INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road. Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Order Number 9510719

Integer sequences associated with trees

Skurnick, Ronald Stuart, Ph.D.

City University of New York, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Integer Sequences Associated With Trees

by

Ronald Skurnick

A dissertation submitted to the Graduate Faculty in Mathematics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1994

This manuscript has been read and accepted for the Graduate Faculty in Mathematics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/16/94 — Joseph Malkevitch
Date — Chair of Examining Committee

9/16/94 — J Dodziuk
Date — Executive Officer

J Dodziuk

Stefan A. Burr

Joseph Malkevitch
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

INTEGER SEQUENCES ASSOCIATED WITH TREES

by

Ronald Skurnick

Advisor: Professor Joseph Malkevitch

The weight $a(v)$ of a vertex v of a tree T is the size (i.e., number of edges) of a maximal subtree T_v of greatest size which starts at v in T. The centroid of T is the set of all vertices of T of minimum weight.

A non-increasing finite sequence $A = \{a_1, a_2, ..., a_n\}$ $(n \geq 2)$ of positive integers is said to be the centroid sequence of a tree if there exists a tree T of order n whose vertices can be labeled $v_1, v_2, ..., v_n$ such that $a(v_i) = a_i$ for each $i = 1, 2, ..., n$. The centroid sequence of a tree of order $n \geq 2$ is characterized and properties of the centroid sequence of a tree are presented.

A caterpillar C is a tree on $n \geq 3$ vertices such that if all the 1-valent vertices of C are removed, the resulting graph is a (possibly degenerate) path. The centroid sequence of a caterpillar is characterized.

The eccentricity $e(v)$ of a vertex v of a connected graph G is the maximum distance in G from v to any vertex of G. The center of G is the set of all vertices of G of minimum eccentricity. Several characterizations of the center of a tree are given.

A non-decreasing finite sequence $E = \{e_1, e_2, \ldots, e_n\}$ $(n \geq 2)$ of positive integers is said to be the eccentricity sequence of a tree if there exists a tree T of order n whose vertices can be labeled $v_1, v_2, \ldots, v_n$ such that $e(v_i) = e_i$ for each $i = 1, 2, \ldots, n$. A new proof of and results related to Lesniak's characterization of the eccentricity sequence of a tree on $n \geq 2$ vertices are presented. Additional results and algorithms concerning centers, centroids, and their associated sequences are also provided. In particular, results are obtained for when a vertex v in a graph G is <u>in</u> the centroid or center of some spanning tree of G, or <u>is</u> the centroid or center of some spanning tree of G.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my thesis advisor, Professor Joseph Malkevitch, for spending countless hours of his spare time sharing with me his great wealth of knowledge in the field of graph theory and preparing me to write this doctoral dissertation. Dr. Malkevitch introduced me to the subject of graph theory via a course that he taught (and I took) at the CUNY Graduate Center, and has since brought to my awareness the existence of numerous open questions that abound in the field of graph theory, some of which have been answered in this thesis.

I would also like to thank Professor Stefan Burr, Professor Jozef Dodziuk, and Professor Harvey Cohn for the roles they have played as members of my supervisory committee. They were, in large part, responsible for ensuring that this doctoral dissertation is "up to snuff". Thanks to Professor Eli Goodman, as well, for the time he invested in teaching me graph theory and combinatorics and for all of the good advice he gave me along the way.

I would like to give thanks to my mother Joan Skurnick, my father Morris Skurnick, my sister Dr. Sara Skurnick, and her husband Dr. Louis Sterling for being the wonderful people that they are and for their love, concern, and constant support. Thank you, Sara, for teaching me the most efficient way to utilize the computer graphics software that was needed to prepare this

manuscript and for all the work you did on the computer in helping me to put this thesis together. Thank you, Louis, for allowing me to use your computer, disks, and laser printer.

Special thanks go out to my friend Yeh Chiung Ying for the enormous amount of time and energy she spent in helping me type up portions of my thesis. Thanks, also, to my friend Dalyoung Jeong for giving me good advice regarding the organization of this dissertation. I would also like to thank Dr. Robert Silverstein for sharing his vast computer know-how with me. I am grateful to my friend Martin Agin for all the help and encouragement that he has provided from the very beginning of this lengthy project. Heart-felt thanks to my high school mathematics teacher and friend Mr. Mickey Blechman for teaching me so much and for motivating me to study higher mathematics.

Last, but not least, I would like to thank my wonderful girl-friend Mia Lee for her tremendous patience, encouragement, support, and assistance in preparing me for my oral exam and in typing portions of this thesis. If this thesis is half as good as you are, Mia, it'll be a masterpiece!

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

1.1 Preliminaries

The focus of the present study is on a special class of graphs called **trees.** Trees are important because of their many applications to a variety of fields, including facility location within a network and cost minimization in the construction of communication and railway networks, as well as the special role they often play within graph theory itself [3]. The concepts of **centroid** and **center,** ideas designed to capture the notion of "central location," as they apply to trees, are discussed in some detail in Chapters 2, 3, and 4. A good portion of the present study deals with the **centroid sequences** of trees which, heretofore, have been unexplored in the literature.

Chapter 2 presents the concept of the centroid of a tree in terms of its branch-weight definition. Several theorems are proved about the centroid of a tree, and an algorithm for locating the centroid of a tree of order $n \geq 3$ is presented. In addition, the centroid sequence of a tree is defined. Necessary and sufficient conditions are established regarding whether a given finite sequence of positive integers is realizable as the centroid sequence of a **caterpillar,** which is a special kind of tree, and, more generally, of a non-trivial tree. A theorem characterizing the **frequency distribution** of the centroid sequence of a tree is discussed. The relationship between the branch-weight definition of the centroid of a tree and P. Slater's definition of the centroid of a connected graph [12], as applied to a tree, is then explored, and several observations are made.

Chapter 3 treats the concept of the center of a tree. Some properties of the center of a tree are discussed, and the **eccentricity sequence** of a tree is examined. Necessary and sufficient conditions for a finite sequence of positive integers to be the eccentricity sequence of a caterpillar and of a non-trivial tree are discussed. A significant difference between the centroid sequence and eccentricity sequence of a tree is considered.

Chapter 4 presents necessary and sufficient conditions for each vertex of a graph G to be (in) the centroid (respectively, center) of some spanning tree T of G. There are also several examples of trees that are noteworthy because they either illustrate a specific concept discussed in Chapter 2 or Chapter 3 or they show the relationship (or lack thereof) between the concepts mentioned in these two chapters. In addition, several theorems are proved about the centroid, centroid sequence, center, and eccentricity sequence of a tree. Chapter 4 concludes with a list of open problems related to the subject matter of the present doctoral dissertation.

As for the remainder of the present chapter, we will define the terms and notation that we will need for this thesis, and state some needed results of graph theory. We will follow the terminology of Gary Chartrand's and Linda Lesniak's *Graphs and*

Digraphs [4], and Fred Buckley's and Frank Harary's *Distance In Graphs* [3].

1.2 Graphs, Plane Graphs, Graph Isomorphism, and Valence of a Vertex

A **graph** G is a finite non-empty set of objects, called **vertices** (singular: **vertex**) or **nodes**, together with a (possibly empty) set of unordered pairs of distinct vertices of G, called **edges**. We shall denote the vertex set of G by V(G) and the edge set of G by E(G). A graph H is a **subgraph** of a graph G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$; in this case, we also say that G is a **supergraph** of H. We denote the fact that H is a subgraph of G by writing $H \subset G$. If W is a non-empty subset of the vertex set V(G) of a graph G, then the subgraph <W> of G **induced** by W is the graph having vertex set W and whose edge set consists of those edges of G incident with two elements of W. A subgraph H of G is called **vertex-induced**, or **induced**, if $H \cong \langle W \rangle$ for some subset W of V(G).

If e = (u, v) is an edge of a graph G, then e is said to **join** vertices u and v, and u and v are **adjacent** vertices. In addition, edge e is **incident** with vertex u and with vertex v.

The cardinality of the vertex set of a graph G is called the **order** of G and is denoted by |V(G)|, while the cardinality of the edge set of G is called the **size** of G and is denoted by |E(G)|. A **(p,q) graph** G is a graph whose order |V(G)| = p and whose size |E(G)| = q.

A graph G is said to be **embeddable in the plane, or planar,** if it can be drawn in the plane so that its edges intersect only at their endpoints. A drawing in the plane of a planar graph G whose edges meet only at their endpoints is called a **planar embedding** of G. A planar embedding of a planar graph is referred to as a **plane graph.**

Example 1.2.1: The graph G of Figure 1.2.1(a) is a planar graph because there exists an embedding G~ of G in the plane, as shown in Figure 1.2.1(b). G~ is a plane graph.

Figure 1.2.1:

Graph G
(a)

Graph G~
(b)

A graph G_1 is **isomorphic** to a graph G_2 if there exists a one-to-one mapping f, called an **isomorphism**, from $V(G_1)$ onto $V(G_2)$ such that f preserves adjacency of vertices; that is, $(u, v) \in E(G_1)$

if and only if $(f(u), f(v)) \in E(G_2)$. If G_1 is isomorphic to G_2, then we say that G_1 and G_2 are isomorphic and write $G_1 \cong G_2$. $\cong$ is an equivalence relation on graphs.

If v is a vertex in a graph G (i.e., $v \in V(G)$), then the **valence** of v in G, denoted **val(v)** (or **val$_G$(v)**), is the number of edges of G incident with v. A vertex of valence 0 in G is called an **isolated** vertex and a vertex of valence 1 in G is called a **1-valent vertex** or an **end-vertex** or a **leaf**. Some authors use the words "valence" and "degree" interchangeably. A fundamental theorem of graph theory, sometimes referred to as "The First Theorem of Graph Theory", follows.

Theorem: Let G be a (p,q) graph, where $V(G) = \{v_1, v_2, ..., v_p\}$. Then,

$$\sum_{i=1}^{p} \text{val}(v_i) = 2q.$$

For a proof, see [2].

The **valence sequence** of a graph G on n vertices is the sequence $S = \{s_1, s_2, ..., s_n\}$ whose elements are the valences of the vertices of G, arranged in non-increasing order.

Example 1.2.2: The valence sequence of the graph of Figure 1.2.1(b) is $S = \{3, 2, 1, 1, 1\}$.

Given a finite sequence S of non-negative integers, when is S realizable as the valence sequence of a graph G? This question is answered by the following theorem of Havel and Hakimi.

Theorem (Havel-Hakimi): A sequence $S = \{s_1, s_2, ..., s_n\}$ $(n \geq 2)$ of non-negative integers with $s_1 \geq s_2 \geq \cdots \geq s_n$ and $s_1 \geq 1$ is graphical if and only if the sequence $S^* = \{s_2 - 1, s_3 - 1, ..., s_{s_1+1} - 1, s_{s_1+2}, ..., s_n\}$ is graphical.

For a proof, see [4, pages 20 and 21].

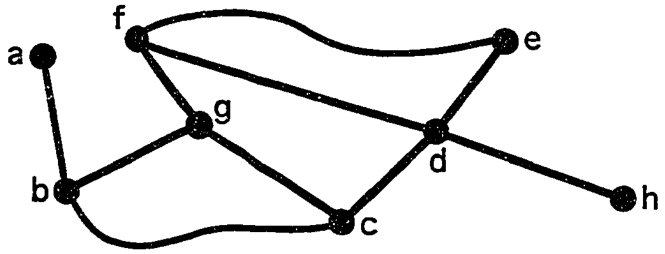
The Havel-Hakimi Theorem provides motivation for much of the work that appears in Chapter 2 concerning the centroid sequence of a tree.

1.3 Paths and Connectedness

Let u and v be (not necessarily distinct) vertices of a graph G. A **u-v walk** of G is a finite, alternating sequence $u = u_0, e_1, u_1, e_2, ..., u_{n-1}, e_n, u_n = v$ of vertices and edges, beginning with vertex u and ending with vertex v, such that $e_i = (u_{i-1}, u_i)$ for $i = 1,2,..., n$. The number n (which is the number of times that edges occur in the sequence) is called the **length** of the walk. Frequently, only the vertices of a walk are indicated (and the edges are omitted) since the edges present are then evident. A trivial walk contains no edges (i.e., $n = 0$). A u-v walk is **open** or **closed** depending on whether $u \neq v$ or $u = v$. A **u-v trail** is a u-v walk in which no edge is repeated, while a **u-v path**, denoted P_{uv}, is a u-v walk in which no vertex is repeated. The trivial u-u path P_{uu} consists of simply the vertex u. Clearly, every path is a trail. Two u-v paths ($u \neq v$) in a graph G are said to be **internally disjoint** if the sets of vertices other than u and v of these two u-v paths are disjoint.

Example 1.3.1: In the graph of Figure 1.3.1, fgcbgcde is an f-e walk, bgcdefdh is a b-h trail, and abcgfde is an a-e path.

Figure 1.3.1:



A non-trivial closed trail of a graph G is known as a **circuit** of G, and a circuit $v_1, v_2, \ldots, v_n, v_1$ ($n \geq 3$) whose n vertices v_i are distinct is called a **cycle.** An **acyclic** graph is one with no cycles. A cycle is called **even** if its length is even; otherwise, it is called **odd.** A cycle of length n is an **n-cycle**; a 3-cycle is usually called a **triangle.** A cycle of a graph G which contains every vertex of G is called a **hamiltonian** cycle of G; a hamiltonian graph is one that contains a hamiltonian cycle. A vertex u is said to be **connected** to a vertex v in a graph G if there exists a u-v path in G. A graph G is **connected** if every pair of its vertices is connected. A graph which is not connected is **disconnected.** A **component** of a graph G is a subgraph of G that is maximal with respect to the property of being connected. A vertex v of a graph G is called a **cut-vertex** of G if the number of components of G-v is $>$ the number of components of G. A connected graph of order $p \geq 3$ with no cut-vertices is called a **cyclic block.** A graph G is **n-connected** ($n \geq 1$) if the removal of fewer than n vertices from G results in neither a disconnected graph nor the **trivial graph** (i.e., the graph consisting of one vertex and no edges). The following theorem of Whitney gives an equivalent formulation of the concept of n-connectedness.

Theorem (Whitney): A non-trivial graph G is n-connected if and only if for each pair u, v of distinct vertices of G there are at least n internally disjoint u-v paths in G.

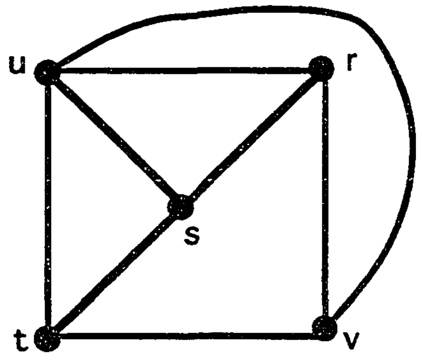
For a proof, see [4, pages 159 and 160].

1.4 Distance and Eccentricity

For a connected graph G, we define the **distance** d(u,v) between two vertices u and v of G as the minimum of the lengths of all u-v paths in G. The vertex set V(G) of G is a metric space under this distance function. The **eccentricity** e(v) of a vertex v of a connected graph G is given by $e(v) = \max_{u \in V(G)} d(u,v)$. The **radius** rad G of G is defined to be $\min_{v \in V(G)} e(v)$, while the **diameter** diam G of G is defined as $\max_{v \in V(G)} e(v)$. A vertex v is a **central** vertex of G if e(v) = rad G and the **center** Z(G) of G consists of its central vertices.

Example 1.4.1: In the graph G of Figure 1.4.1, d(u,v) = 1, e(u) = 1, e(v) = 2, rad G = 1, diam G = 2, and Z(G) = {u}.

Figure 1.4.1:



Graph G

1.5 Trees

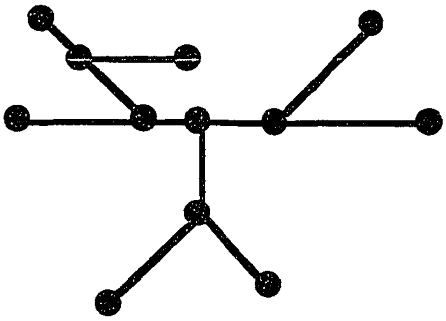
A tree is a connected graph which contains no cycles. The graph T of Figure 1.5.1 is a tree. Observe that $|V(T)| = 12$ and $|E(T)| = 11$, so the order of T is exactly 1 greater than the size of T. The generalization of this observation leads to the following theorem.

Theorem: If T is a tree, then $|V(T)| = |E(T)| + 1$.

For a proof, see [2, page 25].

We generally use the letter T (rather than G, which is used to refer to a more general graph) when referring to a tree.

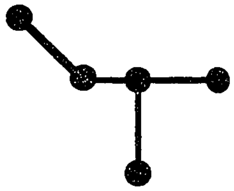
Figure 1.5.1:



Tree T

Every tree T is a planar graph. Every tree on $n \geq 2$ vertices contains at least two 1-valent vertices. A non-trivial **path** P is a tree that contains exactly two 1-valent vertices. If a 1-valent vertex v (as well as the unique edge e of T incident with v) of a tree T on $n \geq 2$ vertices is removed from T, we say that v has been **pruned** from T. We refer to the removal of all the 1-valent vertices of a tree T on $n \geq 3$ vertices as a **pruning** of T. The graph resulting from a pruning of a tree T on $n \geq 3$ vertices is a (possibly trivial) tree T*, and T* is called the **pruned tree** of T. A **star** S is a tree whose pruned tree S* is the trivial tree. The pruned tree T* of the tree T of Figure 1.5.1 is shown in Figure 1.5.2.

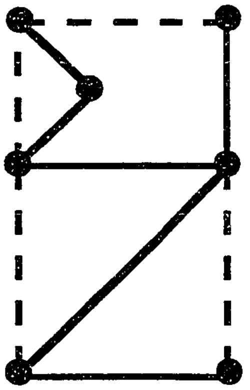
Figure 1.5.2:



Tree T*

A **spanning tree** T of a graph G is a spanning subgraph of G (i.e., T contains all the vertices of G) that is a tree. Note that, by definition, only connected graphs have spanning trees. Figure 1.5.3 shows a connected graph G and one of its spanning trees T (in which the edges of T are indicated with solid line segments, while the edges of G\E(T) are indicated with dotted line segments).

Figure 1.5.3:



Graph G

The following theorem greatly simplifies the study of distance-related questions concerning trees.

Theorem: In a tree, any two vertices are connected by a unique path.

For a proof, see [2, page 25].

So, given any two vertices x and y in a tree T, there is one and only one path P_{xy} joining vertices x and y in T. Therefore, to compute d(x,y) in T, we simply measure the length of path P_{xy}. That is, $d(x,y) = \text{length}(P_{xy})$.

1.6 Centrality in Trees

There are several different notions of centrality within a tree. The **center** of a tree T consists of all vertices of T of minimum eccentricity. A well-known theorem of Jordan characterizes the center of a tree.

Theorem (Jordan): The center of a tree consists of either a single vertex or a pair of adjacent vertices.

For a proof, see [3, page 32].

Remark 1.6.1: When the center of a tree consists of two adjacent vertices u and v, we will say that the center of the tree is the edge (u, v).

Example 1.6.1: The center of the tree T_1 of Figure 1.6.1(a) consists of adjacent vertices x and y, while the center of the tree T_2 of Figure 1.6.1(b) consists of the single vertex z.

Figure 1.6.1:

Tree T_1
(a)

Tree T_2
(b)

The concept of **centroid** was defined originally to apply only to trees. A **branch** at a vertex v of a tree T is a maximal subtree of T containing v as an end-node. Therefore, the number of branches at v is val(v). The **weight** (or **centroid number**) at a vertex v, denoted **a(v)**, of a tree T is the maximum size of any branch at v.

Example 1.6.2: In the tree of Figure 1.6.2, there are 4 branches at vertex v. Also, a(w) = 6 because a branch of T of greatest size at w has size 6.

Definition 1.6.1: A vertex v of a tree T is a **centroid vertex** of T if v has minimum weight, and the **centroid** of T consists of its centroid vertices.

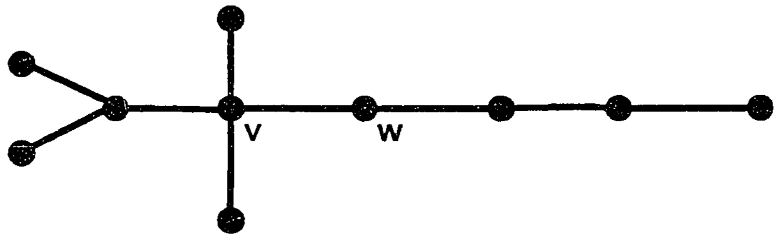
Jordan also proved the following theorem.

Theorem (Jordan): The centroid of a tree consists of either a single vertex or two adjacent vertices.

A proof will be provided in Section 2.1.

Remark 1.6.2: When the centroid of a tree T consists of two adjacent vertices v_1 and v_2, we will say that the centroid of T is the edge (v_1, v_2).

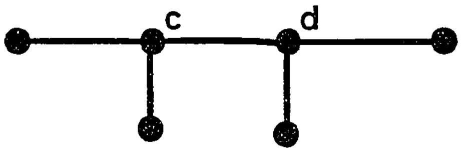
Figure 1.6.2: The centroid of T is vertex v, while the center of T is vertex w.



Tree T

Example 1.6.3: The centroid of the tree T of Figure 1.6.2 consists of the single vertex v, while the centroid of the tree of Figure 1.6.3 consists of the adjacent vertices c and d.

Figure 1.6.3:



Although the centroid and center of a tree may coincide (e.g., in the case of a path P on 5 vertices), they may also differ (as in the case of the tree of Figure 1.6.2). In fact, as we shall see in Section 4.3, given any non-negative integer n, there exists a tree T such that if u is a vertex in the centroid of T and v is a vertex in the center of T, then $d_T(u,v) = n$.

Slater [12] later extended the concept of centroid so that it is now defined for all connected graphs. Much more will be said about this in Section 2.6.

If G is a connected graph, then the **status** s(v) of a vertex v in G is the sum of the distances from v to each other node in G [3]. The **median** M(G) of a graph G is the set of vertices of G of minimum status. The following result of Zelinka [13] shows that the centroid and median of a tree coincide.

Theorem (Zelinka): Vertex v is a centroid vertex of a tree T if and only if v is a median vertex of T.

For a proof, see [13].

Thus, the median of a tree consists of either a single vertex or a pair of adjacent vertices. An algorithm for locating the median (or centroid, by the theorem of Zelinka) of a tree is presented in Section 2.1.

CHAPTER 2 THE CENTROID AND CENTROID SEQUENCE OF A TREE

2.1 The Centroid of a Tree

Recall that a **tree** is a connected graph which contains no cycles. Each tree T has associated with it a **centroid.** As we saw in Chapter 1, the centroid of a tree consists of all of its vertices of minimum weight, and it was proved by Jordan that the centroid of a tree consists of either a single vertex or a pair of adjacent vertices. Jordan obtained this result without any reference to the concept of the weight (or centroid number) of a vertex within a tree.

We now present a proof of Jordan's theorem (see Theorem 2.1.1) in terms of the weights of the vertices of a tree. Our proof differs from that given by Jordan, and requires the following two lemmas concerning the centroid of a tree.

Lemma 2.1.1: Let T be a tree, let v be a vertex of T, and let c be a vertex in the centroid of T. If T_{v^*} is a subtree of T which starts at v such that $\text{size}(T_{v^*}) \geq \text{size}(T_{v\#})$ for all subtrees $T_{v\#}$ of T which start at v, then vertex c lies on T_{v^*}.

Proof: If $v = c$, then the theorem is obviously true. So, assume that v and c are distinct vertices of T. Then, either vertex c lies on T_{v^*}, or it doesn't. If vertex c lies on T_{v^*}, there's nothing to prove. So, suppose that vertex c does not lie on T_{v^*}. Then, c lies on a subtree, say, T_{v+} of T which starts at v such that $T_{v+} \neq T_{v^*}$. Hence, the unique path P_{cv} in T which joins vertices c and v is contained completely in subtree T_{v+}. So, the unique subtree T_c of T which starts at vertex c and contains vertex v is such that $\text{size}(T_c) \geq \text{length}(P_{cv}) + \text{size}(T_{v^*}) \geq 1 + \text{size}(T_{v^*}) > \text{size}(T_{v^*})$. Thus, $a(c) \geq \text{size}(T_c) > \text{size}(T_{v^*}) = a(v)$, i.e., $a(c) > a(v)$, contradicting the hypothesis that vertex c is in the centroid of T. Therefore, vertex c does, in fact, lie on subtree T_{v^*}, and we are done. □

Lemma 2.1.2: Let T be a tree, and let $v_1, v_2 \in V(T)$. Then, v_1 and v_2 are in the centroid of T if and only if they have the same weight in T and are adjacent in T.

Proof: ($\Rightarrow$) Suppose that (distinct) vertices v_1 and v_2 are in the centroid of T. Then, by definition, vertices v_1 and v_2 have the same weight in T. So, all that we have to show is that v_1 and v_2 are adjacent in T. We'll do this by contradiction. Suppose that v_1 and v_2 are not adjacent in T. Let $P_{v_1v_2}$ denote the unique path in T joining vertices v_1 and v_2. Then, there exists a vertex z lying on $P_{v_1v_2}$ with $z \neq v_1$ and $z \neq v_2$ (i.e., z lies strictly between the endpoints v_1 and v_2 of path $P_{v_1v_2}$). By Lemma 2.1.1, if T_z is a subtree of T which emanates from z such that $\text{size}(T_z) \geq \text{size}(T_{\#})$

for all subtrees $T_{\#}$ that emanate from z in T, then vertices v_1 and v_2 must both lie on T_z. But, since vertex z lies between endpoints v_1 and v_2 of path $P_{v_1v_2}$, vertices v_1 and v_2, in fact, lie on two distinct subtrees of T which emanate from z. And, these two subtrees share only vertex z. Hence, assuming the existence of vertex z leads to a contradiction. This implies that no such vertex z exists in T. That is, v_1 and v_2 must be adjacent in T, and we are done. ☐

($\Leftarrow$) Conversely, suppose that vertices v_1 and v_2 have the same weight in T and are adjacent in T. Let T_{v_1} be a subtree of T which starts at v_1 such that $\text{size}(T_{v_1}) \geq \text{size}(T_{\#})$ for all subtrees $T_{\#}$ that start at v_1 in T. Also, let T_{v_2} be a subtree of T which starts at v_2 such that $\text{size}(T_{v_2}) \geq \text{size}(T_{\#\#})$ for all subtrees $T_{\#\#}$ that start at v_2 in T. Then, since $a(v_1) = a(v_2)$, we have $\text{size}(T_{v_1}) = \text{size}(T_{v_2})$. Observe that vertex v_1 lies on T_{v_2}, for if it did not, then the subtree $T_{\sim}$ of T which starts at vertex v_1 and is determined by vertex v_2 would be such that $\text{size}(T_{\sim}) \geq 1 + \text{size}(T_{v_2}) = 1 + a(v_2) = 1 + a(v_1)$. This implies that $\text{size}(T_{\sim}) > a(v_1)$, contradicting the definition of $a(v_1)$. Similarly, vertex v_2 lies on T_{v_1}.

