5-1-2023

Mixing Measures for Trees of Fixed Diameter

Ari Holcombe Pomerance
Macalester College, aripomerance@gmail.com

Mixing Measures for Trees of Fixed Diameter

Ari Holcombe Pomerance

Andrew Beveridge, Advisor

Taryn Flock, Reader

Kristin Heysse, Reader

MACALESTER

Department of Mathematics, Statistics, and Computer Science

May, 2023

Abstract

A mixing measure is the expected length of a random walk in a graph given a set of starting and stopping conditions. We determine the tree structures of order n with diameter d that minimize and maximize for a few mixing measures. We show that the maximizing tree is usually a broom graph or a double broom graph and that the minimizing tree is usually a seesaw graph or a double seesaw graph.

Contents

Chapter 1

Introduction

One way to measure the connectivity of a tree T is by analyzing the expected length of random walks on T. Many different metrics have been established to gauge this information, and they are collectively referred to as *mixing measures*. This paper will primarily focus on two mixing measures: the hitting time (from a source vertex to a target vertex) and the access time (from the stationary distribution to a target vertex).

A *random walk* on a tree T is a sequence of adjacent vertices, in which, after picking a starting vertex, we choose a neighbor at random to traverse next. The *stationary distribution vector* π of T tells us the proportion of time spent at each vertex in a sufficiently long random walk. The *hitting time* $H(s,t)$ from a source vertex s to a target vertex t is the expected length of a random walk starting at s and terminating as soon as t is reached. We also consider the *access time* $H(\pi,t)$ from the stationary distribution to a target t, which is interpreted as the expected length of a random walk starting at a vertex randomly chosen (according to the probabilities from the stationary distribution) and ending at a specific vertex t.

In this project we find the specific tree structures that yield extreme hitting time and access time values among all trees of fixed diameter and order. Namely, we determine which trees T (of diameter d and order n) achieve

- $\min_T \min_i H(\pi, i)$, the smallest minimum access time from π to a vertex of T;

- $\max_T \min_i H(\pi, i)$, the largest minimum access time from π to a vertex of T;

- $\max_T \max_j H(j', j)$, the largest maximum hitting time among every pair of vertices in T;
- $\min_T \max_j H(j', j)$, the smallest maximum hitting time among every pair of vertices in T; and

Previous research has shown that the minimum and maximum hitting/access time values are achieved by stars and paths, respectively, among all trees of fixed order ([2]). This project adds the additional constraint of fixing diameter, and we find similar results. In general, among trees of fixed order and diameter, those that achieve the minimum hitting and access time values are seesaw graphs and double seesaw graphs (Figure 1.1), and those that achieve the maximum such values are broom graphs and double broom graphs (Figure 1.2).

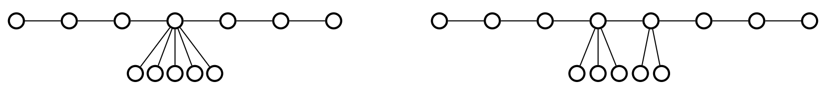

Figure 1.1 The seesaw graph (left) and double seesaw graph (right) are the minimizing trees of fixed order and diameter

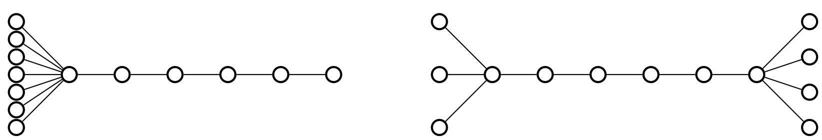

Figure 1.2 The broom (left) and double broom (right) are the maximizing trees of fixed order and diameter

Chapter 2

Preliminaries

2.1 Graphs and trees

A *graph* G consists of a set of vertices V and a set of edges E that connect pairs of vertices. The *order* of G is the number of vertices in V, and the *size* of G is the number of edges in E. We denote the edge between vertices u and v as (u, v). If $(u, v) \in E$, we say u is *adjacent* to v (denoted $u \sim v$). We refer to the set of vertices adjacent to v as the *neighborhood* of v and call the elements of this set the *neighbors* of v. The number of neighbors of v is called the *degree* of v and is denoted as $\deg(v)$. If $\deg(v) = 1$, we say that v is a *leaf*.

A *path* is a sequence $v_1, v_2, \ldots, v_k$ of unique adjacent vertices in V. That is, $(v_i, v_{i+1}) \in E$ for each $1 \leq i < k$ and $v_i = v_j$ if and only if $i = j$. Similarly, a *cycle* is a sequence $v_1, v_2, \ldots, v_k$ of adjacent vertices such that only the first and last vertices are equal. That is, $v_1 = v_k$, and all other vertices are distinct. The *length* of a path or cycle is the number of edges therein. For vertices $u, v \in V$ we say that the *path distance* (or simply *distance*) between u and v, denoted $d(u, v)$, is the length of the shortest path starting at u and ending at v (the shortest (u, v)-path). The *diameter* of G is the length of the longest path among every pair of vertices in V. If G has diameter d, we refer to any path of length d as a *geodesic* of G.

We denote the removal of an edge (u, v) (while leaving all vertices intact) from a graph G as $G - (u, v)$. Likewise, we denote the addition of (u, v) to G as $G + (u, v)$.

This paper will focus on the subset of graphs known as trees. A *tree* T is a graph that is connected and contains no cycles. Equivalently, a graph T is a tree if and only if, for every pair of vertices u, v on T, there exists exactly one (u, v)-path in T. For any $T = (V, E)$, we know that $|E| = |V| - 1$.

2.2 Special trees

Here we introduce a few families of trees that will be important in this paper.

Definition 2.1. A *caterpillar graph* is a graph in which every vertex is either on a central path or is a leaf adjacent to a vertex on the path, as shown in Figure 2.1.

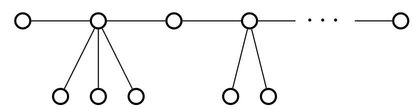

Figure 2.1 A example of a caterpillar graph

Colloquially, we refer to the central path as the *stalk* of the caterpillar and any leaves not on the path as its *legs*.

2.2.1 The seesaw and double seesaw graphs

Definition 2.2. Let $n > d \geq 1$. The *seesaw graph* $S_{n,d}$ is the graph consisting of a path $v_0, v_1, \ldots, v_d$ of length d with $n - d - 1$ leaves attached to $v_{\lfloor d/2 \rfloor}$, as shown in Figure 2.2.

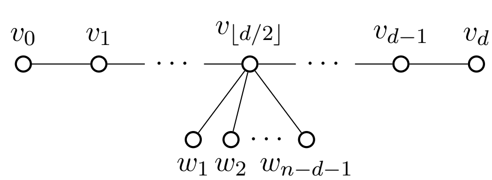

Figure 2.2 The seesaw graph $S_{n,d}$

This is just a graph consisting of a path with some leaves attached to a center vertex on the path. Note that by our definition, $S_{n,d}$ has n vertices and diameter d.

If the diameter is odd, then there are two center vertices, and so we can define a similar graph that splits the leaves between the two centers:

Definition 2.3. Let $j+k+1 > d \geq 1$, where d is odd. The *double seesaw graph* $S_{j,k,d}$ is the graph consisting of a path $v_0, v_1, \ldots, v_d$ with j leaves attached to $v_{(d-1)/2}$ and k leaves attached to $v_{(d+1)/2}$, as shown in Figure 2.3.

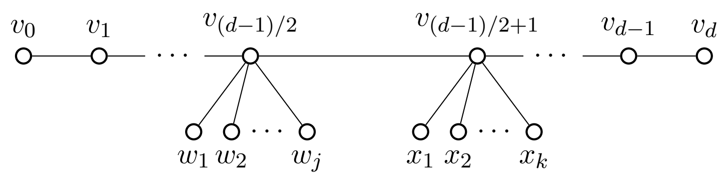

Figure 2.3 The double seesaw graph $S_{j,k,d}$, where d is odd

Note that by our definition, $S_{j,k,d}$ has $j+k+d+1$ vertices and has diameter d.

2.2.2 The broom and double broom graphs

Definition 2.4. Let $n > d \geq 1$.The *broom graph* $B_{n,d}$ is the graph formed by a path graph on d vertices $v_1, v_2, \ldots, v_d$ and attaching $n-d$ leaves $w_1, w_2, \ldots, w_{n-d}$ to vertex v_1, as shown in Figure 2.4.

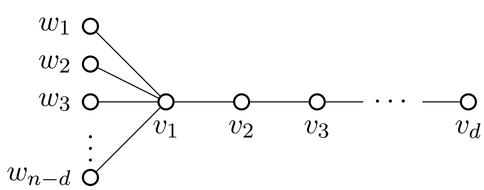

Figure 2.4 The broom graph $B_{n,d}$

Note that by our definition, the graph $B_{n,d}$ has n vertices and diameter d. To continue the broom metaphor, we collectively refer to the vertices $v_1, \ldots, v_d$ in Figure 2.4 as the *handle vertices* and those labeled $w_1, \ldots, w_{n-d}$ as the *straw vertices* (or simply *straws*). A related structure is the double broom.

Definition 2.5. The *double broom graph* $B_{j,k,d}$ is the graph formed by a path graph on $d-1$ vertices $v_1, v_2, \ldots, v_{d-1}$ and attaching j leaves $w_1, w_2, \ldots, w_d$ to v_1 and and k leaves $x_1, x_2, \ldots, x_k$ to the neighbor of the other end, as shown in Figure 2.5.

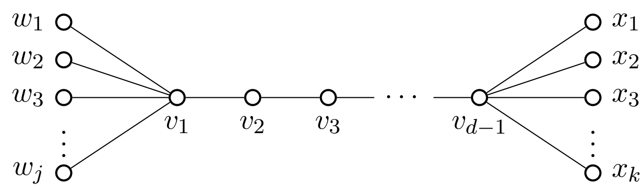

Figure 2.5 The double broom graph $B_{j,k,d}$

Note that by our definition, the graph $B_{j,k,d}$ has $j+k+d-1$ vertices and has diameter d. Naturally, we refer to $v_1, \ldots, v_{d-1}$ as the *handle vertices*, and $w_1, \ldots, w_j$ as the *left straws*, and $x_1, \ldots, x_k$ as the *right straws*.

2.3 Measuring random walks

Let $G = (V, E)$ be a graph of order n with vertex set $\{v_1, v_2, \ldots, v_n\}$. The *probability transition matrix* P of G is the $n \times n$ matrix defined by

$$P_{ij} = \begin{cases} \frac{1}{\deg(v_i)} & v_i \sim v_j \\ 0 & \text{otherwise.} \end{cases}$$

A *random walk* on G is a sequence $(w_1, w_2, \ldots, w_t, \ldots)$ of adjacent vertices where w_{t+1} is chosen uniformly at random from the neighbors of w_t (corresponding to the values in the probability transition matrix). In other words, if (i, j) is an edge on G and $w_t = i$, then the probability that $w_{t+1} = j$ is $P_{ij} = 1/\deg(i)$.

We define the *hitting time* from i to j in G, denoted $H(i, j)$, as the expected number of steps in a random walk starting at i and terminating at j. (This means that $H(i, i) = 0$.) The *stationary distribution vector* of G, denoted π gives the proportion of time spent at each vertex during a random walk on G as the length of the walk approaches ∞. That is, for any vertex i on G, we assign the i-th entry of π (denoted π_i) to be the proportion of time spent at i. For undirected graphs, this entry is proportional to the degree of i; it is given by

$$\pi_i = \frac{\deg(i)}{2|E|}.$$

This is a result of the handshaking lemma, which says that $\sum_{i \in V} \deg(i) = 2|E|$ for any graph G.

For a vertex j on G, we then define the *access time* from the stationary distribution to j, denoted $H(\pi, j)$, as the weighted average of the hitting times to j from every other vertex i, given by the formula

$$H(\pi, j) = \sum_{i \in V} \pi_i H(i, j) = \frac{1}{2|E|} \sum_{i \in V} \deg(i) H(i, j). \tag{2.6}$$

Intuitively, this is the expected time that it takes to walk from i to j, where i is a randomly selected vertex according to the stationary distribution π, and j is a specific vertex on G.

Finally, for a vertex j, a j-*pessimal* vertex j' is a vertex satisfying $H(j', j) = \max_{i \in V} H(i, j)$. This can be thought of as the "worst" starting point for walking to j. Note that sometimes there are multiple j-pessimal vertices. We typically fix one of these j-pessimal vertices to be denoted j'.

2.4 Known results for hitting times on trees

Here we summarize some useful results from [1]. Let $G = (V, E)$ be a tree. We use $G_{u:v} = (V_{u:v}, E_{u:v})$ to denote the subtree rooted at u after the removal of edge (u, v), as illustrated by Figure 2.6. The set $V_{u:v}$ then contains all the vertices that are closer (as measured by path distance) to u than to v.

