \frac{|T_i|}{n} \cdot (1 - \mathcal{P}(T_i)) = 1 - \frac{1}{n} \sum_{i=1}^k |T_i| \cdot \mathcal{P}(T_i).$$

Let m be the number of T_i such that $|T_i|$ is odd. By the induction hypothesis,

$$\mathcal{P}(T, v) = 1 - \frac{1}{n} \left(\frac{m}{2} + \sum_{i=1}^k |T_i| \cdot \frac{1}{2} \right) = 1 - \frac{m}{2n} - \frac{n-1}{2n} = \frac{1}{2} + \frac{1-m}{2n}.$$

This quantity is maximized when m is minimized. If $n = |T|$ is even, then after removing v there are an odd number of vertices and so $m \geq 1$; moreover, we can always achieve the minimum $m = 1$ by taking v to be a leaf (among possibly other options), and so $\mathcal{P}(T) = \frac{1}{2}$ in this case. If n is odd, then we can always achieve the minimum possible value $m = 0$ by taking v to be a leaf (among possibly other options), and so $\mathcal{P}(T) = \frac{1}{2} + \frac{1}{2n}$ in this case. By induction, the result is valid for all trees. □

Remark. It follows from the preceding proof that the set of optimal moves for the first player is precisely the set of vertices v such that $T \setminus v$ has the minimum number of odd-order components (namely, 0 if n is odd and 1 if n is even). In particular, it is always optimal to choose a leaf. In the motivating context of binary search on $\{1, \dots, n\}$, we have that all first moves are equally strong if n is even, while the first player should guess any odd number when n is odd.

4 Optimal strategy versus a random player

In this section, we consider the tree search game when one player is trying to maximize their probability of winning while the other plays uniformly at random. This version of the game is effectively a one-player game, since only one player is playing strategically. For a tree T , let $\mathcal{P}(T)$ be the probability that a player playing with an optimal strategy wins when they have the first move on T against an opponent who chooses vertices uniformly at random. Similarly, let $\mathcal{Q}(T)$ be the probability that a player playing with an optimal strategy wins when they have the second move on T against an opponent who chooses vertices uniformly at random.

It is easy to compute that $\mathcal{P}(S_1) = 1$, $\mathcal{Q}(S_1) = 0$, and $\mathcal{P}(S_2) = \mathcal{Q}(S_2) = \frac{1}{2}$. For the tree on three vertices, a player wins with probability $\frac{1}{3}$ if they choose the middle node (if the choice is incorrect, then the opponent has only one vertex to choose from) and probability $\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{2} = \frac{2}{3}$ if they choose a leaf, so $\mathcal{P}(S_3) = \frac{2}{3}$ and $\mathcal{Q}(S_3) = 1 - \frac{1}{3} \left(\frac{2}{3} + \frac{1}{3} + \frac{2}{3} \right) = \frac{4}{9}$.

For a vertex v of tree T , let $\mathcal{P}(T, v)$ be the probability that the optimal player wins when playing first on T , provided that the optimal player’s first move is v . Since the first player plays optimally,

$$\mathcal{P}(T) = \max_{v \in V(T)} \mathcal{P}(T, v).$$

Since the target vertex is selected uniformly at random,

$$\mathcal{P}(T, v) = \frac{1}{n} + \sum_{T'} \frac{|T'|}{n} \mathcal{Q}(T') = \frac{1}{n} + \frac{1}{n} \sum_{T'} |T'| \mathcal{Q}(T'),$$

where T' ranges over the components of the forest $T \setminus v$. Similarly, let $\mathcal{Q}(T, v)$ be the conditional probability that the optimal player wins when playing second on T , given that the random player’s first move is v . Since the first move is chosen uniformly at random,

$$\mathcal{Q}(T) = \frac{1}{n} \sum_{v \in V(T)} \mathcal{Q}(T, v).$$

Moreover, since the target vertex is selected uniformly at random,

$$\mathcal{Q}(T, v) = \sum_{T'} \frac{|T'|}{n} \mathcal{P}(T') = \frac{1}{n} \sum_{T'} |T'| \mathcal{P}(T'),$$

where T' ranges over the components of $T \setminus v$. Combining the last two formulas gives

$$\mathcal{Q}(T) = \frac{1}{n^2} \sum_{v \in V(T)} \sum_{T'} |T'| \mathcal{P}(T').$$

In the following two subsections, we give exact values for the win probabilities \mathcal{P} and \mathcal{Q} for star graphs and for path graphs, respectively. In a third subsection, we give bounds for general trees.

4.1 Stars

The next theorem exactly computes the probabilities of winning for every star graph.

Theorem 4.1. *For every positive integer n ,*

$$\mathcal{P}(S_n) = \frac{2}{3} - \frac{1}{3n} + \frac{2}{3n} \cdot \frac{\binom{n-1}{(n-1)/2}}{2^{n-1}} \cdot \mathbf{1}_{n \text{ is odd}}$$

and

$$\mathcal{Q}(S_n) = \frac{2}{3} - \frac{2}{3n} + \frac{2}{3n} \cdot \frac{\binom{n}{n/2}}{2^n} \cdot \mathbf{1}_{n \text{ is even}}.$$

Proof. For $n = 1$ the proposed formulas give $\mathcal{P}(S_1) = \frac{2}{3} - \frac{1}{3} + \frac{2}{3} = 1$ and $\mathcal{Q}(S_1) = \frac{2}{3} - \frac{2}{3} = 0$, and for $n = 2$ they give $\mathcal{P}(S_2) = \frac{2}{3} - \frac{1}{6} = \frac{1}{2}$ and $\mathcal{Q}(S_2) = \frac{2}{3} - \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}$, as needed.

For $n \geq 3$, the strategic player may choose either a leaf ℓ or the center vertex c . One has

$$\mathcal{P}(S_n, \ell) = \frac{1}{n} \cdot 1 + \frac{n-1}{n} \cdot \mathcal{Q}(S_{n-1}) > \frac{1}{n} = \mathcal{P}(S_n, c),$$

so it is always optimal to choose a leaf and

$$\mathcal{P}(S_n) = \frac{1}{n} + \frac{n-1}{n} \mathcal{Q}(S_{n-1}).$$

Similarly, considering separately whether the opponent playing randomly chooses the center or a leaf, one has

$$\mathcal{Q}(S_n) = \underbrace{\frac{1}{n} \cdot \frac{n-1}{n} \cdot 1}_{\text{chooses center}} + \underbrace{\frac{n-1}{n} \cdot \frac{n-1}{n} \cdot \mathcal{P}(S_{n-1})}_{\text{chooses leaf}}.$$

It is straightforward to verify that the given formulas satisfy these recurrence equations, so the result holds by induction. □

One may immediately compute the asymptotic win probabilities for stars.