Now, let c be a vertex in the centroid of T. Note that such a vertex c exists in T since the centroid of a tree is non-empty. By Lemma 2.1.1, since $\text{size}(T_{v_1}) \geq \text{size}(T_{\#})$ for all subtrees $T_{\#}$ that start at v_1 in T, vertex c lies on T_{v_1}. Similarly, c also lies on T_{v_2}. So, c lies on both T_{v_1} and T_{v_2}. But, the only vertices of T that lie on both T_{v_1} and T_{v_2} are v_1 and v_2. This implies that either v_1 is in the centroid of T, or v_2 is in the centroid of T, or both v_1 and v_2 are in the centroid of T. Suppose vertex v_1 is in the centroid of T. Then, since $a(v_2) = a(v_1) \leq a(v)$ for all $v \in V(T)$, vertex v_2 is also in the centroid of T. Similarly, if v_2 is in the centroid of T, then v_1 must also be in the centroid of T. Hence, vertices v_1 and v_2 are both in the centroid of T, proving the theorem. □

Theorem 2.1.1 (Jordan): The centroid of a tree consists of either a single vertex or two adjacent vertices.

Proof (not Jordan's): We'll prove this theorem by contradiction. Suppose there exists a tree T whose centroid does not consist of a single vertex and does not consist of a pair of adjacent vertices. As a consequence of the definition of the centroid of a tree, T has a non-empty centroid, so the centroid of T must consist of either a pair of non-adjacent vertices or three or more vertices. But, the centroid of T cannot consist of 2 non-adjacent vertices, by Lemma 2.1.2. So, there is no tree whose centroid consists of 2 non-adjacent vertices. Now, suppose there exists a tree T whose centroid consists of 3 or more vertices. Let c_1, c_2, and c_3 be 3 distinct vertices in the centroid of T (there may or may not be other vertices, as well, in the centroid of T). Then, by Lemma 2.1.2, c_1 must be adjacent to c_2, c_2 must be adjacent to c_3, and c_3 must be adjacent to c_1 in T. This means that vertices c_1 and c_2 must be joined by an edge $e_1 = (c_1, c_2)$, vertices c_2 and c_3 must be joined by an edge $e_2 = (c_2, c_3)$, and vertices c_3 and c_1 must be joined by an edge $e_3 = (c_3, c_1)$, in T. And, since c_1, c_2, and c_3 are distinct vertices of T, e_1, e_2, and e_3 are distinct edges of T. But then, T contains a cycle (i.e., triangle $c_1c_2c_3$), which contradicts the hypothesis that T is a tree. Therefore, the centroid of T cannot consist of 3 or more vertices. Hence, since there is no tree whose centroid consists of a pair of non-adjacent vertices and there is no tree whose centroid consists of 3 or more vertices, we have the result that the centroid of a tree consists of either a single vertex or a pair of adjacent vertices. □

Remark 2.1.1: Each path of odd order is an example of a tree whose centroid consists of a single vertex. Each path of even order is an example of a tree whose centroid consists of a pair of adjacent vertices.

With Jordan's characterization of the centroid of a tree in mind, we now consider the following question: Given a tree T (that is, given a planar embedding of T), how can we locate its centroid? One way to locate the centroid of T is to first compute the weight of each vertex of T. The centroid of T consists of that vertex or that pair of adjacent vertices of T of lowest weight. There is also another more geometric method besides this approach for locating the centroid of a tree of order $n \geq 3$. This new method locates the centroid of a tree T without explicitly requiring the determination of the weight of any vertex of T and is presented below in the form of an algorithm.

Algorithm 2.1.1: (Locating The Centroid of a Tree of Order $n \geq 3$)

Let T be a tree on $n \geq 3$ vertices. The following algorithm provides a method for finding the centroid of T.

(1) Starting from each 1-valent vertex v of T, move to the unique vertex w which is adjacent to v in T, and wiggle edge (v, w).

(2) If every edge of T was wiggled in step (1), then every 1-valent vertex v of T is adjacent to the same vertex w in T, so T is a star. In this case, vertex w is the centroid of T, and we stop here.

(3) If the number of edges of T that are still unwiggled is ≥ 1, then for each vertex u such that u is an endpoint of an edge that has not been wiggled yet, let N_u denote the number of edges (i.e., the size) of the largest completely wiggled subtree T_u of T rooted at u. Locate all vertices y of T such that the number of wiggled edges incident with y, at this time, is (val(y) - 1). [Note: There must be at least 2 such vertices y in T!] Compute N_y for each of these vertices y. Let y_j be any one of these vertices y such that $N_{y_j} \leq N_y$ for all such vertices y.

(4) If, at this stage, every edge of T except one has been wiggled, then, starting from y_j, move to the unique vertex z_j which is adjacent to y_j in T such that edge (y_j, z_j) has not already been wiggled. Do not wiggle (y_j, z_j). In this case, if $N_{y_j} = N_{z_j}$, then edge (y_j, z_j) is the centroid of T. However, if $N_{y_j} < N_{z_j}$, then vertex z_j is

the centroid of T, and we stop here.

(5) If, on the other hand, two or more edges of T have not yet been wiggled, then wiggle edge (y_j, z_j) now.

(6) Go back to step (3).

Proof: (that Algorithm 2.1.1 locates the centroid of a tree of order $n \geq 3$)

If T is a tree of order $n \geq 3$, then the centroid of T does not include any 1-valent vertex of T. Also, T has one or more non-1-valent vertices. Since we know that none of the 1-valent vertices of T is in the centroid of T, we can eliminate the 1-valent vertices of T from consideration as vertices which are possibly in the centroid of T and start with the vertices of T that are adjacent to these 1-valent vertices of T. If there is only one such vertex w which is adjacent to all the 1-valent vertices of T, then T is a star and w is the centroid of T. Otherwise, T is not a star, so T has 2 or more vertices y such that y is adjacent to exactly (val(y) - 1) 1-valent vertices of T. Let y_j be one of these vertices y such that $N_{y_j} \leq N_y$ for each such vertex y. Note that subtree T_{y_j} of T consists of all branches of T emanating from y_j other than a branch B_{y_j} of greatest size which emanates from y_j, since T is not a star. And, since $N_{y_j} \leq N_y$ for each such vertex y, $\text{size}(B_{y_j}) \geq \text{size}(B_y)$ for each such vertex y. Therefore, $a(y_j) \geq a(y)$ for each such vertex y, which implies that y_j is not in the centroid of T unless there is exactly one other vertex y_k amongst these vertices y such that $N_{y_j} = N_{y_k}$ and y_k is adjacent to y_j in T, in which case T is simply a "double-star" (i.e., a tree whose centroid is an edge and whose only non-1-valent vertices are its centroid vertices) and the centroid of T is edge (y_j, y_k). Since y_j is not in the centroid of T if T is not a "double-star," and $(\mathrm{val}(y_j) - 1)$ of the vertices (all lying in T_{y_j}) adjacent to y_j in T have strictly greater weight in T than y_j, the weight a(x) of the only non-1-valent vertex x adjacent to y_j in T must be strictly less than $a(y_j)$. That is, $a(x) < a(y_j)$.

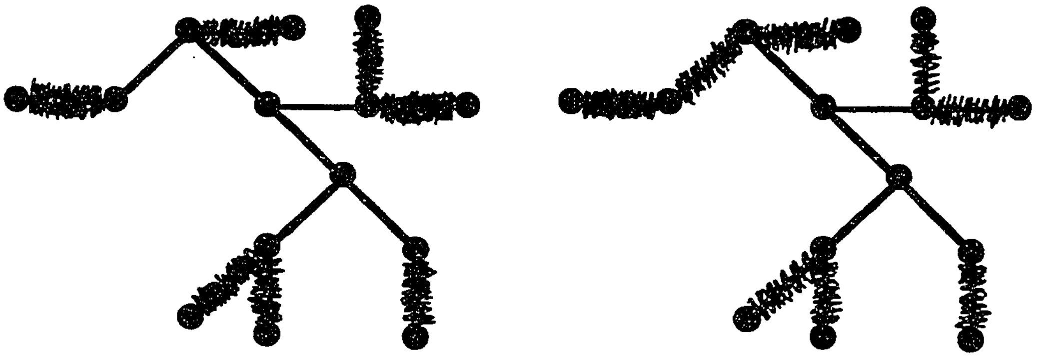
Now that we've eliminated all the 1-valent vertices of T and vertex y_j from consideration as vertices which are possibly in the centroid of T, we can look at the subtree T~~ of T obtained by first pruning all the 1-valent vertices of T (producing subtree T~ of T), and then pruning vertex y_j, which is 1-valent in T~, from T~. T~~ is a tree with one or more vertices. If x is the only vertex of T~~, then x is the centroid of T. Otherwise, T~~ has 2 or more 1-valent vertices. That is, T has 2 or more non-1-valent vertices y^* such that each such vertex y^* is adjacent to exactly 1 vertex of T that is neither a 1-valent vertex of T nor vertex y_j. Let y_j^* be one of these vertices y^* such that $N_{y_j}{}^* \leq N_y{}^*$ for each such vertex y^*. Note that subtree $T_{y_j}{}^*$ consists of all branches of T emanating from y_j^* other than a branch $B_{y_j}{}^*$ of greatest size which emanates from y_j^*, by the way we selected y_j^*. So, in exactly the same way that we argued above concerning vertex y_j, we can show that y_j^* is not in the centroid of T unless there is exactly one other vertex y_k^* in the set of vertices y^* such that $N_{y_j}{}^* = N_{y_k}{}^*$ and y_k^* is adjacent to y_j^* in T, in which case the centroid of T is edge (y_j^*, y_k^*). Again, as we saw for vertex y_j, in case y_j^* is not in the centroid of T, $(\mathrm{val}(y_j^*) - 1)$ vertices (all lying in $T_{y_j}^*$) that are adjacent to y_j^* in T have weights in T that are strictly greater than $a(y_j^*)$. So, the weight $a(x^*)$ of the only vertex x^* which doesn't lie in $T_{y_j}^*$ and is adjacent to y_j^* in T must be strictly less than $a(y_j^*)$. That is, $a(x^*) < a(y_j^*)$.

After i iterations of this procedure, we obtain a subtree T_i of T which is the result of first pruning all the 1-valent vertices of T (to produce T~), then pruning a 1-valent vertex y_j of T~ (to produce T~~), then pruning a 1-valent vertex y_j^* of T~~ (to produce T~~~), where the total number of vertices pruned, one at a time, is (i-1). And, subtree T_i of T contains the centroid of T. Since T contains only a finite number n of vertices, the number of possible iterations of this procedure is finite, so the procedure must end. Note that the $i^{\underline{th}}$ iteration of this procedure finds a vertex v_i of T_i of strictly lower weight in T than a vertex v_{i-1} of T_{i-1} which is a 1-valent vertex of T_{i-1} such that $N_{v_{i-1}} \leq N_{v\#}$ for all 1-valent vertices v# of T_{i-1} and which is adjacent to v_i in T. If T is a star, the procedure terminates after just 1 iteration. In this case, T~ consists of a single vertex which is the centroid of T. If T is not a star, the procedure terminates after some finite number i_* of iterations such that T_{i*} is simply an edge (v_1, v_2). If $N_{v_1} = N_{v_2}$, then $a(v_1) = a(v_2)$, which implies that edge (v_1, v_2) is the centroid of T. If $N_{v_1} < N_{v_2}$, then $a(v_1) > a(v_2)$, which implies that vertex v_2 is the centroid of T. Similarly, if $N_{v_2} < N_{v_1}$, then $a(v_2) > a(v_1)$, which implies that vertex v_1 is the centroid of T. □

In Example 2.1.1 below, the centroid of the tree of Figure 2.1.1 is located by an implementation of Algorithm 2.1.1.

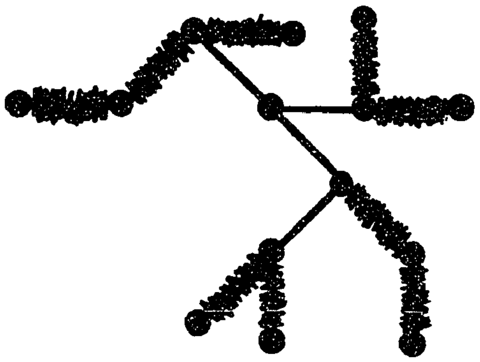
Example 2.1.1: The centroid of the tree of Figure 2.1.1 is located in a sequence of 6 steps by implementing Algorithm 2.1.1. Observe that at each successive step of the procedure, more and more edges of the tree are wiggled until, finally, exactly one unwiggled edge remains in the last step (i.e., step 6). Since, in this last step, $N_{y_j} = 5 < 7 = N_{z_j}$, vertex z_j is the centroid of the tree of Figure 2.1.1.

Figure 2.1.1:



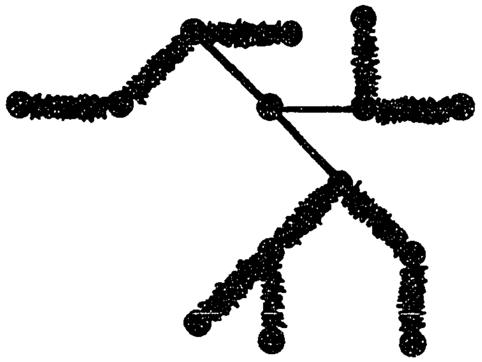
Step 1
Used Algorithm 2.1.1 (1)

Step 2
Used Algorithm 2.1.1 (3), (5)



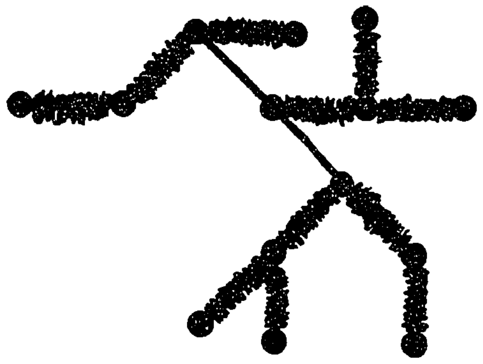
Step 3

Used Algorithm 2.1.1 (3), (5)



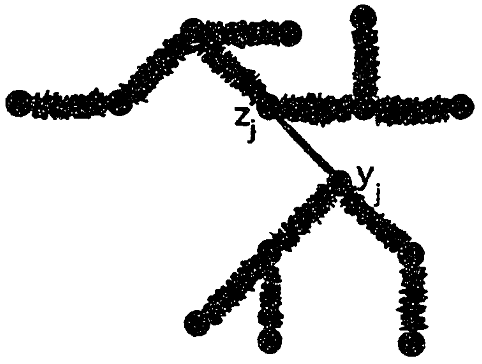
Step 4

Used Algorithm 2.1.1 (3), (5)



Step 5

Used Algorithm 2.1.1 (3), (5)



Step 6

Used Algorithm 2.1.1 (3),
(5), and then (4)

Algorithm 2.1.1 provides another proof of the fact that the centroid of a tree consists of either a single vertex or a pair of adjacent vertices, since the procedure terminates only when either no edges or exactly one edge of the tree remains unwiggled. Additionally, Algorithm 2.1.1 provides a method for locating the <u>median</u> of a tree (which is identical to its centroid, by the theorem of Zelinka [13]) which does not require the determination of the status of each vertex of the tree.

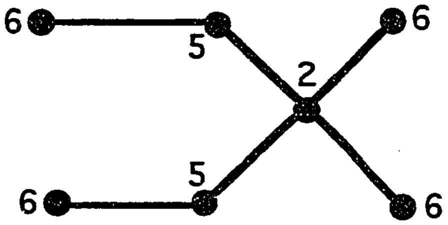
2.2 The Centroid Sequence of a Tree

In Section 1.6, we defined the weight of a vertex v in a tree T. We now define the centroid sequence of a tree.

Definition 2.2.1: Let T be a tree on $n \geq 2$ vertices. We define the **centroid sequence** $A = \{a_1, a_2, ..., a_n\}$ of T to be the sequence of weights of the vertices of T, arranged in non-increasing order (i.e., $a_1 \geq a_2 \geq \cdots \geq a_n$).

Example 2.2.1: The centroid sequence of the tree in Figure 2.2.1 is $A = \{6, 6, 6, 6, 5, 5, 2\}$.

Figure 2.2.1: The weight of each vertex in the tree below is indicated next to that vertex.



An immediate consequence of the definition of the weight of a vertex in a tree is that the weight of each 1-valent vertex in a tree T on $n \geq 2$ vertices is $(n-1)$. Moreover, since each tree T on $n \geq 2$ vertices has at least two 1-valent vertices, the positive integer $(n-1)$ appears at least twice in the centroid sequence A of T.

Remark 2.2.1: The number of times that the positive integer $(n-1)$ appears in the centroid sequence $A = \{a_1, a_2, ..., a_n\}$ of a tree T on $n \geq 2$ vertices is exactly the same as the number of 1-valent vertices of T.

Another property of the centroid sequence of a tree is presented in the following theorem.

Theorem 2.2.1: Let T be a tree on $n \geq 2$ vertices, and let $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ be the centroid sequence of T. Then, $a_{n-1} + a_n = n$.

Proof: Recall that the elements of A are arranged in non-increasing order, so $a_n \leq a_i$ for each $i = 1, 2, ..., n$. So, a_n is the weight of a vertex v_n in the centroid of T. And, by definition, a_n is the size of a branch B_n of T which starts at vertex v_n such that $\text{size}(B_n) \geq \text{size}(B)$ for all branches B of T that start at v_n. Now, let x be any vertex of T other than vertex v_n, and let B_x be a branch of T which starts at vertex x such that $\text{size}(B_x) \geq \text{size}(B\#)$ for all branches B# of T which start at x. According to Lemma 2.1.1, vertex v_n lies on branch B_x. Therefore, a vertex $v_{n-1} \in V(T)\setminus\{v_n\}$ such that $a(v_{n-1}) = a_{n-1} \leq a(v)$ for each vertex $v \in V(T)\setminus\{v_n\}$ must be as close to vertex v_n as possible in T and lie on a branch B~ of T which starts at v_n such that size(B~) ≥ size(B) for all branches B of T that start at v_n. (Note that branch B~ may be different from branch B_n, since B_n is not necessarily the only branch of T which starts at v_n whose size is ≥ the size of each branch of T which starts at v_n). Thus, v_{n-1} must be adjacent to v_n in T and lie on a branch B~ of T which starts at v_n such that size(B~) ≥ size(B) for all branches B of T that start at v_n. So, a_{n-1} is the size of the unique branch B_{n-1} of T which starts at vertex v_{n-1} and contains vertex v_n. Therefore, a_{n-1} = (the sum of the sizes of all branches of T which start at v_n other than branch B~) + 1 = size(T) - size(B~) + 1 = $(n - 1) - a_n + 1 = n - a_n$. That is, $a_{n-1} = n - a_n$, which implies that $a_{n-1} + a_n = n$, and we are done. ☐

Corollary 2.2.1: Let T be a tree on $n \geq 2$ vertices, and let $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ be the centroid sequence of T. Let v_{n-1} be the vertex of T whose weight in T is a_{n-1}, and let v_n be the vertex of T whose weight in T is a_n (i.e., $a(v_{n-1}) = a_{n-1}$ and $a(v_n) = a_n$). Then, vertices v_{n-1} and v_n are adjacent in T. Moreover, if $a_{n-1} = a_n$, then vertices v_{n-1} and v_n are the centroid of T.

Proof: See the proof of Theorem 2.2.1 above. ☐

Remark 2.2.2: If $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ $(n \geq 3)$ is the centroid sequence of a tree, then it cannot be that the last 3 or more elements of A are all equal to each other, by Theorem 2.1.1.

Our next goal is to obtain a characterization of the centroid sequence of a tree. There is a special kind of tree, called a **caterpillar**, whose centroid sequence will be characterized first. This is is the focus of the next section.

2.3 A Characterization of the Centroid Sequence of a Caterpillar

We begin this section with the following definition.

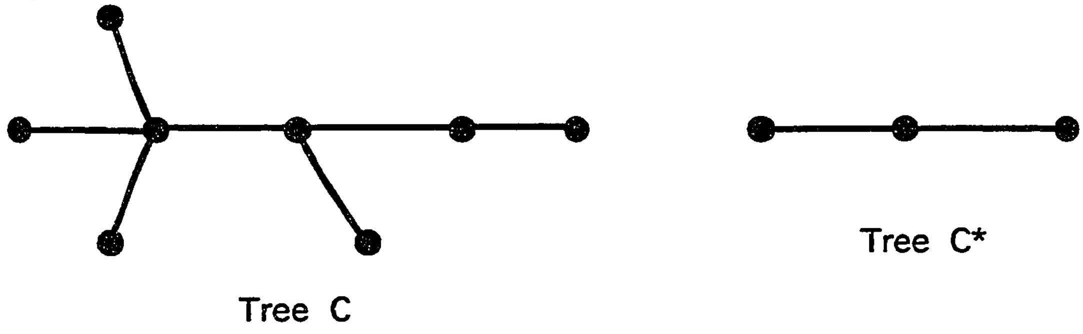
Definition 2.3.1: A **caterpillar** C is a tree on $n \geq 3$ vertices which satisfies the following property:

> If all the 1-valent vertices of C are pruned, the resulting tree is a (possibly degenerate) path.

Remark 2.3.1: A star S on $n \geq 3$ vertices is an example of a caterpillar since the pruned tree S^* of S is a single vertex, which is a degenerate path.

Example 2.3.1: The tree C of Figure 2.3.1(a) is a caterpillar because its pruned tree C* is the path shown in Figure 2.3.1(b).

Figure 2.3.1:


Tree C

(a)

Tree C*

(b)

Definition 2.3.2: Let C be a caterpillar on $n \geq 3$ vertices. Then, the **backbone** of C, denoted by β, is the (possibly trivial) path that remains after all the 1-valent vertices of C have been pruned.

Example 2.3.2: The path C* of Figure 2.3.1(b) is the backbone of the caterpillar C of Figure 2.3.1(a).

Definition 2.3.3: Let T be a tree on $n \geq 2$ vertices, let v be a vertex of T, and let w be a vertex that is adjacent to v in T. Then, the branch B of T which starts at v and **is determined by** w is the unique branch B of T which starts at v and contains vertex w.

The following lemma concerning the weights of the vertices of a tree will be used in the proofs of Theorems 2.3.1 and 2.3.2.

Lemma 2.3.1: Let T be a tree on $n \geq 2$ vertices, let v be a vertex of T, and let a(v) be the weight of v in T.

(i) If v is not in the centroid of T, then there is exactly one vertex w that is adjacent to v in T and whose weight a(w) in T is strictly less than a(v). In addition, the unique branch B_v of greatest size which starts at v in T is determined by vertex w. The weight a(x) in T of each vertex $x \neq w$ that is adjacent to v in T is strictly greater than a(v).

(ii) If v is in the centroid of T, then the weight a(y) of each vertex y that is adjacent to v in T is greater than or equal to a(v). A branch B_v of greatest size which starts at v in T is determined by any vertex z that is adjacent to v in T and whose weight $a(z) \leq a(y)$ for each vertex y that is adjacent to v in T.

Proof: (i) Suppose v is not in the centroid of T. Then, for each vertex u that is adjacent to v in T, $a(u) \neq a(v)$, by Lemma 2.1.2. So, for each vertex u that is adjacent to v in T, either $a(u) > a(v)$ or $a(u) < a(v)$. Let B_v be a branch of greatest size which starts at v in T. Then, $a(v) = \text{size}(B_v)$. Let w be the unique vertex of T which lies on B_v and is adjacent to v in T. Then, w determines branch B_v. Now, if c is a vertex in the centroid of T, then c lies on B_v, by Lemma 2.1.1. Let B_w be a branch of greatest size which starts at w in T. Then, vertex c also lies on B_w, again by Lemma 2.1.1. If w is not in the centroid of T, then B_w is a sub-branch of B_v. And, since (v, w) is an edge of B_v but not of B_w, $\text{size}(B_w) < \text{size}(B_v)$.

Thus, $a(w) < a(v)$. If w is in the centroid of T, then again $a(w) < a(v)$, this time because v is not in the centroid of T, by hypothesis. Hence, in either case, $a(w) < a(v)$.

To prove that branch B_v is the unique branch of greatest size which starts at v in T, let B be any branch of greatest size which starts at v in T. Then, vertex c lies on B, by Lemma 2.1.1. But, vertex c lies on branch B_v which starts at v in T. And, since each vertex of $V(T)\backslash\{v\}$ lies on one and only one branch which starts at v in T, we see that $B = B_v$ (i.e., B and B_v are one and the same branch).

Finally, let $x \neq w$ be a vertex that is adjacent to v in T. Then, since x does not lie on branch B_v, x is not a vertex in the centroid of T. Now, let B_x be a branch of greatest size which starts at x in T. Then, vertex c lies on B_x, by Lemma 2.1.1. And, since vertex c also lies on branch B_v which starts at v and is determined by w, B_v is a subtree of B_x. Also, since (x, v) is an edge of B_x but not of B_v, $\text{size}(B_v) < \text{size}(B_x)$, which implies that $a(v) < a(x)$. Since $x \neq w$ is an arbitrary vertex which is adjacent to v in T, we are done.

(ii) Suppose, now that v is in the centroid of T. Then, the weight a(v) of v in T is ≤ the weight of each vertex of T. In particular, $a(v) \leq a(y)$ for each vertex y that is adjacent to v in T. Let B_v be a branch of greatest size which starts at v in T. If there is another vertex $z \neq v$ in the centroid of T (note that there is at most one more vertex in the centroid of T, by Theorem 2.1.1), then z must lie on B_v, by Lemma 2.1.1. And, vertices v and z must be adjacent in T, by Lemma 2.1.2. Hence, branch B_v is determined by vertex z. And, since vertex z lies on one and only one branch of T which starts at v, B_v is the unique branch of greatest size which starts at v in T in case $z \neq v$ is also in the centroid of T. If, on the other hand, the centroid of T consists of only vertex v, then for each vertex y which is adjacent to v in T, y is not in the centroid of T. So, by (i) above, the unique branch B_y of greatest size which starts at y in T is determined by vertex v. And, size(B_y) = size(T) - (the size of the branch of T which starts at v and that vertex y lies on) + 1. This implies that (the size of the branch of T which starts at v and that vertex y lies on) = size(T) - size(B_y) + 1. Thus, the size of a branch B_v of greatest size which starts at v in T = size(T) - size(B_z) + 1, where z is a vertex that is adjacent to v in T such that size(B_z) = $a(z) \leq a(y)$ = size(B_y) for each vertex y that is adjacent to v in T. That is, a branch of greatest size which starts at v in T is determined by a vertex z that is adjacent to v in T such that $a(z) \leq a(y)$ for each vertex y that is adjacent to v in T. Since there may be more than one vertex adjacent to v in T whose weight in T is $\leq$ the weight in T of each vertex that is adjacent to v in T, there may be more than one branch of greatest size which starts at v in T in case v is the centroid of T. □

Theorem 2.3.1: Let C be a caterpillar on $n \geq 3$ vertices, and let v be a vertex lying on the backbone β of C. Let $A = \{a_1, a_2, ..., a_n\}$ be the centroid sequence of C, and let a(v) be the weight of v in C.

(i) If β consists of only vertex v, then C is the star on n vertices, so the number of 1-valent vertices that are adjacent to v in C is (n-1).

(ii) If length(β) ≥ 1 and v is an endpoint of β, then the number of 1-valent vertices that are adjacent to v in C is (n-1) - a(v).

(iii) If length(β) ≥ 2 and v is a vertex of β which is adjacent to the two distinct vertices w and x which also lie on β and whose weights in C are a(w) and a(x), respectively, then either $a(w) > a(v)$ or $a(x) > a(v)$ or both and the number of 1-valent vertices that are adjacent to v in C is $\max \{a(w) - a(v) - 1,\ a(x) - a(v) - 1\}$.

Proof:

(i) Since $n \geq 3$, C contains at least one vertex v which is not 1-valent (i.e., val(v) $\geq$ 2). Therefore, if we prune all the 1-valent vertices from C, the resulting path β (i.e., the backbone of C) contains at least one vertex v. Now, if β consists of only vertex v, then every vertex of C other than v must be a 1-valent vertex of C. Hence, the number of 1-valent vertices of C that are adjacent to v in C is (n-1), and C is the star on n vertices.