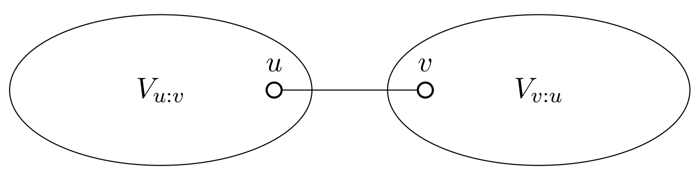

Figure 2.6 The vertex partitions induced by the removal of edge (u, v)

If i and j are adjacent vertices on G, then

$$H(i, j) = \sum_{k \in V_{i:j}} \deg(k) = 2|E| \sum_{k \in V_{i:j}} \pi_k = 2|E|\pi(V_{i:j}). \tag{2.7}$$

A detailed argument proving this result can be found in [1]. However, the intuition behind it is straightforward; if i and j are adjacent, then the entirety of any random walk starting at i and ending at j will take place on the subtree $G_{i:j}$ (i.e., on the "i side" of j), since there is no way to walk from i to the other subtree $G_{j:i}$ without first passing through j. Recall that the time spent at a given vertex during a random walk is given by the stationary distribution vector, and so we can compute $H(i,j)$ by summing over π_k for each $k \in V_{i:j}$ (and then scaling by $2|E|$).

Example 2.8. Let $T = (V, E)$ be the tree shown in Figure 2.7.

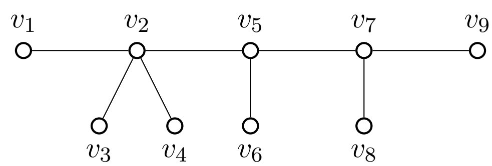

Figure 2.7 A tree on 9 vertices

Consider for example the hitting time $H(v_2, v_5)$ between adjacent vertices. By removing the (v_2, v_5) edge, we get the vertex partition $V_{v_2:v_5} = \{v_1, v_2, v_3, v_4\}$. Now, by Equation 2.7,

$$H(v_2, v_5) = \sum_{k \in V_{v_2:v_5}} \deg(k) = 1 + 4 + 1 + 1 = 7.$$

For vertices i, j, k on G, define $\ell(i,k;j)$ as the shared distance between the (i,j)-path and the (k,j)-path. (See Figure 2.8 for example.)

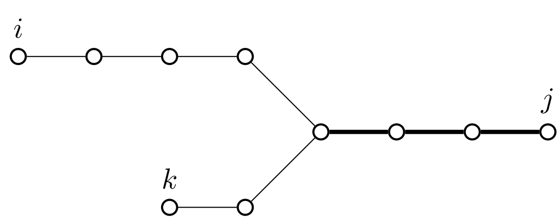

Figure 2.8 The length of the shared path (shown in bold) between the (i,j)-path and (k,j)-path is $\ell(i,k;j) = 3$.

Explicitly, this value is given by the formula

$$\ell(i,k;j) = \frac{1}{2}(d(i,j) + d(k,j) - d(i,k)).$$

The following formula gives the hitting time from i to j for any vertices i and j on G:

$$H(i,j) = \sum_{k \in V} \ell(i,k;j)\deg(k). \tag{2.9}$$

Again, this is stated without a formal proof (which can be found in [1]), however, we demonstrate the intuition behind this formula in the following example.

Example 2.10. Let us revisit our tree on on 9 vertices. This time, consider the hitting time $H(v_1, v_9)$, which is equivalent to the sum of the hitting times along each step of the (v_1, v_9)-path, that is,

$$H(v_1, v_9) = H(v_1, v_2) + H(v_2, v_5) + H(v_5, v_7) + H(v_1, v_9).$$

We can visualize the vertex partitions determined by the edges along this path as a set of circles centered at v_1, as suggested in Figure 2.9.

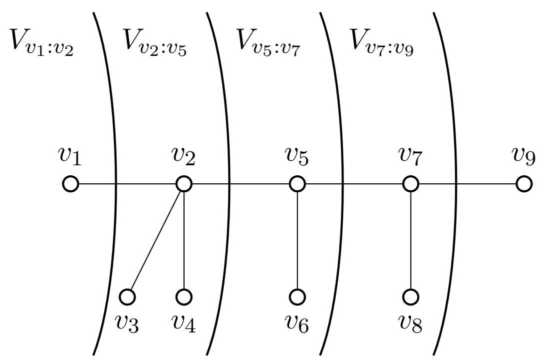

Figure 2.9 The vertex partitions along the (v_1, v_9)-path

It is then a straightforward matter to compute the hitting time using Equation 2.7:

$$H(v_1, v_9) = 1 + 7 + 11 + 15 = 34.$$

As we can see, $V_{v_1:v_2} \subset V_{v_2:v_5} \subset V_{v_5:v_7} \subset V_{v_7:v_9} \subset V$ This means that, in our hitting time formula, the degree of each vertex in $V_{v_1:v_2}$ is counted 4 times, the degree of each vertex in $V_{v_2:v_5}$ (but not in $V_{v_1:v_2}$) is counted 3 times, etc. Indeed, for every $k \in V$, the number of times $\deg(k)$ is counted is equal to $\ell(v_1, k; v_9)$, as claimed by Equation 2.9.

The *barycenter* of G, denoted c, is the vertex (or either of a pair of vertices) on G that minimizes $H(\pi, i)$. In other words, c is a barycenter of G when

$$H(\pi, c) = \min_{i \in V} H(\pi, i).$$

Proposition 2.11 ([1]). *The following statements are equivalent.*

1. *The vertex c is a barycenter of G.*
2. *The vertex c satisfies $H(i, c) \leq H(c, i)$ for all $i \in V$.*
3. $\displaystyle\sum_{k \in V} \pi_k H(k, c) = \min_{i \in V} \sum_{k \in V} \pi_k H(k, i)$.
4. *For every vertex i adjacent to c, $\pi(V_{i:c}) = \displaystyle\sum_{k \in V_{i:c}} \pi_k \leq \frac{1}{2}$.*

Note that the inequality in statement 4 above is equivalent to

$$\sum_{k \in V_{i:c}} \deg(k) \leq |E|.$$

2.5 Summary of main theorems

Throughout this paper we will consider the set of all trees of a given order and diameter when looking for the maximizing and minimizing tree structures.

Definition 2.12. Let $\mathcal{T}_{n,d}$ denote the collection of trees on n vertices with diameter d.

2.5.1 Extreme access times to the barycenter

We first examine the trees of fixed order and diameter that minimize and maximize $H(\pi, c) = \min_{i \in V} H(\pi, i)$, the access time to the barycenter of the tree.

Theorem. *The quantity*

$$\min_{T \in \mathcal{T}_{n,d}} H(\pi, c)$$

is achieved uniquely by the seesaw graph $S_{n,d}$ and is given by the formula

$$\min_{T \in \mathcal{T}_{n,d}} H(\pi, c) = \begin{cases} \frac{1}{6(n-1)} \left(3n + d^3 - 4d - 3\right) & d \text{ is even} \\ \frac{1}{6(n-1)} \left(3n + d^3 - d - 3\right) & d \text{ is odd.} \end{cases}$$

Theorem. *The quantity*

$$\max_{T \in \mathcal{T}_{n,d}} H(\pi, c)$$

is achieved by the double broom graph $B_{\lceil (n-d+1)/2 \rceil, \lfloor (n-d+1)/2 \rfloor, d}$, that is, the double broom on n vertices with diameter d such that half of the leaves ($\pm\frac{1}{2}$) are on each end of the handle. The quantity is given by

$$\begin{aligned} &\max_{T \in \mathcal{T}_{n,d}} H(\pi, c) \\ &= \begin{cases} \dfrac{1}{6(n-1)} \left(d^3 - 3d^2(n+1) \right. \\ \qquad \left. + d\left(3n^2 + 6n - 1\right) - 6n^2 + 3n + 3\right) & n \text{ is odd and } d \text{ is even} \\ \\ \dfrac{1}{6(n-1)} \left(d^3 - 3d^2(n+1) \right. \\ \qquad \left. + d\left(3n^2 + 6n - 1\right) - 6n^2 + 6n\right) & n \text{ is even and } d \text{ is odd} \\ \\ \dfrac{1}{6(n-1)} \left(d^3 - 3d^2(n+1) \right. \\ \qquad \left. + d\left(3n^2 + 6n + 2\right) - 6n^2 + 3n - 3\right) & n \text{ is even and } d \text{ is even} \\ \\ \dfrac{1}{6(n-1)} \left((d-2)\left(d^2 - d(3n+1) + 3n^2\right)\right) & n \text{ is odd and } d \text{ is odd.} \end{cases} \end{aligned}$$

2.5.2 Extreme pessimal hitting times

Next, we examine the trees of fixed order and diameter than minimize and maximize $\max_{j \in V} H(j', j) = \max_{j \in V} \max_{i \in V} H(i, j)$, that is, the greatest hitting time among every pair of vertices in the tree.

Theorem. *The quantity*

$$\max_{T\in\mathcal{T}_{n,d}} \max_{j\in V} H(j',j).$$

is achieved by the broom $B_{n,d}$. The maximizing vertex j is the leaf at the end of the handle, and j' is any straw vertex. The quantity is given by the formula

$$\max_{T\in\mathcal{T}_{n,d}} \max_{j\in V} H(j',j) = 2(d-1)n - d^2 + 2.$$

Theorem. *The quantity*

$$\min_{T\in\mathcal{T}_{n,d}} \max_{j\in V} H(j',j)$$

is achieved by the seesaw graph $S_{n,d}$ for even d and the double seesaw graph $S_{r,r,d}$ where $r = \lfloor (n-d-1)/2 \rfloor$ for odd d. The maximizing vertices j and j' are the leaves at opposite ends of the path of length d. The quantity is given by the formula

$$\min_{T\in\mathcal{T}_{n,d}} \max_{j\in V} H(j',j) = \begin{cases} d(n-1) & \begin{cases} d \text{ is even} \\ d \text{ is odd and } n \text{ is even} \end{cases} \\ d(n-1)+1 & d \text{ is odd and } n \text{ is odd}. \end{cases}$$

2.6 Related Work

Previous work has found similar results for the hitting time, the access time to the barycenter, and several other mixing measures.

Beveridge and Wang [2] found the maximizing and minimizing structures for trees only of fixed order (where the diameter is free). Their results showed that the path graph maximizes hitting times, commute times, $H(\pi, i)$, and the mixing time, while the star graph minimizes these metrics.

Beveridge and Youngblood [3] developed bounds for the best mixing time T_{bestmix} among trees of order n. The path yields this value for even n and a closely related graph, the wishbone graph, yields this value for odd n.

Kemeny's constant $\mathcal{K}$ is a related concept that measures the expected number of steps to get from an arbitrary starting vertex to a randomly chosen target vertex in a random walk. In previous work, Ciardo, Dahl, and Kirkland [5] found upper and lower bounds for $\mathcal{K}$ on trees in terms of the order n and diameter d. The minimizing tree is a specific family of caterpillars and the maximizing trees are broom-stars, which are similar results to our findings for the minimizing and maximizing structures for hitting time and access time to the barycenter.

More generally, Brightwell and Winkler [4] examined maximum hitting times the wider family of graphs (not just trees) of fixed order. Their research found that the maximum hitting time occurs on lollipop graphs (or on closely related graphs).

2.7 Notation

The general proof structure used throughout this paper involves starting with a tree T of order n with diameter d and pruning it (by moving edges and vertices) to increase or decrease a specific mixing measure (while ensuring that this modified graph is still a tree of order n with diameter d). In general, we will label this modified tree as T^* and refer to its properties correspondingly. For example, if $H(\pi, c)$ denotes the access time to the barycenter in T, then we will write the access time to the barycenter in T^* as $H^*(\pi^*, c^*)$.

Chapter 3

Minimizing the access time to the barycenter

Our first goal is to determine which tree in $\mathcal{T}_{n,d}$ has the minimum access time from its stationary distribution to its barycenter. In this chapter we will show that the tree is the seesaw graph (Definition 2.2).

We begin with a useful lemma about hitting times. If we "pluck" a leaf v and reattach it to a vertex t (Figure 3.1), then all the hitting times to t either stay the same or decrease.

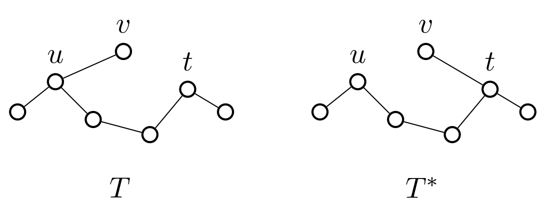

Figure 3.1 The tree obtained by plucking the (u, v) edge and adding a (t, v) edge

Lemma 3.1. *Let T be a tree with vertex set V. Let $t, u, v \in V$, and suppose that v is a leaf adjacent to u. Let $T^* = T - (u, v) + (t, v)$. Then, $H^*(s, t) \leq H(s, t)$ for every vertex $s \in V$.*

Proof. First, we easily observe that the statement holds for both $s = v$ and $s = t$. If $s = v$, then, $H^*(s, t) = 1 < H(s, t)$. And, if $s = t$, then $H^*(s, t) = 0 = H(s, t)$.