Corollary 4.2. *One has*

$$\lim_{n \rightarrow \infty} \mathcal{P}(S_n) = \lim_{n \rightarrow \infty} \mathcal{Q}(S_n) = \frac{2}{3}.$$

n	1	2	3	4	5	6	7
p_n	1	1/2	2/3	7/12	3/5	53/90	37/63
q_n	0	1/2	4/9	1/2	38/75	14/27	386/735

Table 1: The first few values of p_n and q_n

4.2 Paths

In this subsection, we exactly compute the win probabilities \mathcal{P} and \mathcal{Q} for every path graph. Given a positive integer n , define p_n and q_n by

$$p_n = \frac{1}{2} - \frac{1}{n} \cdot \frac{(-2)^n}{n!} + \frac{n+2}{2n} \sum_{j=0}^n \frac{(-2)^j}{j!},$$

$$q_n = \frac{n-1}{2n} - \frac{1}{n} \cdot \frac{(-2)^n}{n!} + \frac{n+3}{2n} \sum_{j=0}^n \frac{(-2)^j}{j!}.$$

We will show that $\mathcal{P}(P_n) = p_n$ is the probability that the player playing strategically beats an opponent playing randomly on a path with n vertices, provided that the optimal player goes first. Similarly, we will show that $\mathcal{Q}(P_n) = q_n$ is the probability that an optimal player beats a random opponent on a path with n vertices, provided that the optimal player goes second.

Because the infinite series $\sum_{j=0}^{\infty} \frac{(-2)^j}{j!}$ converges to e^{-2} , the limits (as n approaches infinity) of p_n and q_n are both $(1 + e^{-2})/2$. For later use, we record the first seven values of p_n and q_n in Table 1. For convenience, define p_0 and q_0 to be $\frac{1}{2}$.

The following lemma shows how to express p_n in terms of q_{n-1} .

Lemma 4.3. *If n is a positive integer, then*

$$p_n = \frac{1}{n} + \frac{n-1}{n} q_{n-1}.$$

Proof. The case $n = 1$ is easy, so we may assume that $n \geq 2$. Using the definitions of p and q , we have

$$\begin{aligned} np_n &= \frac{n}{2} - \frac{(-2)^n}{n!} + \frac{n+2}{2} \sum_{j=0}^n \frac{(-2)^j}{j!} \\ &= \frac{n}{2} + \frac{n}{2} \cdot \frac{(-2)^n}{n!} + \frac{n+2}{2} \sum_{j=0}^{n-1} \frac{(-2)^j}{j!} \\ &= 1 + \frac{n-2}{2} - \frac{(-2)^{n-1}}{(n-1)!} + \frac{n+2}{2} \sum_{j=0}^{n-1} \frac{(-2)^j}{j!} \\ &= 1 + (n-1)q_{n-1}, \end{aligned}$$

as desired. □

Our next lemma expresses q_n as an average involving previous values of p .

Lemma 4.4. *If n is a positive integer, then*

$$q_n = \frac{1}{n} \sum_{k=1}^n \left(\frac{k-1}{n} p_{k-1} + \frac{n-k}{n} p_{n-k} \right).$$

Proof. The case $n = 1$ is easy. For $n \geq 2$, we can simplify the desired equation as follows:

$$n^2 q_n = 2 \sum_{k=1}^{n-1} k p_k.$$

By telescoping sums, it is sufficient to verify the difference equation

$$(k+1)^2 q_{k+1} - k^2 q_k = 2k p_k.$$

Plugging in the definition of q_k and q_{k+1} and then using the definition of p_k , we have

$$\begin{aligned} (k+1)^2 q_{k+1} - k^2 q_k &= k + \frac{k+4}{2} \cdot \frac{(-2)^{k+1}}{k!} - \frac{(-2)^{k+1}}{k!} + k \cdot \frac{(-2)^k}{k!} + (k+2) \sum_{j=0}^k \frac{(-2)^j}{j!} \\ &= k - 2 \cdot \frac{(-2)^k}{k!} + (k+2) \sum_{j=0}^k \frac{(-2)^j}{j!} \\ &= 2k p_k, \end{aligned}$$

which is the difference equation. □

The next lemma gives good upper and lower bounds on q_n .

Lemma 4.5. *If n is a positive integer, then*

$$n q_n \leq \frac{1 + e^{-2}}{2} n - \frac{1}{8}.$$

Furthermore, if $n \geq 3$, then

$$\frac{1 + e^{-2}}{2} n - \frac{3}{8} \leq n q_n \leq \frac{1 + e^{-2}}{2} n - \frac{1}{4}.$$

Proof. Using Table 1, we can check the cases $n \leq 7$, so we may assume that $n \geq 8$. By Taylor’s theorem applied to the exponential function,

$$\left| \sum_{j=0}^n \frac{(-2)^j}{j!} - e^{-2} \right| \leq \frac{2^{n+1}}{(n+1)!}.$$

Plugging this bound into the definition of q_n gives

$$\left| n q_n - \frac{1 + e^{-2}}{2} n + \frac{1 - 3e^{-2}}{2} \right| \leq \frac{n+3}{2} \cdot \frac{2^{n+1}}{(n+1)!} + \frac{2^n}{n!} = (n+2) \frac{2^{n+1}}{(n+1)!}.$$

m	n	$mq_m + nq_n$	$(m + n)q_{m+n}$
1	1	0	1
1	2	1	4/3
2	2	2	2

Table 2: Small cases in the proof of Lemma 4.6

Since $n \geq 8$, the right side $(n + 2)2^{n+1}/(n + 1)!$ is at most $\frac{1}{25}$. Hence

$$\frac{1 + e^{-2}}{2}n - \frac{1 - 3e^{-2}}{2} - \frac{1}{25} \leq nq_n \leq \frac{1 + e^{-2}}{2}n - \frac{1 - 3e^{-2}}{2} + \frac{1}{25}.$$

Since $(1 - 3e^{-2})/2$ is between 0.29 and 0.3, we are done. □

Since $q_2 > q_3$, the sequence $\{q_n\}_{n \geq 1}$ is not increasing, though the sequence $\{q_n\}_{n \geq 3}$ is strictly increasing. Our next lemma shows that the sequence $\{q_n\}_{n \geq 1}$ satisfies an increasing-like property, namely, the sequence $\{nq_n\}_{n \geq 1}$ is superadditive.

Lemma 4.6. *If m and n are nonnegative integers, then*

$$mq_m + nq_n \leq (m + n)q_{m+n}.$$

Proof. By symmetry, we may assume that $m \leq n$. The case $m = 0$ is trivial, so we may assume that $m \geq 1$. The cases $n \leq 2$ are verified in Table 2. Hence we may further assume that $n \geq 3$. Using Lemma 4.5 twice, we have

$$mq_m \leq \frac{1 + e^{-2}}{2}m - \frac{1}{8} \quad \text{and} \quad nq_n \leq \frac{1 + e^{-2}}{2}n - \frac{1}{4}.$$

Adding these two inequalities and using Lemma 4.5 again, we have

$$mq_m + nq_n \leq \frac{1 + e^{-2}}{2}(m + n) - \frac{3}{8} \leq (m + n)q_{m+n}. \quad \square$$

Lemma 4.3 expressed p_n in terms of q_{n-1} . The next lemma shows that p_n can be expressed as a maximum involving all previous q .