(ii) Now, if length(β) $\geq$ 1, then β contains at least 2 distinct vertices. Suppose v is an endpoint of β. Then, there is a vertex (say, x) which lies on β, is distinct from v, and is adjacent to v on β. And, since x lies on β, val(x) in C must be ≥ 2. Now, since v is an endpoint of β, all vertices of C that are adjacent to v in C, with the sole exception of vertex x, must be 1-valent vertices of C. So, <u>the</u> branch B of greatest size which starts at v in C is determined by vertex x (because B is the only branch which starts at v in C such that size(B) > 1). Hence, a(v) = size(B) [i.e., the weight of vertex v in C = size(B)]. And, since C is a caterpillar on $n \geq 3$ vertices, size(C) = n - 1. Since branch B contains a(v) of the (n-1) edges of C, the set of branches of C other than B which start at v contain a total of (n-1) - a(v) edges. But, each branch of C which starts at v, other than branch B, is of size 1. Therefore, there are exactly (n-1) - a(v) 1-valent vertices adjacent to vertex v in C.

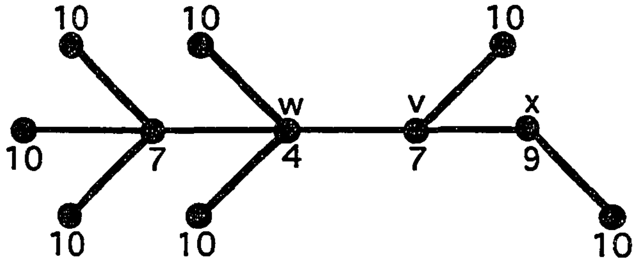
(iii) If length(β) $\geq$ 2, then there is a vertex v lying on β which is adjacent to two other vertices w and x which also lie on β (i.e., v lies between w and x on β). Let's consider the weights in C of vertices w and x relative to the weight of vertex v in C.

Is it possible that the following two inequalities hold simultaneously in C: $a(w) \leq a(v)$ <u>and</u> $a(x) \leq a(v)$? If $a(w) \leq a(v)$ <u>and</u> $a(x) \leq a(v)$, then vertex v cannot be in the centroid of C, by Theorem 2.1.1. Therefore, there is exactly one vertex, say, v~, that is adjacent to v in C whose weight in C is strictly less than that of v, and each vertex other than v~ that is adjacent to v in C is such that its weight in C is strictly greater than a(v), by

Lemma 2.3.1. Hence, it is impossible that $a(w) \le a(v)$ <u>and</u> $a(x) \le a(v)$ hold simultaneously in C.

Can it be that, in C, $a(w) \le a(v) < a(x)$ (or $a(x) \le a(v) < a(w)$)? This <u>is</u> possible, as illustrated by Figure 2.3.2.

<u>Figure 2.3.2</u>: In caterpillar C below, the weight of each vertex in C is indicated next to that vertex. Note particularly vertices w, v, and x which all lie on the backbone β of C.



Caterpillar C

In Figure 2.3.2, the branch B_x of greatest size which emanates from x in C is determined by v (note that $a(v) < a(x)$) and the branch B_v of greatest size which emanates from v in C is determined by w (note that $a(w) \le a(v)$). Observe that $\text{size}(B_x) = a(x)$ and $\text{size}(B_v) = a(v)$. Furthermore, B_v is a subtree of B_x. In fact, $B_x = B_v \cup (v, x) \cup$ {all branches of length 1 emanating from v in C} $\cup \{x\}$. Thus, $a(x) = a(v) + 1 +$ (the number of 1-valent vertices adjacent to v in C), which implies that (the number of 1-valent vertices adjacent to v in C) $= a(x) - a(v) - 1$. Therefore, if $a(w) \leq a(v) < a(x)$, then the number of 1-valent vertices that are adjacent to v in C is $a(x) - a(v) - 1$. If, instead, $a(x) \leq a(v) < a(w)$, then by exactly the same kind of argument, the number of 1-valent vertices that are adjacent to v in C is $a(w) - a(v) - 1$. So, in case vertex v is adjacent to 2 vertices w and x on β, one of whose weights in C is $> a(v)$ and the other of whose weights in C is $\leq a(v)$, the number of 1-valent vertices of C that are adjacent to v in C is given by max $\{a(w) - a(v) - 1,\ a(x) - a(v) - 1\}$.

Finally, can we have $a(w) > a(v)$ and $a(x) > a(v)$ in C? The answer is yes, provided that v is the centroid of C. In this case, $a(v) + \min\{a(w), a(x)\} = n$ [because $a(v) = a_n$ and $\min\{a(w), a(x)\} = a_{n-1}$, since either vertex w or vertex x (or both) determine(s) a branch of greatest size which emanates from v in C, and because $a_{n-1} + a_n = n$]. Suppose, without loss of generality, that $a(w) \leq a(x)$. Then, the branch B_x of greatest size emanating from x in C is determined by v, the branch B_w of greatest size emanating from w in C is determined by v, and a branch B_v of greatest size emanating from v in C is determined by w. Hence, B_v is a subtree of B_x. So, $B_x = B_v \cup (v, x) \cup \{$all branches of length 1 emanating from v in C$\} \cup \{x\}$, which implies that (the number of 1-valent vertices adjacent to v in C) $= a(x) - a(v) - 1$. If, on the other hand, we had assumed that $a(x) \leq a(w)$, then our result would have been: (the number of 1-valent vertices adjacent to v in C) = $a(w) - a(v) - 1$. So, in case $a(w) > a(v)$ and $a(x) > a(v)$, vertex v is the centroid of C and the number of 1-valent vertices that are adjacent to v in C is given by:

$$\max\{a(w) - a(v) - 1,\ a(x) - a(v) - 1\}.$$

Hence, we have shown that if v is adjacent to both w and x on β, then either $a(w) > a(v)$ or $a(x) > a(v)$ or both in C, and the number of 1-valent vertices adjacent to v in C is $\max\{a(w) - a(v) - 1, a(x) - a(v) - 1\}$. This completes the proof. □

The task of deciding whether a given finite sequence A of positive integers is realizable as the centroid sequence of a caterpillar can be accomplished by applying the statement of the following theorem.

Theorem 2.3.2 (Characterization of the Centroid Sequence of a Caterpillar):

The finite sequence $A = \{a_1, a_2, \ldots, a_n\}$ $(n \geq 3)$ of positive integers, arranged in non-increasing order (i.e., $a_1 \geq a_2 \geq \cdots \geq a_n$), is realizable as the centroid sequence of a caterpillar C if and only if all of the following conditions are satisfied:

(i) $a_1 \leq n - 1$;

(ii) $a_{n-1} + a_n = n$;

(iii) for each $a_i \in A$ such that $a_i < n - 1$, there is at most one $a_j \in A$ with $a_i = a_j$ but $i \neq j$.

Remark 2.3.2: The caterpillar of Figure 2.3.2 provides an example of a caterpillar which satisfies condition (iii). This caterpillar has $n = 11$ vertices, and its centroid sequence is $A = \{10, 10, 10, 10, 10, 10, 10, 9, 7, 7, 4\}$. Note that each positive integer strictly less than 10 that appears in A occurs at most twice in A, and "7" occurs in A exactly twice.

Proof (of Theorem 2.3.2):

(⇒) Suppose that $A = \{a_1, a_2, ..., a_n\}$ $(n \geq 3)$ is realizable as the centroid sequence of a caterpillar C. Since C is a non-trivial tree, conditions (i) and (ii) are clearly satisfied. So, if we can prove that condition (iii) also holds, we will be done. Condition (iii) states that each positive integer $k < n - 1$ which appears in A occurs at most twice in A.

We shall use the method of proof by contradiction to prove condition (iii). So, suppose that it is not the case that each positive integer $k < n - 1$ which appears in A occurs at most twice in A. Then, there is a positive integer $j < n - 1$ that appears in A such that j occurs three or more times in A. Observe that j cannot be the weight in C of a vertex in the centroid of C, by Theorem 2.1.1. Therefore, j must be the weight in C of 3 or more distinct vertices of C, all of which lie on the backbone β of C and none of which are in the centroid of C. Let x, y, and z be any three of these 3 or more vertices of C each of whose weights in C is j. Then, $a(x) = a(y) = a(z) = j$. Let c be a vertex in the centroid of C. Observe that c also lies on β, $a(c) < j$, and c is distinct from x, y, and z. So, if we draw caterpillar C so that its backbone β is horizontal, vertex c divides β into two sides -- a left side and a right side. Therefore, at least 2 of the vertices x, y, and z must lie on the same side of c on β. Suppose, without loss of generality, that vertices x and y both lie to the right of c on β. Note that the weights of the vertices to the right of c on β must appear in strictly increasing order, from left to right, by Lemma 2.3.1. But, then, it is impossible that both x and y lie to the right of c on β, because x and y both have the same weight j in C. Hence, assuming that there is a positive integer $j < n - 1$ that appears in A such that j occurs 3 or more times in A leads to a contradiction. So, each positive integer $k < n - 1$ which appears in A occurs at most 2 times in A. This proves condition (iii), and we are done. $\square$

($\Leftarrow$) Conversely, suppose conditions (i), (ii), and (iii) hold. Let S be the subsequence of A consisting of all elements of A that are < n-1. Note, in particular, that $a_n \in S$ since $a_n < n-1$ when $n \geq 3$. Since $|S| < \infty$ and since no positive integer occurring in S appears more than twice, the elements of S can be arranged (i.e., ordered) in such a way that those elements appearing to the left of a_n in S appear in strictly decreasing order and those elements appearing to the right of a_n in S appear in strictly increasing order. (Observe that there may be several different such orderings of the elements of S.)

Claim: There exists a caterpillar C whose centroid sequence is A and whose backbone β contains exactly $|S|$ vertices whose weights in C are simply the elements of S arranged in the same order as they appear in S.

Proof of Claim: We will prove the existence of such a caterpillar C by actually constructing C. First of all, note that conditions (i) and (ii) are necessary conditions for A to be the centroid sequence of a tree on n vertices. Condition (iii) allows for the construction of a backbone β for C. Before we begin the construction of C, we should note that since C is to be a caterpillar on $n \geq 3$ vertices, it will have at least two 1-valent vertices. That is, if A is to be the centroid sequence of C, then we must have $a_1 = n - 1 = a_2$. But, condition (i) states only that $a_1 \leq n - 1$. We will prove that conditions (i), (ii), and (iii) together imply that $a_1 = n - 1 = a_2$. Condition (i) implies that $a_2 \leq n - 1$ since the elements of A are arranged in non-increasing order. Condition (ii) implies that $a_{n-1} \geq n/2$. For $n \geq 3$, a_1 and a_{n-1} are distinct elements of A (although they could represent the same positive integer). So, for $n \geq 3$, if $a_{n-1} = n - 1$, then $a_1 = a_2 = \dots = a_{n-1} = n - 1$, and we are done. If $a_{n-1} \neq n-1$, how can we minimize the number of times that the integer (n-1) appears in A? The answer to this question is: Minimize a_{n-1} and maximize the number of times positive integers $< (n-1)$ appear in A. Recall that according to condition (iii), each positive integer $< (n-1)$ can appear at most twice in A.

If n is even, then the smallest possible value for a_{n-1} is $n/2$. Consider the sequence $A = \{a_1, a_2, \dots, a_{n-1}, a_n\}$. $|A| = n$, so the number of elements of the subsequence A~ = $\{a_2, a_3, \ldots, a_{n-1}\}$ of A is (n-2). And, we would like A~ to contain as many positive integers that are < (n-1) as possible. So, suppose $a_{n-1} = n/2$ and $a_2 \leq n - 2$. Note that the number of integers in the set $X = \{n/2, (n/2) + 1, (n/2) + 2, \ldots, n-2\}$ is $(n/2) - 1$. In addition, observe that every element of X may appear at most twice in A~, with the exception of n/2 which may appear only once in A~ [because $a_{n-1} = n/2$ implies that $a_n = n/2$ (by condition (ii)), so n/2 already appears twice in A]. So, if each element of X appears in A~ as many times as it possibly can, then $2((n/2) - 1) - 1 = n - 2 - 1 = n - 3$ elements of A~ will be accounted for. But, |A~| = n - 2. This means that even if we allow every possible positive integer < (n-1) to appear in A~ with its greatest possible frequency of occurrence, there will still be one element of A~ (namely, a_2) which is unaccounted for. This implies that it must be the case that $a_2 = n - 1$. By condition (i), $a_1 = n - 1$, as well.

If, on the other hand, n is odd, then the smallest possible value for a_{n-1} is (n+1)/2. In this case, $a_n = (n-1)/2$ (by condition (ii)). As before, if we would like A~ = $\{a_2, a_3, \ldots, a_{n-1}\}$ to contain as many positive integers that are < (n-1) as possible, we let $a_{n-1} = (n+1)/2$ and $a_2 \leq n-2$. Note that the number of integers in the set $Y = \{(n+1)/2, (n+1)/2 + 1, \ldots, n-2\}$ is $(n-1)/2 - 1$. And, this time, every element of Y may appear at most 2 times in A~ (because $a_{n-1} \neq a_n$). So, if each element of Y appears twice in A~, then $2((n-1)/2 - 1) = n - 1 - 2 = n - 3$ elements of A~ will be accounted for. But, since |A~| = n - 2, one element of A~ (namely,

a_2) will still be unaccounted for. This implies that $a_2 = n - 1$. So, even in case every possible positive integer $< (n-1)$ appears in A~ twice, it must still be that $a_2 = n - 1$. By condition (i), $a_1 = n - 1$, as well. Hence, whether n is even or odd, $a_1 = n - 1 = a_2$, which is a necessary condition for A to be the centroid sequence of a tree.

Now, we'll proceed with the construction of a caterpillar C whose centroid sequence is A. We now know that the positive integer $(n-1)$ appears at least twice in A. Suppose the number of times that $(n-1)$ appears in A is i, where $i \geq 2$. Then, there are $(n-i)$ elements of A, each of which is $< (n-1)$. That is, if S is the subsequence of A consisting of all elements of A that are $< (n-1)$, then $|S| = n - i$. Since $n \geq 3$, $a_n < n-1$, so $a_n \in S$. If we arrange (i.e., order) the elements of S so that those elements appearing to the left of a_n in S appear in strictly decreasing order and those elements appearing to the right of a_n in S appear in strictly increasing order, then we can draw a path β (which will be the backbone of C) with $|S| = n - i$ vertices and assign the elements of S to be the weights in C of these vertices, in the same order, from left to right along β, that the elements of S have been arranged in above. Note that such an ordering of the elements of S is possible (and may not be unique) because each positive integer appearing in S occurs at most twice. And, the weights in C of the vertices along β can appear in the same order as they now appear in S, by Lemma 2.3.1. Now, we know from Theorem 2.3.1 that each vertex v lying on β will have a specific number of 1-valent vertices attached to it in C, depending on its weight and the weight (respectively, weights) of the vertex (respectively, vertices) that is (respectively, are) adjacent to v on β. The question that we must address is: Is the number of times (namely, i) that (n-1) occurs in A the same as the number of 1-valent vertices needed to complete the construction of the caterpillar C, starting with β as backbone for C? We'll use Theorem 2.3.1 to compute the total number of 1-valent vertices needed to construct C from β. There are 3 cases to be considered.

First, if β consists of only a single vertex v, then C must be a star and the weight of v in C must be 1. So, in this case, the number of 1-valent vertices needeed to construct C is (n-1). So, it must be that $i = n - 1$. And, this, indeed, is the case if $a_n = 1$ because then $a_{n-1} = n - 1$. Thus, $a_1 = a_2 = \ldots = a_{n-1} = n - 1$. This implies that the positive integer (n-1) occurs (n-1) times in A. Hence, $i = n - 1$.

Second, if length(β) ≥ 1, then β has 2 endpoints (call them q and r, respectively) and possibly some vertices which are not endpoints between q and r. If q and r are the only vertices of β, and a(q) and a(r) are the weights of q and r, respectively, in C, then the number of 1-valent vertices needed to construct C is $((n-1) - a(q)) + ((n-1) - a(r)) = 2n - 2 - a(q) - a(r)$. And, in this case, one of a(q) and a(r) is a_n and the other is a_{n-1} because these are the only 2 vertices of C whose weights in C will be $< n-1$. So, the number of 1-valent vertices needed to construct C, in this case, is $2n - 2 - a(q) - a(r) = 2n - 2 - (a(q) + a(r)) =$ $2n - 2 - (a_{n-1} + a_n) = 2n - 2 - n = n - 2$. And, if q and r are the only vertices of β, then $|S| = 2$, which implies that the number of times that the positive integer (n-1) occurs in A is also (n-2) [i.e., $i = n - 2$].

Finally, if there are vertices other than q and r lying on β, then let c be the vertex whose weight in C will be a_n and let d be the vertex whose weight in C will be a_{n-1}. Note that q, r, c, and d may not all be distinct (although $q \neq r$ and $c \neq d$). Let's re-label the vertices of β from left to right as follows, where we assume, without loss of generality, that q is the left endpoint of β and r is the right endpoint of β: $b_1 = q, b_2, b_3, \ldots, b_{n-i} = r$. Then, the number of 1-valent vertices needed to construct C in this case is:

$$((n-1) - a(q)) + ((n-1) - a(r)) + \sum_{j=1}^{n-i-1} (|b_j - b_{j+1}| - 1) - (a_{n-1} - a_n - 1)$$

$$= 2n - 2 - a(q) - a(r) + [(a(q) + a(r)) - (a_n + a_{n-1})] - (n - i - 2)$$

$$= 2n - 2 - a(q) - a(r) + a(q) + a(r) - a_n - a_{n-1} - n + i + 2$$

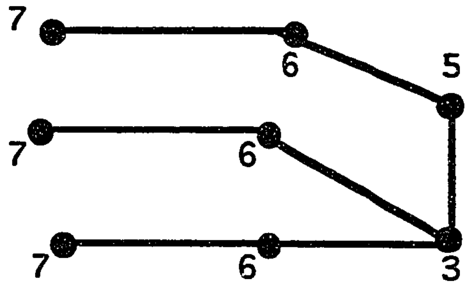
$$= n - a_n - a_{n-1} + i = n - (a_n + a_{n-1}) + i = n - n + i = i.$$

And, i is precisely the number of times that the positive integer (n-1) occurs in A.

So, in each of the 3 cases discussed above, the number i of times that (n-1) occurs in A is the same as the number i of 1-valent vertices neeeded to construct C from β. Hence, in each case, if we attach the appropriate number of 1-valent vertices to each vertex of β (as determined by Theorem 2.3.1), we will use a total of i 1-valent vertices, each having weight (n-1) in C. That is, the (n-i) vertices of β, each of weight < (n-1) in C, will require the attachment of a total of i 1-valent vertices, each of weight (n-1) in C, to construct C. It is clear that C is a caterpillar since C was constructed from path β by attaching i 1-valent vertices to the vertices of β. And, for each vertex of C, the element of A assigned to that vertex is, in fact, its weight in C, by construction, and because the number i of times that (n-1) occurs in A is exactly the same as the number i of 1-valent vertices needed to attach to the vertices of β according to the specifications of Theorem 2.3.1. So, C is a caterpillar whose centroid sequence is A and whose backbone β contains exactly |S| vertices whose weights in C are simply the elements of S arranged in the same order as they appear in S, by the way we constructed C. ☐

Example 2.3.3: A = {7, 7, 7, 6, 6, 6, 5, 3} is not realizable as the centroid sequence of a caterpillar because the number "6" appears in A more than two times, which violates condition (iii) of Theorem 2.3.1 above. Note, however, that A is realizable as the centroid sequence of the tree of Figure 2.3.3.

Figure 2.3.3:



Remark 2.3.3: Note that if either condition (i) or condition (ii) of Theorem 2.3.1 is violated, then $A = \{a_1, a_2, \ldots, a_n\}$ cannot be realized as the centroid sequence of a caterpillar, since each of these two conditions is a necessary condition for a sequence of $n \geq 3$ positive integers to be the centroid sequence of a tree (see Section 2.4).

Algorithm 2.3.1: Construction of a Caterpillar from a Finite Sequence $A = \{a_1, a_2, \ldots, a_n\}$ $(n \geq 3)$ of Positive Integers which is Known to be the Centroid Sequence of a Caterpillar

(i) Let S be the subsequence of A which consists of all elements of A that are strictly less than (n-1). Observe that since $n \geq 3$, $a_n \in S$.

(ii) Order the elements of S in such a way that those elements of S that are on the left side of a_n appear in strictly decreasing order, from left to right, and those elements of S that are on the right side of a_n appear in strictly increasing order, from left to right. (Note that there may be several ways to do this, with each way leading to the construction of a caterpillar whose centroid sequence is A which is not isomorphic to any caterpillar obtained from a different ordering of the elements of S.)

(iii) Draw a path β of order $|S|$. Label the $|S|$ vertices of β with the elements of S, from left to right, in exactly the same order as these elements appear in S in step (ii) above.

(iv) For each vertex v which is an endpoint of β, let a(v) be the element of S that was used to label v in step (iii) above. Attach ((n-1) - a(v)) 1-valent vertices to vertex v.

(v) For each vertex v of β which is not an endpoint of β, let w and x be the 2 distinct vertices of β that are adjacent to v on β. Let a(w) be the element of S that was used to label w in step (iii) above, and let a(x) be the element of S that was used to label x in step (iii) above. Attach max $\{a(w) - a(v) - 1, a(x) - a(v) - 1\}$ 1-valent vertices to vertex v.

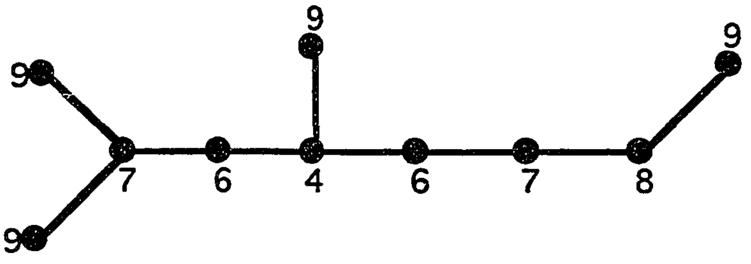
(vi) Label each 1-valent vertex of the resulting graph C with any element of A\S (since each element of A\S equals (n-1)).

(vii) The graph C of step (vi) above is a caterpillar whose centroid sequence is A since the weight in C of each vertex of C which lies on β is given by the element of S that was used to label that vertex in step (iii) above, the weight in C of each 1-valent vertex of C is (n-1) (i.e., an element of A\S), the number of vertices of C that lie on β is |S|, and the number of 1-valent vertices of C is |A\S|.

<u>Example 2.3.4</u>: Suppose we are given the sequence A = {9, 9, 9, 9, 8, 7, 7, 6, 6, 4}. Notice that |A| = 10 and that A satisfies conditions (i), (ii), and (iii) of Theorem 2.3.2. Therefore, A is realizable as the centroid sequence of a caterpillar. Now, the elements of A that are strictly less than 9 can be ordered as follows: S = {7, 6, 4, 6, 7, 8} (see step (ii) of Algorithm 2.3.1 above). In Figure 2.3.4, we construct a caterpillar C whose

centroid sequence is A by first drawing a path β of order |S| = 6 and labeling the vertices of β, from left to right, with the elements of S, in exactly the order in which they appear in S, from left to right. We then attach two 1-valent vertices to the left endpoint (whose label is "7") of β, no 1-valent vertices to the vertex labeled "6" that is adjacent to the left endpoint of β, one 1-valent vertex to the vertex of β which is labeled "4", no 1-valent vertices to the vertex labeled "6" which is to the right of the vertex labeled "4" on β, no 1-valent vertices to the vertex labeled "7" which is adjacent to the right endpoint of β (whose label is "8"), and one 1-valent vertex to the right endpoint of β (whose label is "8"). Finally, we label each of the four 1-valent vertices of the resulting graph with a "9". This produces a caterpillar C whose centroid sequence is A. Note that the label given to each vertex of C is the weight of that vertex in C.

Figure 2.3.4: The caterpillar C constructed in Example 2.3.4 above. The weight of each vertex of C is indicated next to that vertex.



Caterpillar C

A question that arises naturally at this point is:
If $A = \{a_1, a_2, ..., a_n\}$ is the centroid sequence of a caterpillar C, can A also be the centroid sequence of another caterpillar C~ which is not isomorphic to C? The next theorem answers this question.

Theorem 2.3.3: Let $A = \{a_1, a_2, ..., a_n\}$ $(n \geq 2)$ be the centroid sequence of a caterpillar C. Let S = the subsequence of A consisting of all elements of A that are $\neq (n-1)$, and let i = the number of positive integers $\neq a_n$ appearing in S exactly one time. Then:

(i) if $i = 0$, there is exactly one caterpillar (namely, C), up to isomorphism, whose centroid sequence is A;

(ii) if $i \geq 1$, there are exactly 2^{i-1} distinct caterpillars, up to isomorphism, such that A is the centroid sequence of each of these caterpillars.

Proof:

(i) Note that if $i = 0$, then a_n is the only positive integer in A that may appear exactly once in S. So, every positive integer k such that $a_n < k < n-1$ appears either exactly twice in A or exactly zero times (i.e., not at all) in A. Hence, there is no way to re-arrange the weights of the vertices lying on the backbone β of C to construct a new caterpillar C~ which is not isomorphic to C. Thus, C is the unique caterpillar, up to isomorphism, whose centroid sequence is A.

(ii) We will prove this part by induction on the number i of positive integers $\neq a_n$ that appear in S exactly one time. If i = 1, then there is exactly one postive integer $k \neq a_n$ that appears in S exactly one time. Observe that we can view the vertex v_n of C whose weight in C is a_n as a "fixed" vertex of β which separates the "left side" of β from the "right side" of β. Then, we can think of each vertex of C whose weight in C is k, where $a_n < k < n-1$, as a vertex of β which can be moved from one side of a_n to the other along β. Suppose, without loss of generality, that the unique vertex of C whose weight is k (in case i = 1) lies to the right of the vertex whose weight is a_n on β. Since the set of weights, other than k, of the vertices to the right of the vertex whose weight is a_n is the same as the set of weights of the vertices to the left of the vertex whose weight is a_n, and since the set of weights of the vertices along β to the right of a_n must be arranged in strictly increasing order, from left to right, and the the set of weights of the vertices along β to the left of a_n must be arranged in strictly increasing order, from right to left, we see that k is the only weight along β that can be re-located (i.e., moved from one side of a_n to the other). But, if we move k (and the vertex of β whose weight in C is k) from its original position along β, we can only re-locate k to the identical position on the left side of a_n that k used to occupy on the right side of a_n. So, this "new" arrangement of the weights along β is actually just the mirror-image of the original arrangement of the weights along β. That is, the caterpillar corresponding to this re-location of k is isomorphic to C. Since there is no other way to re-arrange the weights along β to render a new caterpillar which is not isomorphic to C, we see that when i = 1, there is one and only one caterpillar, up to isomorphism, whose centroid sequence is A. So, when i = 1, there is $2^{1-1} = 2^0 = 1$ caterpillar whose centroid sequence is A, verifying the statement of the theorem for the case i = 1.