Now, let $s \in V \setminus \{t, v\}$. We have

$$H(s,t) = \sum_{k \in V} \ell(s,k;t) \deg(k)$$

and

$$H^*(s,t) = \sum_{k \in V} \ell^*(s,k;t) \deg^*(k).$$

Now, since v is a leaf, we know that $\ell(s,k;t) = \ell^*(s,k;t)$ for all $k \neq v$. Furthermore, $\deg(k) = \deg^*(k)$ for all $k \neq t, u$. So, the terms of the two sums are identical, except for those corresponding to vertices t, u, and v. We can then write

$$\begin{aligned} H^*(s,t) = \quad & H(s,t) \\ & + \ell^*(s,t;t) \deg^*(t) - \ell(s,t;t) \deg(t) \\ & + \ell^*(s,u;t) \deg^*(u) - \ell(s,u;t) \deg(u) \\ & + \ell^*(s,v;t) \deg^*(v) - \ell(s,v;t) \deg(v). \end{aligned}$$

First, observe that $\ell(s,t;t) = \ell^*(s,t;t) = 0$ (since the (t,t)-path has length zero). So,

$$\ell^*(s,t;t) \deg^*(t) - \ell(s,t;t) \deg(t) = 0 - 0 = 0.$$

Next, since $\ell^*(s,u;t) = \ell(s,u;t)$ and $\deg^*(u) = \deg(u) - 1$, we have

$$\begin{aligned} \ell^*(s,u;t) \deg^*(u) - \ell(s,u;t) \deg(u) &= \ell(s,u;t) \left(\deg(u) - 1 - \deg(u)\right) \\ &= -\ell(s,u;t). \end{aligned}$$

Finally, consider the quantity $\ell^*(s,v;t)$. Since v is a leaf adjacent to t in T^*, the (v,t)-path and (s,t)-path have no edge in common. Therefore,

$$\ell^*(s,v;t) \deg^*(v) - \ell(s,v;t) \deg(v) = -\ell(s,v;t) \deg(v) = -\ell(s,v;t).$$

Returning to our equation, we now have

$$H^*(s,t) = H(s,t) - \ell(s,u;t) - \ell(s,v;t).$$

Thus, $H^*(s,t) \leq H(s,t)$. ■

Next, we prove a well-known result about hitting times on the the path P_{d+1} (that is, the path of diameter d).

Lemma 3.2. *Let G be a path graph with vertices labeled $v_0, v_1, \ldots, v_d$. Then, the hitting time between v_i and v_j in G is given by*

$$H(v_i, v_j) = \begin{cases} j^2 - i^2 & 0 \le i < j \le d, \\ (d-j)^2 - (d-i)^2 & 0 \le j < i \le d. \end{cases}$$

Proof. First, let $0 < k \le d$. We claim that $H(v_0, v_k) = k^2$. This can be proven by induction. For the base case, we have that $H(v_0, v_1) = 1 = 1^2$, since v_0 is a leaf adjacent to v_1. Now, suppose for our hypothesis that

$$H(v_0, v_{k-1}) = (k-1)^2.$$

Since G is a path graph, any (v_0, v_k)-walk must conclude with the edge (v_{k-1}, v_k) for all $k > 0$. Therefore,

$$\begin{aligned} H(v_0, v_k) &= H(v_0, v_{k-1}) + H(v_{k-1}, v_k) \\ &= (k-1)^2 + H(v_{k-1}, v_k). \end{aligned}$$

We have that

$$\begin{aligned} H(v_{k-1}, v_k) &= \sum_{i \in V_{v_{k-1}:v_k}} \deg(i) \\ &= 1 + 2(k-1) \\ &= 2k - 1. \end{aligned}$$

Returning to the previous equation, we get

$$\begin{aligned} H(v_0, v_k) &= (k-1)^2 + 2k - 2 \\ &= k^2 - 2k + 1 + 2k - 1 \\ &= k^2. \end{aligned}$$

Now, let $0 < i < j \le d$. By a similar process, we claim that

$$H(v_i, v_j) = j^2 - i^2.$$

Again, since G is a path, we have that

$$\begin{aligned} H(v_i, v_j) &= H(v_0, v_j) - H(v_0, v_i) \\ &= j^2 - i^2. \end{aligned}$$

Finally, let $0 < j < i \leq d$. We claim that $H(v_i, v_j) = (d-j)^2 - (d-i)^2$. By the same reasoning used to show that $H(v_0, v_k) = k^2$, we also argue that $H(v_d, v_k) = (d-k)^2$ (for any $0 \leq k < d$). Now,

$$\begin{aligned} H(v_i, v_j) &= H(v_d, v_j) - H(v_d, v_i) \\ &= (d-j)^2 - (d-i)^2, \end{aligned}$$

as desired. ∎

We can then use this lemma to show that the access time to a center vertex of any path is at least as large as the access time to any vertex to its left.

Lemma 3.3. *Let G be a path graph with vertices labeled $v_0, v_1, \ldots, v_c, v_{c+1}, \ldots, v_d$, where $c < \lfloor \frac{d}{2} \rfloor$ (as shown in Figure 3.2). Then, $H(\pi, v_c) \geq H(\pi, v_{c+1})$.*

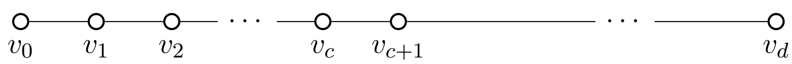

Figure 3.2 A path G of diameter d

Proof. Let G be the path graph shown in Figure 3.2 with vertex set V and edge set E. Since G is a path, every walk from a vertex v_k with $k < c+1$ to v_{c+1} must include v_c. Likewise, every walk from a vertex v_k with $k > c$ to v_c must include v_{c+1}. So, we have

$$H(v_k, v_{c+1}) = H(v_k, v_c) + H(v_c, v_{c+1})$$

for all $k < c+1$, and

$$H(v_k, v_c) = H(v_k, v_{c+1}) + H(v_{c+1}, v_c)$$

for all $k > c$. Then, observe that

$$\begin{aligned} H(\pi, v_c) &= \sum_{v_k \in V} \pi_{v_k} H(v_k, v_c) \\ &= \sum_{k=0}^{c-1} \pi_{v_k} H(v_k, v_c) + \sum_{k=c+1}^{d} \pi_{v_k} H(v_k, v_c) \\ &= \sum_{k=0}^{c-1} \pi_{v_k} H(v_k, v_c) + \sum_{k=c+1}^{d} \pi_{v_k} H(v_k, v_{c+1}) + \sum_{k=c+1}^{d} \pi_{v_k} H(v_{c+1}, v_c). \end{aligned}$$

Likewise,

$$\begin{aligned} H(\pi, v_{c+1}) &= \sum_{v_k \in V} \pi_{v_k} H(v_k, v_{c+1}) \\ &= \sum_{k=0}^{c} \pi_{v_k} H(v_k, v_{c+1}) + \sum_{k=c+2}^{d} \pi_{v_k} H(v_k, v_{c+1}) \\ &= \sum_{k=0}^{c} \pi_{v_k} H(v_k, v_c) + \sum_{k=0}^{c} \pi_{v_k} H(v_c, v_{c+1}) + \sum_{k=c+2}^{d} \pi_{v_k} H(v_k, v_{c+1}). \end{aligned}$$

Since $H(v_k, v_k) = 0$ for any k, we get the difference

$$\begin{aligned} H(\pi, v_c) - H(\pi, v_{c+1}) &= \sum_{k=c+1}^{d} \pi_{v_k} H(v_{c+1}, v_c) - \sum_{k=0}^{c} \pi_{v_k} H(v_c, v_{c+1}) \\ &= \sum_{k=c+1}^{d} \pi_{v_k} \left((d-c)^2 - (d-c-1)^2\right) \\ &\qquad - \sum_{k=0}^{c} \pi_{v_k} \left((c+1)^2 - c^2\right) \\ &= \sum_{k=c+1}^{d} \pi_{v_k} (2d - 2c - 1) - \sum_{k=0}^{c} \pi_{v_k} (2c+1), \end{aligned}$$

by Lemma 3.2. Since $c < \lfloor \frac{d}{2} \rfloor$, it follows that $c \leq \frac{d-1}{2}$. So,

$$\begin{aligned} H(\pi, v_c) - H(\pi, v_{c+1}) &\geq \sum_{k=c+1}^{d} \pi_{v_k} \left(2d - 2\frac{d-1}{2} - 1\right) \\ &\qquad - \sum_{k=0}^{c} \pi_{v_k} \left(2\frac{d-1}{2} + 1\right) \\ &= \sum_{k=c+1}^{d} \pi_{v_k} d - \sum_{k=0}^{c} \pi_{v_k} d. \end{aligned}$$

Note that $\pi_{v_k} = \deg(v_k)/2|E|$ for any k, so we have

$$H(\pi, v_c) - H(\pi, v_{c+1}) \geq \frac{1}{2|E|} \left(\sum_{k=c+1}^{d} \deg(v_k) d - \sum_{k=0}^{c} \deg(v_k) d \right).$$

And, since G is a path, we can further simplify to get

$$\begin{aligned} H(\pi, v_c) - H(\pi, v_{c+1}) &\geq \frac{1}{|E|} \left(\sum_{k=c+1}^{d-1} d - \sum_{k=1}^{c} d \right) \\ &= \frac{d}{|E|}(d - c - 1 - c) \\ &= \frac{d}{|E|}(d - 2c - 1) \\ &\geq \frac{d}{|E|} \left(d - 2\frac{d-1}{2} - 1 \right) \\ &= 0. \end{aligned}$$

Therefore, since their difference is at least zero, it follows that $H(\pi, v_c) \geq H(\pi, v_{c+1})$. ■

We are now equipped to prove the main theorem of this chapter. Recall that we denote the barycenter of a tree as c.

Theorem 3.4. *The quantity*

$$\min_{T \in \mathcal{T}_{n,d}} H(\pi, c)$$

is achieved uniquely by the seesaw graph $S_{n,d}$.

Proof. Let $T = (V, E)$ be a tree of order n, and diameter d. We will assume that $d < n - 1$, since if $d = n - 1$, then T can only be a path, and we are done. Furthermore, suppose that $T \neq S_{n,d}$. To prove that $H(\pi, c)$ is not at a minimum, we will demonstrate a process by which we can relocate the vertices and edges of T to decrease its value.

Step 1. First, fix a path P in T with vertices $v_0, v_1, \ldots, v_d$ such that (v_{i-1}, v_i) is an edge for $1 \leq i \leq d$. Note that we will not remove the leaves v_0, v_d of P (and thus are guaranteed to maintain diameter d). Since $T \neq S_{n,d}$, there exists a leaf $v \in V \setminus \{v_0, v_d\}$ that is not adjacent to c. Suppose instead that $v \sim u$. Create a new tree T^* by plucking v and attaching it to c, i.e.,

$$T^* = T - (u, v) + (c, v).$$

The vertex c is a barycenter of T. We claim that c is a barycenter of T^*. From Proposition 2.11, we know that a vertex c is a barycenter of a tree of size $|E|$ if and only if

$$\sum_{k \in V_{i:c}} \deg(k) \leq |E| \tag{3.5}$$

for each $i \sim c$. Consider the vertex sets $V_{i:c}$ in T^* for each vertex i adjacent to c. These are identical to the vertex sets of T, except for: (a) the set $V_{j:c}$ containing u loses one vertex (namely, v); and (b) we have one new set $V_{v:c}$. For set $V_{j:c}$, the sum of degrees decreases by 2, so it still satisfies inequality Equation 3.5. The set $V_{v:c}$ has a degree sum of 1, which is also less than $|E|$. Thus, c is in fact a barycenter of T^*.

Next, we show that $H^*(\pi^*, c) < H(\pi, c)$. Since

$$H(\pi, c) = \sum_{k \in V} \pi_k H(k, c)$$

and

$$H^*(\pi^*, c) = \sum_{k \in V} \pi^*_k H(k, c),$$

we can show this termwise. Let $k \in V$, and consider the term from each sum that corresponds to k. Ignoring c (since $H(c, c) = 0$), we have three cases.

1. Suppose $k = v$. In this case, we have $\pi_v = \pi^*_v = 1/2|E|$ since $\pi_k = \deg(k)/2|E|$. In T, the hitting time $H(v, c) > 1$, since $v \not\sim c$. But, in T^*, we have $H^*(v, c) = 1$, since v is adjacent only to c. Therefore,

$\pi_v^* H^*(v, c) < \pi_v H(v, c)$.

2. Suppose $k = u$. Then, we have $\pi_u = \deg(u)/2|E|$ and $\pi_u^* = (\deg(u) - 1)/2|E|$. So, $\pi_u^* < \pi_u$. Also, by Lemma 3.1, we know that $H^*(u, c) \leq H(u, c)$. It then follows that $\pi_u^* H^*(u, c) < \pi_u H(u, c)$.