Lemma 4.7. *If n is a positive integer, then*

$$p_n = \max_{1 \leq k \leq n} \left(\frac{1}{n} + \frac{k - 1}{n}q_{k-1} + \frac{n - k}{n}q_{n-k} \right).$$

Proof. By considering $k = 1$ and using Lemma 4.3, we see that the maximum is at least p_n . On the other hand, for every k , by Lemmas 4.3 and 4.6, we have

$$\frac{1}{n} + \frac{k - 1}{n}q_{k-1} + \frac{n - k}{n}q_{n-k} \leq \frac{1}{n} + \frac{n - 1}{n}q_{n-1} = p_n.$$

Hence the maximum is at most p_n . □

Finally, we are ready to prove our formulas for the probabilities of winning on a path.

Theorem 4.8. *If n is a positive integer, then $\mathcal{P}(P_n) = p_n$ and $\mathcal{Q}(P_n) = q_n$.*

Proof. We may assume that the vertices of the path in order are $1, 2, \dots, n$. Recall that $\mathcal{P}(P_n)$ is $\max_k \mathcal{P}(P_n, k)$. For every vertex k , we have

$$\mathcal{P}(P_n, k) = \frac{1}{n} + \frac{k-1}{n} \mathcal{Q}(P_{k-1}) + \frac{n-k}{n} \mathcal{Q}(P_{n-k}).$$

Hence Lemma 4.7 provides the correct recurrence relation for $\mathcal{P}(P_n)$: it is given by the maximum over all vertices k from 1 to n . Similarly, recall that $\mathcal{Q}(P_n)$ is the average of $\mathcal{Q}(P_n, k)$ over all vertices k . For every vertex k , we have

$$\mathcal{Q}(P_n, k) = \frac{k-1}{n} \mathcal{P}(P_{k-1}) + \frac{n-k}{n} \mathcal{P}(P_{n-k}).$$

Hence Lemma 4.4 provides the correct recurrence relation for $\mathcal{Q}(P_n)$: it is given by the average over all vertices k from 1 to n . The theorem follows by induction on n . □

4.3 General trees

In this subsection, we bound the probabilities $\mathcal{P}(T)$ and $\mathcal{Q}(T)$ of winning for every tree. We start with an intriguing conjecture that paths and stars have the extreme win probabilities.

Conjecture 4.9. *For every tree T on n vertices,*

$$\mathcal{P}(P_n) \leq \mathcal{P}(T) \leq \mathcal{P}(S_n) \quad \text{and} \quad \mathcal{Q}(P_n) \leq \mathcal{Q}(T) \leq \mathcal{Q}(S_n).$$

We have confirmed the conjecture for $n \leq 20$. We are able to prove half of the conjecture: stars have the largest probabilities of winning.

Theorem 4.10. *If T is a tree with n vertices, then*

$$\mathcal{P}(T) \leq \mathcal{P}(S_n) \quad \text{and} \quad \mathcal{Q}(T) \leq \mathcal{Q}(S_n).$$

Remark. Before embarking on the proof, we observe one complication that may help explain why this semi-random version is more complicated than the fully optimal play. Unlike the situation in which both players play optimally, it is *not* the case that there is always a leaf among the optimal moves. Consider the tree T formed by starting from the path P_5 and adding two edges incident to each endpoint; see Figure 6. Then for the central vertex c one has

$$\mathcal{P}(T, c) = \frac{1}{9} + \frac{8}{9} \cdot \frac{9}{16} \approx 0.611,$$

while for a leaf ℓ one has

$$\mathcal{P}(T, \ell) = \frac{1}{9} + \frac{8}{9} \cdot \frac{12601}{23040} \approx 0.597.$$

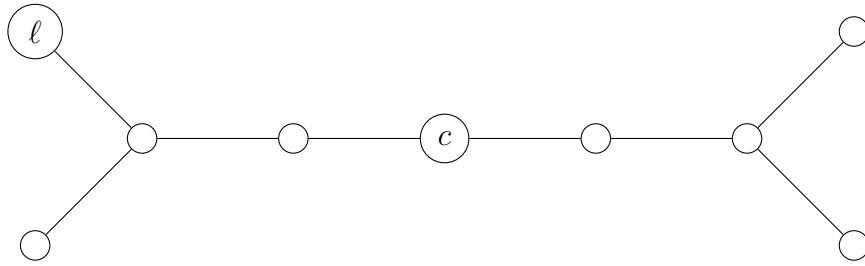


Figure 6: A tree for which every leaf is a suboptimal move

As a first step toward the proof of Theorem 4.10, we study the sequence $\{Q(S_n)\}_{n \geq 1}$. This sequence alternately increases and decreases, but we prove a loose monotonicity property on the values as they drift upward toward $\frac{2}{3}$. Namely, the next lemma shows that the sequence $\{nQ(S_n)\}_{n \geq 1}$ is superadditive.

Lemma 4.11. *If m and n are positive integers, then*

$$mQ(S_m) + nQ(S_n) \leq (m + n)Q(S_{m+n}).$$

Proof. If k is a positive even number, then $\binom{k}{k/2}/2^k$ is at most $\frac{1}{2}$. Hence, by using Theorem 4.1 twice, we have

$$mQ(S_m) \leq \frac{2}{3}m - \frac{2}{3} + \frac{2}{3} \cdot \frac{1}{2} = \frac{2}{3}m - \frac{1}{3} \quad \text{and} \quad nQ(S_n) \leq \frac{2}{3}n - \frac{2}{3} + \frac{2}{3} \cdot \frac{1}{2} = \frac{2}{3}n - \frac{1}{3}.$$

Adding these two inequalities and using Theorem 4.1 again, we conclude

$$mQ(S_m) + nQ(S_n) \leq \frac{2}{3}m - \frac{1}{3} + \frac{2}{3}n - \frac{1}{3} = \frac{2}{3}(m + n) - \frac{2}{3} \leq (m + n)Q(S_{m+n}). \quad \square$$

Proof of Theorem 4.10. The proof is by strong induction on n . The base case $n = 1$ is easy. For $n \geq 2$, assume the result is true for every tree with fewer than n vertices. We will prove the result for tree T .