In case i = 2, there are exactly 2 distinct positive integers $j_1 \neq a_n$ and $j_2 \neq a_n$ that each appear exactly one time in S. So, <u>either</u> both j_1 and j_2 appear on the same side of a_n on β <u>or</u> one of them appears to the right of a_n on β and the other appears to the left of a_n on β. And, since the weights to the right of a_n on β strictly increase, from left to right, and the weights to the left of a_n on β strictly increase, from right to left, there is one and only one way that both j_1 and j_2 can appear on the same side of a_n, up to isomorphism, and there is one and only one way that j_1 and j_2 can appear on opposite sides of a_n, up to isomorphism. Hence, there are exactly 2 non-isomorphic caterpillars whose centroid sequence is A. That is, when i = 2, there are $2^{i-1} = 2^{2-1} = 2$ non-isomorphic caterpillars such that A is the centroid sequence of each of these caterpillars, verifying the statement of the theorem for the case i = 2. Now, we introduce our induction hypothesis. Suppose that for i = k $(k \geq 1)$, there are exactly 2^{k-1} non-isomorphic caterpillars such that A is the centroid sequence of each of these caterpillars. Then, we must show that for i = k+1, there are exactly $2^{(k+1)-1} = 2^k$ non-isomorphic caterpillars, each having A as its centroid sequence. When $i = k+1$, there are exactly $(k+1)$ distinct positive integers, each $\neq a_n$, that each appear exactly once in S. Now, if we remove one of these $(k+1)$ positive integers (say, z), from A and replace it in A with an extra occurrence of the positive integer $(n-1)$, there will be exactly k distinct positive integers, each $\neq a_n$, that each appear exactly one time in S# (where S# is the sequence obtained from S by removing z from S). And, according to our induction hypothesis, there are 2^{k-1} non-isomorphic caterpillars such that A# (i.e., the sequence obtained from A by removing z and replacing it with an extra occurrence of $(n-1)$) is the centroid sequence of each of these caterpillars. Note that the order in which the elements of S# appear along the backbone of any one of these 2^{k-1} non-isomorphic caterpillars differs from the order in which these elements appear along the backbone of any of the other $(2^{k-1} - 1)$ non-isomorphic caterpillars. Let C~ be any one of these 2^{k-1} non-isomorphic caterpillars, and let β~ be the backbone of C~. Now, re-introduce the integer z that had been removed from A and eliminate the extra occurrence of $(n-1)$ in A which was used to replace z. Since z appears exactly once in S, z can be added to exactly one side of a_n along β~. Note that the right side of β~ and the left side of β~ are <u>not</u> mirror images of each other because, by hypothesis, $k \geq 1$ (i.e., there is at least one positive integer $x \neq a_n$ that appears exactly once along β~ and, therefore, appears on exactly one side of a_n along β~). Also, the integer z does not already appear along β~ because, by hypothesis, the integer z occurs exactly once in S and z was not used along the backbone of any of the 2^{k-1} non-isomorphic caterpillars discussed above. Hence, there are exactly 2 ways to add the integer z along $\beta\sim$ -- either place z on the right side of a_n, or place z on the left side of a_n. Since the right side of $\beta\sim$ and the left side of $\beta\sim$ are not mirror images of each other, these 2 ways to add z along $\beta\sim$ are distinct and, therefore, give rise to 2 non-isomorphic caterpillars. And, since the weights to the right of a_n must strictly increase, from left to right, and those to the left of a_n must strictly increase, from right to left, in order for these weights to be the weights of the vertices lying on the backbone of a caterpillar, there is exactly one way to add z to the right of a_n, and exactly one way to add z to the left of a_n. So, for each caterpillar C~ of the 2^{k-1} non-isomorphic caterpillars whose backbones each contain $k \geq 1$ elements of S# (other than a_n) that appear exactly one time in S#, there are exactly 2 non-isomorphic caterpillars that can be constructed by the addition of z to S# (thereby recovering S). That is, for each caterpillar C~ in the collection of the 2^{k-1} non-isomorphic caterpillars mentioned above, there are exactly 2 non-isomorphic caterpillars $C_1\sim$ and $C_2\sim$ that can be constructed from C~ by placing the positive integer z on $\beta\sim$. Thus, since there are 2^{k-1} such non-isomorphic caterpillars C~, and each of these gives rise to 2 non-isomorphic caterpillars $C_1\sim$ and $C_2\sim$ (when z is placed on $\beta\sim$), there are exactly $2^{k-1} \cdot 2 = 2^k$ non-isomorphic caterpillars when $i = k+1$ (i.e., when there are exactly $(k+1)$ positive integers, each $\neq a_n$, that each appear exactly one time in

S) whose centroid sequence is A. This completes the proof by induction. □

Since the centroid sequence A of a caterpillar C *may* also be the centroid sequence of a caterpillar C~ which is not isomorphic to C, can A also be the centroid sequence of a tree T which is not a caterpillar? The following theorem shows that A *may*, in fact, be the centroid sequence of a non-caterpillar tree T, and sets down the conditions under which this is possible.

Theorem 2.3.4: Let $A = \{a_1, a_2, ..., a_n\}$ $(n \geq 3)$ be the centroid sequence of a caterpillar C and let $A_* = \{a_i \in A \mid a_i \neq n - 1\}$. Also, let $a_z \in A_*$ be such that $a_z \geq a_i$ for all $a_i \in A_*$. Then, A is also the centroid sequence of a non-caterpillar tree T if and only if:

(i) $|A_*| \geq 4$; and

(ii) there is a caterpillar C~ whose centroid sequence is A and such that there is a pair of adjacent vertices, say, v_r and v_s, respectively, along the backbone B~ of C~, whose weights a_r and a_s, respectively, in C~ satisfy the following conditions:

a) $a_r, a_s \in A_* \setminus \{a_z\}$; and

b) $|a_r - a_s| > n - a_z$.

Proof: (⇒) Suppose that $|A_*| \geq 4$ and that there is a caterpillar C~ whose centroid sequence is A such that there is a pair of adjacent vertices v_r and v_s along the backbone B~ of C~ whose weights satisfy conditions (ii)a) and (ii)b). Let the pair of adjacent vertices v_r and v_s along B~ be chosen so that $|a_r - a_s| \geq |a_i - a_j|$ for each pair of adjacent vertices v_i (whose weight in C~ is a_i) and v_j (whose weight in C~ is a_j) along B~ such that $a_i, a_j \in A_* \backslash \{a_z\}$. Note that $a_r \neq a_s$, by condition (ii)b), since $n - a_z > 0$. Assume, without loss of generality, that $a_r > a_s$. Then, since C~ is a caterpillar, the number of 1-valent vertices adjacent to vertex v_s in C~ is $a_r - a_s - 1$, by Theorem 2.3.1. Now, since $a_z \geq a_i$ for all $a_i \in A_*$, the vertex v_z whose weight in C~ is a_z must be an endpoint of backbone (i.e., path) B~. Therefore, the number of 1-valent vertices of C~ adjacent to vertex v_z in C~ is $n - a_z - 1$. By condition (ii)b), $a_r - a_s = |a_r - a_s| > n - a_z \Rightarrow a_r - a_s - 1 > n - a_z - 1 \Rightarrow a_r - a_s - 1 \geq n - a_z$, since a_r, a_s, n, and a_z are all positive integers. That is, the number of 1-valent vertices adjacent to vertex v_s in C~ is ≥ the number of 1-valent vertices adjacent to vertex v_z in C~ plus 1. So, we can remove from C~ $(n-a_z)$ of the 1-valent vertices (and the $(n-a_z)$ edges incident with them) that are adjacent to vertex v_s and also remove from C~ the branch T_* that starts at the unique vertex (say, v_x) which is adjacent to v_z along B~ and is determined by vertex v_z, except for vertex v_x itself (i.e., don't remove vertex v_x), and then attach the $(n-a_z)$ 1-valent vertices (with their edges) to vertex v_x and also attach branch T_* (minus vertex v_x) to vertex v_s. These removals and subsequent attachments produce a tree (say, T~) whose centroid sequence A is the same as that of C~. There are now 3 cases to be considered.

(1) If $v_x = v_r$, then T~ is not a caterpillar because, by the adjacency of vertices v_r and v_s along B~ and conditions (i) and (ii)a), there is a vertex v_t, with $v_t \neq v_r$ and $v_t \neq v_z$, which is adjacent to v_s along B~. This proves the theorem in this case.

(2) If $v_x = v_s$, then T~ ≡ C~, so T~ is a caterpillar. Now, by the adjacency of v_s and v_r along B~ and conditions (i) and (ii)a), there is a vertex v_t, with $v_t \neq v_s$ and $v_t \neq v_z$, such that v_t is adjacent to v_r along B~. Since $a_r > a_s$, we have $a_t > a_r$, because $a_t \leq a_r$ implies that a_r is in the centroid of T~, contradicting the fact that $a_r > a_s$. So, $a_z \geq a_t > a_r > a_s$.

If $a_z > a_t$, there is a caterpillar C~~ whose backbone B~~ contains the following ordered sequence of vertices, starting from endpoint v_z of B~~: v_z, v_t, v_s, v_r. Observe that the number of 1-valent vertices that are adjacent to vertex v_s in C~~ is $a_t - a_s - 1 > a_r - a_s - 1 \geq n - a_z$, since $a_t > a_r$. So, we can remove from C~~ n-a_z of the 1-valent vertices (and the (n-a_z) edges incident with them) that are adjacent to vertex v_s and also remove from C~~ branch T** that starts at vertex v_t and is determined by vertex v_z, except for vertex v_t itself (i.e., don't remove vertex v_t), and then attach the (n-a_z) 1-valent vertices (with their edges) to vertex v_t and also attach branch T** (minus vertex v_t) to vertex v_s. These removals and subsequent attachments produce a tree T~~ which is not a caterpillar since vertices v_z, v_t, v_s, and v_r do not all lie on a single path in T~~. And, by the way we constructed T~~, A (the centroid sequence of C~, T~, and C~~) is also the centroid sequence of T~~. Hence, T~~ is the required tree which is not a caterpillar.

If, however, $a_z = a_t$, then we can remove from C~ (n-a_z) of the 1-valent vertices (and the (n-a_z) edges incident with them) that are adjacent to v_s and also remove from C~ the branch T*** that starts at vertex v_r and is determined by vertex v_t, except for vertex v_r itself (i.e., don't remove vertex v_r), and then attach the (n-a_z) 1-valent vertices (with their edges) to vertex v_r and also attach branch T*** (minus vertex v_r) to vertex v_s. The resulting graph is a tree T which is not a caterpillar, since vertices v_z, v_s, v_r, and v_t do not lie on a single path in T. In addition, the centroid sequence A of T is the same as that of C~, by the way we constructed T. Hence, we are done in case (2).

(3) If $v_x \neq v_r$ and $v_x \neq v_s$, then B~ contains one of the following two ordered sequences of vertices, starting from endpoint v_z of B~: either v_z, v_x, v_s, v_r or v_z, v_x, v_r, v_s.

If B~ contains the ordered sequence of vertices v_z, v_x, v_s, v_r, then there are again at least (n-a_z) 1-valent vertices adjacent to vertex v_s in C~. So, we can remove from C~ (n-a_z) 1-valent vertices (and the (n-a_z) edges incident with them) that are adjacent to v_s and also remove from C~ branch T**** that starts at vertex v_x and is determined by vertex v_z, except for vertex v_x itself (i.e., don't remove vertex v_x), and then attach the (n-a_z) 1-valent vertices (with their edges) to vertex v_x and also attach branch T**** (minus vertex v_x) to vertex v_s. This construction produces a tree T~~~ which is not a caterpillar, and we are done in this case.

If, on the other hand, B~ contains the ordered sequence of vertices v_z, v_x, v_r, v_s, then $a_z > a_x > a_r > a_s$. If there is a vertex (say, v_y) other than v_r which is adjacent to v_s along B~, then, in a manner similar to that presented in case (1), case (2), and the second paragraph of case (3) above, vertex v_z can be removed from B~ and attached directly to vertex v_s, producing a non-caterpillar tree T~. If, however, v_s is an endpoint of B~, then a new caterpillar C~~~ can be constructed whose backbone B~~~ contains the following ordered sequence of vertices, starting with endpoint v_z of B~~~: v_z, v_r, v_s, v_x. Hence, C~~~ is a case (1) type caterpillar, and we can construct from C~~~ a non-caterpillar tree T~~~~, just as we did in case (1) above, completing the proof in this direction. ☐

($\Leftarrow$) Conversely, suppose that $A = \{a_1, a_2, ..., a_n\}$ is the centroid sequence of a non-caterpillar tree T. Let v_n be the vertex of T whose weight is a_n, and let v_{n-1} be the vertex of T whose weight is a_{n-1}. Then, v_n is in the centroid of T, and v_{n-1} is adjacent to v_n in T. Thus, (v_n, v_{n-1}) is an edge of T. Let P be a longest path in T which contains edge (v_n, v_{n-1}). Since T is not a caterpillar, and by our choice of P, P contains at least 3 internal (i.e., non-endpoint) vertices. Furthermore, at least one of these ≥ 3 internal vertices of P is at distance ≥ 2 from each endpoint of P and is adjacent to a non-1-valent vertex of T which does not lie on P. Let v_i be an internal vertex of P which is at distance ≥ 2 from each endpoint of P and which is adjacent to a non-1-valent vertex (say, v_j) of T which does not lie on P whose weight a_j (of vertex v_j) in T is $\leq$ the weight of any other non-1-valent vertex of T which does not lie on P and is adjacent to an internal vertex of P which is at distance $\geq$ 2 from each endpoint of P. Note that we can think of the subpath P* of P whose vertex set is the set of all internal vertices of P as a kind of "backbone" of T (although, strictly speaking, the term "backbone" has been defined only for caterpillars). That is, if T were a caterpillar (although, by hypothesis, T is a non-caterpillar), then for each internal vertex v* of P*, the number of 1-valent vertices adjacent to v* in T would be determined by the weight a** in T of a vertex v** of P* which is adjacent to v* along P* and has greatest weight in T, as well as by the weight a* in T of vertex v* itself. Specifically, the number of 1-valent vertices adjacent to v* in T would be $a^{**} - a^{*} - 1$. Now, since T is not a caterpillar, a** determines, instead, the sum of the sizes of all branches of T that start at v* other than the two branches that are determined by the two vertices that are adjacent to v* along P*. So, since vertex v_i is an internal vertex of path P*, there is a vertex (say, v_k) adjacent to v_i along P* whose weight a_k in T, together with the weight a_i in T of vertex v_i, determines the sum of the sizes of all branches that start at v_i other than those two branches determined by the two vertices that are adjacent to v_i along P*. So, the sum of the sizes of these (val(v_i) - 2) branches that start at v_i in T is $a_k - a_i - 1$. (Note that $a_k > a_i$, by our choice of vertex v_k.)

We now consider all possibilities for the size of a_j relative to the sizes of a_i and a_k. If $a_i < a_j < a_k$, then we can place vertex v_j between vertices v_i and v_k along P and all other non-1-valent vertices of T can be placed on P in such a way that the weights of the vertices along the resulting path P~ strictly increase as we move away from vertex v_n along P~, with the sole exception that the weight a_n in T of vertex v_n may be equal to the weight a_{n-1} in T of vertex v_{n-1}. Note that by the minimality of a_j, no vertex of T which did not originally lie on P has a weight in T which is strictly between a_i and a_j. Hence, after each non-1-valent vertex of T is placed along P (to form P~) in the manner mentioned above and labeled with its weight in T, and then each vertex of this newly formed backbone P~ is joined to an appropriate number of 1-valent vertices determined by the labeled weights of these vertices of P~, we obtain a caterpillar C~ in which vertices v_i and v_j are adjacent. Note that the centroid sequence of C~ is the same as that of C and of T (namely, $A = \{a_1, a_2, ..., a_n\}$), since the set of weights of the vertices of C~ that are $< n-1$ is the same as A_*. So, $|A_*| \geq 4$. Also, since $a_k - a_i - 1 \geq n - a_j$, we have $|a_j - a_i| = a_j - a_i \geq n - a_k + 1$, which implies that $|a_j - a_i| > n - a_k \geq n - a_z$, where a_z is the weight in C~ of the vertex of greatest weight which lies along the backbone of C~. Additionally, note that $a_i < a_j < a_k$, which implies that

a_i, $a_j \in A_* \backslash \{a_z\}$. So, in case $a_i < a_j < a_k$, the theorem is proved, since caterpillar C~ satisfies conditions (i) and (ii).

Observe that it is impossible that $a_j = a_i$, since $a_j = a_i$ implies that edge (v_j, v_i) is the centroid of T, contradicting our choice of path P. It is also impossible that $a_j < a_i$, because either vertex v_i is in the centroid of T, in which case $a_i \leq a_x$ for all $a_x \in A$, or vertex v_i is not in the centroid of T, in which case there is a vertex (say, v_m, where $v_m \neq v_k$) which is adjacent to v_i along P such that $a_m < a_i$ (where a_m is the weight of vertex v_m in T), since the unique branch of T which starts at v_i and contains the centroid of T is the only branch starting at v_i which is determined by a vertex whose weight in T is strictly lower than that of v_i. Thus, $a_j \geq a_i$, contradicting the hypothesis that $a_j < a_i$.

If $a_j = a_k$, then no non-1-valent vertex of T which did not originally lie on P has a weight which is strictly between a_i and a_k, by the minimality of a_j. Therefore, $|a_k - a_i| = a_k - a_i = a_j - a_i > n - a_k \geq n - a_z$. So, if we place all of the non-1-valent vertices of T, labeled with their weights in T, on P in such a way that the weights of the vertices along this new path P~~ strictly increase as we move away from v_n along P~~ (with the one exception that it is possible that $a_{n-1} = a_n$), and then join each vertex of P~~ to an appropriate number of 1-valent vertices determined by the labeled weights of these vertices of P~~, we obtain a caterpillar C~~ whose centroid sequence is A . And, C~~ satisfies conditions (i) and (ii), since vertices v_i and v_k are adjacent along P~~ and since $a_k \leq a_z$ implies that a_i, $a_k \in A_* \backslash \{a_z\}$ (because we can take $j = z$ if

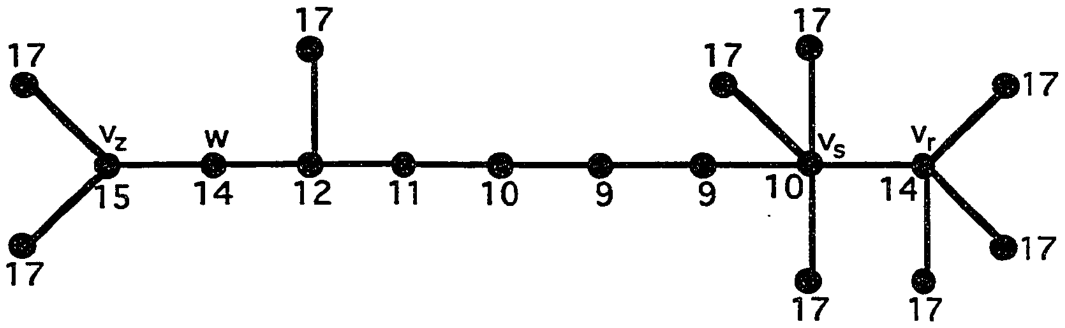
$a_k = a_z$), and we are done. Finally, if $a_j > a_k$, again no non-1-valent vertex of T which did not originally lie on P has a weight in T which is strictly between a_i and a_k, by the minimality of a_j. Hence, $|a_k - a_i| = a_k - a_i > n - a_j \geq n - a_z$. Thus, as before, we can construct a caterpillar C~~~ whose centroid sequence is A which satisfies conditions (i) and (ii), since v_i and v_k are adjacent along P~~~ and since $a_j > a_k$ implies that $a_i, a_k \in A_*\backslash\{a_z\}$.

Therefore, since there is a caterpillar whose centroid sequence is A such that this caterpillar satisfies conditions (i) and (ii), regardless of the size of a_j relative to the sizes of a_i and a_k, the theorem is proved. □

Example 2.3.5: Theorem 2.3.4 may be applied as follows. Let $A = \{17, 17, 17, 17, 17, 17, 17, 17, 17, 15, 14, 14, 12, 11, 10, 10, 9, 9\}$. Then, A is realizable as the centroid sequence of a caterpillar, by Theorem 2.3.2. Also, $A_* = \{15, 14, 14, 12, 11, 10, 10, 9, 9\}$, so $|A_*| = 9 \geq 4$ and $a_z = 15$. Observe that A is the centroid sequence of the caterpillar C~ of Figure 2.3.5(a). Note that vertices v_r and v_s are adjacent along the backbone B~ of C~, $a_r = 14$ and $a_s = 10$. So, $a_r, a_s \in A_*\backslash\{a_z\}$ and $|a_r - a_s| = |14 - 10| = 4 > 3 = 18 - 15 = n - a_z$. Hence, by Theorem 2.3.4, there is a tree T which is not a caterpillar such that A is also the centroid sequence of T. Note that T is obtained from C~ by removing the branch B_w (except for vertex w) of C~ which starts at vertex w and is determined by vertex v_z (whose weight in C~ is $a_z = 15$) and removing 3 of the 1-valent vertices that are adjacent to vertex v_s

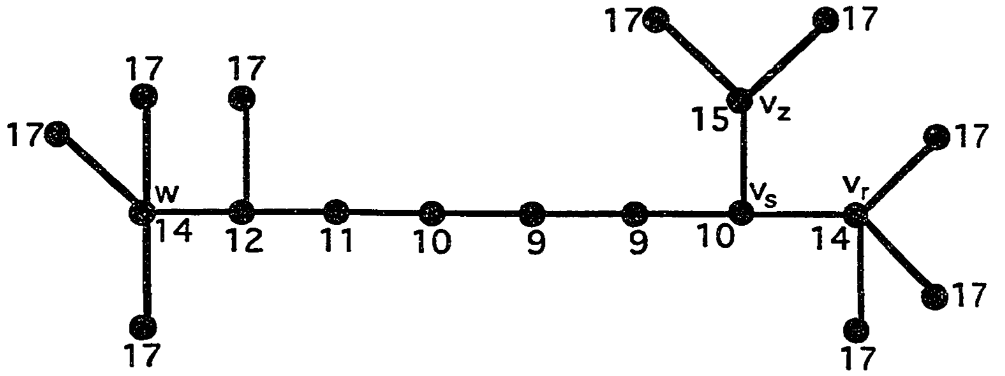
in C~, and then attaching branch B_w (minus vertex w) to vertex v_s and attaching the three 1-valent vertices that were removed to vertex w. See Figure 2.3.5(b).

Figure 2.3.5: In (a), the weight in C~ of each vertex of caterpillar C~ is indicated next to that vertex. In (b), the weight in T of each vertex of tree T is indicated next to that vertex.



Caterpillar C~

(a)



Tree T

(b)

We have just looked at several results concerning the centroid sequence of a caterpillar. What can be said about the centroid sequence of a tree? We will look into the issue of characterizing the centroid sequence of a tree in Section 2.4.

2.4 A Characterization of the Centroid Sequence of a Tree

The task of characterizing the centroid sequence of a tree is more difficult than that of characterizing the centroid sequence of a caterpillar. To facilitate the characterization of the centroid sequence of a tree, we introduce the following definitions.

Definition 2.4.1 : Let v be a vertex of a tree T on $n \geq 2$ vertices, and let $a(v)$ = the weight of v in T. We define the **complementary weight** of v in T, denoted $b(v)$, by the following equation:

$$b(v) = n - a(v).$$

Remark 2.4.1: Observe that $b(v)$ may be interpreted geometrically as the sum of the sizes of all branches of T that start at v except for one specific branch B of greatest size which starts at v, plus 1.

Proof: $b(v) = n - a(v)$
$= (n - 1) - a(v) + 1$
$= \text{size}(T) - \text{size}(B) + 1.$ $\square$

Definition 2.4.2: Let T be a tree on $n \geq 2$ vertices, and let $A = \{a_1, a_2, \ldots, a_n\}$ be the centroid sequence of T. We define the **complementary sequence** of T (and of A) to be the sequence $B = \{b_1, b_2, \ldots, b_n\}$, where $b_i = n - a_i$ for each $i = 1, 2, \ldots, n$.

Example 2.4.1: Suppose T is a tree whose centroid sequence is $A = \{8, 8, 8, 8, 7, 7, 7, 7, 2\}$ (see Figure 3.2.2). Since A contains 9 elements, $n = 9$. Hence, the complementary sequence of T (and of A) is $B = \{1, 1, 1, 1, 2, 2, 2, 2, 7\}$.

Remark 2.4.2: Since the centroid sequence A of a tree T is arranged in non-increasing order, the complementary sequence B of T (and of A) is arranged in non-decreasing order.

We now develop some new vocabulary that will be used in the statement and proof of Theorem 2.4.1.

Definition 2.4.3: Let $B = \{b_1, b_2, \ldots, b_n\}$ be a sequence of positive integers arranged in non-decreasing order (i.e., $b_1 \leq b_2 \leq \cdots \leq b_n$), and let $b_i \in B$. We say that b_i is **satisfiable** if there exists a (possibly empty) subsequence B_i of $B \backslash \{b_{n-1}, b_n\}$ such that $\sum_{b_k \in B_i} b_k = b_i - 1$. We call B_i the **satisfaction subsequence** for b_i, and the elements of B_i are said to **satisfy** b_i. The element $b_i \in B$ is said to be **satisfied** when the elements of B_i have all been crossed out from B.

Example 2.4.2: Suppose $B = \{1, 1, 1, 1, 2, 3, 4\}$. Observe that $3 \in B$ is satisfiable because, for example, the sum of the elements of the subsequence $\{1, 1\}$ of B is $1 + 1 = 2 = 3 - 1$. In this case, $\{1, 1\}$ is the satisfaction subsequence for 3 and the elements 1 and 1 satisfy 3. 3 is satisfied when the two elements of $\{1, 1\}$ have been crossed out from B, as follows: $B = \{1, \cancel{1}, \cancel{1}, 1, 2, 3, 4\}$.

Definition 2.4.4 : Let $B = \{b_1, b_2, \ldots, b_n\}$ be a sequence of positive integers arranged in non-decreasing order (i.e., $b_1 \leq b_2 \leq \cdots \leq b_n$). We call B a **satisfiable sequence** if:

(i) each element $b_i \in B$ can be satisfied in such a way that once an element b_k of the satisfaction subsequence $B_i \subset B$ for b_i has been crossed out, b_k cannot be used again to satisfy any element $b_j \in B$ with $j \neq i$; and

(ii) $\bigcup_{i=1}^{n} B_i = B \backslash \{b_{n-1}, b_n\}$.

Example 2.4.3: Let's look once again at the sequence $B = \{1, 1, 1, 1, 2, 3, 4\}$ from Example 2.4.2 above. Note that $|B| = 7$ and $3 + 4 = 7$. Additionally, each of the 1's in B is automatically satisfied and requires no crossing out of other elements of B. We can cross out the first 1 for the 2, the second and third 1's for the

3, and the fourth 1 and the 2 for the 4. That is, each element of B can be satisfied in such a way that each element of B\{3, 4} which is crossed out along the way is not used again and such that the only non-crossed out elements of B are the 3 and the 4. Therefore, B is a satisfiable sequence.

Before characterizing the centroid sequence of a tree in Theorem 2.4.1, we first state and prove the following two lemmas.

Lemma 2.4.1: Let T be a tree on $n \geq 2$ vertices, and let v and w be distinct, adjacent vertices of T such that $a(v) \leq a(w)$. Then, the unique branch of T emanating from v determined by w has size $(n - a(w))$.

Proof: First, suppose v is not in the centroid of T. Since w is adjacent to v and $a(v) \leq a(w)$, we must, in fact, have $a(v) < a(w)$, by Lemma 2.3.1. So, the branch B_{vw} emanating from v determined by w is a not a branch of greatest size which starts at v in T. Hence, $\text{size}(B_{vw}) < a(v)$. Now, since $a(w) > a(v)$, branch B_{wv} which emanates from w and is determined by v is the unique branch of greatest size which starts at w in T, and thus, $\text{size}(B_{wv}) = a(w)$. Hence, the sum of the sizes of all branches of T other than B_{wv} which emanate from w is $(n-1) - a(w)$. But, $\text{size}(B_{vw})$ = (the sum of the sizes of all branches of T other than B_{wv} which emanate from w) + (length of edge (w, v)) = $(n-1) - a(w) + 1 = n - a(w)$.