3. Finally, suppose $k \neq u, v$. Then, $\pi_k^* = \pi_k$, and $H^*(k, c) \leq H(k, c)$ by Lemma 3.1. Thus, $\pi_k^* H^*(k, c) \leq \pi_k H(k, c)$.

So, since every term of $H^*(\pi^*, c)$ is no greater than its analogous term in $H(\pi, c)$, it follows that $H^*(\pi^*, c) \leq H(\pi, c)$.

By repeating this process, we arrive at a final T^* the takes the form shown on the left in Figure 3.3. That is, T^* contains a path P on vertices $v_0, v_1, \ldots, v_d$ and a barycenter $c^* = c$ with a cluster of leaves attached. Either $c^* = v_k$ for some $0 \leq k \leq d$, or v_k is connected to c^* via another path $w_1, w_2, \ldots, w_j$. If c is on the path P, then we move ahead to Step 3. Otherwise, we want to move vertices to place the barycenter on P.

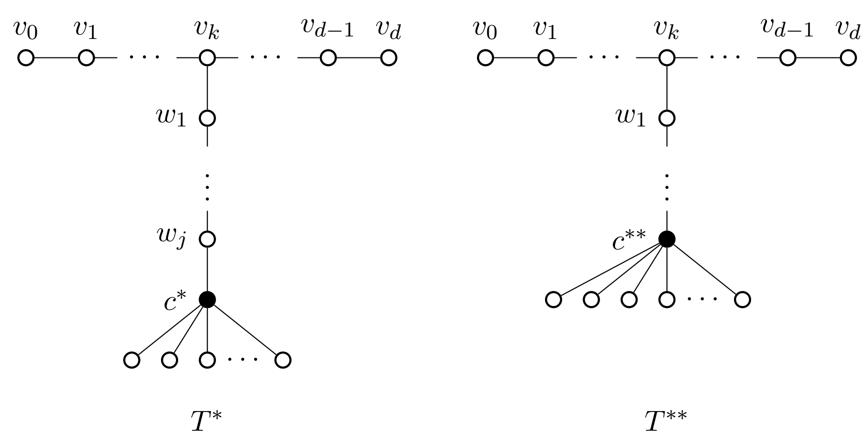

Figure 3.3 Moving the leaves (and the barycenter) toward P

Step 2. To make T^{**}, take the cluster of leaves adjacent to c^* in T^*, and move them all so that they are adjacent to w_j, i.e., one edge closer to the path P, as shown in Figure 3.3. To show that $w_j = c^{**}$ must be the barycenter of T^{**}, compare its vertex sets to those of c^* in T^*. Every vertex set V consisting solely of a leaf has degree sum 1, so those satisfy the condition from Equation 3.5. If c^{**} is still not on P, then there is one other vertex set (the one that contains the path P); its degree sum decreases by 2, so it still satisfies Equation 3.5. Alternatively, if c^{**} is on P, then the vertex set $V_{c^{**}:c^*} = P$ splits into to two smaller vertex sets: $\{v_0, v_1, \ldots, v_{k-1}\}$ and $\{v_{k+1}, \ldots, v_d\}$. So, Equation 3.5 is still satisfied.

We claim that $H^{**}(\pi^{**}, c^{**}) < H^*(\pi^*, c^*)$. Again, we show this termwise. Let $k \in V$.

1. Suppose k is a leaf adjacent to c^* in T^*. Then, k is a leaf adjacent to c^{**} in T^{**}. So, $H^{**}(k, c^{**}) = H^*(k, c^*) = 1$.

2. Suppose k is not a leaf adjacent to c^* in T^* and that $k \neq c^*, c^{**}$. Then, the degree of k is unchanged, so $\pi_k^{**} = \pi_k^*$. Furthermore,

$$\begin{aligned} H^*(k, c^*) &= H^*(k, c^{**}) + H^*(c^{**}, c^*) \\ &= H^{**}(k, c^{**}) + H^*(c^{**}, c^*), \end{aligned}$$

since every walk from k to c^* must pass through c^{**}. Therefore, $H^{**}(k, c^{**}) \leq H^*(k, c^*)$.

3. Finally, consider $k = c^*$ and $k = c^{**}$ as a pair. The corresponding hitting times scaled by degree (in each of T^* and T^{**}) are given in the table below, where

$$x = \sum_{i \in V_{w_j:c^*}} \deg^*(i)$$

(per Equation 2.7).

k	$\deg^*(k)H^*(k, c^*)$	$\deg^{**}(k)H^{**}(k, c^{**})$
c^*	0	1
$c^{**} = w_j$	$2x$	0

Since x is certainly at least 1, we conclude that

$$\pi_{c^*}^{**} H^{**}(c^*, c^{**}) \leq \pi_{c^{**}}^* H^*(c^{**}, c^*).$$

Putting it all together, we have

$$\begin{aligned} H^{**}(\pi^{**}, c^{**}) &= \sum_{k \in V} \pi_k^{**} H^{**}(k, c^{**}) \\ &= \sum_{k \neq c^*, c^{**}} \pi_k^{**} H^{**}(k, c^{**}) + \pi_{c^*}^{**} H^{**}(c^*, c^{**}) \\ &\leq \sum_{k \neq c^*, c^{**}} \pi_k^{*} H^{*}(k, c^{*}) + \pi_{c^{**}}^{*} H^{*}(c^{**}, c^*) \\ &= H^*(\pi^*, c^*), \end{aligned}$$

as desired.

By repeating this process, we arrive at a final T^{**} that takes the form of a caterpillar with stalk length d and the remaining vertices attached as leaves to c^{**}, as shown on the left in Figure 3.4. This proves that the minimum value of $H(\pi, c)$ occurs for some tree of this form. The only question remaining is where on the stalk to put the cluster of leaves (and the barycenter) so as to minimize $H(\pi, c)$. We claim that attaching the cluster of leaves to the center vertex on P gives the minimum value. If d is even, then the center vertex is simply $v_{d/2}$. If d is odd, either $v_{\lfloor d/2 \rfloor}$ or $v_{\lceil d/2 \rceil}$ will suffice as the center vertex. (Either choice will yield the same value of $H(\pi, c)$.) Arbitrarily, we will choose $v_{\lfloor d/2 \rfloor}$ to be the center vertex.

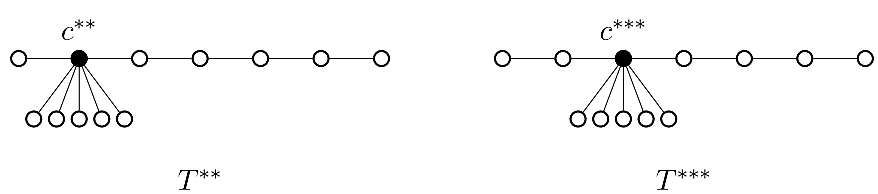

Figure 3.4 Moving the leaves toward the center of P

Step 3. Now, suppose we get a tree T^{**} by the process above but that the leaves are not attached to the center vertex. That is, a path P with vertices $v_0, v_1, \ldots, v_d$ and $n - d - 1$ leaves attached to vertex v_k. Let T^{***} be the same tree but with the leaf cluster moved one step toward the center (as shown in Figure 3.4). First, observe that the barycenter becomes the new leaf-adjacent vertex, since the degree sums of the subtrees rooted are each no greater than $|E|$.

Next, since the hitting times from vertices in the leaf cluster to the barycenter are unchanged, we only need to compare the hitting times from vertices along the path. By Lemma 3.3, it then follows that $H^{***}(\pi^{***}, c^{***}) \leq H^{**}(\pi^{**}, c^{**})$.

By repeating this process, we arrive at a tree on n vertices of diameter d, in which all vertices not on the diameter are leaves attached to the center vertex. This gives us the minimum value of $H(\pi, c)$. ■

Having identified the unique tree $S_{n,d}$ achieving $\min_{T\in\mathcal{T}_{n,d}} H(\pi, c)$, we calculate the value of this access time.

Lemma 3.6. *We have*

$$\min_{T\in\mathcal{T}_{n,d}} H(\pi, c) = \begin{cases} \frac{1}{6(n-1)}\left(3n + d^3 - 4d - 3\right) & d \text{ is even} \\ \frac{1}{6(n-1)}\left(3n + d^3 - d - 3\right) & d \text{ is odd.} \end{cases}$$

Proof. It is straightforward to compute the formula for the access time to the barycenter in $S_{n,d}$ using Equation 2.6. Label the vertices of $S_{n,d}$ as shown in Figure 2.2. Let v_r be the vertex directly to the left of c and v_s be the vertex directly to the right of c. If d is even, then we have $c = v_{d/2}$, $r = d/2 - 1$, and $s = d/2 + 1$; if d is odd, then $c = v_{(d-1)/2}$, $r = (d-1)/2 - 1$, and $s = (d-1)/2 + 1$. Then, we have

$$\begin{aligned} H(\pi, c) &= \frac{1}{2|E|} \sum_{k\in V} \deg(k) H(k, c) \\ &= \frac{1}{2(n-1)} \left(1 \cdot H(v_0, c) + \sum_{i=1}^{r} 2 \cdot H(v_i, c) \right. \\ &\qquad \left. + \sum_{i=s}^{d} 2 \cdot H(v_i, c) + 1 \cdot H(v_d, c) + \sum_{i=1}^{n-d-1} 1 \cdot H(w_i, c) \right). \end{aligned}$$

Now, we can use the path hitting time formula given by Lemma 3.2 to compute $H(v_i, c)$ for each $0 \leq i \leq d$ (since the leaves of c do not affect hitting times on the path). Furthermore, the hitting time $H(w_i, c)$ is simply 1 for each $1 \leq i \leq n - d - 1$. These yield the following result:

$$\min_{T\in\mathcal{T}_{n,d}} H(\pi, c) = \begin{cases} \frac{1}{6(n-1)}\left(3n + d^3 - 4d - 3\right) & d \text{ is even} \\ \frac{1}{6(n-1)}\left(3n + d^3 - d - 3\right) & d \text{ is odd,} \end{cases}$$

as desired. ■

Chapter 4

Maximizing the access time to the barycenter

We now describe a process by which we can increase $H(\pi, c)$ in order to find the tree of order n and diameter d that yields the *maximum* value of $H(\pi, c)$. We claim that the double broom (Definition 2.5) maximizes $H(\pi, c)$.

First, we prove the following lemma.

Lemma 4.1 (Broom Lemma)**.** *For a fixed leaf ℓ on the diameter, the quantity*

$$\max_{T \in \mathcal{T}_{n,d}} H(\pi, \ell)$$

is achieved by the broom graph $B_{n,d}$, where ℓ is the leaf at the end of the broom's handle.

Proof. We prove this by double induction over n, the order of the tree, and d, the diameter of the tree.

Base case. Let $T \in \mathcal{T}_{n,d}$, where $d = 3$. Fix a path with vertices v_0, v_1, v_2, v_3 such that $v_i \sim v_{i+1}$ for $0 \leq i < 3$. Then, since T is a tree of diameter 3, each vertex not on the path must be a leaf attached to v_1 or a leaf attached to v_2. Let a denote the number of leaves attached to v_1 (*including* v_0) and b denote the number of leaves attached to v_2 (*excluding* v_3). Note that $a + b = n - 3$ and $1 \leq a \leq n - 3$, as shown in Figure 4.1.

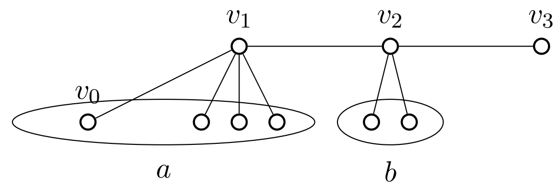

Figure 4.1 The graph T of diameter 3. In this example $a = 4$, and $b = 2$. Since $n = 9$, this satisfies the equation $a + b = 6 = n - 3$.

We can then use Equation 2.7 to find hitting times for every pair of adjacent vertices in T:

- $H(\ell, v_1) = 1$ for all leaves $\ell \sim v_1$;
- $H(v_1, v_2) = 2a + 1$;
- $H(\ell, v_2) = 1$ for all leaves $\ell \sim v_2$; and
- $H(v_2, v_3) = 2a + 2b + 3$.

This leads to the following hitting times to our target vertex v_3:

- $H(\ell, v_1) = 1 + (2a + 1) + (2a + 2b + 3) = 4a + 2b + 5$ for all $\ell \sim v_1$;
- $H(v_1, v_3) = (2a + 1) + (2a + 2b + 3) = 4a + 2b + 4$;
- $H(\ell, v_3) = 1 + (2a + 2b + 3) = 2a + 2b + 4$ for all $\ell \sim v_2$; and
- $H(v_2, v_3) = 2a + 2b + 3$.