First we prove the bound on $\mathcal{P}(T)$. Let v be a vertex of T . We have

$$\mathcal{P}(T, v) = \frac{1}{n} + \sum_{T'} \frac{|T'|}{n} Q(T') = \frac{1}{n} + \frac{1}{n} \sum_{T'} |T'| Q(T'),$$

where T' ranges over the components of $T \setminus v$. By the induction hypothesis and Lemma 4.11,

$$\mathcal{P}(T, v) \leq \frac{1}{n} + \frac{1}{n} \sum_{T'} |T'| Q(S_{|T'|}) \leq \frac{1}{n} + \frac{n-1}{n} Q(S_{n-1}).$$

As shown in the proof of Theorem 4.1, the right side equals $\mathcal{P}(S_n)$, so $\mathcal{P}(T, v) \leq \mathcal{P}(S_n)$. We conclude that

$$\mathcal{P}(T) = \max_{v \in V(T)} \mathcal{P}(T, v) \leq \mathcal{P}(S_n).$$

Next we prove the bound on $\mathcal{Q}(T)$. Let v be a vertex of T . We have

$$\mathcal{Q}(T, v) = \sum_{T'} \frac{|T'|}{n} \mathcal{P}(T') = \frac{1}{n} \sum_{T'} |T'| \mathcal{P}(T'),$$

where T' ranges over the components of $T \setminus v$. By the induction hypothesis,

$$\mathcal{Q}(T, v) \leq \frac{1}{n} \sum_{T'} |T'| \mathcal{P}(S_{|T'|}).$$

Given a positive integer m , let $C(m)$ be $\binom{m-1}{(m-1)/2} / 2^{m-1}$ if m is odd and 0 otherwise. Theorem 4.1 says that

$$m\mathcal{P}(S_m) = \frac{2}{3}m - \frac{1}{3} + \frac{2}{3}C(m).$$

Plugging this formula into our inequality for $\mathcal{Q}(T, v)$ gives

$$\mathcal{Q}(T, v) \leq \frac{1}{n} \sum_{T'} \left(\frac{2}{3}|T'| - \frac{1}{3} + \frac{2}{3}C(|T'|) \right) = \frac{2(n-1)}{3n} - \frac{1}{3n} \deg(v) + \frac{2}{3n} \sum_{T'} C(|T'|).$$

Given a vertex w that is a neighbor of v , recall from the proof of Lemma 2.1(c) that the tree $T_{v,w}$ is the component of $T \setminus v$ that contains w . We can rewrite our previous inequality as

$$\mathcal{Q}(T, v) \leq \frac{2(n-1)}{3n} - \frac{1}{3n} \deg(v) + \frac{2}{3n} \sum_{\substack{w \in V(T) \\ w \sim v}} C(|T_{v,w}|).$$

Averaging over all vertices v gives the bound

$$\mathcal{Q}(T) = \frac{1}{n} \sum_{v \in V(T)} \mathcal{Q}(T, v) \leq \frac{2(n-1)^2}{3n^2} + \frac{2}{3n^2} \sum_{\substack{v, w \in V(T) \\ v \sim w}} C(|T_{v,w}|).$$

Given an edge $\{v, w\}$ of T , we claim that $C(|T_{v,w}|) + C(|T_{w,v}|)$ is at most $1 + C(n-1)$. Recall that $|T_{v,w}| + |T_{w,v}|$ is n . If either $|T_{v,w}|$ or $|T_{w,v}|$ is 1, then the other is $n-1$, so the claim is true with equality. Otherwise, both $|T_{v,w}|$ and $|T_{w,v}|$ are greater than 1, in which case $C(|T_{v,w}|)$ and $C(|T_{w,v}|)$ are each at most $\frac{1}{2}$, so the claim is again true. Plugging our claim into our inequality for $\mathcal{Q}(T)$ gives

$$\mathcal{Q}(T) \leq \frac{2(n-1)^2}{3n^2} + \frac{2(n-1)}{3n^2} (1 + C(n-1)) = \frac{2}{3} - \frac{2}{3n} + \frac{2(n-1)}{3n^2} C(n-1).$$

It is straightforward to verify that $(n-1)C(n-1) = nC(n+1)$, so

$$\mathcal{Q}(T) \leq \frac{2}{3} - \frac{2}{3n} + \frac{2}{3n} C(n+1).$$

By Theorem 4.1, the right side is $\mathcal{Q}(S_n)$, which completes the proof. □

The next theorem makes progress toward the lower bounds of Conjecture 4.9. In particular, we show that the player playing optimally wins with probability at least $\frac{9}{16}$ (asymptotically).

Theorem 4.12. *Let T be a tree on n vertices. If $n \geq 4$, then*

$$\mathcal{P}(T) > \frac{9}{16} \quad \text{and} \quad \mathcal{Q}(T) \geq \frac{9}{16} - \frac{5}{16n}.$$

Proof. Define the function δ by $\delta(1) = \frac{5}{16}$, $\delta(2) = -\frac{1}{4}$, $\delta(3) = \frac{3}{16}$, $\delta(4) = -\frac{1}{24}$, $\delta(5) = \frac{1}{16}$, and $\delta(k) = 0$ for $k \geq 6$. We will prove, for every positive integer n , the more precise inequalities

$$\mathcal{P}(T) \geq \frac{9}{16} + \frac{1}{8n} + \frac{\delta(n)}{n} \quad \text{and} \quad \mathcal{Q}(T) \geq \frac{9}{16} - \frac{5}{16n} + \frac{\delta(n+1)}{n}.$$

The proof is by strong induction on n . The base cases $n \leq 4$ are easy to check. For $n \geq 5$, assume the inequalities hold for every tree with fewer than n vertices. We will prove the inequalities for tree T .

First we bound $\mathcal{P}(T)$. Let v be a leaf of T . By the induction hypothesis,

$$\begin{aligned} \mathcal{P}(T) &\geq \mathcal{P}(T, v) \\ &= \frac{1}{n} + \frac{n-1}{n} \mathcal{Q}(T \setminus v) \\ &\geq \frac{1}{n} + \frac{n-1}{n} \left(\frac{9}{16} - \frac{5}{16(n-1)} + \frac{\delta(n)}{n-1} \right) \\ &= \frac{9}{16} + \frac{1}{n} - \frac{9}{16n} - \frac{5}{16n} + \frac{\delta(n)}{n} \\ &= \frac{9}{16} + \frac{1}{8n} + \frac{\delta(n)}{n}, \end{aligned}$$

as desired.