Suppose, now, that v is a vertex in the centroid of T. Then, for each vertex w which is adjacent to v in T, $a(v) \leq a(w)$. If w lies on a branch B_{vw} emanating from v which is not of greatest size, then $a(w) > a(v)$. In this case, the proof proceeds exactly as it did in case v was not in the centroid of T, and, again, $\text{size}(B_{vw}) = n - a(w)$. However, if w determines a branch B~ of greatest size which emanates from v (note that there may be more than one such branch B~ of greatest size emanating from v in T since v is in the centroid of T), then the value of a(w) must be the same as that of a_{n-1} (i..e., $a(w) = a_{n-1}$ as positive integers) because the branch B# of greatest size which emanates from w must be determined by v and, hence, $\text{size}(B\#) = a(w) = (n - 1) - (a(v) - 1) = n - a(v) = n - a_n = a_{n-1}$ (note that $a(v) = a_n$ as positive integers since v is in the centroid of T). So, $\text{size}(B\sim) = a(v) = a_n = n - a_{n-1} = n - a(w)$, and we are done. □

Lemma 2.4.2: Let T be a tree on $n \geq 2$ vertices, and let $v \in V(T)$. Let $W = \{w_1, w_2, ..., w_k\}$ be the set of all vertices that are adjacent to v in T, where $k = \text{val}(v)$, and let $w_i \in W$ be any vertex of W such that $a(w_i) \leq a(w_j)$ for all $j = 1, 2, ..., k$. Then,

$$\sum_{w_j \in W \setminus \{w_i\}} a(w_j) - a(v) - 1 = n \cdot (\text{val}(v) - 2).$$

Proof: Note that if v is a 1-valent vertex of T, then $|W| = 1$, which implies that $W \setminus \{w_i\} = \emptyset$. So, in this case, the statement of the theorem requires that $0 - (n-1) - 1 = n \cdot (1-2)$. That is, $-n + 1 - 1 = -n = n \cdot (-1)$, so the statement is true for each 1-valent vertex of

T. Now, suppose $\mathrm{val}(v) \geq 2$. Let $W = \{w_1, w_2, \ldots, w_k\}$ be the set of all vertices of T that are adjacent to v in T. Let $w_i \in W$ be any vertex of W such that $a(w_i) \leq a(w_j)$ for all $j = 1, 2, \ldots, k$. Then, we know that $a(v) < a(w_j)$ for each $w_j \in W \backslash \{w_i\}$, since each vertex of T that is adjacent to v, with the possible exception of w_i, must have strictly higher weight in T than v. So, by Lemma 2.4.1, for each vertex $w_j \in W \backslash \{w_i\}$, the size of the unique branch B~ of T which starts at v and is determined by w_j is $(n - w_j)$ [i.e., size(B~) = $n - w_j$]. Now, the branch B of T which starts at v and is determined by w_i is a branch of greatest size which starts at v, so size(B) = a(v). Hence, since size(T) = n - 1, we have

$$\sum_{w_j \in W \backslash \{w_i\}} (n - a(w_j)) + a(v) = n - 1.$$

Noting that $|W| = k = \mathrm{val}(v)$, we have

$$n \cdot (\mathrm{val}(v) - 1) - \sum_{w_j \in W \backslash \{w_i\}} a(w_j) + a(v) = n - 1. \text{ Thus,}$$

$$\sum_{w_j \in W \backslash \{w_i\}} a(w_j) - a(v) - 1 = n \cdot (\mathrm{val}(v) - 2). \qquad \square$$

We now have all the tools that we need in order to characterize the centroid sequence of a tree in Theorem 2.4.1 below.

Theorem 2.4.1 (Characterization of the Centroid Sequence of a Tree):

Let $A = \{a_1, a_2, \ldots, a_n\}$ $(n \geq 2)$ be a sequence of positive integers arranged in non-increasing order (i.e., $a_1 \geq a_2 \geq \cdots \geq a_n$), and let

$B = \{b_1, b_2, \ldots, b_n\}$ be the complementary sequence of A. Then, A is realizable as the centroid sequence of a tree T if and only if the following two conditions are satisfied:

(i) $b_{n-1} + b_n = n$; and

(ii) B is a satisfiable sequence.

Proof: (⇒) Suppose that $A = \{a_1, a_2, \ldots, a_n\}$ $(n \geq 2)$ is realizable as the centroid sequence of a tree T. Then, $|A| = n = |V(T)|$. Since A is the centroid sequence of T, $a_{n-1} + a_n = n$. So, $b_{n-1} + b_n = (n - a_{n-1}) + (n - a_n) = 2n - (a_{n-1} + a_n) = 2n - n = n$, proving condition (i). To prove condition (ii), let a_i be the weight of vertex v_i of tree T for each $i = 1, 2, \ldots, n$. Let $v_i \in V(T)$, and let $W\backslash\{w_i\}$ be the set of all vertices that are adjacent to v_i in T, with the exception of any one vertex w_i of lowest weight that is adjacent to v_i in T. Then, by Lemma 2.4.2, we know that

$$\sum_{w \in W\backslash\{w_i\}} a(w) - a_i - 1 = n\cdot(\mathrm{val}(v_i) - 2),$$

where a(w) is the weight of vertex w in T. Solving this equation for $-a_i$, we find that

$$\begin{aligned} -a_i &= n\cdot(\mathrm{val}(v_i) - 2) - \sum_{w \in W\backslash\{w_i\}} a(w) + 1 \\ &= n\cdot(\mathrm{val}(v_i) - 1) - \sum_{w \in W\backslash\{w_i\}} a(w) + 1 - n \\ &= \sum_{w \in W\backslash\{w_i\}} (n - a(w)) + 1 - n \\ &= \sum_{w \in W\backslash\{w_i\}} b(w) + 1 - n \end{aligned}$$

(where $b(w) = n - a(w)$ is the complementary weight of vertex w in T).

So, $n - a_i = \sum_{w \in W\setminus\{w_i\}} b(w) + 1 \Rightarrow b_i = \sum_{w \in W\setminus\{w_i\}} b(w) + 1$

or

$$(*) \qquad \sum_{w \in W\setminus\{w_i\}} b(w) = b_i - 1$$

(where b_i is the complementary weight of vertex v_i in T). That is, the sum of the complementary weights of the vertices of $W\setminus\{w_i\}$ is 1 less than the complementary weight of vertex v_i. Since $v_i \neq v_j$ implies that $W\setminus\{w_i\} \cap W\setminus\{w_j\} = \emptyset$, for each vertex $v_i \in V(T)$, the complementary weight $b_i \in B$ of vertex v_i can be associated with a satisfaction subsequence B_i of B, which is determined by the set of vertices $W\setminus\{w_i\}$ adjacent to v_i in T, with the exception of any one vertex w_i of lowest weight which is adjacent to v_i in T, in such a way that $b_k \in B_i$. Thus, $b_k \notin B_j$ if $j \neq i$ (although an element of B which represents the same positive integer as that represented by b_k may be an element of B_j). Thus, if $B = \{b_1, b_2, ..., b_n\}$ is the complementary sequence of A, then each $b_i \in B$ (corresponding to vertex v_i of T) can be satisfied by crossing out from B the elements of the subsequence B_i of $B\setminus\{b_{n-1}, b_n\}$ determined by the vertices of $W\setminus\{w_i\}$ that are adjacent to v_i in T. (Note that if $b_i = 1$, then v_i is a 1-valent vertex of T, so $W\setminus\{w_i\} = \emptyset$.) So, we can rewrite equation (*) as follows:

$$(**) \qquad \sum_{b_k \in B_i} b_k = b_i - 1.$$

And, since $\bigcup_{i=1}^{n} W\setminus\{w_i\} = V(T)\setminus\{v_{n-1}, v_n\}$, we have $\bigcup_{i=1}^{n} B_i = B\setminus\{b_{n-1}, b_n\}$

(note that the symbol $\cup$ means "disjoint union" here). Hence, B is a satisfiable sequence, proving condition (ii), and completing the proof in this direction. $\square$

($\Leftarrow$) Conversely, suppose that conditions (i) and (ii) hold. By condition (i), we have $n = b_{n-1} + b_n = (n-a_{n-1}) + (n-a_n) = 2n - (a_{n-1} + a_n)$. That is, $a_{n-1} + a_n = n$, which is a necessary condition for A to be the centroid sequence of a tree on n vertices. By condition (ii), each element $b_i \in B$ can be satisfied in such a way that once an element b_k of the satisfaction subsequence B_i of $B\backslash\{b_{n-1}, b_n\}$ for b_i has been crossed out, b_k cannot be used again to satisfy any element $b_j \in B$ with $j \neq i$, and

$\bigcup_{i=1}^{n} B_i = B\backslash\{b_{n-1}, b_n\}$. So, for each $b_i \in B$,

$$\sum_{b_k \in B_i} b_k = b_i - 1 \quad \Rightarrow \quad \sum_{a_k \in A_i} (n - a_k) = n - a_i - 1$$

[where $A_i \subset A$ is defined as $A_i = \{a_k = (n-b_k) \mid b_k \in B_i\}$]

$$\Rightarrow \quad n \cdot |A_i| - \sum_{a_k \in A_i} a_k = n - a_i - 1 \quad \Rightarrow$$

$$(*) \quad \sum_{a_k \in A_i} a_k - a_i - 1 = n \cdot (|A_i| - 1).$$ Therefore, for each $a_i \in A$,

there is a subsequence A_i of $A\backslash\{a_{n-1}, a_n\}$ whose elements satisfy equation (*) above and such that $a_k \in A_i$ implies that $a_k \notin A_j$ if $j \neq i$ (although the positive integer represented by a_k may appear in both A_i and A_j). In addition,

$\bigcup_{i=1}^{n} A_i = A\backslash\{a_{n-1}, a_n\}$ (because of the corresponding property of the union of the B_i's). Now, if we draw n vertices, label each vertex v_i ($i = 1,2,\dots,n$) with a distinct element a_i of A, and join a pair of these labeled vertices (labeled, say, a_i and a_j, where $i \neq j$) with an edge if and only if either: (i) $a_i \in A_j$, or (ii) $a_j \in A_i$, or (iii) $\{a_i, a_j\} = \{a_{n-1}, a_n\}$, we obtain a tree T on n vertices.

Claim: $A = \{a_1, a_2, \dots, a_n\}$ is the centroid sequence of the tree T constructed above.

Proof of Claim: Note that for each vertex v_i of T (whose label is a_i), equation (*) above holds. And, each $a_k \in A_i$ is the label of a vertex v_k which is adjacent to vertex v_i in T. In fact, by our construction of T, $|A_i| = \mathrm{val}(v_i) - 1$, so for each $a_i \in A$, equation (*) can be rewritten as

$$(**) \quad \sum_{a_k \in A_i} a_k - a_i - 1 = n\cdot(\mathrm{val}(v_i) - 2),$$ which, by Lemma 2.4.2, is

a necessary condition for A to be the centroid sequence of T. To show that A is, indeed, the centroid sequence of T, we will show that for each vertex v_i of T, the label a_i assigned to v_i in the construction of T is the weight of v_i in T. Note that since $n \geq 2$, T has at least two 1-valent vertices. For each 1-valent vertex v_i (whose label is a_i) of T, $A_i = \emptyset$ and $a_i = n - 1$ (by the way we constructed T and by equation (**)). Thus, for each 1-valent vertex v_i of T, the label $a_i = n - 1$ of v_i coincides with the weight of v_i in T. Now, by pruning all the 1-valent vertices from T, we obtain a new (possibly degenerate) tree T~. If T~ consists of just a single vertex (say, v_j), then T is a star and the weight of v_j in T is 1. Also, by (**),

$$a_j = \sum_{a_k \in A_i} a_k - n\cdot(\mathrm{val}(v_j) - 2) - 1 = (n-1)(n-2) - n\cdot(n - 1 - 2) -1$$

$= n^2 - 3n + 2 - n^2 + 3n -1 = 1$. That is, $a_j = 1$, so a_j coincides with the weight of vertex v_j in case T is a star, and we are done. Otherwise, T~ is a non-degenerate tree, so T~ has at least two 1-valent vertices. Let v_r be a 1-valent vertex of T~ whose label a_r is $\geq$ the label of any other 1-valent vertex of T~. Then, v_r is adjacent to $\mathrm{val}(v_r) - 1$ 1-valent vertices in T. Thus, the weight of v_r in T is $(n - 1) - (\mathrm{val}(v_r) - 1) = n - \mathrm{val}(v_r)$. Also, by (**),

$$a_r = \sum_{a_k \in A_i} a_k - n\cdot(\mathrm{val}(v_r) - 2) - 1 =$$

$(n-1)(\mathrm{val}(v_r) - 1) - n\cdot(\mathrm{val}(v_r) - 2) - 1 = n - \mathrm{val}(v_r)$. That is, $a_r = n - \mathrm{val}(v_r)$, so a_r coincides with the weight of vertex v_r in T. Next, prune vertex v_r from T~ to obtain another (possibly degenerate) tree T~~. As before, if T~~ is a degenerate tree, it can be shown that the weight in T of the sole vertex (say, v_s) of T~~ coincides with the label a_s assigned to vertex v_s in T. And, if $|T{\sim\sim}| \geq 2$, T~~ has at least two 1-valent vertices. As we did above, we can show, using Lemma 2.4.2 and equation (**), that if v_t is a 1-valent vertex of T~~ such that a_t, the label assigned to vertex v_t in T, is $\geq$ the label assigned to any other 1-valent vertex of T~~, then a_t = the weight of vertex v_t in T (because we already know that the label a_k of each vertex v_k such that $a_k \in A_t$ coincides with the weight of vertex v_k in T, since v_k is a 1-valent vertex of either T or T~). By iterating this procedure, we can verify that for each vertex v_i ($i = 1, 2, \ldots, n$) of T, the label a_i assigned to v_i

coincides with the weight of v_i in T. Hence, A is realizable as the centroid sequence of tree T, and the proof is complete. ☐

Theorem 2.4.1 gives rise to the following algorithm.

Algorithm 2.4.1: (Determining Whether a Given Finite Sequence of Positive Integers is Realizable as the Centroid Sequence of a Tree)

1) Given a finite sequence $A = \{a_1, a_2, ..., a_n\}$ $(n \geq 2)$ of positive integers, construct the sequence $B = \{b_1, b_2, ..., b_n\}$, where $b_i = n - a_i$ for each $i = 1, 2, ..., n$.

2) If $b_{n-1} + b_n \neq n$, stop here because A is not the centroid sequence of a tree. Otherwise, $b_{n-1} + b_n = n$, and we move on to step 3) below.

3) For each element $b_i \in B$ with $b_i > 1$, locate a subset B_i of elements of $B \setminus \{b_{n-1}, b_n\}$ such that the sum of the elements of B_i is $(b_i - 1)$.

4) After locating subset B_i described in step 3) above, cross out the elements of B_i from B so that these elements will not be used again.

5) If steps 3) and 4) above can be carried out successfully for each $b_i \in B$ with $b_i > 1$ in such a way that at the end of this procedure, b_{n-1} and b_n are the only elements of B that have not been crossed out, then A is realizable as the centroid sequence of a tree. If, on the other hand, steps 3) and 4) above cannot be carried out successfully for each $b_i \in B$, or if steps 3) and 4) cannot be carried out in such a way that all the elements of

$B\backslash\{b_{n-1}, b_n\}$ will be crossed out by the end of the procedure, then A is not realizable as the centroid sequence of a tree.

Example 2.4.4: Is A = {7, 7, 7, 6, 6, 6, 4, 4} realizable as the centroid sequence of a tree? Here, B = {1, 1, 1, 2, 2, 2, 4, 4}. Note that although 4 + 4 = 8 (= n), there is no way to cross out the first six entries of B such that each entry of B is satisfied. Thus, A is not realizable as the centroid sequence of a tree.

Example 2.4.5: Is the sequence A = {9, 9, 9, 9, 9, 8, 7, 7 ,7, 3} realizable as the centroid sequence of a tree? Observe, first of all, that A is not realizable as the centroid sequence of a caterpillar since the number "7" appears three times in A (see Theorem 2.3.1). To determine whether A can be realized as the centroid sequence of a tree, we first consider the complementary sequence B = {1, 1, 1, 1, 1, 2, 3, 3, 3, 7} of A, obtained by subtracting each entry of A from 10 (because 10 is the value of n here). Since we can cross out the first 1 to satisfy the 2, then cross out the second and third 1's to satisfy the first 3, then cross out the fourth and fifth 1's to satisfy the second 3, then cross out the 2 to satisfy the third 3, then cross out the first and second 3's to satisfy the 7, and since the last two entries of B add up to n = 10 and they are the only non-crossed out entries of B at the end of this "crossing out procedure", we conclude that A is, indeed, realizable as the centroid sequence of a tree.

Remark 2.4.3: The problem of finding a satisfaction subsequence B_i of $B\setminus\{b_{n-1}, b_n\}$ for each $b_i \in B$ such that $B_i \cap B_j = \emptyset$ if $i \neq j$ and

$$\bigcup_{i=1}^{n} B_i = B\setminus\{b_{n-1}, b_n\}$$ is a variation of the so-called "bin packing" problem [5]. Observe that for the sequence $B = \{1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 8, 10, 11\}$, the "best fit" and "worst fit" bin packing algorithms, working from left to right in B, both fail to show that B is a satisfiable sequence, even though B is, indeed, a satisfiable sequence. Also, for the sequence $B = \{1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 9, 10, 11\}$, the "best fit" and "worst fit" bin packing algorithms, working from right to left in B, both fail to show that B is a satisfiable sequence, even though B is, in fact, a satisfiable sequence. Hence, it is not clear just how difficult it is to decide, in general, whether a given finite sequence of positive integers is a satisfiable sequence. We shall leave the determination of the computational complexity of Algorithm 2.4.1 as an open problem (see Section 4.5).

Given a finite sequence A of positive integers that is known to be realizable as the centroid sequence of a tree, can we actually construct a tree T whose centroid sequence is A? Algorithm 2.4.2 below shows that, in fact, we can.

Algorithm 2.4.2: (Construction of a Tree From A Finite Sequence of $n \geq 2$ Positive Integers which is Known to be Realizable as the Centroid Sequence of a Tree)

(i) Let $A = \{a_1, a_2, \ldots, a_n\}$ $(n \geq 2)$ be the centroid sequence of a tree. Then, the complementary sequence $B = \{b_1, b_2, \ldots, b_n\}$ of A is a satisfiable sequence. So, for each $b_i \in B$, there is a satisfaction subsequence B_i of $B \backslash \{b_{n-1}, b_n\}$ such that $i \neq j$ implies that

$$B_i \cap B_j = \emptyset \quad \text{and} \quad \bigcup_{i=1}^{n} B_i = B \backslash \{b_{n-1}, b_n\}.$$

(ii) To construct a tree T on n vertices whose centroid sequence is A, first draw a pair of vertices and join them with an edge. Label one of these vertices b_n and the other b_{n-1}. Let B_n be the satisfaction subsequence for b_n and let B_{n-1} be the satisfaction subsequence for b_{n-1}. Draw $|B_n|$ vertices and join each of them via a single edge to the vertex labeled b_n. Label each of these $|B_n|$ new vertices with an element of B_n, using each element of B_n as a label once and only once. Then, draw $|B_{n-1}|$ vertices and join each of them via a single edge to the vertex labeled b_{n-1}. Label each of these $|B_{n-1}|$ new vertices with an element of B_{n-1}, using each element of B_{n-1} as a label once and only once.

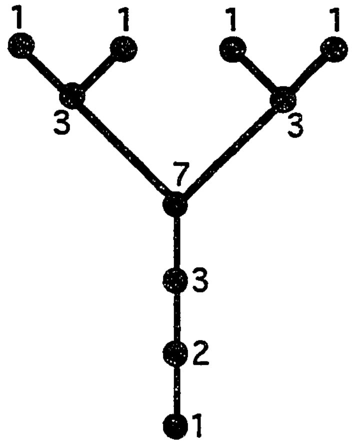
(iii) Iterate this procedure for each vertex that is drawn. That is, if v is a vertex that has already been drawn and the label given to v is b_i, draw $|B_i|$ vertices and join each of them via a single edge to vertex v. Label each of these $|B_i|$ new vertices with an element of B_i, using each element of B_i as a label once and only once.

(iv) The graph which results after completing steps (ii) and (iii) above is a tree T on n vertices such that for each vertex v of T, the label assigned to that vertex is the complementary weight b(v) of v in T. This implies that B is the complementary sequence of T. Thus, A is the centroid sequence of T.

Example 2.4.6: We can apply Algorithm 2.4.2 to construct the tree T of Figure 2.4.1 from the sequence A = {9, 9, 9, 9, 9, 8, 7, 7 ,7, 3}, which was shown to be realizable as the centroid sequence of a tree in Example 2.4.5 above. Recall that the complementary sequence of A is B = {1, 1, 1, 1, 1, 2, 3, 3, 3, 7}.

We start the construction of T by drawing two vertices and joining them with an edge. Label one of these vertices "7" and the other "3". Then, draw two new vertices next to the vertex labeled "7" and join each new vertex via an edge to the vertex labeled "7". Label each of these two new vertices "3". Now, draw a single vertex next to the first vertex that was labeled "3" and join this new vertex via an edge to the first vertex that was labeled "3". Label the new vertex "2". Then, draw a single vertex next to the vertex labeled "2" and join the new vertex via an edge to the vertex labeled "2". Label this new vertex "1". Now, draw a pair of vertices next to one of the vertices labeled "3" in the fourth sentence of this paragraph, and draw another pair of vertices next to the other vertex labeled "3" in the fourth sentence of this paragraph. Join each of the first pair of new vertices via an edge to the vertex labeled "3" that each of these two new vertices has been drawn next to. Label each of these two new vertices "1". Similarly, join each of the second pair of new vertices via an edge to the vertex labeled "3" that each of these two new vertices has been drawn next to. Label each of these two new vertices "1", as well. The resulting graph is the tree T of Figure 2.4.1. Observe that each vertex of T has been labeled with its complementary weight in T. Hence, B is the complementary sequence of T, which implies that A is the centroid sequence of T.

Figure 2.4.1: In the tree below, the complementary weight of each vertex is indicated next to that vertex.



Tree T

In the next section, we continue our study of the centroid sequence of a tree by considering its frequency distribution.

2.5 The Frequency Distribution of the Centroid Sequence of a Tree

In this section, we will relate the concept of a **partition** of a positive integer n to the **frequency distribution** of the centroid sequence of a tree. Specifically, we will see which partitions of a positive integer n can be realized as the frequency distribution of the centroid sequence of a tree T on n vertices. To do this, we start with the following definitions.

Definition 2.5.1: Let n be a positive integer. A **partition** P of n is an (unordered) collection of positive integers whose sum is n. (Note that a positive integer k appearing in a partition P of n may occur more than once in P.)

Example 2.5.1: The partitions of 5 are: {5}, {4, 1}, {3, 2}, {3, 1, 1}, {2, 2, 1}, {2, 1, 1, 1}, and {1, 1, 1, 1, 1}.

Definition 2.5.2: Let $K = \{k_1, k_2, ..., k_n\}$ $(n \geq 1)$ be a sequence of n not-necessarily distinct positive integers. Then, the **frequency distribution of K, denoted f(K)**, is defined to be the unique unordered collection of positive integers that satisfies the following two conditions:

(i) $y \in f(K)$ if and only if there exists a positive integer $x \in K$ which occurs exactly y times in K;

and

(ii) if $y_1, y_2 \in f(K)$ such that $y_1 = y_2$, then there must be distinct positive integers x_1 and x_2 each of which occurs exactly y_1 times in K.

Example 2.5.2: Suppose $K = \{1, 2, 1, 3, 2, 2, 4\}$. Then, $f(K) = \{2, 3, 1, 1\}$ because "1" occurs twice in K, "2" occurs three times in K, "3" occurs once in K, and "4" occurs once in K. Note that since f(K) is, by definition, an unordered collection of numbers, we could, for example, have counted the number of 3's of K first, and then the number of 1's of K, etc., when we were computing f(K).

With these definitions in place, we can now specify which partitions of a positive integer $n \geq 3$ arise as the frequency distribution of the centroid sequence of a tree on n vertices.

Theorem 2.5.1: Let $n \geq 3$ be a positive integer, and let $P = \{p_1, p_2, ..., p_k\}$ $(k \geq 1)$ be a partition of n, where the elements of P have been ordered so that $p_1 \geq p_2 \geq \cdots \geq p_k$. Then, P is realizable as the frequency distribution of the centroid sequence $A = \{a_1, a_2, ..., a_n\}$ of a tree T (i.,e., $P = f(A)$) if and only if:

(i) $2 \leq |P| = k \leq [(n+1)/2]$, where $[(n+1)/2]$ is the greatest integer which is $\leq (n+1)/2$; and

(ii) <u>either</u> $1 \in P$ <u>or</u> $1 \notin P$ but $2 \in P$ and the sequence $B\sim = \{b\sim_1, b\sim_2, \ldots, b\sim_n\}$ which consists of p_1 1's, p_2 2's, p_3 3's, ..., p_{k-1} (k-1)'s, and $p_k = 2$ (n/2)'s is a satisfiable sequence.

<u>Proof</u>: (⇒) Suppose that P is the frequency distribution of the centroid sequence $A = \{a_1, a_2, \ldots, a_n\}$ of a tree T (i.e., $P = f(A)$). Then, $|V(T)| = n$. By the theorem of Jordan (Theorem 2.1.1), the centroid of T consists of either a single vertex or a pair of adjacent vertices.

If the centroid of T consists of a single vertex v, then the weight $a(v) = a_{n-1}$ of v in T is < the weight of any other vertex of T. This implies that the positive integer $1 \in P$, proving condition (ii). Now, since $n \geq 3$, by hypothesis, A contains at least 2 elements besides a_{n-1}, each of which is $> a_{n-1}$. Thus,

(1) $$|P| \geq 2.$$

Also, since the centroid of T consists of a single vertex v, we have $1 \leq a_n \leq [(n-1)/2]$. Therefore, $[(n+2)/2] \leq a_{n-1} \leq n-1$, which implies that

(2) $$|P| \leq [(n+1)/2],$$

where $[(n+1)/2]$ is the number of positive integers from $[(n+2)/2]$ to $(n-1)$, inclusive, plus 1 (where the "1" counts the positive integer represented by a_n). Combining inequalities (1) and (2), we have

(3) $$2 \leq |P| \leq [(n+1)/2],$$

proving condition (i), and finishing the proof in case the centroid of T consists of a single vertex.