We can use these to derive a formula $J(a, b) = H(\pi, v_3)$ for arbitrary a and b:

$$
\begin{aligned}
J(a,b) &= \frac{1}{2(n-1)}(a(4a + 2b + 5) + (a + 1)(4a + 2b + 4) \\
&\qquad + b(2a + 2b + 4) + (b + 2)(2a + 2b + 3)) \\
&= \frac{1}{2(n-1)}\left(8a^2 + 8ab + 17a + 4b^2 + 13b + 10\right).
\end{aligned}
$$

Suppose $a \leq n - d$, meaning that there is at least one leaf $\ell \neq v_3$ that could be plucked from v_2 and reattached to v_1. Let T^* be that tree, i.e.,

$T^* = T - (v_2, \ell) + (v_1, \ell)$. We have

$$\begin{aligned} H^*(\pi, v_3) - H(\pi, v_3) &= J(a+1, b-1) - J(a, b) \\ &= \frac{1}{2(n-1)}(8a+8) > 0, \end{aligned}$$

which means that $H^*(\pi, v_3) > H(\pi, v_3)$. Thus, we can repeatedly pluck leaves from v_2 and reattach them to v_1, thereby increasing $H(\pi, v_3)$, until $a = n - d$ and $b = 0$. It then follows that the tree in $\mathcal{T}_{n,3}$ that maximizes $H(\pi, v_3)$ is $B_{n,3}$.

Inductive hypothesis. Let $n' < n$ and $d' < d$. Now, assume for our hypothesis that, among trees on n' vertices with diameter d', the maximum value of $H(\pi, \ell)$ is achieved by $B_{n',d'}$.

Inductive step. We want to show that $B_{n,d}$ maximizes $H(\pi, \ell)$ in $\mathcal{T}_{n,d}$. Let $T \in \mathcal{T}_{n,d}$ and fix a path with vertices $v_0, v_1, v_2, \ldots, v_{d-1}, v_d$. To derive a general formula for the hitting time to v_d, let $k \neq v_d$ be in the vertex set of T, denoted V_T. Then,

$$\begin{aligned} H(k, v_d) &= H(k, v_{d-1}) + H(v_{d-1}, v_d) \\ &= H(k, v_{d-1}) + \sum_{j \in V_{v_{d-1}:v_d}} \deg(j) \\ &= H(k, v_{d-1}) + \sum_{j \in V_T} \deg(j) - \deg(v_d) \\ &= H(k, v_{d-1}) + 2(n-1) - 1 \\ &= H(k, v_{d-1}) + 2n - 3, \end{aligned}$$

since the degree sum of V_T is $2(n-1)$ by the handshaking lemma.

Assume that T maximizes $H(\pi, v_d)$. Let

$$S = T - \{v_d, u_1, \ldots, u_b\}$$

be the subtree (of order $n - 1 - b$ and diameter $d - 1$) induced by removing the leaf v_d and the other b leaves adjacent to v_{d-1} from T. Let H_S denote the value of $H(\pi, v_{d-1})$ for S. From Equation 2.6, we get

$$2|E_S|H_S = \sum_{k \in V_S} \deg(k) H(k, v_{d-1}), \tag{4.2}$$

where $|E_S|$ denotes the number of edges in S. This follows from the fact that, for any $k \in V_S$, the hitting time $H(k, v_{d-1})$ is equivalent in S and T. Let H_T denote the value of $H(\pi, v_d)$ for T. We can express H_T in terms of H_S by separately summing over the vertices in V_S, the single vertex v_{d-1} and the remaining b leaves in V_T:

$$\begin{aligned} H_T &= \frac{1}{2|E_T|} \sum_{k \in V_T} \deg_T(k) H(k, v_d) \\ &= \frac{1}{2|E_T|} \Bigg(\sum_{k \in V_S} \deg_S(k) \big(H(k, v_{d-1}) + H(v_{d-1}, v_d) \big) \\ &\qquad + \deg_T(v_{d-1}) H(v_{d-1}, v_d) + \sum_{i=1}^{b} \big(H(u_i, v_d) \big) \Bigg) \\ &= \frac{1}{2|E_T|} \Bigg(\sum_{k \in V_S} \deg_S(k) H(k, v_{d-1}) + (n - b - 1) H(v_{d-1}, v_d) \\ &\qquad + (b+2) H(v_{d-1}, v_d) + b H(v_{d-1}, v_d) + b \Bigg). \end{aligned}$$

We can then use Equation 4.2 to get

$$H_T = \frac{1}{2|E_T|} \left(2|E_S| H_S + (n + b + 1) H(v_{d-1}, v_d) + b \right).$$

Noting that $|E_T| = n - 1$ and $|E_S| = n - b - 2$ and that $H(v_{d-1}, v_d) = 2n - 3$, we arrive at

$$H_T = \frac{1}{2(n-1)} \left(2(n - b - 2) H_S + (n + b + 1)(2n - 3) + b \right).$$

Now, since S is a tree of diameter $d - 1$ and order less than n, our inductive hypothesis tells us that H_S is maximized for $S = B_{n-1-b,d-1}$. If $b > 0$ then moving the b leaves to be adjacent to v_1 further increases the access time from the stationary distribution to v_d. Therefore, we should set $b = 0$. Thus, the maximum value of H_T occurs for $T = B_{n-1,d-1} + (v_{d-1}, v_d)$. This is equivalent the broom on $n - 1$ vertices of diameter $d - 1$ with an additional vertex attached to the end of its handle, and thus T is equivalent the broom graph $B_{n,d}$. ■

We now derive an exact formula for the access time to the end of the handle of the graph $B_{n,d}$ in terms of n and d.

Lemma 4.3. *Let $G = B_{n,d}$ be a broom, and let v_d be the leaf at the end of the handle of G. Then,*

$$H(\pi, v_d) = \frac{1}{6(n-1)}\left((12d-12)n^2 + (-12d^2+15)n + 4d^3 - 4d - 3\right).$$

Proof. Let $G = B_{n,d}$. Label the straw vertices $w_1, w_2, \ldots, w_{n-d}$ and the handle vertices $v_1, v_2, \ldots, v_d$ (as shown in Figure 2.4). Recall from Equation 2.7 that for any adjacent vertices i, j, we have

$$H(i,j) = \sum_{k \in V_{i:j}} \deg(k).$$

Using this formula, we can calculate the hitting times between each pair of adjacent vertices in our graph, as shown in Figure 4.2.

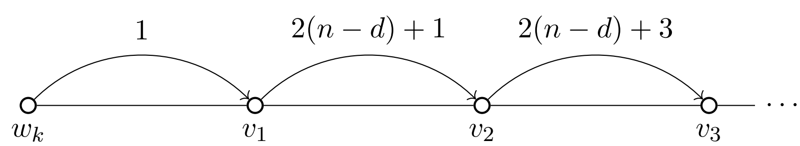

Figure 4.2 The hitting times between adjacent vertices on $B_{n,d}$, where w_k stands for any straw vertex

(Note that w_k can be any straw on G; the hitting time $H(w_k, v_1)$ is 1 for all $1 \le k \le n-d$.) Then, we can use addition to get the hitting time between each vertex and v_d. Now,

$$\begin{aligned}
H(\pi, v_d) &= \frac{1}{2|E|} \sum_{v_k \in V} \deg(v_k) H(v_k, v_d) \\
&= \frac{1}{2|E|} \Bigg((n-d) \cdot 1 \cdot H(v_0, v_d) + (n-d+1) H(v_1, v_d) \\
&\qquad\qquad + 2 \sum_{k=1}^{d-1} H(v_k, v_d) \Bigg) \\
&= \frac{1}{2(n-1)} \Bigg((n-d)(d^2 + (d-1) \cdot 2(n-d-1)) \\
&\quad + (n-d+1)(d^2 - 1^2 + (d-1) \cdot 2(n-d-1)) \\
&\qquad + 2 \sum_{k=2}^{d-1} (d^2 - k^2 + (d-k) \cdot 2(n-d-1)) \Bigg) \\
&= \frac{1}{6(n-1)} \left((12d - 12)n^2 + (-12d^2 + 15)n + 4d^3 - 4d - 3 \right),
\end{aligned}$$

which gives us our formula. ∎

With these lemmas in hand, we are now ready to show that a double broom is the maximizing structure for the access time to the barycenter c.

Theorem 4.4. *The quantity*

$$\max_{T \in \mathcal{T}_{n,d}} H(\pi, c)$$

is achieved by a double broom.

Proof. Let T be a tree of order n and diameter d. Suppose that T is not a double broom. To show that $H(\pi, c)$ is not at a maximum, we will demonstrate a process by which we can relocate the vertices and edges of T to increase its value.

Step 1. Find a geodesic (a path of length d) in T, and split it into two segments at the vertex b on the geodesic that is closest to the barycenter c. Let d_1 and d_2 be the lengths of these segments, where $d_1 \geq d_2$ (and $d_1 + d_2 = d$), and let $\delta = d(b, c)$. We want c to be on a geodesic, so if $\delta = 0$, then we can skip ahead to Step 2. Otherwise, label the vertex sets of the subtrees rooted at c as $V_0, V_1, \ldots, V_k$, where $\deg(c) = k + 1$ and V_0 contains the split geodesic.

To "broomify" a vertex set V_i, we extend a path until the diameter is maximized, then place any remaining vertices as straws at the end of the broom, as shown in Figure 4.3.

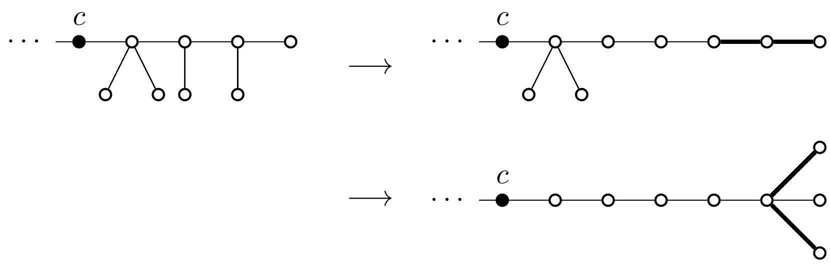

Figure 4.3 The broomification of a vertex set: first, extend the path until the diameter is maximized (top); second, place any remaining vertices not on the path as far away from c as possible (bottom)

To define the process explicitly, suppose V_i contains m vertices. If $m+\delta \leq d_2$, then replace V_i with a path on m vertices rooted at c. Otherwise, replace it with the broom $B_{m,d_2-\delta-1}$. By Lemma 4.1, this maximizes $H(\pi, c)$ within each vertex set, and thus increases $H(\pi, c)$ for all of T.

Now, for each $1 \leq i \leq k$, broomify the vertex set V_i. This process will result in one of the following cases.

1. At least one of the vertex sets has diameter $d_2 - \delta - 1$ (Figure 4.4). In this case, c is now on a path of length $d_1 + \delta + d_2 - \delta = d$, and thus c is on a geodesic, as desired.

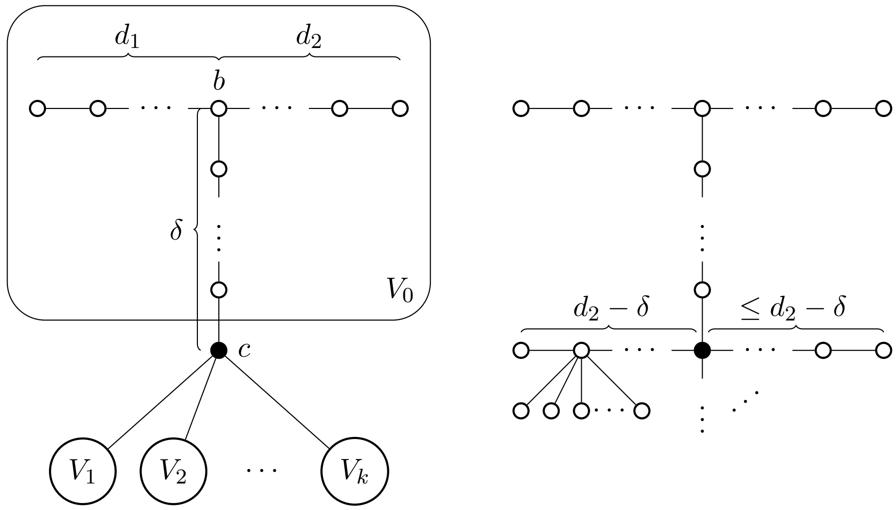

Figure 4.4 On the left is a tree with a fixed geodesic of length $d = d_1 + d_2$, where $d_1 \geq d_2$. It has barycenter c with a distance of δ from the geodesic and vertex sets V_0 (containing the geodesic) and $V_1, \ldots, V_k$, where $k = deg(c)+1$. Case 1 in the proof of Theorem 4.4: By broomifying each of $V_1, \ldots, V_k$, we get the graph on the right, in which every subtree $V_1, \ldots, V_k$ is a broom of diameter $d_2 - \delta - 1$ or a path of diameter no greater than $d_2 - \delta - 1$.