Next we bound $\mathcal{Q}(T)$. By the recurrence formula for \mathcal{Q} , we have

$$\begin{aligned} \mathcal{Q}(T) &= \frac{1}{n^2} \sum_v \sum_{T'} |T'| \mathcal{P}(T') \\ &\geq \frac{1}{n^2} \sum_v \sum_{T'} \left(\frac{9}{16} |T'| + \frac{1}{8} + \delta(|T'|) \right) \\ &= \frac{9(n-1)}{16n} + \frac{n-1}{4n^2} + \frac{1}{n^2} \sum_v \sum_{T'} \delta(|T'|) \\ &= \frac{9}{16} - \frac{5}{16n} - \frac{1}{4n^2} + \frac{1}{n^2} \left(\frac{5}{16} \ell_1(T) - \frac{1}{4} \ell_2(T) + \frac{3}{16} \ell_3(T) - \frac{1}{24} \ell_4(T) + \frac{1}{16} \ell_5(T) \right) \\ &= \frac{9}{16} - \frac{5}{16n} + \frac{1}{48n^2} \left(15\ell_1(T) - 12\ell_2(T) + 9\ell_3(T) - 2\ell_4(T) + 3\ell_5(T) - 12 \right). \end{aligned}$$

To prove the desired bound on $\mathcal{Q}(T)$, it suffices to prove that the expression in parentheses is nonnegative. For short, write ℓ_k for $\ell_k(T)$. If T is the unique exceptional tree from Lemma 2.1(g) (shown in Figure 1), then $\ell_1 = 3$, $\ell_2 = 2$, $\ell_3 = 0$, $\ell_4 = 2$, and $\ell_5 = 3$, which means

$$15\ell_1 - 12\ell_2 + 9\ell_3 - 2\ell_4 + 3\ell_5 = 45 - 24 + 0 - 4 + 9 = 26.$$

Hence we may assume that T is not the exceptional tree. By Lemma 2.1(g), we have $\ell_1 + \ell_3 \geq \ell_2 + \ell_4$, so

$$15\ell_1 - 12\ell_2 + 9\ell_3 - 2\ell_4 + 3\ell_5 \geq 13\ell_1 - 10\ell_2 + 7\ell_3 + 3\ell_5.$$

If $\ell_1 = 2$, then T is a path (Lemma 2.2(a)), so $\ell_2 = \ell_3 = 2$ (Lemma 2.2(b)), which means

$$13\ell_1 - 10\ell_2 + 7\ell_3 + 3\ell_5 \geq 13\ell_1 - 10\ell_2 + 7\ell_3 = 26 - 20 + 14 = 20.$$

If $\ell_1 \geq 4$, then because $\ell_1 \geq \ell_2$ (Lemma 2.1(f)), we have

$$13\ell_1 - 10\ell_2 + 7\ell_3 + 3\ell_5 \geq 13\ell_1 - 10\ell_2 \geq 3\ell_1 \geq 12.$$

Hence we may assume that $\ell_1 = 3$. If $\ell_2 \leq 2$, then

$$13\ell_1 - 10\ell_2 + 7\ell_3 + 3\ell_5 \geq 13\ell_1 - 10\ell_2 = 39 - 10\ell_2 \geq 39 - 20 = 19.$$

Hence we may assume that $\ell_2 = 3$. If $\ell_3 \geq 1$, then

$$13\ell_1 - 10\ell_2 + 7\ell_3 + 3\ell_5 \geq 13\ell_1 - 10\ell_2 + 7\ell_3 = 39 - 30 + 7\ell_3 \geq 39 - 30 + 7 = 16.$$

Hence we may assume that $\ell_3 = 0$. By Lemma 2.2(e), the only tree with $\ell_1 = 3$, $\ell_2 = 3$, and $\ell_3 = 0$ is the spider $S_{2,2,2}$, shown in Figure 5. For this tree, $\ell_5 = 3$, which means

$$13\ell_1 - 10\ell_2 + 7\ell_3 + 3\ell_5 = 39 - 30 + 0 + 9 = 18.$$

In every case, we have $15\ell_1 - 12\ell_2 + 9\ell_3 - 2\ell_4 + 3\ell_5 - 12 \geq 0$. This completes the proof. □

5 Two random players

In this section, we consider the tree search game when both players play randomly. Given a tree T , let $\mathcal{P}(T)$ be the probability that the first player wins on T when both players choose vertices uniformly at random from among the vertices that could be the target. For example, $\mathcal{P}(S_1) = 1$, $\mathcal{P}(S_2) = \frac{1}{2}$, and $\mathcal{P}(S_3) = \frac{5}{9}$.

For every vertex v of a tree T , let $\mathcal{P}(T, v)$ be the conditional probability that the first player wins on T given that they select v first. Since the first contestant plays uniformly at random,

$$\mathcal{P}(T) = \frac{1}{n} \sum_{v \in V(T)} \mathcal{P}(T, v).$$

Moreover, since the target vertex is selected uniformly at random,

$$\mathcal{P}(T, v) = \frac{1}{n} + \sum_{T'} \frac{|T'|}{n} (1 - \mathcal{P}(T')) = 1 - \frac{1}{n} \sum_{T'} |T'| \mathcal{P}(T'),$$

where T' ranges over the components of $T \setminus v$. Combining these two formulas gives

$$\mathcal{P}(T) = 1 - \frac{1}{n^2} \sum_{v \in V(T)} \sum_{T'} |T'| \mathcal{P}(T').$$

We first exactly compute the probability of winning for every star.

Theorem 5.1. *If n is a positive integer, then*

$$\mathcal{P}(S_n) = \frac{1}{2} + \frac{1}{2n^2} \cdot \mathbf{1}_{n \text{ is odd}}.$$

Proof. The proof is by induction on n . The base cases $n = 1$ and $n = 2$ are easy to check. For $n \geq 3$, assume the result is true for $n - 1$.

If c is the center of the star, then $\mathcal{P}(S_n, c) = \frac{1}{n}$. If ℓ is a leaf, then

$$\mathcal{P}(S_n, \ell) = \frac{1}{n} + \frac{n-1}{n} \left(1 - \mathcal{P}(S_{n-1})\right) = 1 - \frac{n-1}{n} \mathcal{P}(S_{n-1}).$$

Taking the weighted average of the center and the $n - 1$ leaves, we have

$$\mathcal{P}(S_n) = \frac{1}{n} \cdot \frac{1}{n} + \frac{n-1}{n} \left(1 - \frac{n-1}{n} \mathcal{P}(S_{n-1})\right) = 1 - \frac{1}{n} + \frac{1}{n^2} - \frac{(n-1)^2}{n^2} \mathcal{P}(S_{n-1}).$$

By the induction hypothesis,

$$\begin{aligned} \mathcal{P}(S_n) &= 1 - \frac{1}{n} + \frac{1}{n^2} - \frac{(n-1)^2}{n^2} \left(\frac{1}{2} + \frac{1}{2(n-1)^2} \cdot \mathbf{1}_{n \text{ is even}}\right) \\ &= \frac{1}{2} + \frac{1}{2n^2} - \frac{1}{2n^2} \cdot \mathbf{1}_{n \text{ is even}} \\ &= \frac{1}{2} + \frac{1}{2n^2} \cdot \mathbf{1}_{n \text{ is odd}}, \end{aligned}$$

which completes the induction. □

We next exactly compute the probability of winning for a path.