If the centroid of T consists of a pair of adjacent vertices v_1 and v_2, then $a_{n-1} = a_n < a_i$ for each $a_i \in A\backslash\{a_{n-1}, a_n\}$. (Note that $A\backslash\{a_{n-1},a_n\}$ is not empty since $n \geq 3$.) Hence, the positive integer $2 \in P$. Also,

(4) $$|P| \geq 2.$$

Observe that, in this case, n must be even, since $a_{n-1} = a_n$ and $a_{n-1} + a_n = n$. So, $a_n = n/2 = a_{n-1}$. Therefore,

(5) $$|P| \leq n/2 = [(n+1)/2]$$

(since n is even), where n/2 is the number of positive integers from n/2 to (n-1), inclusive. Combining inequalities (4) and (5), we obtain

(6) $$2 \leq |P| \leq [(n+1)/2],$$

proving condition (i) in this case. Now, if any positive integer occurring in A appears exactly once in A, then the positive integer $1 \in P$, satisfying condition (ii), and we are done. If not, then the smallest positive integer which occurs in P is 2 (i.e., $1 \notin P$ but $2 \in P$), so $p_k = 2 \leq p_i$ for each $p_i \in P$, where p_k counts the number of times the positive integer n/2 appears in P. To complete the proof, we must show that the sequence $B\sim = \{b\sim_1, b\sim_2, ..., b\sim_n\}$ which consists of p_1 1's, p_2 2's, p_3 3's, ..., p_{k-1} (k-1)'s, and $p_k = 2$ (n/2)'s is a satisfiable sequence. Now, by hypothesis, P is the frequency distribution of the centroid sequence $A = \{a_1, a_2, ..., a_n\}$ of T. This implies that P is also the frequency distribution of the complementary sequence $B = \{b_1, b_2, ..., b_n\}$ of T. So, if $B = B\sim$ (as multi-sets), then $B\sim$ is a satisfiable sequence since it is the complementary sequence of tree T, and we are done.

If, however, B $\neq$ B~ (as multi-sets), then note that since B is a satisfiable sequence, the sequence B#, obtained by leaving each "1" of B intact and leaving each of the two occurrences of "n/2" in B intact, and replacing each appearance of the second smallest positive integer that occurs in B by a "2", and by replacing each appearance of the third smallest positive integer that occurs in B by a "3", and by replacing each appearance of the fourth smallest integer that occurs in B by a "4", etc., is also a satisfiable sequence since at each stage s of the crossing-out procedure used to verify that B is a satisfiable sequence, each element of B which is $\leq$ the element $b_s \in$ B being satisfied at stage s and which has not yet been crossed out at stage s is realizable as the sum of one or more elements of B# which are $\leq$ the element b_s# $\in$ B# being satisfied at stage s and which have not yet been crossed out at stage s. That is, B satisfiable $\Rightarrow$ B# satisfiable.

Now, B~ can be obtained from B# by leaving each "1" and each "n/2" of B# intact, and by replacing each appearance of a positive integer > 1 which occurs most frequently in B# by a "2", and by replacing each appearance of a positive integer > 1 which occurs second-most frequently in B# by a "3", etc. (where "ties" for "most frequently", "second-most frequently", etc., are settled randomly). If B# = B~ (as multi-sets), then we are done. Otherwise, B# $\neq$ B~ (as multi-sets). But, observe that B# satisfiable $\Rightarrow$ B~ satisfiable because each element b~ $\in$ B~ such that 1 < b~ < n/2 can be satisfied by another element b~~ $\in$ B~ such that b~~ = b~ - 1 (since the number of elements of B~ that are equal to b~~ is ≥ the number of elements of B~ that are equal to b~). And, at the stage t of the crossing-out procedure (used to verify that B# is a satisfiable sequence) where the first "n/2" is being satisfied, B# contains a subset of non-crossed out elements whose sum is (n/2) - 1. This implies that at the same stage t of the crossing-out procedure for B~, there is a subset of non-crossed out elements of B~ whose sum is also (n/2) - 1, since each non-crossed out element of B# at stage t is realizable as a sum of one or more non-crossed out elements of B~ at stage t. This implies that the second "n/2" of B~ can also be satisfied, since the sum of the non-crossed out elements of B~ at stage t (not including the first "n/2") is (n-2). So, after the second "n/2" of B~ has been satisfied, the only non-crossed out elements of B~ that remain are b_{n-1} and b_n. Thus, B~ is a satisfiable sequence. Hence, B satisfiable ⇒ B# satisfiable ⇒ B~ satisfiable, completing the proof. □

Example 2.5.3: B = {1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 4, 4, 4, 4, 9, 9} is a satisfiable sequence. Therefore,
B# = {1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 3, 3, 3, 3, 9, 9}
is a satisfiable sequence. Hence,
B~ = {1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 9, 9}
is a satisfiable sequence.

(⇐) Conversely, suppose that conditions (i) and (ii) hold. By (i), 2 ≤ |P| ≤ [(n+1)/2].

Claim: If $1 \in P$ (which is one of the two possibilities provided by condition (ii)), then P is the frequency distribution of the centroid sequence $A = \{a_1, a_2, ..., a_n\}$ of a tree T whose complementary sequence $B = \{b_1, b_2, ..., b_n\}$ consists of p_1 1's, p_2 2's, ..., p_{k-1} (k-1)'s , and $p_k = 1$ $(n - (k - 1))$.

Proof of Claim: We need to show that B really is the complementary sequence of a tree T on n vertices. Observe that a necessary condition for B to be the complementary sequnce of a tree T on n vertices is that $b_{n-1} + b_n = n$. And, according to our construction of B, $b_{n-1} = k - 1$ and $b_n = n - (k - 1)$. This implies that $b_{n-1} + b_n = (k - 1) + n - (k - 1) = n$. Another necessary condition for B to be the complementary sequence of a tree T on n vertices is that the elements of B must be arranged in non-decreasing order. Note that, by our construction of B, we need only check that $b_{n-1} \le b_n$. Now, by condition (i), $2 \le |P| \le [(n+1)/2]$. This implies that $1 \le b_{n-1} = k - 1 \le [(n+1)/2] - 1 = [(n-1)/2]$. Thus, $[n/2] + 1 = [(n+2)/2] \le b_n = n - (k - 1) \le n - 1$, which implies that $b_{n-1} \le b_n$ (actually, $b_{n-1} < b_n$), since $n \ge 3$, by hypothesis. To show that B is the complementary sequence of a tree T on n vertices, we will show that B is a satisfiable sequence. Note that for each positive integer $i > 1$ which appears in B, the number of times p_{i-1} that the positive integer (i-1) appears in B is $\ge$ the number of times p_i that i appears in B (i.e., $p_{i-1} \ge p_i$ for each $i = 2, 3, ..., k$). Therefore, each element b_s of $B \backslash \{b_n\}$ which is > 1 (of which there is at least 1, since $n \ge 3$) can be satisfied by a single element $b_t \in B \backslash \{b_{n-1}, b_n\}$ such that $b_t = b_s - 1$ since the number of occurrences in B of the positive integer represented by $b_t \geq$ the number of occurrences in B of the positive integer represented by b_s. Also, b_n is satisfied by exactly the sum S of the elements of B that have not already been crossed out, because this sum is:

$S = (1p_1 - 1p_2) + (2p_2 - 2p_3) + (3p_3 - 3p_4) + \ldots +$
$((k-2)p_{k-2} - (k-2)p_{k-1}) + ((k-1)p_{k-1} - (k-1)\cdot 1)$

[Note that by our construction of B, each set of parentheses in the sum S contains a non-negative integer.]

$= p_1 + p_2 + p_3 + \ldots + p_{k-1} - (k-1)$

$$= \sum_{j=1}^{k} p_j \quad - p_k - (k-1) = n - 1 - (k-1) = n - (k-1) - 1 = b_n - 1.$$

Hence, B is, indeed, a satisfiable sequence, and B is the complementary sequence of a tree T on n vertices whose centroid sequence A is obtained from B by subtracting each element of B from n. Note that two vertices u and v of T are adjacent on T if and only if either $b_u \in B_v$, or $b_v \in B_u$, or the complementary numbers of vertices u and v in T are b_{n-1} and b_n, not necessarily in that order. And, by our construction of B, P is the frequency distribution of B, and, therefore, of A, as well (i.e., P = f(B) = f(A)).

If, on the other hand, $1 \notin P$ but $2 \in P$ and the sequence $B\sim = \{b\sim_1, b\sim_2, \ldots, b\sim_n\}$ which consists of p_1 1's, p_2 2's, p_3 3's, ..., p_{k-1} (k-1)'s, and $p_k = 2$ (n/2)'s is a satisfiable sequence, then, since $b\sim_{n-1} + b\sim_n = n/2 + n/2 = n$ and the elements of B~ are arranged in non-decreasing order (note that, by condition (i),

$1 \leq b\sim_{n-2} = k - 1 \leq [(n-1)/2] < n/2 = b\sim_{n-1} = b\sim_n$), B~ is the complementary sequence of a tree T on n vertices whose centroid sequence A is obtained from B~ by subtracting each element of B~ from n. As before, two vertices x and y of T are adjacent in T if and only if either $b\sim_x \in B_y$, or $b\sim_y \in B\sim_x$, or the complementary numbers of vertices x and y in T are $b\sim_{n-1}$ and $b\sim_n$, not necessarily in that order. And, by the definition of B~ here, P is the frequency distribution of B~, and, hence, also of A (i.e., P = f(B~) = f(A)), and we are done. ☐

Remark 2.5.1: Suppose P is a partition of a positive integer $n \geq 3$ such that $1 \notin P$ but $2 \in P$. If P is realizable as the frequency distribution of the centroid sequence $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ of a tree T of order n, then the centroid of T must consist of a pair of adjacent vertices (since $1 \notin P$ but $2 \in P$). Therefore, $a_{n-1} = a_n$. And, since $a_{n-1} + a_n = n$, we have $a_{n-1} = n/2 = a_n$. This implies that n must be even, since each element of A is a positive integer. Compare Theorem 4.4.13.

In effect, Theorem 2.5.1 states that if P is a partition of a positive integer $n \geq 3$ and $1 \in P$, then it is the number of elements of P and not the size of these elements that determines whether P can be realized as the frequency distribution of the centroid sequence of a tree T on n vertices. However, if $1 \notin P$ but $2 \in P$, then both the number and size of the elements of P play a role in determining whether P can be realized as the frequency distribution of the centroid sequence of a tree T on n vertices.

Theorem 2.5.1 can be applied as follows. Given a sequence $S = \{s_1, s_2, ..., s_n\}$ of $n \geq 3$ positive integers, we can compute f(S) and check whether f(S) satisfies the conditions of Theorem 2.5.1. If it does not, then S is not realizable as the centroid sequence of a tree. If it does, then S may or may not be realizable as the centroid sequence of a tree, so another "litmus test" (e.g., Theorem 2.4.1) must be applied to S to determine whether S is realizable as the centroid sequence of a tree.

Example 2.5.4: Let $S = \{5, 5, 4, 4, 3, 2\}$. Can S be realized as the centroid sequence of a tree T? Observe that since S has 6 elements, $n = 6$. And, $f(S) = \{2, 2, 1, 1\}$ is a partition of $n = 6$ whose elements are arranged in non-increasing order. Now, by condition (i) of Theorem 2.5.1, a necessary condition for f(S) to be the frequency distribution of the centroid sequence of a tree is that $2 \leq |f(S)| \leq [(n+1)/2] = [(6+1)/2] = 3$. That is, it must be the case that $2 \leq |f(S)| \leq 3$. However, in our example, $|f(S)| = 4$. Hence, we conclude that f(S) is not realizable as the frequency distribution of the centroid sequence of a tree. This implies that S cannot be realized as the centroid sequence of a tree.

Remark 2.5.2: We could have also shown that S is not realizable as the centroid sequence of a tree by simply noticing that $s_{n-1} + s_n \neq n$ (i.e., $3 + 2 \neq 6$).

In the next section, we will see how the work of Peter Slater has extended the concept of centroid from the rather limited realm of trees to the much broader world of connected graphs.

2.6 Slater's Extension of the Concept of Centroid to Connected Graphs

Up to this point, we have considered the concept of centroid only as it applies to trees. Thanks to the work of Peter Slater in 1975 [12], we can now talk about the centroid of any connected graph G.

Definition 2.6.1: Let G be a connected graph with $n \geq 2$ vertices and let v be a vertex of G. The **Slater number** c(v) associated with vertex v in G is defined as follows:

$$c(v) = \min_{w \in V(G)\setminus\{v\}} f(v,w),$$

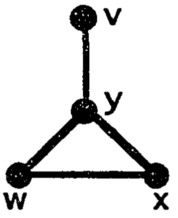
where f(v,w) = (the number of vertices of G closer to v than to w) - (the number of vertices of G closer to w than to v).

Remark 2.6.1: Each vertex v of a connected graph G is closer to itself than to any other vertex of G.

Example 2.6.1: In the connected graph G of Figure 2.6.1, f(v,w) = 1 - 2 = -1.

Example 2.6.2: The Slater numbers of the vertices of the connected graph G of Figure 2.6.1 are: c(v) = 1 - 3 = -2, c(w) = 1 - 2 = -1, c(x) = 1 - 2 = -1, and c(y) = 2 - 1 = 1.

Figure 2.6.1:



Graph G

Note that since a tree T is, by definition, a connected graph, each vertex v of T has a Slater number c(v) associated with it. It turns out that there is a rather simple relationship between the weight a(v) and the Slater number c(v) of a vertex v in a tree T.

Theorem 2.6.1: Let T be a tree with $n \geq 2$ vertices and let v be any vertex of T. If a(v) is the weight of v in T and c(v) is the Slater number of v in T, then $c(v) = n - 2 \cdot a(v)$.

Proof: If v is a vertex of T, a vertex $v\sim$ ($\neq v$) of T that minimizes the quantity f(v,w) in the definition of the Slater number c(v) of v (i.e., a vertex $v\sim \in V(T)/\{v\}$ satisfying $f(v,v\sim) = \min_{w \in V(T)\setminus\{v\}} f(v,w)$) is a vertex that is adjacent to v in T and which determines a branch B of T which emanates from v such that $\text{size}(B) \geq \text{size}(B_\#)$ for all branches $B_\#$ which emanate from v in T. The reason that such a vertex v~ minimizes the quantity f(v,w) is that f(v,v~) = (the number of vertices on all of the branches of T emanating from v other than branch B, counting vertex v itself just one time) - (the number of vertices on branch B minus 1, since we don't count vertex v here). Each vertex $w \in V(T)\setminus\{v\}$ which differs from that (respectively, those) vertex (respectively, vertices) v~ described above will produce a value of f(v,w) which is > f(v,v~).

$$\text{Now, } c(v) = \min_{w \in V(T)\setminus\{v\}} f(v,w) = f(v,v\sim)$$

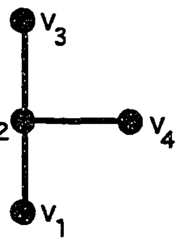
$$= (n - a(v)) - a(v) = n - 2\cdot a(v). \quad \square$$

Example 2.6.3: The weight $a(v_i)$ and Slater number $c(v_i)$ of each vertex v_i $(i = 1, 2, 3, 4)$ of the tree T of Figure 2.6.2 are indicated below. Note that for each i, $c(v_i) = n - 2\cdot a(v_i) = 4 - 2\cdot a(v_i)$.

$c(v_1) = 1 - 3 = -2$; $a(v_1) = 3$. $c(v_2) = 3 - 1 = 2$; $a(v_2) = 1$.

$c(v_3) = 1 - 3 = -2$; $a(v_3) = 3$. $c(v_4) = 1 - 3 = -2$; $a(v_4) = 3$.

Figure 2.6.2:



Tree T

Remark 2.6.2: Observe that Theorem 2.6.1 implies that for each vertex v of a non-trivial tree T, $c(v) \equiv n \pmod 2$, since c(v), n, and a(v) are all integers in case T is a non-trivial tree. This means that for each vertex v of a non-trivial tree T, the Slater number c(v) of v in T is of the same parity as the order n of T.

Let $A = \{a_1, a_2, \ldots, a_n\}$ $(n \geq 2)$ be the centroid sequence of a tree T on n vertices. Since the weight a(v) of a vertex $v \in V(T)$ is related to its Slater number c(v) by the equation $c(v) = n - 2 \cdot a(v)$, we can write down a new sequence $C = \{c_1, c_2, \ldots, c_n\}$, where $c_i = n - 2a_i$ for each $i = 1, 2, \ldots, n$. We call C the **Slater sequence** of T since each $c_i \in C$ $(i = 1, 2, \ldots, n)$ is the Slater number of the vertex $v_i \in V(T)$ whose weight in T is a_i. Observe that since A is arranged in non-increasing order, C is arranged in non-decreasing order.

The following four corollaries follow immediately from Theorem 2.6.1.

Corollary 2.6.1: If $C = \{c_1, c_2, \ldots, c_{n-1}, c_n\}$ is the Slater sequence of a tree T, then $c_{n-1} + c_n = 0$.

Proof: $c_{n-1} + c_n = (n - 2 \cdot a_{n-1}) + (n - 2 \cdot a_n) = 2n - 2(a_{n-1} + a_n)$ $= 2n - 2n = 0$. □

Corollary 2.6.2: Let T be a tree with $n \geq 3$ vertices and let $C = \{c_1, c_2, \ldots, c_{n-1}, c_n\}$ be the Slater sequence of T. If i and j are positive integers such that $1 \leq i, j \leq n-1$ and $i \neq j$, then $c_i + c_j < 0$.

Proof: If i and j are positive integers such that $1 \leq i, j \leq n-1$ and $i \neq j$, then $a_i + a_j \geq a_{n-2} + a_{n-1} > a_n + a_{n-1} = n$. This implies that $c_i + c_j = (n - 2 \cdot a_i) + (n - 2 \cdot a_j) = 2n - 2(a_i + a_j) < 2n - 2(a_n + a_{n-1}) = 2n - 2n = 0$. That is, $c_i + c_j < 0$, and we are done. ☐

Corollary 2.6.3: Let T be a tree on $n \geq 3$ vertices, and let $C = \{c_1, c_2, \ldots, c_{n-1}, c_n\}$ be the Slater sequence of T. Then, C contains at least one negative integer.

Proof: From Corollary 2.6.1, we know that $c_{n-1} = -c_n$. If $c_n \neq 0$ (i.e., if $c_n > 0$, since $c_n \geq c_{n-1}$), we are done, since $c_n \neq 0$ implies that $c_{n-1} < 0$. So, suppose $c_n = 0$. Then, $c_{n-1} = 0$. Let $c_k \in C$ with $k \neq n-1$ and $k \neq n$. Note that such an element c_k exists in C because $n \geq 3$. And, $c_k + c_{n-1} < 0$, by Corollary 2.6.2. But, $c_{n-1} = 0$ implies that $c_k < 0$, and the proof is complete. ☐

Corollary 2.6.4: If $C = \{c_1, c_2, \ldots, c_n\}$ is the Slater sequence of a tree T, then:

(i) the number of positive integers contained in C is at most 1;

and

(ii) $c_n \geq 0$.

Proof: (i) Suppose that C contains 2 or more positive integers. Then, since the elements of C are arranged in non-decreasing order, $c_{n-1} > 0$ and $c_n > 0$. But, this contradicts Corollary 2.6.1, which states that $c_{n-1} + c_n = 0$.

(ii) Suppose $c_n < 0$. By Corollary 2.6.1, $c_{n-1} + c_n = 0$ implies that $c_{n-1} > 0$, contradicting the fact that $c_{n-1} \leq c_n$. Hence, $c_n \geq 0$. ☐

The following application of Slater's work involves extending the branch-weight definition of the centroid number (i.e., weight) of a vertex of a tree T to apply to a vertex of any connected graph G. Here's our definition.

Definition 2.6.2: Let G be a connected graph on n vertices, and let v be a vertex of G. Then the **branch-weight centroid number** a(v) of v in G may be defined in terms of the Slater number c(v) of v in G as follows: $a(v) = (n - c(v))/2$.

The following theorem relates the concepts of Definition 2.6.2 to those of Definition 2.6.1.

Theorem 2.6.2: Let G be a connected graph on n vertices, and let v be a vertex of G. Then, the branch-weight centroid number a(v) of v in G is given by the equation: $a(v) = x_{u,v} + (1/2)\cdot y_{u,v}$, where $x_{u,v}$ = the number of vertices of G that are closer to u than to v for some vertex $u \in V(G)\backslash\{v\}$ such that $f(v,u) = \min_{w \in V(T)\backslash\{v\}} f(v,w)$, and $y_{u,v}$ = the number of vertices of G that are equidistant from v and any vertex u satisfying the definition of $x_{u,v}$ above.

Proof: $a(v) = (n - c(v))/2 = (n - f(v,u))/2$ [for some $u \in V(G)\setminus\{v\}$ that satisfies the definition of $x_{u,v}$ above]

$= (n - x_{v,u} + x_{u,v})/2$ [where $x_{v,u}$ is the number of vertices of G closer to v than to u, and $x_{u,v}$ is the number of vertices of G closer to u than to v]

$= [(x_{v,u} + x_{u,v} + y_{u,v}) - x_{v,u} + x_{u,v}]/2 = (2\cdot x_{u,v} + y_{u,v})/2$

$= x_{u,v} + (1/2)\cdot y_{u,v}$.

That is, $a(v) = x_{u,v} + (1/2)\cdot y_{u,v}$, and we are done. ☐

We have the following immediate corollary of Theorem 2.6.2.

Corollary 2.6.5: If G is a connected graph on $n \geq 2$ vertices and v is a vertex of G, then a(v) is either a positive integer or half of a positive integer.

Proof: If G is a tree, a(v) must be a positive integer because $y_{u,v} = 0$ for a tree. If G is a connected graph that is not a tree, then $a(v) = x_{u,v} + (1/2)\cdot y_{u,v}$. Since $x_{u,v} \in Z^+$ and $y_{u,v} \in Z^+ \cup \{0\}$, where Z^+ = the set of positive integers, we see that a(v) is either a positive integer or half of a positive integer. ☐

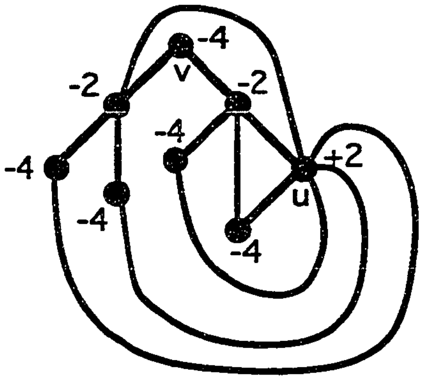
Remark 2.6.3: Note that if G is a non-trivial connected graph which is not a tree, it is not necessarily the case that for each vertex v of G, the Slater number c(v) of v in G is of the same parity as the order n of G, since a(v) may not be an integer. Compare Remark 2.6.2.

In the proof of Theorem 2.6.1, we saw that for a vertex v of a tree T, a vertex u $\in$ V(T)\{v} that minimizes f(v,u) in Slater's definition of c(v) must be adjacent to v in T. The following remark points out the fact that the existence of cycles within a connected graph G makes the calculation of the Slater numbers of the vertices of G more complex than the calculation of the Slater numbers of the vertices of a tree.

Remark 2.6.4: If G is a connected graph on $n \geq 3$ vertices which is not a tree and v $\in$ V(G), then a vertex u $\in$ V(G)\{v} that minimizes f(v,u) in Slater's definition of c(v) need not be adjacent to v in G.

Proof: The proof here is by example. Observe that in graph G of Figure 2.6.3, f(v,u) $\leq$ f(v,w) for each w $\in$ V(G), but vertex u is not adjacent to vertex v in G. If G were a tree, on the other hand, u would have to be adjacent to v in G. ☐

Figure 2.6.3: The Slater number of each vertex of graph G (at the top of page 102) is indicated next to that vertex.



Graph G

Slater called the set of all vertices v of a connected graph G for which $c(v) \geq c(x)$ for all $x \in V(G)$ the **security center**, denoted **C(G)**, of G. Observe that vertex u is the security center of graph G of Figure 2.6.3. Slater proved the following theorem [12].

Theorem (Slater): If T is a tree, then the security center C(G) of G is the centroid of G.

Therefore, the centroid of a tree T consists of all of its vertices of maximum Slater number. Slater also defined a function g on the vertex set V(G) of a connected graph G, as follows.

Definition 2.6.3: Let G be a connected graph, and let v be a vertex of G. Define the function $g: V(G) \to Z$ by the following equation.

$$g(v) = \sum_{w \in V(G)\setminus\{v\}} f(v,w),$$

where Z is the set of integers and f(v,w) is defined in Definition 2.6.1.

Slater then went on to define the **security centroid**, denoted $C_1(G)$, of a connected graph G to be the set of all vertices of G at which function g takes its maximum value. F. Buckley and F. Harary have referred to the work of Peter J. Slater with respect to Slater's extension of the concept of a centroid so that it is now defined for all connected graphs (see [3, pages 34 and 35]). Buckley and Harary have mistakenly written that: "The centroid of a graph G is the set of all nodes for which g(u) is maximum." If this were correct, then the centroid of any tree T, which, by the above theorem of Slater, coincides with its security center, would also coincide with its security centroid. That is, for any tree T, the concepts of centroid, security center, and security centroid would all coincide if Buckley's and Harary's statement were correct. However, Slater showed that for any natural number n, there is a tree T_n whose security centroid consists of n independent vertices (in fact, no three of which are on one path) [12]. Hence, by Theorem 2.1.1 (Jordan), the security centroid cannot coinicide with the centroid of such trees T_n with $n \geq 3$. Therefore, it is suggested that what Buckley and Harary probably should have written is: "The centroid of a connected graph G is the set of all nodes u of G for which c(u) is maximum." (Note that Slater uses the notation f(u) rather than c(u) in his paper). In other words, the centroid of a connected graph G coincides with its <u>security center</u> C(G).

CHAPTER 3 THE CENTER AND ECCENTRICITY SEQUENCE OF A TREE

3.1 The Center of a Tree

The concept of the **center** of a graph has been explored in the literature much more extensively than the concept of the centroid of a graph. One reason for this may be that, until Peter Slater's paper of 1975 [12], the concept of centroid had been defined only for trees. The concept of center, on the other hand, was defined originally to be applicable to any connected graph. In this section, we will look at some fundamental results concerning the center of a tree.

It was already known to Jordan in 1869 that the center of a tree consists of either a single vertex or a pair of adjacent vertices [8]. His proof of this fact made use of the following observation:

> If you prune all the 1-valent vertices from a tree T on $n \geq 3$ vertices, the resulting tree T* has the same center as the original tree T.

One should note that if the word "center" is replaced by the word "centroid" in the above statement, the new statement obtained is <u>not</u> <u>true</u> in general. So, while the center of a tree T can be located by simply pruning all the 1-valent vertices of T, and then of T*, etc., until only a single vertex or a single edge and its endpoints remain, the centroid of T can be located only by carefully taking account of which 1-valent vertices have already been pruned and which one should be pruned next (see Algorithm 2.1.1).

Some of the ideas developed in the present section were known to D. Konig in 1950 [9, 10]. However, the material of this section, in its entirety, was discovered independently of Konig's work. The statements and proofs provided below of the theorems that were known to Konig are different from those given by Konig.

The following theorems provide several different characterizations of the center of a tree.

Theorem 3.1.1: Let T be a tree on $n \geq 3$ vertices, and let $v \in V(T)$. Then, v is the center of T if and only if two longest internally disjoint paths in T emanating from v have the same length.