2. Otherwise, every vertex set $V_1, \ldots, V_k$ is a path of diameter less than $d_2 - \delta - 1$ (Figure 4.5).

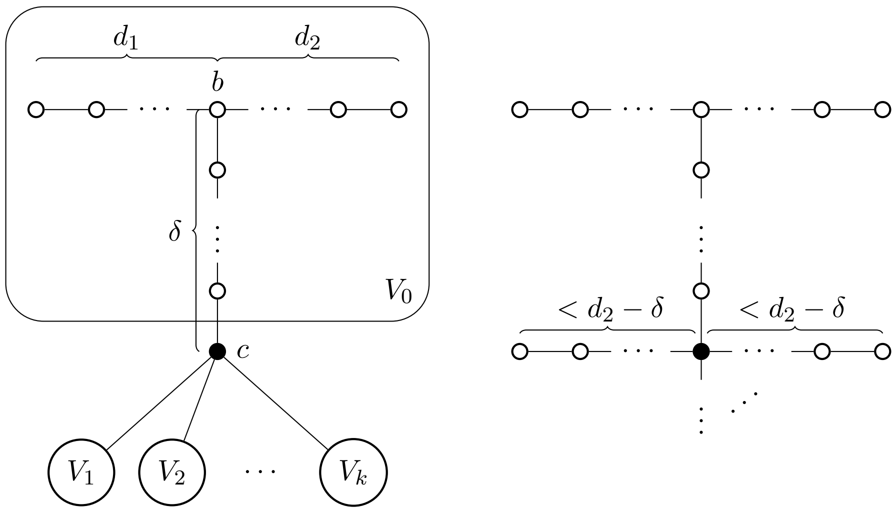

Figure 4.5 Case 2 in the proof of Theorem 4.4: By broomifying each of $V_1, \ldots, V_k$, we get the graph on the right, in which every subtree $V_1, \ldots, V_k$ is a path of diameter (strictly) less than $d_2 - \delta - 1$.

In this case, we can repeatedly pluck the leaf from the shortest path and attach it to the longest path, until it has length $d_2-\delta-1$. Naturally, this increases $H(\pi, c)$ since the increase in hitting time from adding the leaf to the long path outweighs the decrease in hitting time from removing the leaf from the short path. (This is a result of Lemma 3.2.)

Note that this is in fact possible, i.e., that there are enough vertices to construct such a path of length $d_2 - \delta - 1$. This is because V_0 contains at most half of the vertices in T (since c is the barycenter) and $\delta > 0$.

By Lemma 3.2, this will increase the value of $H(\pi, c)$ for T, and so c is now on a geodesic.

Step 2. Now that c is on a geodesic, broomify V_0, the subtree rooted at c of diameter $d_1 + \delta - 1$, by repeatedly plucking leaves (except the one on the geodesic) and attaching them as straws. Since c is on a geodesic, we define this process as follows (so as to ensure the diameter of T remains unchanged). Suppose V_0 contains m vertices. Then, replace it with the broom $B_{m,d_1+\delta-1}$. Now, do the same for the remaining vertex sets, but keep their diameters no greater than $d_2 - \delta - 1$ instead. By Lemma 4.1, this process increases $H(\pi, c)$.

At this point we have a broom-star, meaning a tree consisting of a central vertex c and attached vertex sets $V_0, \ldots, V_k$, each of which forms a broom by itself. Specifically, we desire the double broom, a subfamily of broom-stars in which there are no more than two vertex sets (of size > 1) adjacent to c.

Step 3. Fix V_0, which has diameter $d_1+\delta-1$ and the vertex set of diameter $d_2-\delta-1$ that contains the most vertices (without loss of generality, suppose this set is V_1). Proceed by repeatedly plucking leaves from the remaining vertex sets $V_2, \ldots, V_k$ and reattaching them as straws in V_1. Using the formula in Lemma 4.3, the increase in access time in V_1 outweighs the decrease in access time in any other vertex set, and so the net effect on $H(\pi, c)$ is positive. We continue this process until we arrive at a double broom. ∎

We've now shown that $H(\pi, c)$ is maximized by some sort of double broom. Our next step is to show that placing half of the straws ($\pm\frac{1}{2}$ if we have an odd number of straws) on each side of the double broom yields the maximum value (Theorem 4.7). We maximize this access time in two cases: where n is odd and where n is even.

Lemma 4.5. *Let $T \in \mathcal{T}_{n,d}$ be a double broom, where $n = 2k+1$ is odd. The value $H(\pi, c)$ is maximized by the symmetric double broom.*

Proof. Let T be a double broom on $2k+1$ vertices with diameter d. Consider the components T_1 and T_2 of $T - c$. Since c is a barycenter, T_1 and T_2 contain k vertices each. Let $H_i(\pi, c)$ denote the access time $H(\pi, c)$ for the subgraph $T_i + c$ (that is, T_i with the edge to c added back). We know that each $T_i + c$ has $k+1$ vertices and thus k edges, so we have

$$H_i(\pi, c) = \frac{1}{2k} \sum_{j \in T_i} \deg(j) H(j, c).$$

So, for the entire graph T, we have

$$\begin{aligned} H(\pi, c) &= \frac{1}{4k}\left(\sum_{j\in T_1} \deg(j)H(j,c) + \sum_{j\in T_2} \deg(j)H(j,c)\right) \\ &= \frac{1}{4k}\left(2k \cdot H_1(\pi,c) + 2k \cdot H_2(\pi,c)\right) \\ &= \frac{1}{2}\left(H_1(\pi,c) + H_2(\pi,c)\right). \end{aligned}$$

Suppose $d = d_1 + d_2$, where d_i is the diameter of $T_i + c$. This means that $T_i + c$ is the broom B_{k+1,d_i}, where c is the $(k+1)$-st vertex. By Lemma 4.3, we have that

$$H_i(\pi, c) = \frac{1}{6(k-1)}\left((12d_i - 12)k^2 + (-12d_i^2 + 15)k + 4d_i^3 - 4d_i - 3\right).$$

We can write $H(\pi, c)$ in terms of k, d_1, and $d_2 = d - d_1$. Since n and d are fixed, we only care about the terms that depend on d_1 (and can ignore the leading constant), which gives us

$$4d_1^3 - 12nd_1^2 + (12n^2 - 4)d_1 + 4(d-d_1)^3 - 12n(d-d_1)^2 + (12n^2-4)(d-d_1).$$

Taking the derivative with respect to d_1, we get

$$12(2n-d)(2d_1-d),$$

which equals zero at $d_1 = \frac{d}{2}$. Therefore, this is the critical point that maximizes $H(\pi, c)$ on the interval $[0, d]$. Thus, $H(\pi, c)$ is maximized by the symmetric double broom, meaning the double broom with half of the straw vertices on the left and half on the right. ∎

With some slight modifications, the proof also works if T has an even number of vertices.

Lemma 4.6. *Let $T \in \mathcal{T}_{n,d}$ be a double broom, where $n = 2k$ is even. The value $H(\pi, c)$ is maximized by the off-by-one symmetric double broom.*

Proof. Let T be a double broom on $2k$ vertices with diameter d. Consider the components T_1 and T_2 of $T - c$. This time, since c is a barycenter one of the components contains k vertices, and the other contains $k-1$. Without loss of generality, assume T_1 contains k vertices. Now, this means that

$$H_1(\pi, c) = \frac{1}{2k} \sum_{j \in T_1} \deg(j) H(j, c),$$

and

$$H_2(\pi, c) = \frac{1}{2(k-1)} \sum_{j \in T_2} \deg(j) H(j, c).$$

Now, for the entire graph T, we get

$$\begin{aligned} H(\pi, c) &= \frac{1}{2(2k-1)} \left(2k \cdot H_1(\pi, c) + 2(k-1) \cdot H_2(\pi, c)\right) \\ &= \frac{1}{2(n-1)} \left(n \cdot H_1(\pi, c) + (n-2) \cdot H_2(\pi, c)\right) \end{aligned}$$

Now, letting d_i be the diameter of $T_i + c$, we know that $T_1 + c$ is the broom $B_{\frac{n}{2}+1,d_1}$ and $T_2 + c$ is $B_{\frac{n}{2},d_2}$. By Lemma 4.3, we then get

$$\begin{aligned} H(\pi, c) = \frac{1}{2(n-1)} \Bigg(\frac{4d^3}{3} - 2d^2(n + 2d_1) + d \left(n^2 + 4nd_1 + 4d_1^2 - \frac{4}{3} \right) \\ -2n^2 - 4nd_1^2 + 4nd_1 + n - 4d_1^2 + 4d_1 - 1 \Bigg). \end{aligned}$$

Taking the derivative with respect to d_1, we arrive at the much nicer expression

$$-4(d - n - 1)(d - 2d_1 + 1),$$

which equals zero at $d_1 = \frac{d+1}{2}$. Therefore, this is the critical point that maximizes $H(\pi, c)$ on the interval $d_1 \in [0, d]$. ∎

We summarize these results in the following theorem.

Theorem 4.7. *The quantity*

$$\max_{T \in \mathcal{T}_{n,d}} H(\pi, c)$$

is achieved by the double broom graph $B_{\lceil (n-d+1)/2 \rceil, \lfloor (n-d+1)/2 \rfloor, d}$*, that is, the double broom on* n *vertices with diameter* d *such that half of the leaves* ($\pm\frac{1}{2}$) *are on each end of the handle.*

Proof. By the Double Broom Theorem (4.4), we know that $H(\pi, c)$ is maximized by a double broom. Then, it follows from Lemmas 4.6 and 4.5 that the double broom with half $\pm\frac{1}{2}$ of the leaves as left straws and half $\mp\frac{1}{2}$ of the leaves as right straws achieves the maximum value. ■

We now calculate the access time to the barycenter for the symmetric (and nigh symmetric) double broom. The arithmetic is a little messy, but the logic is not too difficult.

Lemma 4.8. *We have*

$$\max_{T\in\mathcal{T}_{n,d}} H(\pi,c)$$
$$= \begin{cases} \frac{1}{6(n-1)}\big(d^3-3d^2(n+1) \\ \qquad +d\left(3n^2+6n-1\right)-6n^2+3n+3\big) & n \text{ is odd and } d \text{ is even} \\ \frac{1}{6(n-1)}\big(d^3-3d^2(n+1) \\ \qquad +d\left(3n^2+6n-1\right)-6n^2+6n\big) & n \text{ is even and } d \text{ is odd} \\ \frac{1}{6(n-1)}\big(d^3-3d^2(n+1) \\ \qquad +d\left(3n^2+6n+2\right)-6n^2+3n-3\big) & n \text{ is even and } d \text{ is even} \\ \frac{1}{6(n-1)}\left((d-2)\left(d^2-d(3n+1)+3n^2\right)\right) & n \text{ is odd and } d \text{ is odd.} \end{cases}$$

Proof. Noting that a double broom consists of two brooms whose handle ends are the barycenter $v_{\lfloor d/2\rfloor}$ of the double broom, we return to the formula given in Lemma 4.3. Let

$$J(n,d)=\frac{1}{6(n-1)}\left((12d-12)n^2+(-12d^2+15)n+4d^3-4d-3\right)$$

denote the access time $H(\pi,v_d)$ to the end of the handle in the broom $B_{n,d}$. We can sum the access times of the smaller brooms to find the access time of the larger double broom. We just need to be careful when scaling by the number of edges in each graph. Consider four cases.