Theorem 5.2. *If n is an integer such that $n \geq 3$, then*

$$\mathcal{P}(P_n) = \frac{1}{2} + \frac{1}{6n}.$$

Proof. For convenience, define $\mathcal{P}(P_0)$ to be $\frac{1}{2}$. Define the function δ by $\delta(0) = -\frac{1}{6}$, $\delta(1) = \frac{1}{3}$, $\delta(2) = -\frac{1}{6}$, and $\delta(k) = 0$ for $k \geq 3$. We will prove, for every nonnegative integer n , that $n\mathcal{P}(P_n) = \frac{1}{2}n + \frac{1}{6} + \delta(n)$. The proof is by strong induction on n . The base cases $n \leq 2$ are easy to check. For $n \geq 3$, assume the result is true for every nonnegative integer less than n . We will prove the result for n .

We may assume that the vertices of the path in order are $1, 2, \dots, n$. By the induction hypothesis, for every vertex k , we have

$$\begin{aligned} \mathcal{P}(P_n, k) &= 1 - \frac{k-1}{n}\mathcal{P}(P_{k-1}) - \frac{n-k}{n}\mathcal{P}(P_{n-k}) \\ &= 1 - \frac{1}{n}\left(\frac{1}{2}(k-1) + \frac{1}{6} + \delta(k-1)\right) - \frac{1}{n}\left(\frac{1}{2}(n-k) + \frac{1}{6} + \delta(n-k)\right) \\ &= \frac{1}{2} + \frac{1}{6n} - \frac{\delta(k-1)}{n} - \frac{\delta(n-k)}{n}. \end{aligned}$$

Averaging over all vertices gives

$$\begin{aligned} \mathcal{P}(P_n) &= \frac{1}{n} \sum_{k=1}^n \mathcal{P}(P_n, k) \\ &= \frac{1}{2} + \frac{1}{6n} - \frac{1}{n^2} \sum_{k=1}^n \delta(k-1) - \frac{1}{n^2} \sum_{k=1}^n \delta(n-k) \\ &= \frac{1}{2} + \frac{1}{6n}. \end{aligned} \quad \square$$

We again conjecture that stars and paths have the extreme win probabilities (though now stars are the lower bound and paths are the upper bound).

Conjecture 5.3. *For every tree T on n vertices, $\mathcal{P}(S_n) \leq \mathcal{P}(T) \leq \mathcal{P}(P_n)$.*

We have confirmed the conjecture for $n \leq 20$. Conjecture 5.3 implies that the probability of the first player winning is always at least $\frac{1}{2}$ and approaches $\frac{1}{2}$ on every sequence of trees as $n \rightarrow \infty$.

The next theorem makes progress toward Conjecture 5.3. We show that both players win with probability between $\frac{13}{30}$ and $\frac{17}{30}$ (for $n \geq 2$).

Theorem 5.4. *Let T be a tree with n vertices. If $n \geq 2$, then*

$$\frac{13}{30} < \mathcal{P}(T) < \frac{17}{30}.$$

Proof. Define the function δ by $\delta(1) = \frac{4}{15}$, $\delta(2) = -\frac{1}{6}$, $\delta(3) = \frac{1}{15}$, $\delta(4) = -\frac{1}{30}$, $\delta(5) = \frac{2}{15}$, and $\delta(k) = 0$ for $k \geq 6$. Define the function Δ by $\Delta(1) = \frac{1}{3}$, $\Delta(2) = 0$, $\Delta(3) = \frac{2}{45}$, and $\Delta(k) = 0$ for $k \geq 4$. We will prove, for every positive integer n , the more precise inequality

$$\frac{13}{30} + \frac{3}{10n} + \frac{\delta(n)}{n} \leq \mathcal{P}(T) \leq \frac{17}{30} - \frac{1}{6n} + \frac{4}{15n^2} + \frac{\Delta(n)}{n}.$$

The proof is by strong induction on n . The base cases $n \leq 5$ are easy to check. For $n \geq 6$, assume the inequality holds for every tree with fewer than n vertices. We will prove the inequality for tree T .

First we prove the upper bound on $\mathcal{P}(T)$. By the induction hypothesis,

$$\begin{aligned} \mathcal{P}(T) &= 1 - \frac{1}{n^2} \sum_v \sum_{T'} |T'| \mathcal{P}(T') \\ &\leq 1 - \frac{1}{n^2} \sum_v \sum_{T'} \left(\frac{13}{30} |T'| + \frac{3}{10} + \delta(|T'|) \right) \\ &= 1 - \frac{13n(n-1)}{30n^2} - \frac{6(n-1)}{10n^2} - \frac{1}{n^2} \sum_v \sum_{T'} \delta(|T'|) \\ &= \frac{17}{30} - \frac{1}{6n} + \frac{3}{5n^2} - \frac{1}{n^2} \sum_v \sum_{T'} \delta(|T'|) \\ &= \frac{17}{30} - \frac{1}{6n} + \frac{3}{5n^2} - \frac{1}{n^2} \left(\frac{4}{15} \ell_1(T) - \frac{1}{6} \ell_2(T) + \frac{1}{15} \ell_3(T) - \frac{1}{30} \ell_4(T) + \frac{2}{15} \ell_5(T) \right) \\ &= \frac{17}{30} - \frac{1}{6n} + \frac{3}{5n^2} - \frac{1}{30n^2} \left(8\ell_1(T) - 5\ell_2(T) + 2\ell_3(T) - \ell_4(T) + 4\ell_5(T) \right). \end{aligned}$$

We claim that $8\ell_1(T) - 5\ell_2(T) + 2\ell_3(T) - \ell_4(T) + 4\ell_5(T) \geq 10$. For short, write ℓ_k for $\ell_k(T)$. If T is the unique exceptional tree from Lemma 2.1(g) (shown in Figure 1), then $\ell_1 = 3, \ell_2 = 2, \ell_3 = 0, \ell_4 = 2,$ and $\ell_5 = 3$, which means

$$8\ell_1 - 5\ell_2 + 2\ell_3 - \ell_4 + 4\ell_5 = 24 - 10 + 0 - 2 + 12 = 24.$$

Hence we may assume that T is not the exceptional tree. By Lemma 2.1(g), we have $\ell_1 + \ell_3 \geq \ell_2 + \ell_4$, which means

$$8\ell_1 - 5\ell_2 + 2\ell_3 - \ell_4 + 4\ell_5 \geq 7\ell_1 - 4\ell_2 + \ell_3 + 4\ell_5.$$

If $\ell_1 = 2$, then T is a path (Lemma 2.2(a)), so $\ell_2 = \ell_3 = \ell_4 = \ell_5 = 2$ (Lemma 2.2(b)), which means

$$7\ell_1 - 4\ell_2 + \ell_3 + 4\ell_5 = 14 - 8 + 2 + 8 = 16.$$

If $\ell_1 \geq 4$, then because $\ell_1 \geq \ell_2$ (Lemma 2.1(f)), we have

$$7\ell_1 - 4\ell_2 + \ell_3 + 4\ell_5 \geq 7\ell_1 - 4\ell_2 \geq 3\ell_1 \geq 12.$$