Proof: ($\Rightarrow$) Suppose that v is the center of T. Note that since $n \geq 3$, v cannot be a 1-valent vertex of T, for if it were 1-valent, then the unique vertex w adjacent to v in T would be such that $e(w) < e(v)$, contradicting the hypothesis that v is the center of T. So, $\text{val}(v) \geq 2$. Hence, there are at least two branches emanating from v in T, and, thus, at least two internally disjoint paths emanating from v in T. Let P_1 and P_2 be two longest internally disjoint paths in T emanating from v, where P_1 lies on branch B_1 emanating from v, and P_2 lies on branch B_2 emanating from v. Suppose, without loss of generality, that $\text{length}(P_1) = a > b =$

length (P_2), where a and b are positive integers and $b \geq 1$. Then, $e(v) = a$. Now, consider the unique vertex $v\sim$ which is adjacent to v on path P_1. Note that $d(v\sim,x) \leq a - 1$ for each vertex $x \neq v$ lying on branch B_1. Also, $d(v\sim,y) \leq b + 1$ for each vertex y lying on branch B_2 (note that v is one of these vertices y). And, $d(v\sim, z) \leq b + 1$ for each vertex z of T lying on a branch B emanating from v such that $B \neq B_1$ and $B \neq B_2$. (Note that such branches B do not exist in case v is 2-valent in T.) That is, $d(v\sim,u) \leq \max(a-1, b+1)$ for each vertex $u \in V(T)$. And, since $b < a$, $b + 1 \leq a$. Thus, $e(v\sim) \leq a$. But, this contradicts the fact that v is <u>the</u> center of T, since $e(v) = a$. Hence, assuming that $\text{length}(P_1) = a > b = \text{length}(P_2)$ leads to a contradiction. Note that assuming that $\text{length}(P_2) > \text{length}(P_1)$ leads to a similar contradiction. So, $\text{length}(P_1) = \text{length}(P_2)$ (i.e., $a = b$), and we are done. ☐

$(\Leftarrow)$ Suppose, conversely, that two longest internally disjoint paths P_1 and P_2 in T emanating from v have the same length, where P_1 lies on branch B_1 emanating from v, and P_2 lies on branch B_2 emanating from v. Say, $\text{length}(P_1) = a = \text{length}(P_2)$. Then, $e(v) = a$. Let $w \neq v$ be a vertex of T. If w lies on branch B_1, then there are one or more vertices of P_2 at distance $\geq (a+1)$ from w in T. So, in this case, $e(w) \geq a + 1$. Similarly, if w lies on branch B_2, then there are one or more vertices of P_1 at distance $\geq a+1$ from w in T. So, again, $e(w) \geq a + 1$. Finally, if w lies on a branch B emanating from v such that $B \neq B_1$ and $B \neq B_2$, then there are one or more vertices of P_1 and P_2 at distance $\geq a+1$ from w in T. Hence, $e(w) \geq a + 1$ in this case, as well. So, for each vertex $w \neq v$ of T, $e(w) > e(v)$, which implies that v is the center of T. ☐

Theorem 3.1.2: Let T be a tree on $n \geq 2$ vertices, and let (v, w) be an edge of T. Then, (v, w) is the center of T if and only if a longest path P_1 emanating from v and a longest path P_2 emanating from w have the same length.

Proof: ($\Rightarrow$) Suppose edge (v, w) is the center of T. If $n = 2$, the result is obvious. Otherwise, if $n \geq 3$, $e(v) = e(w) < e(x)$ for each $x \in V(T)\setminus\{v,w\}$. In particular, $e(v) = e(w)$ means that a longest path P_1 emanating from v has the same length as a longest path P_2 emanating from w, and we are done. ☐

($\Leftarrow$) Conversely, suppose that a longest path P_1 emanating from v and a longest path P_2 emanating from w have the same length. Then, by definition, $e(v) = e(w)$. If $n = 2$, the result is clear. Otherwise, for $n \geq 3$, if we can show that for any vertex $x \in V(T)\setminus\{v,w\}$, $e(v) = e(w) < e(x)$, we will have proved that edge (v, w) is the center of T. Note that path P_1 must pass through vertex w, for if it did not, there would be a path P# emanating from w, obtained by juxtaposing edge (w, v) with path P_1 at vertex v, such that length(P#) = 1 + length(P_1) > length(P_2), contradicting the defintiton of P_2. Similarly, path P_2 must pass through vertex v. Now, let $x \in V(T)\setminus\{v, w\}$. Then, vertex x lies either on the unique branch B_{vw} in T emanating from v determined by w, or on a branch $B \neq B_{vw}$ of T which emanates from v. If x lies on branch B_{vw}, then, since $x \neq v$ and $x \neq w$, $d(x,w) \geq 1$, and the unique path P_{xw} in T joining vertices x and w is such that $E(P_{xw}) \cap E(P_2) = \emptyset$ and $V(P_{xw}) \cap V(P_2) = \{w\}$. Hence, if we juxtapose paths P_{xw} and P_2 at vertex w, we obtain a new path P~ in T such that length(P~) $\geq 1 +$ length(P_2), which implies that $e(x) \geq 1 + e(w)$. Thus, $e(x) > e(w) = e(v)$. On the other hand, if x lies on a branch $B \neq B_{vw}$ of T emanating from v, then $d(x,v) \geq 1$ and the unique path P_{xv} in T joining vertices x and v is such that $E(P_{xv}) \cap E(P_1) = \emptyset$ and $V(P_{xv}) \cap V(P_1) = \{v\}$. Thus, if we juxtapose paths P_{xv} and P_1 at vertex v, we obtain a new path P~~ in T such that length(P~~) $\geq 1 +$ length(P_1), which implies that $e(x) \geq 1 + e(v)$. Hence, $e(x) > e(v) = e(w)$, and the proof is complete. □

As we saw in Lemma 2.1.1, if v is any vertex of a tree T, then a branch of T of greatest size which starts at v must contain the centroid of T. It is, therefore, not surprising that a longest path which starts at v in T must contain the center of T.

Lemma 3.1.1: Let T be a tree on $n \geq 2$ vertices, and let v be any vertex of T. If w is a vertex of T such that $d(v,w) = e(v)$, then the unique path P_{vw} in T joining vertices v and w must contain the center of T.

Proof: If $n = 2$, the result is obvious. So, suppose $n \geq 3$. We'll prove the theorem by considering the following 2 separate cases:
(i) the case in which the center of T is a single vertex, and
(ii) the case in which the center of T is an edge.

Case (i): Suppose that the center of T is a single vertex c. If $v = c$, the result is immediate. So, consider any vertex $v \neq c$ of T. Since $v \neq c$, v lies on a unique branch B of T emanating from c. Let $w \neq v$ be a vertex of T such that $d(v,w) = e(v)$. Note that $w \neq c$ since w must be a 1-valent vertex of T. If w lies on a branch $B\sim \neq B$ of T emanating from c, then the unique path P_{vw} in T joining vertices v and w contains vertex c, and we are done. So, suppose instead that w lies on branch B. Then, branch B is the unique branch of T emanating from c which contains path P_{vw}. Now, since vertex c does not lie on path P_{vw}, there is a unique vertex x of path P_{vw} such that $d(c,x) < d(c,y)$ for each vertex $y \neq x$ lying on P_{vw}. Note that $x \neq w$, since w must be a 1-valent vertex of T. Also, $d(c,x)$ is, by definition, the distance in T from vertex c to path P_{vw}, and $d(c,x) \geq 1$.

At this point, we would like to determine an upper bound on the length of path P_{vw}. Recall that c is the center of T. Thus, by Theorem 3.1.1, two longest internally disjoint paths emanating from c must have the same length L. Now, $d(c,v) + d(c,w) = d(c,x) + d(x,v) + d(c,x) + d(x,w) = 2 \cdot d(c,x) + \text{length}(P_{vw})$. Therefore, $\text{length}(P_{vw}) = d(c,v) + d(c,w) - 2 \cdot d(c,x) \leq d(c,v) + L - 2$. [Note that vertex x may coincide with vertex v (i.e., $x = v$), in which case $d(x,v) = 0$.] But, since c is the center of T, there is a vertex z lying on a branch $B^* \neq B$ of T emanating from c such that $d(c,z) = e(c) = L$. So, $\text{length}(P_{vz}) = d(v,c) + d(c,z) = d(c,v) + L$. This implies that $\text{length}(P_{vz}) > \text{length}(P_{vw})$, contradicting the hypothesis that $d(v,w) = e(v)$. So, assuming that vertex w lies on branch B leads to a contradiction. This implies that w must lie on a branch $B\sim \neq B$ of T emanating from c. Thus, P_{vw} must contain the center c of T.

Case (ii): Suppose, now, that the center of T is an edge (c_1, c_2). If $v = c_1$, then let $w \neq v$ be a vertex of T such that $d(v,w) = e(v)$. Consider the unique path P_{vw} in T joining vertices v and w. Since $v = c_1$, vertex c_1 lies on (in fact, is an endpoint of) path P_{vw}. So, it remains to show only that vertex c_2 also lies on path P_{vw}. Suppose that c_2 does not lie on path P_{vw}. Then, if we join edge (c_2, c_1) to path P_{vw} at vertex v ($= c_1$), we obtain a new path P_{c_2w} joining vertices c_2 and w in T. And, $\text{length}(P_{c_2w}) = 1 + \text{length}(P_{vw}) = 1 + e(v)$. That is, there is a path P_{c_2w} in T emanating from c_2 whose length is $> e(v)$. This implies that $e(c_2) > e(v) = e(c_1)$, contradicting the hypothesis that edge (c_1, c_2) is the center of T. Hence, vertex c_2 must, in fact, lie on path P_{vw} in case $v = c_1$. Similarly, we can show that if $v = c_2$ and $x \neq v$ is a vertex of T such that $d(v,x) = e(v)$, then vertex c_1 must lie on the unique path P_{vx} in T joining vertices v ($= c_2$) and x.

Finally, suppose $v \neq c_1$ and $v \neq c_2$. Then, one of c_1 or c_2 is strictly closer to v in T than the other. Assume, without loss of generality, that c_1 is closer to v than c_2 is. Then, v lies on a unique branch B of T emanating from c_1 that is not determined by

c_2. Let $w \neq v$ be a vertex of T such that $d(v,w) = e(v)$. Note that $w \neq c_1$ and $w \neq c_2$, since w must be a 1-valent vertex of T. If w lies on branch $B_{c_1c_2}$ of T emanating from c_1 determined by c_2, then the unique path P_{vw} in T joining vertices v and w contains edge (c_1, c_2), and we are done. Otherwise, w lies on a branch $B\sim \neq B_{c_1c_2}$ of T emanating from c_1 which is not determined by c_2. If $B\sim = B$, then the unique path P_{vw} in T joining vertices v and w is contained completely by branch B. In this case, the proof proceeds almost the same way that it did in Case(i) above, because any longest path P emanating from c_1 in T must pass through c_2. So, P is contained completely by branch $B_{c_1c_2}$. Therefore, w cannot lie on branch B, so, in fact, $B\sim \neq B$. So, w must lie on a branch $B\sim$ of T emanating from c_1 such that $B\sim \neq B_{c_1c_2}$ and $B\sim \neq B$. Hence, v and w lie on two different branches which emanate from c_1, neither of which contains c_2. This implies that the unique path P_{vw} in T joining vertices v and w contains c_1 but does not contain c_2. Note that path P_{vw} is obtained by joining path P_{vc_1}, which lies completely on branch B, with path P_{wc_1}, which lies completely on branch $B\sim$, at vertex c_1, since $V(P_{vc_1}) \cap V(P_{wc_1}) = \{c_1\}$ and $E(P_{vc_1}) \cap E(P_{wc_1}) = \varnothing$. So, $\text{length}(P_{vw}) = \text{length}(P_{vc_1}) + \text{length}(P_{wc_1}) = d(v,c_1) + d(w,c_1)$. Now, since w does not lie on $B_{c_1c_2}$, $d(c_1,w) < e(c_1)$. Let z be a vertex of T lying on $B_{c_1c_2}$ such that $d(c_1,z) = e(c_1)$. Then, since P_{vc_1} and P_{zc_1} lie on different branches of T which emanate from c_1, we can join these 2 paths at vertex c_1 to obtain a new path P_{vz} joining vertices v and z in T. And, $\text{length}(P_{vz}) = \text{length}(P_{vc_1}) + \text{length}(P_{zc_1}) = d(v,c_1) + d(z,c_1) =$

$d(v,c_1) + e(c_1) > d(v,c_1) + d(w,c_1) = \text{length}(P_{vw})$. That is, $\text{length}(P_{vz}) > \text{length}(P_{vw})$, contradicting the hypothesis that $d(v,w) = e(v)$. Hence, w cannot lie on a branch B~ of T, emanating from c_1, which is different from $B_{c_1c_2}$. So, w must lie on $B_{c_1c_2}$. This implies that P_{vw} must contain the center (c_1, c_2) of T, and we are done. ☐

Lemma 3.1.1 will be needed for the proofs of Theorems 3.1.3 and 3.1.4 below. These two theorems make reference to the concept of a **diameter** of a connected graph, which can be defined in terms of **geodesics** within a connected graph.

Definition 3.1.1: Let G be a connected graph, and let u and v be vertices of G. A **u-v geodesic** of G is a path P_{uv} in G joining vertices u and v of minimum length.

Definition 3.1.2: Let G be a connected graph. A **diameter** D of G is a longest geodesic of G. That is, D is a path in G whose endpoints are, say, x and y such that $\text{length}(D) = d(x,y)$ and $d(x,y) \geq d(u,v)$ for each pair of vertices u and v of G.

Recall that in a tree T, for each pair of distinct vertices $u, v \in V(T)$, there is one and only one path P_{uv} in T joining vertices u and v. Hence, P_{uv} is a u-v geodesic of T. Therefore, in a tree T, a diameter D is simply any longest path in T. It is not difficult to see that the endpoints of a diameter of a non-trivial tree T must both be 1-valent vertices. We are now ready to characterize the center of a tree T in terms of a diameter D of T.

Theorem 3.1.3: Let T be a tree on $n \geq 3$ vertices, let v be a vertex of T, and let D be a diameter of T. Then, v is the center of T if and only if D has even length and v is the midpoint of D (i.e., v is the unique vertex of the path D that divides D into two subpaths of equal length).

Proof: ($\Rightarrow$) Suppose v is the center of T. Let the endpoints of diameter D be vertices x and y. Since D is a diameter of T, D is a longest path in T. Thus, $d(x,y) = e(x)$, which implies that D must contain vertex v. That is, vertex v lies on D. Note that vertex v divides D into two subpaths D_1 and D_2. Let $\text{length}(D_1) = d_1$ and $\text{length}(D_2) = d_2$, where, without loss of generality, $d_1 \geq d_2$. Then, $e(v) \geq d_1$. Since v is the center of T, two longest internally disjoint paths in T emanating from v must have the same length. Hence, there exists a path P in T such that $\text{length}(P) \geq 2d_1$. Now, D is a diameter of T and $\text{length}(D) \leq 2d_1$. And, since $\text{length}(D) \geq \text{length}(P)$, we have $2d_1 \geq \text{length}(D) \geq \text{length}(P) \geq 2d_1$. Therefore, $\text{length}(D) = \text{length}(P) = 2d_1$, which implies that $d_1 = d_2$. Thus, v is the midpoint of D, and D has even length (namely, $2d_1$). ☐

($\Leftarrow$) Conversely, suppose D has even length and v is the midpoint of D. Then, v has 2 internally disjoint paths (call them P_1 and P_2, respectively) of equal length emanating from it in T (namely, the 2 halves of D). If P_1 and P_2 are two longest internally disjoint paths emanating from v in T, then v is <u>the</u> center of T, by Theorem 3.1.1, and we are done. So, suppose they are not. That is, suppose that there are 2 longest internally disjoint paths $P_1\sim$ and $P_2\sim$ emanating from v in T such that either length($P_1\sim$) > length(P_1) = length(P_2), or length ($P_2\sim$) > length(P_1) = length(P_2), or both. Then, the path $P_\#$ obtained by joining paths $P_1\sim$ and $P_2\sim$ at vertex v is such that length($P_\#$) > length(D), contradicting the hypothesis that D is a diameter of T. Hence, P_1 and P_2 are two <u>longest</u> internally disjoint paths emanating from v in T, and since length(P_1) = length(P_2), vertex v is <u>the</u> center of T. □

<u>Theorem 3.1.4</u>: Let T be a tree on $n \geq 2$ vertices, let (v, w) be an edge of T, and let D be a diameter of T. Then, (v, w) is the center of T if and only if D has odd length and (v, w) is the "mid-edge" of D [i.e., (v, w) is the unique edge of the path D whose removal (from D) divides D into two subpaths of equal length].

<u>Proof</u>: (⇒) Suppose edge (v, w) is the center of T. Let the endpoints of diameter D be vertices x and y. Since D is a diameter of T, D is a longest path in T. Thus, d(x,y) = e(x), which implies that D must contain edge (v, w) (the center of T). That is, (v, w) is an edge of path D. Note that the removal of edge (v, w) divides D into two subpaths D_1 and D_2 such that length(D_1) = d_1 and length(D_2) = d_2, where we can assume without loss of generality

that $D_1 = P_{vx}$, $D_2 = P_{wy}$, and $d_1 \geq d_2$. So, $e(v) = e(w) \geq d_1 + 1$, since path D_1 does not contain vertex w. Since (v, w) is the center of T, a longest path P_1 emanating from v in T must have the same length as a longest path P_2 emanating from w in T. This implies that there exists a path P in T such that length(P) $\geq$ $(d_1 + 1) + (d_1 + 1) - 1 = 2d_1 + 1$ (since w must lie on P_1 and v must lie on P_2). Now, D is a diameter of T and length(D) $\leq 2d_1 + 1$. And, since length(D) $\geq$ length(P), we have $2d_1 + 1 \geq$ length(D) $\geq$ length(P) $\geq 2d_1 + 1$. Therefore, length(D) = length(P) = $2d_1 + 1$, which implies that $d_1 = d_2$. Hence, edge (v, w) is the "mid-edge" of D and D has odd length (namely, $2d_1 + 1$). □

($\Leftarrow$) Conversely, suppose D has odd length and (v, w) is the "mid-edge" of D. Then, v has a path P_1 (containing w) of length (length(D) + 1)/2 emanating from it in T, and w has a path P_2 (containing v) of length (length(D) + 1)/2 emanating from it in T. If P_1 is a longest path in T emanating from v, and P_2 is a longest path in T emanating from w, then edge (v, w) is the center of T, by Theorem 3.1.2. Suppose not. That is, suppose that there is a longest path $P_1\sim$ emanating from v and a longest path $P_2\sim$ emanating from w in T such that either length($P_1\sim$) > length(P_1) = length(P_2), or length($P_2\sim$) > length(P_2) = length(P_1), or both. Suppose, without loss of generality, that length($P_1\sim$) > length(P_1). Note that vertex w must lie on $P_1\sim$, for if it did not, then path P# obtained by joining paths $P_1\sim$ and P_1 at vertex v would be such that length(P#) = length($P_1\sim$) + length(P_1) > 2·length(P_1) =

2(length(D) + 1)/2 = length(D) + 1, contradicting the fact that D is a diameter of T. So, vertex w lies on P_1~. Consider path P## obtained by joining path P_1~ and path P_2 at vertex v. [Note that $V(P_1\sim) \cap V(P_2) = \{v, w\}$ and $E(P_1\sim) \cap E(P_2) = \{(v, w)\}$.] Length (P##) = length(P_1~) + length(P_2) - 1 > length(P_1) + length(P_2) - 1 = 2(length(D) + 1)/2 - 1 = length(D). That is, length(P##) > length(D), again contradicting the definition of D. Hence, P_1 is a longest path emanating from v in T, P_2 is a longest path emanating from w in T, and since length(P_1) = length(P_2), edge (v, w) is the center of T. □

The theory developed in the present section shows how intimately the center of a tree T is related to "eccentric" (i.e., longest) paths within T. This relationship can be quantified by examining and analyzing the eccentricity sequence of T, which is the subject of the next section.

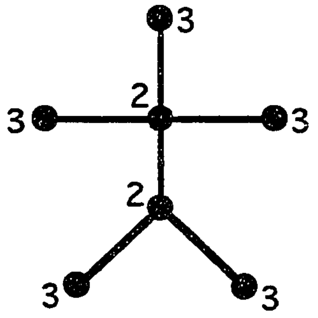
3.2 The Eccentricity Sequence of a Tree

In Section 1.4, we defined the eccentricity e(v) of a vertex v in a connected graph G. In the present section, we shall restrict our attention to the eccentricities of the vertices of a tree T.

Definition 3.2.1: Let T be a tree on $n \geq 2$ vertices. Then, the eccentricity sequence $E = \{e_1, e_2, \ldots, e_n\}$ of T is the sequence of eccentricities of the vertices of T, arranged in non-decreasing order (i.e., $e_1 \leq e_2 \leq \cdots \leq e_n$).

Example 3.2.1: The eccentricity sequence E if the tree of Figure 3.2.1 is E = {2, 2, 3, 3, 3, 3, 3}.

Figure 3.2.1: The eccentricity of each vertex in the tree below is indicated next to that vertex.



Much of the research concerning the eccentricity sequences of connected graphs was carried out by Linda Lesniak in her paper of 1975 [11]. Amongst other things, Lesniak characterized the eccentricity sequences of connected graphs, in general, and of trees, in particular. What follows was developed independently of Lesniak's work. The proof given below differs from that provided by Lesniak in her paper.

Theorem 3.2.1 (Lesniak): Let $E = \{e_1, e_2, \ldots, e_n\}$ $(n \geq 3)$ be a non-decreasing finite sequence of positive integers. Then, E is the eccentricity sequence of a tree T if and only if:

(i) either $e_n = 2 \cdot e_1$ and $e_1 < e_2$ or $e_n = 2 \cdot e_1 - 1$ and $e_1 = e_2 < e_3$;

<u>and</u>

(ii) each positive integer i such that $e_1 + 1 \leq i \leq e_n$ appears in E at least twice.

<u>Proof</u>: (⇒) There is exactly 1 tree of order 2, up to isomorphism, and its eccentricity sequence is $E = \{1, 1\}$. Now, suppose that $|E| \geq 3$ and E is the eccentricity sequence of a tree T. Since E contains $n \geq 3$ elements, the order of T is n. Now, T may be a **central** tree (i.e., a tree with exactly one central node) or a **bicentral** tree (i.e., a tree whose center consists of a pair of adjacent nodes). We'll consider each of these two possibilities separately.

Suppose, first, that T is a central tree, and let $c \in V(T)$ be the center of T. Then, $e(c) = e_1$. We have already shown in Theorem 3.1.3 that for each diameter D of T, D must have even length and c must be the midpoint of D. Since T has at least one diameter D, there are at least two vertices $x, y \in V(T)$ (namely, the endpoints of D) such that $d(x,y) = \text{length}(D) \geq d(u,v)$ for each pair of vertices $u, v \in V(T)$. Moreover, $d(x,y)$ is even and $d(x,y) = d(x,c) + d(y,c)$. Since D is a diameter of T, $d(x,c) = d(y,c)$, and x and y are both eccentric nodes for c. Hence, $d(x,y) = e(x) = e(y) = 2 \cdot e(c)$. Since $e(c) = e_1$, we have $e(x) = e(y) = 2e_1$. And, since $e(x) = e(y) \geq e(v)$ for each $v \in V(T)$, $e_n = e_{n-1} = e(x) = e(y) = 2e_1$. That is, $e_n = e_{n-1} = 2e_1$. Also, since T is, by hypothesis, a central tree, $e(c) = e_1 < e_j$ for each $j = 2, \ldots, n$. So, in particular, $e_1 < e_2$, completing the proof of condition (i). As for condition (ii), note that diameter D of T is a longest path of T whose length is even and whose midpoint is c. Thus, D contains an odd number of vertices, and the eccentricity sequence E_D of D is a subsequence of E, where $E_D = \{e_1, e_1 + 1, e_1 + 1, e_1 + 2, e_1 + 2, ..., 2e_1, 2e_1\}$. Since every positive integer i such that $e_1 + 1 \leq i \leq 2e_1 = e_n$ appears exactly twice in E_D, every such positive integer i appears at least twice in E (since E_D is contained in E), and we are done in case T is a central tree.

Now, suppose that T is a bicentral tree, and let $(c_1, c_2) \in E(T)$ be the center of T. Then, $e(c_1) = e(c_2) < e(v)$ for each vertex $v \in V(T)\setminus\{c_1,c_2\}$. Suppose, without loss of generality, that $e(c_1) = e_1$ and $e(c_2) = e_2$. For each diameter D~ of T (of which there is at least one), D~ must have odd length and (c_1,c_2) must be the "mid-edge" of D~. This implies that there are at least two vertices $r, s \in V(T)$ (namely, the endpoints of D~) such that $d(r,s) = \text{length}(D\sim) \geq d(u,v)$ for each pair of vertices $u, v \in V(T)$. In addition, $d(r,s)$ is odd and we can assume, without loss of generality, that c_1 is closer to r than to s in T. This implies that $d(r,s) = d(r,c_2) + d(s,c_1) - 1$. Since D~ is a diameter of T, $d(r,c_2) = d(s,c_1)$, r is an eccentric node for c_2, and s is an eccentric node for c_1. Therefore, $d(r,s) = e(r) = e(s) = d(r,c_2) + d(s,c_1) - 1 = e(c_2) + e(c_1) - 1 = 2\cdot e(c_1) - 1$, since $e(c_1) = e(c_2)$. Since $e(c_1) = e_1$ and $e(c_2) = e_2$, we have $e(r) = e(s) = 2e_1 - 1$. And, since $e(r) = e(s) \geq e(v)$ for each $v \in V(T)$, $e_n = e_{n-1} = e(r) = e(s) = 2e_1 - 1$. That is, $e_n = e_{n-1} = 2e_1 - 1$. Also, since T is, by hypothesis, a bicentral tree, $e_1 = e(c_1) = e(c_2) = e_2 < e_k$ for each $k = 3, ..., n$. So, in particular, $e_1 = e_2 < e_3$, and we have finished the proof of condition (i). To prove condition (ii), recall that diameter $D_\sim$ of T is a longest path of T whose length is odd and whose "mid-edge" is (c_1, c_2). Hence, $D_\sim$ contains an even number of vertices, and the eccentricity sequence $E_{D\sim}$ of $D_\sim$ is a subsequence of E, where $E_{D\sim} = \{e_1, e_1, e_1 + 1, e_1 + 1, e_1 + 2, e_1 + 2, \ldots, 2e_1 - 1, 2e_1 - 1\}$. Since every positive integer i such that $e_1 + 1 \leq i \leq 2e_1 - 1 = e_n$ appears exactly twice in $E_{D\sim}$, every such positive integer i appears at least twice in E (since $E_{D\sim}$ is a subsequence of E), completing the proof in case T is a bicentral tree. Since T is either central or bicentral, we are done. □

($\Leftarrow$) Conversely, suppose that conditions (i) and (ii) hold. To show the existence of a tree T such that E is the eccentricity sequence of T, we shall actually construct T. There are 2 cases to consider: either e_n is even, or e_n is odd.