Case 1. (n is odd and d is even.) For the symmetric double broom of order n and diameter d, we have

$$\begin{aligned}H(\pi,c) &= H(\pi,v_{d/2})\\ &= \frac{1}{2(|E|)}\left(2\frac{|E|}{2}\cdot J\left(\frac{n-1}{2},\frac{d}{2}\right)+2\frac{|E|}{2}\cdot J\left(\frac{n-1}{2},\frac{d}{2}\right)\right)\\ &= J\left(\frac{n-1}{2},\frac{d}{2}\right)\\ &= \frac{1}{6(n-1)}\left(d^3-3d^2(n+1)+d\left(3n^2+6n-1\right)-6n^2+3n+3\right).\end{aligned}$$

Case 2. (n is even and d is odd.) For the symmetric double broom of order n and diameter d, we have

$$\begin{aligned}H(\pi,c) &= H(\pi,v_{(d-1)/2})\\ &= \frac{1}{2(|E|)}\left(2\frac{|E|-1}{2}\cdot J\left(\frac{n}{2},\frac{d-1}{2}\right)\right.\\ &\qquad\left.+2\frac{|E|+1}{2}\cdot J\left(\frac{n}{2}+1,\frac{d+1}{2}\right)\right)\\ &= \frac{1}{2(n-1)}\left((n-2)\cdot J\left(\frac{n}{2},\frac{d-1}{2}\right)\right.\\ &\qquad\left.+n\cdot J\left(\frac{n}{2}+1,\frac{d+1}{2}\right)\right)\\ &= \frac{1}{6(n-1)}\left(d^3-3d^2(n+1)+d\left(3n^2+6n-1\right)-6n^2+6n\right).\end{aligned}$$

Case 3. (n is even and d is even.) For the off-by-one symmetric double broom of order n and diameter d, we have

$$\begin{aligned}H(\pi,c) &= H(\pi,v_{d/2})\\ &= \frac{1}{2(|E|)}\left(2\frac{|E|+1}{2}\cdot J\left(\frac{n}{2}+1,\frac{d}{2}\right)+2\frac{|E|-1}{2}\cdot J\left(\frac{n}{2},\frac{d}{2}\right)\right)\\ &= \frac{1}{2(n-1)}\left(n\cdot J\left(\frac{n}{2}+1,\frac{d}{2}\right)+(n-2)\cdot J\left(\frac{n}{2},\frac{d}{2}\right)\right)\\ &= \frac{1}{6(n-1)}\left(d^3-3d^2(n+1)+d\left(3n^2+6n+2\right)-6n^2+3n-3\right).\end{aligned}$$

Case 4. (n is odd and d is odd.) For the off-by-one symmetric double broom of order n and diameter d, we have

$$\begin{aligned}
H(\pi, c) &= H(\pi, v_{d/2}) \\
&= \frac{1}{2(|E|)}\left(2\frac{|E|}{2}\cdot J\left(\frac{n+1}{2}+1, \frac{d-1}{2}\right)\right. \\
&\qquad\qquad \left. +2\frac{|E|}{2}\cdot J\left(\frac{n+1}{2}, \frac{d+1}{2}\right)\right) \\
&= \begin{array}{l} \frac{1}{2}\left(J\left(\frac{n+1}{2}+1, \frac{d-1}{2}\right)\right. \\ \qquad \left. +J\left(\frac{n+1}{2}, \frac{d+1}{2}\right)\right) \end{array} \\
&= \frac{1}{6(n-1)}\left((d-2)\left(d^2-d(3n+1)+3n^2\right)\right).
\end{aligned}$$

These four cases give our piecewise formula for the maximum access time to the barycenter. ■

Chapter 5

Maximizing the pessimal hitting time

We now turn our attention to finding

$$\max_{T\in\mathcal{T}_{n,d}} \max_{j\in V} H(j',j) = \max_{T\in\mathcal{T}_{n,d}} \left(\max_{j\in V} \max_{i\in V} H(i,j) \right),$$

which is the largest possible hitting time between two vertices among all tress of order n and diameter d. We claim that the graph that achieves this is the broom $B_{n,d}$ (Defintion 2.4).

We begin with a proof about hitting times on the general caterpillar graph.

Lemma 5.1. *Let T be a caterpillar with vertices v_0 and v_d at each end of the stalk. Let T^* be the tree resulting from taking a leaf $u \neq v_d$ such that $u \not\sim v_1$ and moving u one vertex closer to v_0 on the stalk. Then, $H^*(v_0,v_d) = H(v_0,v_d) + 2$.*

Proof. Let T be a caterpillar of diameter d. Label the vertices on the stalk as $v_0, v_1, v_2, \ldots, v_d$, where $v_i \sim v_{i+1}$ for $0 \leq i < d$. Let u be a leaf (other than v_0 or v_d) adjacent to v_j on the stalk. Assume $j \geq 2$. Consider the new tree $T^* = T - (v_j,u) + (v_{j-1},u)$. Noting that $\ell(v_0,v_k;v_d) = d-k$ for any

$0 \leq k \leq d$, we have

$$\begin{aligned}
H^*(v_0, v_d) &= \sum_{k \in V} \ell^*(v_0, k; v_d) \deg^*(k) \\
&= H(v_0, v_d) + \ell^*(v_0, v_j; v_d) \deg^*(v_j) - \ell(v_0, v_j; v_d) \deg(v_j) \\
&\quad + \ell^*(v_0, v_{j-1}; v_d) \deg^*(v_{j-1}) - \ell(v_0, v_{j-1}; v_d) \deg(v_{j-1}) \\
&\quad + \ell^*(v_0, u; v_d) - \ell(v_0, u; v_d) \\
&= H(v_0, v_d) + (d-j)(\deg(v_j) - 1) - (d-j) \deg(v_j) \\
&\quad + (d-(j-1))(\deg(v_{j-1}) + 1) - (d-(j-1)) \deg(v_{j-1}) \\
&\quad + (d-(j-1)) - (d-j) \\
&= H(v_0, v_d) - (d-j) + (d-j+1) + (d-j+1) - (d-j) \\
&= H(v_0, v_d) + 2,
\end{aligned}$$

as desired. ∎

We now prove that the broom is the maximizing structure.

Theorem 5.2. *The quantity*

$$\max_{T \in \mathcal{T}_{n,d}} \max_{j \in V} H(j', j).$$

is achieved by the broom $B_{n,d}$. The maximizing vertex j is the leaf at the end of the handle, and j' is any straw vertex.

Proof. Let T be a tree on n vertices with diameter d, and let z be a vertex in T. We are looking for the z that maximizes $H(z', z)$ among every vertex on T, so we can safely assume that z is a leaf (since otherwise we could move it farther from z' to increase $H(z', z)$). By the same reasoning, we know that z' must be a leaf as well.

Step 1. If z and z' are already on a geodesic (i.e., if $d(z', z) = d$), then we can skip to Step 2. Otherwise, our goal is to pluck and reattach leaves in order to "push" z' and z away from each other until $d(z', z) = d$. To do this, we first fix a path of length d in T and then define a process $*$ as follows.

Suppose $d(z', z) < d$, i.e., that z', z are not on a geodesic. Choose any leaf y from T other than z', z, or the two leaves on the fixed diameter. We know that such a leaf exists since otherwise the hitting time between the leaves on the diameter would necessarily be larger than $H(z', z)$. Let $a = \ell(z', y; z)$ denote the shared distance between the (z', z)- and (y, z)-paths. Now, to produce T^* from T, pluck the vertex y, attach it to z, and relabel it $y := z^*$. Explicitly, $T^* = T - (y_0, y) + (z, y)$, where y_0 is the neighbor of y. (See Figure 5.1.)

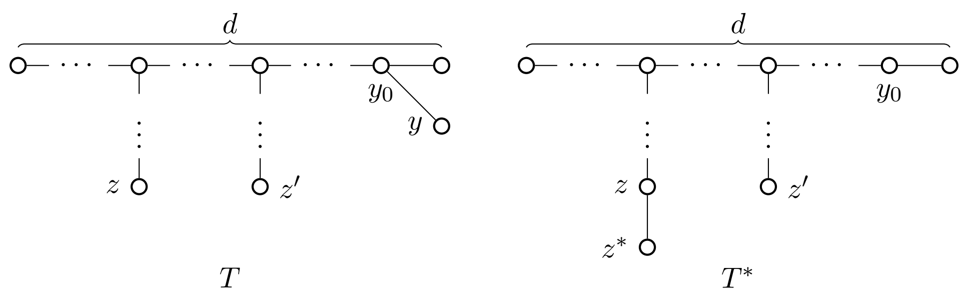

Figure 5.1 Moving z and z' apart

We claim that the hitting time $H^*(z', z^*)$ in T^* is greater than $H(z', z)$ in T. To show this, observe that

$$H^*(z', z^*) = H^*(z', z) + H^*(z, z^*),$$

since every walk from z' to z^* in T^* must first pass through z. Noting that $\ell(z', y_0; z) = \ell(z', y; z) = a$, we can use Equation 2.9 to get

$$\begin{aligned} H^*(z', z) &= \sum_{k \in V} \ell^*(z', k; z) \deg^*(k) \\ &= H(z', z) + \ell^*(z', z^*; z) \deg^*(y) - \ell(z', y; z) \deg(y) \\ &\qquad + \ell^*(z', y_0; z) \deg^*(y_0) - \ell(z', y_0; z) \deg(y_0) \\ &= H(z', z) + 0 - a + a \cdot (\deg(y_0) - 1) - a \cdot \deg(y_0) \\ &= H(z', z) - 2a. \end{aligned}$$

And since $z \sim z^*$, we can use Equation 2.7 to get

$$H^*(z, z^*) = \sum_{k \in V_{z:z^*}} \deg^*(k) = 2(n-1) - \deg(y) = 2n - 3.$$

Therefore,

$$\begin{aligned} H^*(z', z^*) &= (H(z', z) - 2a) + (2n - 3) \\ &= H(z', z) + 2n - (2a + 3). \end{aligned}$$

Thus, $H^*(z', z^*) < H(z', z)$ as long as $a < n - \frac{3}{2}$. To show this must be true, observe that $d(z', z) \leq d - 1$. Then, since $z' \neq y$, we know that the shared distance a is at most $d(z', z) - 1 = d - 2$. Because $d \leq n - 1$, we have $a \leq n - 3$.

We repeat this process until either there are no remaining leaves (other than z', z^*, and the fixed diameter leaves), or until $d(z', z^*) = d$. In fact, if we run out of leaves to pluck, the distance between z' and z^* must equal d, since otherwise we know there exist vertices i, j such that $H(i, j) > H(z', z^*)$.

Step 2. Now, we can "caterpillarize" the graph by fixing the (z', z^*)-path and replacing every subtree rooted on (and containing none of the edges of) that path with a star (of the same number of vertices) centered on the path, as demonstrated in Figure 5.2. In other words, we simplify the graph's structure by turning every vertex not on the (z', z^*)-path into a leaf attached to the path. This does not affect the value of $H^*(z', z^*)$, since the shared path distance $\ell(z', k; z^*)$ remains fixed for any k in a subtree. So, the resulting graph is a caterpillar with $d(z', z^*) = d$.

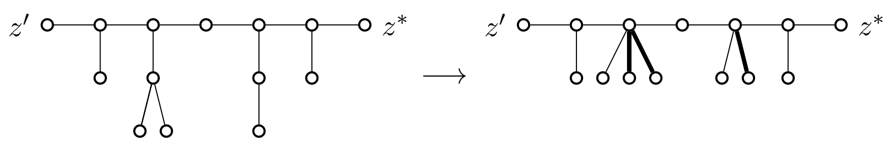

Figure 5.2 The caterpillarization of a tree: push vertices "up" so that they are adjacent to the path

Step 3. We continue to maximize $H(z', z^*)$ by defining a new process $**$ as follows. Suppose T^* is a caterpillar and not a broom. Pluck a leaf $y \neq z', z^*$ adjacent to vertex y_0 on the stalk and reattach it the vertex x_0 on the stalk that is one edge closer to z' (as shown in Figure 5.3).

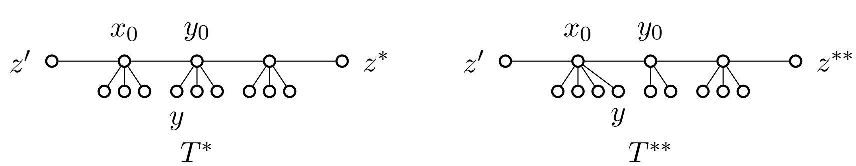

Figure 5.3 Making a caterpillar a broom

Explicitly, $T^{**} = T^* - (y_0, y) + (x_0, y)$. Now, the new hitting time is $H^*(z, z^*) + 2$ by Lemma 5.1. Thus, $H^{**}(z', z^{**}) > H^*(z', z^*)$, so we can continue with this process until we reach a final T^{**}, which is the broom $B_{n,d}$. ■

We now calculate the corresponding hitting time on the broom from a straw to the end of the handle.

Lemma 5.3. *We have*

$$\max_{T \in \mathcal{T}_{n,d}} \max_{j \in V} H(j', j) = 2(d-1)n - d^2 + 2.$$

Proof. The formula for this hitting time is easily computed using the adjacent hitting time formula in Equation 2.7. Using the vertex labeling in Figure 2.4, for any straw vertex w_i (i.e., $1 \leq i \leq n-d$), we have

$$\begin{aligned} H(w_i, v_d) &= H(w_i, v_1) + H(v_1, v_2) + H(v_2, v_3) + \cdots + H(v_{d-1}, v_d) \\ &= 1 + (2(n-d)+1) + (2(n-d)+3) + \cdots + (2(n-d) + 2d - 3) \\ &= 1 + 2(d-1)(n-d) + d^2 - 2d + 1 \\ &= 2(d-1)n - d^2 + 2, \end{aligned}$$

as desired. ■

Chapter 6

Minimizing the pessimal hitting time

Finally, we find the tree structure that yields the opposite result:

$$\min_{T \in \mathcal{T}_{n,d}} \max_{j \in V} H(j', j) = \min_{T \in \mathcal{T}_{n,d}} \left(\max_{j \in V} \max_{i \in V} H(i, j) \right),$$

the smallest possible maximum hitting time between two vertices. Once again, we claim that the trees that achieve this value are caterpillars—specifically seesaws and double seesaws (Definitions 2.2 and 2.3).

We will use the following formula for the hitting time across the stalk of a generic caterpillar.

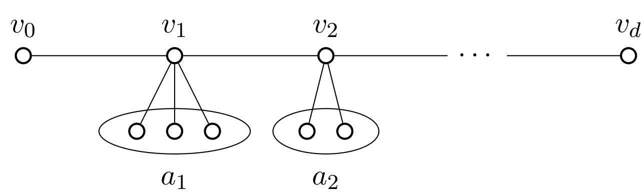

Figure 6.1 A caterpillar graph with stalk $v_0, v_1, \dots, v_d$. The value a_i denotes the number of non-stalk leaves adjacent to vertex v_i.