Hence we may assume that $\ell_1 = 3$. If $\ell_2 \leq 2$, then

$$7\ell_1 - 4\ell_2 + \ell_3 + 4\ell_5 \geq 7\ell_1 - 4\ell_2 = 21 - 4\ell_2 \geq 21 - 8 = 13.$$

Hence we may assume that $\ell_2 = 3$. If $\ell_3 \geq 1$, then

$$7\ell_1 - 4\ell_2 + \ell_3 + 4\ell_5 \geq 7\ell_1 - 4\ell_2 + \ell_3 = 21 - 12 + \ell_3 \geq 21 - 12 + 1 = 10.$$

Hence we may assume that $\ell_3 = 0$. By Lemma 2.2(e), the only tree with $\ell_1 = 3, \ell_2 = 3,$ and $\ell_3 = 0$ is the spider $S_{2,2,2}$, shown in Figure 5. For that tree, $\ell_5 = 3$, which means

$$7\ell_1 - 4\ell_2 + \ell_3 + 4\ell_5 = 21 - 12 + 0 + 12 = 21.$$

In every case, we have proved the claim $8\ell_1(T) - 5\ell_2(T) + 2\ell_3(T) - \ell_4(T) + 4\ell_5(T) \geq 10$.

Plugging the claim into our previous inequality for $\mathcal{P}(T)$ gives

$$\begin{aligned} \mathcal{P}(T) &\leq \frac{17}{30} - \frac{1}{6n} + \frac{3}{5n^2} - \frac{1}{30n^2} \left(8\ell_1(T) - 5\ell_2(T) + 2\ell_3(T) - \ell_4(T) + 4\ell_5(T) \right) \\ &\leq \frac{17}{30} - \frac{1}{6n} + \frac{3}{5n^2} - \frac{10}{30n^2} \\ &= \frac{17}{30} - \frac{1}{6n} + \frac{4}{15n^2}, \end{aligned}$$

which is the desired upper bound.

Next we prove the lower bound on $\mathcal{P}(T)$. By the induction hypothesis,

$$\begin{aligned} \mathcal{P}(T) &= 1 - \frac{1}{n^2} \sum_{v \in V(T)} \sum_{T'} |T'| \mathcal{P}(T') \\ &\geq 1 - \frac{1}{n^2} \sum_v \sum_{T'} \left(\frac{17}{30} |T'| - \frac{1}{6} + \frac{4}{15|T'|} + \Delta(|T'|) \right) \\ &= 1 - \frac{17n(n-1)}{30n^2} + \frac{2(n-1)}{6n^2} - \frac{4}{15n^2} \sum_v \sum_{T'} \frac{1}{|T'|} - \frac{1}{n^2} \sum_v \sum_{T'} \Delta(|T'|) \\ &= \frac{13}{30} + \frac{17}{30n} - \frac{4}{15n^2} \sum_v \sum_{T'} \frac{1}{|T'|} + \frac{n-1}{3n^2} - \frac{1}{n^2} \sum_v \sum_{T'} \Delta(|T'|) \\ &= \frac{13}{30} + \frac{17}{30n} - \frac{4}{15n^2} \sum_v \sum_{T'} \frac{1}{|T'|} + \frac{n-1}{3n^2} - \frac{1}{n^2} \left(\frac{1}{3} \ell_1(T) + \frac{2}{45} \ell_3(T) \right) \\ &= \frac{13}{30} + \frac{17}{30n} - \frac{4}{15n^2} \sum_v \sum_{T'} \frac{1}{|T'|} + \frac{1}{3n^2} \left(n-1 - \ell_1(T) - \frac{2}{15} \ell_3(T) \right). \end{aligned}$$

By Lemma 2.1(c) and Lemma 2.1(d), the expression in parentheses is nonnegative:

$$\ell_1(T) + \frac{2}{15} \ell_3(T) = \frac{1}{2} \ell_1(T) + \frac{1}{2} \ell_{n-1}(T) + \frac{2}{15} \ell_3(T) \leq \frac{1}{2} \sum_{k=1}^{n-1} \ell_k(T) = \frac{1}{2} \cdot 2(n-1) = n-1.$$

Hence

$$\mathcal{P}(T) \geq \frac{13}{30} + \frac{17}{30n} - \frac{4}{15n^2} \sum_v \sum_{T'} \frac{1}{|T'|}.$$

We claim that the double sum can be bounded as follows:

$$\sum_v \sum_{T'} \frac{1}{|T'|} \leq n.$$

Given a vertex v and one of its neighboring vertices w , recall from the proof of Lemma 2.1(c) that the tree $T_{v,w}$ is the component of $T \setminus v$ that contains w . We can rewrite the inequality as

$$\sum_{\substack{v,w \\ v \sim w}} \frac{1}{|T_{v,w}|} \leq n.$$

For every edge $\{v, w\}$ of T , we have

$$\frac{1}{|T_{v,w}|} + \frac{1}{|T_{w,v}|} = \frac{|T_{v,w}| + |T_{w,v}|}{|T_{v,w}| \cdot |T_{w,v}|} = \frac{n}{|T_{v,w}| \cdot |T_{w,v}|} \leq \frac{n}{n-1}.$$

Hence we can sum over every edge:

$$\sum_{\substack{v,w \\ v \sim w}} \frac{1}{|T_{v,w}|} \leq (n-1) \cdot \frac{n}{n-1} = n,$$

which establishes the claimed bound on the double sum.

Plugging this bound on the double sum into our previous inequality for $\mathcal{P}(T)$ gives

$$\mathcal{P}(T) \geq \frac{13}{30} + \frac{17}{30n} - \frac{4}{15n^2} \sum_v \sum_{T'} \frac{1}{|T'|} \geq \frac{13}{30} + \frac{17}{30n} - \frac{4}{15n^2} \cdot n = \frac{13}{30} + \frac{3}{10n},$$

which is the desired lower bound. □

6 Other questions

The main question arising from our work is to resolve Conjectures 4.9 and 5.3, namely, to show that the path and star have the extreme probabilities of winning among all trees of fixed order in both the semi-random and all-random models. In this section we outline a number of other possible directions of future research.

6.1 Other models

The rules of the game studied in this paper can be varied or extended in numerous ways. We mention a few explicit cases that are similar in spirit and might lead in interesting directions.

Guessing edges

Rather than choosing a vertex, players might alternate choosing *edges*, with the goal being to isolate the randomly selected target vertex. This version of the game is reminiscent of the random process called *cutting down trees*, which was introduced by Meir and Moon [4] and is an active area of research (see, e.g., [1]).

Searching on general graphs

In [3], the binary search algorithm on trees was extended to a model on general connected graphs: upon querying a vertex v , the player either is informed that v is the target or is given an edge out of v that lies on a shortest path from v to the

target. One could extend the two-player version of the game to this model. On some graphs, and unlike for trees, the first player may be at a marked disadvantage in this generalization: for example, on the complete graph K_n , the first player wins with probability $\frac{1}{n}$.