Lemma 6.1. *Let T be a caterpillar of diameter d with stalk vertices labeled $v_0, v_1, \dots, v_d$. For each $0 \le i \le d$, let a_i denote the number of leaves (not on the stalk) adjacent to v_i (see Figure 6.1). Then, the following formulas yield the* *hitting times from each end of the stalk to the other*:

$$H(v_0, v_d) = d + 2\sum_{i=1}^{d-1}(d-i)(a_i+1),$$

and

$$H(v_d, v_0) = d + 2\sum_{i=1}^{d-1} i(a_i+1).$$

Proof. This result follows easily from Equation 2.9. Let k be a vertex in T such that either $k = v_i$ or k is a non-stalk leaf adjacent to v_i for some $0 \le i \le d$. Then, the shared path distance is $\ell(v_0, k; v_d) = d - i$. Using this, we have

$$\begin{aligned} H(v_0, v_d) &= \sum_{k\in V} \ell(v_0, k; v_d)\deg(k) \\ &= d + \sum_{i=1}^{d-1}(d-i)(a_i+2) + \sum_{i=1}^{d-1} a_i(d-i) \\ &= d + 2\sum_{i=1}^{d-1}(d-i)(a_i+1). \end{aligned}$$

Similarly, the shared path $\ell(v_d, k; v_0)$ is simply equal to i. Thus,

$$H(v_d, v_0) = d + 2\sum_{i=1}^{d-1} i(a_i+1),$$

as desired. ■

Now, we show that the seesaw or double seesaw (depending on the parity of d) minimize the maximum pessimal hitting time.

Theorem 6.2. *The quantity*

$$\min_{T\in\mathcal{T}_{n,d}} \max_{j\in V} H(j', j)$$

is achieved by the seesaw graph $S_{n,d}$ for even d. For odd d, let $r = \lfloor (n-d-1)/2 \rfloor$. The quantity is then achieved by the double seesaw graph $S_{r,r,d}$ if n is even and by the double seesaw graph $S_{r+1,r,d}$ if n is odd. The maximizing vertices j and j' are the leaves at opposite ends of the path of length d.

Proof. Let T be a tree on n vertices with diameter d, and let z be a vertex on T such that $H(z', z) = \max_{i \in V} H(i', i)$. First, we show that we can reduce T to a caterpillar where z' and z are at opposite ends of the stalk (geodesic). Suppose that they are not already on a geodesic. Then, by our assumption, we know that $\max\{H(v_0, v_d), H(v_d, v_0)\} < H(z', z)$, where v_0 and v_d are leaves at opposite ends of a geodesic. As demonstrated before in Figure 5.2, we can then caterpillarize T along the path, that is, pluck every vertex not on the path and attach it as a leaf to the nearest path vertex. By Equation 2.9, this has no affect on $H(v_0, v_d)$ nor on $H(v_d, v_0)$, and so now the larger of them is the new pessimal hitting time.

Now, T is a caterpillar with the (z', z)-path as its stalk, where $H(z', z) = \max_{i \in V} H(i', i)$. We can then label the vertices on the stalk from left to right as $v_0, v_1, v_2, \ldots, v_d$, where $v_i \sim v_{i+1}$ for $0 \le i < d$ and $z' = v_0$, $z = v_d$. Consider three cases based on the parity of d and n.

Case 1. Assume d is even. Now, if there exist two leaves (other than z' and z) adjacent to some stalk vertices v_i, v_j with $i < d/2 < j$, let the process $*$ denote the action of plucking those leaves and reattaching them to vertices v_{i+1} and v_{j-1}, as demonstrated in Figure 6.2.

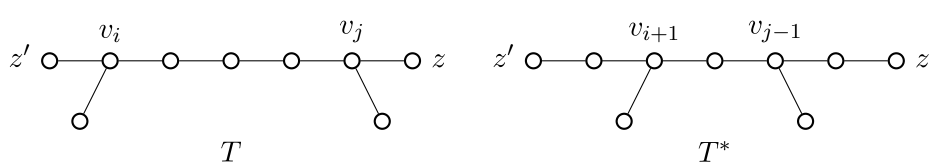

Figure 6.2 Pushing pairs of leaves towards the center

By Lemma 5.1, the new hitting time $H^*(z', z)$ remains unchanged (since the two leaf moves cancel each other out). We repeat this process—pushing pairs of leaves from the left and right towards the center vertex $v_{d/2}$—until either there are no leaves (other than z') on the left side of $v_{d/2}$ or no leaves (other than z) on the right side of $v_{d/2}$. In fact, any remaining leaves must be on the left side of $v_{d/2}$ (i.e., adjacent to v_i for some $i < d/2$), since otherwise $H^*(z, z')$ would be greater than $H^*(z', z)$.

If there are no leaves on the left side, then we are done. Otherwise, we can pluck all the remaining leaves from the left side and reattach them to $v_{d/2}$ in one fell swoop. By the formula in Lemma 6.1, this decreases $H(z', z)$ while increasing $H(z, z')$. At this point, both hitting times will be equal, and any further moving of vertices would increase either $H(z', z)$ or $H(z, z')$.

Since our goal is to decrease the *maximum* hitting time, we are now done.

Case 2. Assume d is odd and n is even (so the number of edges in T is odd). The process is similar to that in Case 1, but now we are pushing leaves to the central *pair* of vertices, $v_{(d-1)/2}$ and $v_{(d+1)/2}$. If there exist two leaves (other than z' and z) adjacent to some stalk vertices v_i, v_j with $i < (d-1)/2 < (d+1)/2 < j$, let the process $*$ denote the action of plucking those leaves and reattaching them to vertices v_{i+1} and v_{j-1}. Again, by Lemma 5.1, the hitting time remains unchanged. We repeat this process until there are no remaining leaves on the right side of $v_{(d+1)/2}$.

At this point, if there are at least two remaining leaves on the left side, we can proceed by plucking them and reattaching one to each of $v_{(d-1)/2}$ and $v_{(d+1)/2}$, thereby decreasing $H(z', z)$. Note again that Lemma 6.1 ensures that $H(z', z)$ is still at least as large as $H(z, z')$. Since we assumed an odd number of edges in T, there are an even number of edges not on the diameter. Thus, repeating this process will yield a final T with half of the leaves attached to $v_{(d-1)/2}$ and the other half attached to $v_{(d+1)/2}$.

Case 3. Assume d is odd and n is odd (so the number of edges in T is even). We proceed using the same steps as in Case 2, until we reach a T with exactly one remaining leaf on the left side of $v_{(d-1)/2}$. We pluck this leaf and reattach it to $v_{(d-1)/2}$, which decreases the hitting time $H(z', z)$ as much as possible without making it less than $H(z, z')$. ■

This proves that the seesaw and double seesaw achieve minimum values for $\max_{j \in V} H(j', j)$, but they are not the unique structures to do so. In fact, any symmetric caterpillar graph on n vertices of diameter d will yield the same value of $\max_{j \in V} H(j', j)$.

Corollary 6.3. *Let T be a caterpillar graph of order n and diameter d with stalk $v_0, v_1, \ldots, v_d$. Let L denote the number of "left" leaves, that is, leaves adjacent v_i for some $1 \leq i < d/2$ and R the number of "right" leaves, adjacent to v_j for some $d/2 < j \leq d-1$. If $R - 1 \leq L \leq R + 1$, then*

$$\max_{j \in V} H(j', j) = \min_{T \in \mathcal{T}_{n,d}} \max_{j \in V} H(j', j).$$

That is, if half the leaves ($\pm\frac{1}{2}$) are on each side of T, then T minimizes the maximum pessimal hitting time.

Proof. This can be proven essentially by working backwards from the seesaw (or double seesaw) graph. Let T be a seesaw or double seesaw of order n, diameter d. By Theorem 6.2, we know that T minimizes $\max_{j\in V} H(j'j)$ and that the vertices that yield this hitting time are v_0 and v_d. If d is even, we can pluck pairs of leaves from $v_{d/2}$ and push them apart (i.e., move one leaf to the left and the other leaf to the right along the stalk). Likewise, if d is odd, we can pluck pairs of leaves, one each from $v_{(d-1)/2}$ and $v_{(d+1)/2}$, and push the former to the left and the latter to the right. By the formula in Lemma 6.1, this has no effect on the value of $H(v_0, v_d)$ nor $H(v_d, v_0)$. Thus, $\max_{j\in V} H(j', j)$ remains at a minimum. ■

We now compute the formula for the maximum pessimal hitting time in $\mathcal{T}_{n,d}$. We will prove this for the seesaw and double seesaw, but note that the value is the same for any symmetric caterpillar by Corollary 6.3.

Lemma 6.4. *We have*

$$\min_{T\in\mathcal{T}_{n,d}} \max_{j\in V} H(j', j) = \begin{cases} d(n-1) & \begin{cases} d \text{ is even} \\ d \text{ is odd and } n \text{ is even} \end{cases} \\ d(n-1)+1 & d \text{ is odd and } n \text{ is odd.} \end{cases}$$

Proof. First, consider the case where d is even, and so $S_{n,d}$ is the minimizing tree structure. We use the vertex labeling in Figure 2.2, where the $n-d-1$ leaves are adjacent to $v_{d/2}$. Observe that $\ell(v_0, v_i; v_d) = d - i$ for all $0 \le i \le d$, and $\ell(v_0, w_i; v_d) = d/2$ for all $1 \le i \le n-d-1$. Then, we can use Equation 2.9 to get

$$\begin{aligned} H(v_0, v_d) &= \sum_{i=0}^{d} \ell(v_0, v_i; v_d) \deg(v_i) + \sum_{i=1}^{n-d-1} \ell(v_0, w_i; v_d) \deg(w_i) \\ &= d + 2\sum_{i=1}^{d/2-1} (d-i) + \left(d - \frac{d}{2}\right)(n-d+1) \\ &\qquad + 2\sum_{i=d/2+1}^{d-1} (d-i) + (n-d-1)\frac{d}{2} \\ &= d(n-1). \end{aligned}$$

Next, consider the case where d is odd and n is even. We're now examining the double seesaw of diameter d with $(n-d-1)/2$ leaves attached to each of $v_{(d-1)/2}$ and $v_{(d-1)/2+1}$ (as shown in Figure 2.3). Via a slight modification to the above arithmetic, we find that $H(v_0, v_d) = d(n-1)$ in this case too.

Finally, consider the case where d is odd and n is odd. Then, we have $(n-d)/2$ leaves attached to $v_{(d-1)/2}$ and $(n-d-2)/2$ leaves attached to $v_{(d-1)/2+1}$ Here, we simply add 1 to the hitting time to get $H(v_0, v_d) = d(n-1)+1$. ■

Chapter 7

Future work

This paper provides some first steps into the world of extreme hitting times for trees on fixed diameter, though there is much more work that could be conducted on this subject. We have found the minimizing and maximizing structures for the access time to the barycenter and for the maximum (pessimal) hitting time. One next step could be to explore which trees yield minimum and maximum for $\max_{i \in V} H(\pi, i)$. Rather than the barycenter, which is the easiest vertex to walk to, this problem considers the vertex that is the hardest to walk to. Intuitively, one expects the (double) seesaw and (double) broom to minimize and maximize (based on the results in this paper). However, this is yet to be proven.

Future work could also explore the same problem but for other mixing measures. The *mixing time* $H(i, \pi)$ is the expected number of steps before a random walk "reaches" the stationary distribution. Here are a few examples of mixing time-related measures.

- The *pessimal mixing time* is given by $T_{\text{mix}} = \max_{i \in V} H(i, \pi)$.
- The *best mixing time* is given by $T_{\text{bestmix}} = \min_{i \in V} H(i, \pi)$.
- The *reset time* is given by $T_{\text{reset}} = \sum_{i \in V} \pi_i H(i, \pi)$.

Previous work has found their minimizing and maximizing structures for trees of fixed order ([2]), but finding the extreme structures for trees of fixed order *and* diameter is an open problem.

Bibliography

[1] Andrew Beveridge. Centers for random walks on trees. *SIAM Journal on Discrete Mathematics*, 23(1):300–318, 2009.

[2] Andrew Beveridge and Meng Wang. Exact mixing times for random walks on trees. *Graphs & Combinatorics*, 29(4):757–772, 2013.

[3] Andrew Beveridge and Jeanmarie Youngblood. The best mixing time for random walks on trees. *Graphs & Combinatorics*, 32(6):2211 – 2239, 2016.

[4] Graham Brightwell and Peter Winkler. Maximum hitting time for random walks on graphs. *Random Structures & Algorithms*, 1(3):263–276, 1990.

[5] Lorenzo Ciardo, Geir Dahl, and Steve Kirkland. On Kemeny's constant for trees with fixed order and diameter. *Linear and Multilinear Algebra*, 70(12):2331–2353, 2022.