Misère search

In the misère version of the game, the target vertex is poisoned, and the player who finds it loses. For example, the one-vertex tree is a first player loss, the two-vertex tree is a fair game, and the first player wins on the three-vertex tree with probability $\frac{2}{3}$ if they choose the central vertex and with probability $\frac{1}{3}$ if they choose a leaf.

Continuous models

The following continuous model of binary search also admits a two-player version: a target point is selected uniformly at random in the real interval $[a, b]$, and a player wins if they pick a point within some fixed distance $\varepsilon > 0$ of the target; non-winning guesses are informed whether the guess is too large or too small. Higher-dimensional analogues are also possible: a guess might consist of a hyperplane, with the goal being to bound the target point in a region of sufficiently small volume; or a guess might consist of a point and a hyperplane through that point, with the goal of choosing a point within ε of the target.

6.2 Probability distributions

In the models where one or both players play randomly, the set of probabilities of a first-player win lies in some nontrivial interval $[a, b]$. As the order n of the trees goes to infinity, is the set of these probabilities dense in some interval? Is there a limiting distribution of probabilities of winning, where (for example) we think of drawing a vertex-labeled tree uniformly at random?

6.3 Random vs. random as a random process

The all-random game (Section 5) involves no strategy, so we can view it as a purely random process. Namely, we are given a tree T , one vertex of which has been randomly selected as the target. At each step, a vertex v is chosen at random; if v is not the target, then the tree is cut down to the subtree of $T \setminus v$ in which the target lies. The *stopping time* is the number of steps it takes to hit the target. Our problem can be restated as bounding the probability that the stopping time is odd.

For the star S_n , the stopping time has expected value about $n/3$. For the path P_n , the stopping time has expected value about $2 \ln n$. Does every tree with n vertices have expected stopping time between those of the path and the star? For a given

tree, is the probability distribution of the stopping time unimodal? Answering these questions might help resolve Conjecture 5.3.

6.4 Fixed target

The all-random game (Section 5) has three sources of randomness: the target vertex, the first player, and the second player. What if instead the target vertex were fixed (so that effectively we have a rooted tree)? Does the first player always win with probability at least $\frac{1}{2}$? After looking at small trees (up to five vertices), we tentatively conjecture that the answer is yes. If the target is a leaf, then a simple induction shows that the game is fair (win probability exactly $\frac{1}{2}$).

6.5 Relation with the gold grabber game

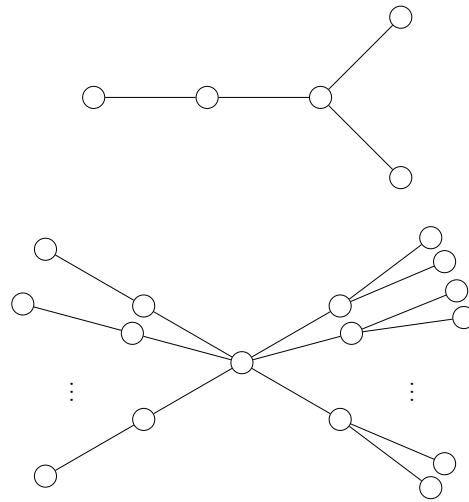
In [7], Seacrest and Seacrest consider the following (non-random, complete information) game on trees: some coins of various values are distributed on the vertices of a tree, and players take turns selecting a leaf and collecting the coins on the chosen leaf. Their main theorem is that on a tree with an even number of vertices, the first player can always guarantee to acquire at least half the total value of coins. Our game could be rephrased in similar language, with the target having a coin of positive value and all other vertices having coins of value 0. In this rephrasing, the key difference between our game and the game considered by Seacrest and Seacrest is that the players do not know the distribution of the coins. Are there interesting variations of the binary search game involving more elaborate distributions of weight?

6.6 Other constraints on limb numbers

Lemma 2.1 gives a number of constraints on the limb numbers $\ell_i(T)$, which are used for the bounds in several results in Sections 4 and 5. These relations are not exhaustive; for example, the proof of Lemma 2.1(g) may be extended to show that

$$\ell_1(T) + \ell_3(T) + \cdots + \ell_{2k-1}(T) \geq \ell_2(T) + \ell_4(T) + \cdots + \ell_{2k}(T)$$

for every tree T of order at least $4k$. On the other hand, for any positive integers a and b , one can construct a tree T such that $\ell_2(T) = a$ and $\ell_3(T) = b$: if $a = b = 1$ then take T to be the spider $S_{2,1,1}$, shown at the top of Figure 7, while if $a + b > 2$ then take T to be the bottom tree in Figure 7 with a branches on the left and b branches on the right. What else can one say about the constraints satisfied by the tuples (ℓ_1, \dots, ℓ_k) ? For example, are the tuples (ℓ_1, \dots, ℓ_k) the set of lattice points in a polytope (at least for n sufficiently large)?

Figure 7: Trees with specified values for ℓ_2 and ℓ_3

Acknowledgments

We first encountered a version of this problem on the Art of Problem Solving forum [2]. We thank Stephen Merriman for proposing the initial version of this problem. We thank Zachary Abel and Alison Miller for performing some helpful computer calculations. We are especially grateful to Simon Rubinstein-Salzedo for asking about playing randomly, for formulating the tree version of the problem, and for writing a useful computer program. We thank the two referees for their numerous helpful comments. We used the SageMath computer algebra system [6] to test our conjectures. The second author’s research was supported in part by a ORAU Powe Award and a Simons Collaboration Grant (634530).

References

- [1] L. Addario-Berry, N. Broutin and C. Holmgren, Cutting down trees with a Markov chainsaw, *Ann. Appl. Probab.* **24** (6) (2014), 2297–2339.
- [2] Art of Problem Solving forum, “I’m thinking of a number,” <https://artofproblemsolving.com/community/c6h11224>, 2004.
- [3] E. Emamjomeh-Zadeh, D. Kempe and V. Singhal, “Deterministic and Probabilistic Binary Search in Graphs,” 48th Annual ACM Symposium on Theory of Computing (STOC), pp. 519–532, 2016.
- [4] A. Meir and J.W. Moon, Cutting down random trees, *J. Austral. Math. Soc.* **11** (1970), 313–324.

- [5] K. Onak and P. Parys, “Generalization of Binary Search: Searching in Trees and Forest-Like Partial Orders,” 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 379–388, 2006.
- [6] The Sage Developers, *SageMath, the Sage Mathematics Software System* (Version 9.1), 2020.
- [7] D. E. Seacrest and T. Seacrest, Grabbing the gold, *Discrete Math.* **312** (2012), 1804–1806.
- [8] D. B. West, *Introduction to Graph Theory*, 2nd edition, Pearson, 2001.

(Received 28 Aug 2020; revised 16 Dec 2021)