If e_n is even, construct a path P of length e_n. Note that if x and y are the endpoints of P, then $e(x) = e(y) = e_n$ (in P). So, the eccentricity sequence E_P of path P is $E_P = \{e_n/2, e_n/2 + 1, e_n/2 + 1, e_n/2 + 2, e_n/2 + 2, \ldots, e_n, e_n\}$. Note that since e_n is even, $e_n = 2e_1$, by condition (i). This implies that $e_1 = e_n/2$. Also, $e_1 < e_2$, by condition (i), and for each positive integer i such that $e_1 + 1 \leq i \leq e_n$, i appears at least twice in E, by condition (ii). So, $e_2 = e_1 + 1 = e_n/2 + 1$, and $e_3 = e_n/2 + 1$. In fact, every positive integer i such that $e_1 + 1 \leq i \leq e_n$ appears exactly twice in E_P and at least twice in E. Moreover, the positive integer represented by $e_n/2 = e_1$ appears exactly once in both E_P and E. Note, particularly, that $m \in Z^+$ appears in E if and only if m appears in E_P. Since E_P is the eccentricity sequence of path P, we can add 1-valent vertices (so-called "leaves") to the vertices of P (if necessary) in order to produce T from P. So, if $E = E_P$, then $T = P$, and we are done. However, if E_P is strictly contained in E, then for one or more positive integers i such $e_1 + 1 \leq i \leq e_n$, i occurs in E three or more times, while i occurs in E_P exactly two times. So, if i occurs $k \geq 3$ times in E, we add (k-2) 1-valent vertices to P at a vertex v of P such that $e(v) = i - 1$ (in P) [note that P contains at least one such vertex v] to obtain a tree T# in which the eccentricity of each vertex that was also a vertex of P remains unchanged and in which the number of vertices of eccentricity i is k. Iterating this procedure of adding 1-valent vertices to the vertices of our original path P where necessary, we never alter the eccentricity of any vertex that was already present at a previous stage of the procedure, and we finally produce a tree T whose eccentricity sequence is E, completing the proof in this case (in which e_n is even). Note that at each stage of the above procedure, we produce a tree which consists of path P together with a number of 1-valent vertices that have been added to one or more of the vertices of P. In particular, T is such a tree. Recall that we call this type of tree a **caterpillar**. Observe that if all the 1-valent vertices of T are pruned, the resulting tree is simply a path $P^* \neq P$, since each 1-valent vertex added to P to produce T was added to an internal vertex (i. e., non-endpoint) of

P. Note further that no claim has been made that T is the only tree whose centroid sequence is E.

Now, we consider the case in which e_n is odd. In this case, construct a path P~ of length e_n. The eccentricity sequence $E_{P\sim}$ of path P~ is $E_{P\sim} = \{(e_n + 1)/2, (e_n + 1)/2, (e_n + 1)/2 \ + 1, (e_n + 1)/2 \ + 1, \ldots, e_n, e_n\}$. Since e_n is odd, $e_n = 2e_1 - 1$, by condition (i). Thus, $e_1 = (e_n + 1)/2$. Also, $e_1 = e_2 < e_3$, by condition (i), and for each positive integer i such that $e_1 + 1 \leq i \leq e_n$, i appears in E at least twice, by condition (ii). So, $e_2 = (e_n + 1)/2$, and $e_3 = (e_n + 1)/2 \ + 1$. And, every positive integer i such that $e_1 + 1 \leq i \leq e_n$ appears exactly twice in $E_{P\sim}$ and at least twice in E. Additionally, the positive integer represented by $(e_n + 1)/2 = e_1 = e_2$ appears exactly twice in both $E_{P\sim}$ and E. Note, in particular, that $j \in Z^+$ appears in E if and only if j appears in $E_{P\sim}$. So, since $E_{P\sim}$ is the eccentricity sequence of path P~, we can, exactly as we did for path P in the preceding paragraph, add suitable numbers of 1-valent vertices to appropriately selected vertices of P~ to produce a tree T~ whose eccentricity sequence is E. And, T~ is a caterpillar and is not necessarily the only tree whose eccentricity sequence is E. ☐

The next corollary follows immediately from the proof of the converse direction of Theorem 3.2.1.

Corollary 3.2.1: If $E = \{e_1, e_2, \ldots, e_n\}$ $(n \geq 3)$ is the eccentricity sequence of a tree T, then E is the eccentricity sequence of a caterpillar C, as well.

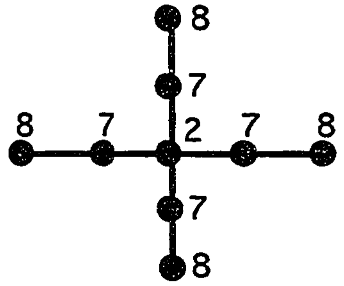
Proof: First, note that T is a tree on $n \geq 3$ vertices. If T is itself a caterpillar, then C = T, and we are done. Otherwise, if T is not a caterpillar, then the eccentricity sequence E of T satisfies conditions (i) and (ii) of Theorem 3.2.1. So, just as we did in the proof of the converse $(\Leftarrow)$ direction of Theorem 3.2.1, we can construct a path P (if e_n is even) or a path P~ (if e_n is odd) of length e_n. Then, we can, as before, add suitable numbers of 1-valent vertices to appropriately selected vertices of P (or P~) to produce a caterpillar C whose eccentricity sequence E is the same as that of T. ☐

Thus, the eccentricity sequence of a tree can always be realized by a caterpillar.

Remark 3.2.1: The centroid sequence of a tree is not, in general, realizable by a caterpillar, as the next example shows.

Example 3.2.2: The centroid sequence $A = \{8, 8, 8, 8, 7, 7, 7, 7, 2\}$ of the tree T of Figure 3.2.2 is not realizable as the centroid sequence of a caterpillar, by Theorem 2.3.1, since the number "7" appears in A more than two times. Note, however, that the centroid sequence of the pruned tree T* of T, obtained by pruning all the 1-valent vertices of T, is realizable by a caterpillar, since T* is, itself, a caterpillar. More generally, for each positive integer r, there is a tree S_r whose centroid sequence is not realizable by a caterpillar. See Theorem 4.3.3.

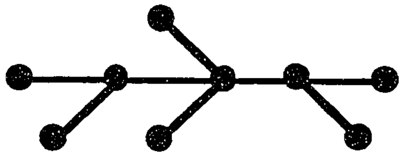
Figure 3.2.2: The figure depicts a tree T whose centroid sequence cannot be realized by a caterpillar. The weight of each vertex of T is indicated next to that vertex.



Tree T

Remark 3.2.2: Observe that the eccentricity sequence of the tree T of Figure 3.2.2 is E = {2, 3, 3, 3, 3, 4, 4, 4, 4}, and E can be realized by a caterpillar. See Figure 3.2.3.

Figure 3.2.3: A caterpillar that realizes the eccentricity sequence E of the tree T of Figure 3.2.2.



So, while caterpillars alone generate the set of all possible eccentricity sequences of a tree, they clearly do not generate the set of all possible centroid sequences of a tree.

Given a finite sequence $E = \{e_1, e_2, ..., e_n\}$ of positive integers which is known to be the eccentricity sequence of a tree T, we may like to know, in light of Corollary 3.2.1, how many non-isomorphic caterpillars there are whose eccentricity sequence is E. Theorem 3.2.2 below provides the solution to this problem.

Theorem 3.2.2: Let $E = \{e_1, e_2, ..., e_n\}$ be the eccentricity sequence of a tree T. Let x_k be the number of times that the positive integer k appears in E, and let N be the number of non-isomorphic caterpillars whose eccentricity sequence is E.

(i) If $e_1 < e_2$, then $N = [(\prod_{k=e_1+2}^{e_n} (x_k - 1) + 1)/2]$;

(ii) If $e_1 = e_2$, then $N = [(\prod_{k=e_1+1}^{e_n} (x_k - 1) + 1)/2]$,

where $[t]$ denotes the greatest integer $\leq t$.

Proof: By Corollary 3.2.1, we know that E is the eccentricity sequence of at least one caterpillar C. Let D be a diameter of C. Then, the eccentricities in C of the vertices of D are all the positive integers from e_1 to e_n, where each positive integer k such that $e_1 + 1 \leq k \leq e_n$ is the eccentricity of two distinct vertices of D, and the positive integer represented by e_1 is the eccentricity of either one or two distinct vertices of D, respectively, depending on whether the length of D is even or odd, respectively.

Case (i): If length(D) is even, then $e_1 < e_2$, and the center of C consists of a single vertex c. Since C is a caterpillar, each 1-valent vertex v of C which does not lie on D is such that e(v) = the eccentricity of 2 other distinct vertices of C that lie on D, and these 2 vertices of D are equidistant from the center (vertex c) of C. Note that if v is a 1-valent vertex of C which does not lie on D, then e(v) = e(v~) + 1, where v~ is the unique vertex of D which is adjacent to v in C. Now, since vertex c is the only vertex lying on D whose eccentricity (e_1) is $\neq$ the eccentricity of any other vertex lying on D, the only 1-valent vertices v of C not lying on D that cannot be pruned from C and then attached to a vertex w $\neq$ v~ of D without changing the original eccentricity of v in C are those v whose eccentricity in C is equal to $e_1 + 1$. So, each 1-valent vertex v of C that doesn't lie on D and whose eccentricity e(v) is such that $e_1 + 2 \leq e(v) \leq e_n$ can be pruned from C and attached to a new vertex w $\neq$ v~ of D such that d(v~,c) = d(w,c). Therefore, the number N of non-isomorphic caterpillars whose centroid sequence is E is obtained by calculating the number of different ways that the 1-valent vertices v of C not lying on D such that $e_1 + 2 \leq e(v) \leq e_n$ can be attached, each via a single edge, to a vertex of D without changing the original eccentricity of v in C. The number N of non-isomorphic caterpillars whose eccentricity sequence is E is, therefore,

$$(*) \qquad N = \left[\left(\prod_{k=e_1+2}^{e_n} (x_k - 1) + 1\right)/2\right],$$

where x_k = the number of times that the positive integer k appears in E. The reason for this is as follows. For each k such that $e_1 + 2 \leq k \leq e_n$, there are $(x_k - 2)$ 1-valent vertices v of C that do not lie on D whose eccentricity in C is k, since there are 2 distinct vertices of eccentricity k which lie on D. And, the attachments of 1-valent vertices of unequal eccentricities to the vertices of D are independent of one another (i.e., the attachment of a 1-valent vertex of eccentricity e_* is unrelated to the attachment of a 1-valent vertex of eccentricity e_{**} if $e_* \neq e_{**}$). Observe that if there are $(x_k - 2)$ 1-valent vertices v of C that do not lie on D whose eccentricity in C is k, then there are $(x_k - 1)$ ways to attach the $(x_k - 2)$ vertices v to one of the two possible attaching locations (i.e., vertices) of D (i.e., from attaching none of them at that location to attaching all $(x_k - 2)$ of them at that location). However, because of the symmetrical arrangement of the vertices of equal eccentricity along D,

$$\prod_{k=e_1+2}^{e_n} (x_k - 1)$$

counts each non-isomorphic caterpillar whose eccentricity sequence is E twice, except for exactly one of these caterpillars, provided that $(x_k - 2)$ is even for each k such that $e_1 + 2 \leq k \leq e_n$, in which exactly $(x_k - 2)/2$ 1-valent vertices v are attached at each of the two possible attachment sites along D for each k such that $e_1 + 2 \leq k \leq e_n$. The result (*) follows from this discussion.

Case (ii): If length(D) is odd, then $e_1 = e_2$, and the center of C consists of a pair of adjacent vertices, say, c_1 and c_2, which both lie on D. Observe that in this case, each positive integer k such that $e_1 \leq k \leq e_n$ is the eccentricity in C of 2 distinct vertices of C that lie on D. Since C is a caterpillar, each vertex v of C which does not lie on D is a 1-valent vertex of C which is adjacent to a vertex v~ of C which lies on D such that e(v) = e(v~) + 1 in C. Let x_k denote the number of times that the positive integer k appears in E. Then, there are exactly $(x_k - 2)$ 1-valent vertices of C which do not lie on D whose eccentricity in C is k. Since the positive integer (k-1) is the eccentricity in C of exactly 2 distinct vertices of C that lie on D, there are $(x_k - 1)$ ways to attach one of these $(x_k - 2)$ 1-valent vertices of C of eccentricity k which do not lie on D to one of the 2 vertices of D of eccentricity (k-1) in C without altering the eccentricity of any of these vertices. By the symmetrical arrangement of the eccentricities in C of the vertices of D and the fact that each positive integer from e_1 to e_n appears as the eccentricity in C of exactly 2 distinct vertices of D, there are

$$[(\prod_{k=e_1+1}^{e_n} (x_k - 1) \quad + 1)/2]$$

non-isomorphic caterpillars whose eccentricity sequence is E in case length(D) is odd, since

$$\prod_{k=e_1+1}^{e_n} (x_k - 1)$$

counts each non-isomorphic caterpillar whose eccentricity sequence is E twice except for exactly one of these caterpillars provided that $(x_k - 2)$ is even for each k such that $e_1 + 1 \leq k \leq e_n$. □

Example 3.2.3: Suppose $E = \{3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5\}$. It can be verified that E is realizable as the eccentricity sequence of a tree by applying Theorem 3.2.1. Then, E is also realizable as the eccentricity sequence of a caterpillar, by Corollary 3.2.1. Let C be a caterpillar whose eccentricity sequence is E, and let D be a diameter of C. Then, since $e_n = 5$, length(D) = 5, which is odd. Therefore, in order to determine the number N of non-isomorphic caterpillars whose eccentricity sequence is E, we can apply part (ii) of Theorem 3.2.2. Note that $e_1 = 3$, $e_1 + 1 = 4$, and $e_1 + 2 = e_n = 5$. Also, $x_4 = 4$ and $x_5 = 7$. Hence, $N = [((3)(6) + 1)/2] = [19/2] = [9.5] = 9$. That is, there are exactly 9 non-isomorphic caterpillars whose eccentricity sequence is E.

The following theorem establishes the conditions under which the eccentricity sequence of a caterpillar is also realizable as the eccentricity sequence of a tree which is not a caterpillar.

Theorem 3.2.3: Let $E = \{e_1, e_2, ..., e_n\}$ be the eccentricity sequence of a caterpillar C. Then, E is also the eccentricity sequence of a tree T which is not a caterpillar if and only if there is a pair of consecutive integers each of which appears in E three or more times.

Proof: (⇒) Suppose that E is also the eccentricity sequence of a tree T which is not a caterpillar, and let D be a diameter of T. Then, length(D) ≥ 4. Let vertices x and y be the endpoints of D. Since T is not a caterpillar, there is a path P of length 2 in T which starts at a vertex v of D such that $d(v,x) \geq 2$ and $d(v,y) \geq 2$ and which includes no other vertex of D and no edge of D. Let t be the vertex of P that is adjacent to v, and let u be the vertex of P such that $d(u,v) = 2$. Now, since D is a diameter of T, the eccentricities in T of the vertices of D are all the positive integers from e_1 to e_n, where each positive integer k such that $e_1 + 1 \leq k \leq e_n$ is the eccentricity of two distinct vertices of D, and the positive integer represented by e_1 is the eccentricity of either one or two distinct vertices of D, respectively, depending on whether the length of D is even or odd, repectively. Also, $e(t) = e(v) + 1$ and $e(u) = e(v) + 2$, so neither vertex t nor vertex u is in the center of T. That is, $e(t) > e_1$ and $e(u) > e_1$, so the positive integer represented by e(t) is also the eccentricity in T of 2 distinct vertices of D, and the positive integer represented by e(u) is also the eccentricity in T of 2 other distinct vertices of D. Hence, the positive integer represented by e(t) appears at least 3 times in E, and the positive integer represented by e(u) appears at least 3 times in E. And, since $e(t) = e(v) + 1$ and $e(u) = e(v) + 2$, the positive integers represented by e(t) and e(u) are consecutive integers, so we are done in this direction. □

(⇐) Conversely, suppose that there is a pair of consecutive integers each of which appears in E three or more times. By hypothesis, E is the eccentricity sequence of a caterpillar C. Let D be a diameter of C, and let vertices x and y be the endpoints of D. Then, the eccentricities in C of the vertices of D are all the positive integers from e_1 to e_n, where each positive integer k such that $e_1 + 1 \leq k \leq e_n$ is the eccentricity of two distinct vertices of D, and the positive integer represented by e_1 is the eccentricity of either one or two distinct vertices of D, respectively, depending on whether the length of D is even or odd, repectively. Now, since there is a pair of consecutive integers i and $j = i + 1$ each of which appears 3 or more times in E, there are at least 2 distinct 1-valent vertices, say, r and s, in T, neither of which lies on D, such that $e(r) = i$ and $e(s) = j = i + 1$. Let vertex r be adjacent to, say, vertex R which lies on D, and let vertex s be adjacent to, say, vertex S which lies on D. Note that since $i \neq j$, vertices R and S are distinct. Then, $e(r) = i = e(R) + 1$ and $e(s) = j = i + 1 = e(S) + 1$. This implies that $e(S) = i$, so $e(S) = e(r)$. Therefore, we can prune 1-valent vertex s from C and then attach s to vertex r via an edge to obtain a new tree T whose eccentricity sequence is the same as that of C (namely, E). Observe that in tree T, there is a path P (namely, the path from R to r to s) of length 2 which starts at vertex R on D and includes no other vertex of D and no edge of D. And, since $e(r) > e(R)$ and $e(s) > e(R)$, the positive integer i represented by $e(r)$ is the eccentricity in T of a pair of distinct vertices of D, and the positive integer $j = i + 1$ represented by $e(s)$ is the eccentricity in T of another pair of distinct vertices of D. Since one vertex from each of these two pairs of vertices lies on one side of R along D, and the other vertex from each of these two pairs of vertices lies on the other side of R along D, we see that, in T, $d(R,x) \geq 2$ and $d(R,y) \geq 2$, where x and y are the endpoints of D. Hence, T is a tree which is not a caterpillar, and the proof is complete. □

The next chapter contains more results concerning the centroid, centroid sequence, center, and eccentricity sequence of a tree.

CHAPTER 4 SPANNING TREES, THE CENTROID, AND THE CENTER

4.1 Spanning Trees of a Connected Graph and the Centroid

Recall from Section 1.5 that a **spanning tree** T of a connected graph G is a spanning subgraph of G that is a tree. Note that if a graph G~ is not connected, then G~ does not have any spanning trees.

We consider in this section the following questions: Are there necessary and/or sufficient conditions for each vertex of a connected graph G to be <u>in</u> the (or to <u>be</u> the) centroid of some spanning tree of G? Theorems 4.1.1 and 4.1.2 below provide answers to these questions. The following pair of lemmas is needed for the proofs of these two theorems.

<u>Lemma 4.1.1</u>: A vertex c in a tree T on $n \geq 2$ vertices is one of a pair of adjacent vertices in the centroid of T if and only if the size of a branch of greatest size emanating from c is exactly one greater than the sum of the sizes of all the other branches emanating from c.

Proof: ($\Rightarrow$) Suppose c is one of a pair of adjacent vertices in the centroid of T. Let $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ be the centroid sequence of T, where $n \geq 2$ is the order of T. Then, the weight of c in T is either a_n or a_{n-1}. And, since the centroid of T consists of a pair of adjacent vertices, $a_n = a_{n-1} = n/2$ (since $a_{n-1} + a_n = n$). That is, a branch B of greatest size emanating from c in T has size n/2. Since the size of T is (n-1), the sum of the sizes of all branches other than B which emanate from c in T is $(n-1) - n/2 = (2n - 2 - n)/2 = (n-2)/2$. And, since $n/2 = (n-2)/2 + 1$, the theorem follows. □

($\Leftarrow$) Suppose, conversely, that T is of order $n \geq 2$ and that the size of a branch B of greatest size emanating from c in T is exactly one greater than the sum of the sizes of all branches other than B emanating from c in T. Let size(B) = k. Then, $k + (k-1) = n-1$, where (n-1) is the size of T. That is, $2k = n$, so $k = n/2$. Thus, size(B) = n/2, so the weight a(c) of c in T is n/2.

Claim: Vertex c is in the centroid of T.

Proof of Claim: Suppose not. Let $v \neq c$ be a vertex in the centroid of T. Since c is not in the centroid of T, $a(v) < a(c) = n/2$. If $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ is the centroid sequence of T, then either $a(v) = a_n$ or $a(v) = a_{n-1}$. And, since $a_{n-1} + a_n = n$ and $a_{n-1} \geq a_n$, it must be that $a(v) = a_n$ (since $a(v) < n/2$), which implies that $a_{n-1} > n/2$. But, $a_n < n/2$ and $a_{n-1} > n/2$ together imply that $a_i \neq n/2$ for each $a_i \in A$, since A is arranged in non-increasing order. This contradicts the fact that $a(c) = n/2$. Hence, assuming that c is not in the centroid of T leads to a contradiction. So, vertex c is in the centroid of T, as claimed.

To finish the proof, we must show that c is one of a pair of adjacent vertices in the centroid of T. Since c is in the centroid of T, either $a(c) = a_n$ or $a(c) = a_{n-1}$. And, since $a(c) = n/2$, we have $a_n = a_{n-1} = n/2$. This implies that there is a vertex $c\sim \neq c$ of T which is also in the centroid of T. Since T is a tree, the centroid of T consists of exactly the 2 vertices c and c~, which must be adjacent in T, by Theorem 2.1.1. □

<u>Lemma 4.1.2</u>: A vertex c in a tree T on $n \geq 2$ vertices is the centroid of T if and only if the size of a branch of greatest size emanating from c is ≤ the sum of the sizes of all the other branches emanating from c.

<u>Proof</u>: (⇒) Suppose vertex c is the centroid of T. Let $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ be the centroid sequence of T, where $n \geq 2$. Then, the weight of c in T is a_n, and $a_n < a_{n-1}$. Since $a_{n-1} + a_n = n$, we have $a_n + a_n < n$. Thus, $a_n < n/2$. That is, the size of a branch B of greatest size emanating from c in T is $< n/2$ (i.e., $\text{size}(B) < n/2$). Since $\text{size}(T) = n - 1$, the sum of the sizes of all branches other than B which emanate from c in T is $> (n-1) - n/2 = (2n - 2 - n)/2 = (n-2)/2$. This means that if n is even, $\text{size}(B) \leq (n-2)/2$, and the sum of the sizes of all branches other than B which emanate from c in T is $\geq n/2$. Also, if n is odd, $\text{size}(B) \leq (n-1)/2$, and the sum of the sizes of all branches other

than B which emanate from c in T is $\geq (n-1)/2$. So, if c is the centroid of T, the size of a branch B of greatest size emanating from c is $\leq$ the sum of the sizes of all branches other than B emanating from c, with equality possible if and only if the order n of T is odd. ☐

($\Leftarrow$) Conversely, let T be of order $n \geq 2$, and suppose that the size of a branch B of greatest size emanating from a vertex c of T is $\leq$ the sum of the sizes of all branches other than B emanating from c. Let size(B) = k. Since size(T) = n-1, we have size(B) = a(c) = $k \leq (n-1)/2$. That is, $a(c) \leq (n-1)/2$.

Claim: Vertex c is the centroid of T.

Proof of Claim: Suppose not. Let $v \neq c$ be a vertex in the centroid of T. Then, $a(v) \leq a(c) \leq (n-1)/2$. Let $A = \{a_1, a_2, \ldots, a_{n-1}, a_n\}$ be the centroid sequence of T. Then, either $a(v) = a_n$ or $a(v) = a_{n-1}$. Since $a_{n-1} + a_n = n$ and $a_{n-1} \geq a_n$, it must be that $a(v) = a_n$ (since $a(v) \leq (n-1)/2$ implies that $a_{n-1} \geq (n+1)/2$). But, $a(v) = a_n \leq (n-1)/2$ and $a_{n-1} \geq (n+1)/2$ together imply that $a_i \geq (n+1)/2$ for each $a_i \in A/\{a_n\}$, since A is arranged in non-increasing order. This contradicts the fact that $a(c) \leq (n-1)/2$, since $c \neq v$ implies that the weight a(c) in T of vertex c is an element of $A \backslash \{a_n\}$. Hence, assuming that c is not the centroid of T leads to a contradiction. So, vertex c is the centroid of T, and we are done. ☐

Theorem 4.1.1: Let G be a graph on $n \geq 3$ vertices. Then, each vertex of G is in the centroid of some spanning tree of G if and only if G is 2-connected.

Proof: ($\Rightarrow$) Suppose G is 2-connected. Let v be a vertex of G. Then, for each vertex $w \in V(G)\backslash\{v\}$, there is a cycle C_w in G that both v and w lie on. Let C be a cycle of maximal size in G which contains vertex v (i.e., if C~ is any cycle of G containing v, then size(C) $\geq$ size(C~)). Note that since C is a cycle, order(C) ≥ 3. Let the two vertices of C that are adjacent to v in C be called x and y, respectively.

Claim: For each vertex $z \in V(G)\backslash\{v,x,y\}$,

(i) there exists a path P_{xz}, joining vertices x and z in G, which does not contain vertices v and y; and

(ii) there exists a path P_{yz}, joining vertices y and z in G, which does not contain vertices v and x.

Proof of Claim: If z lies on C, the result is obvious. So, suppose z is a vertex of G which does not lie on C. Since G is 2-connected, there are 2 internally disjoint paths in G joining vertices x and z. If at least one of these 2 paths contains neither v nor y, then we are done. If not, then one of these paths (say, P_1) contains v but not y, and the other (say, P_2) contains y but not v (because if either P_1 or P_2 contains both v and y, then the other contains neither v nor y). Since P_1 is a path in G from x to z which contains v but does not contain y, let a be the last vertex of C which lies on P_1 as we traverse P_1 from x to z. Note that it is possible that $a = v$.

If $a \neq v$, there is a path $P\sim$ from x to a along C which contains neither v nor y. And, by our choice of a, there is a path P_a in G from a to z which contains no vertex of $C\backslash\{a\}$. So, if we join paths $P\sim$ and P_a at vertex a, we obtain a new path $P_{\#}$ from x to z which contains neither v nor y, proving part (i) of the claim.

If, on the other hand, $a = v$, then there is a subpath P_v of P_1 in G from v to z which contains no vertex of $C\backslash\{v\}$. Now, we turn our attention to path P_2. Since P_2 is a path in G from x to z which contains y but not v, let b be the last vertex of C which lies on P_2 as we traverse P_2 from x to z. Note that it is possible that $b = y$.

If $b \neq y$, there is a path $P\sim\sim$ from x to b along C which contains neither y nor v. And, by our choice of b, there is a path P_b in G from b to z which contains no vertex of $C\backslash\{b\}$. So, if we join paths $P\sim\sim$ and P_b at vertex b, we obtain a new path $P_{\#\#}$ from x to z which contains neither y nor v, again proving part (i) of the claim (in case $b \neq y$).

Finally, we consider the case in which $b = y$. Recall that we are simultaneously assuming that $a = v$, for, as we have seen already, if $a \neq v$, we are done with the proof of the first part of the claim. If $b = y$, then there is a subpath P_y of P_2 in G from y to z which contains no vertex of $C\backslash\{y\}$. Note that paths P_v and P_y have only vertex z in common since P_v is a subpath of P_1, P_y is a subpath of P_2, and P_1 and P_2 are two internally disjoint paths that join vertices x and z in G. Consider the cycle C_* in G obtained by removing edge (v, y) from C and adding paths P_v and P_y. Since $\text{length}(P_v) \geq 1$ and $\text{length}(P_y) \geq 1$, $\text{size}(C_*) = (\text{size}(C) - 1) + \text{length}(P_v) + \text{length}(P_y) \geq (\text{size}(C) - 1) + 1 + 1 = \text{size}(C) + 1$. That is, $\text{size}(C_*) > \text{size}(C)$, contradicting our choice of C as a maximal cycle of G which contains vertex v. (Note that C_* contains vertex v!) So, it cannot be that a = v and b = y. Hence, either $a \neq v$ or $b \neq y$ or both, in which case there is a path, say, P_{xz}, in G from x to z which contains neither v nor y. This completes the proof of part (i) of the claim. Part (ii) of the claim is proven in exactly the same way as part (i).

Now, construct a spanning tree T of G as follows. First, recall that the order of G is n, and $n \geq 3$. Then, the size of any spanning tree of G is (n-1). To begin the construction of T, draw cycle C (mentioned above) in red ink. (The spanning tree T will be drawn in blue ink.) Then, starting at vertex x, using blue ink, draw edge (x, v) as well as each edge incident with vertex x except for those that have an endpoint on $C\backslash\{x\}$, one by one, such that the total number of edges drawn does not exceed $[(n-1)/2]$. If the number of edges drawn in blue is $[(n-1)/2]$, stop here for a moment. If, however, the number of edges drawn in blue is $< [(n-1)/2]$, choose an endpoint other than x and v of any edge drawn previously in blue and draw each edge incident with that endpoint except for those that have an endpoint on C, one by one, such that the total number of edges drawn in blue does not exceed $[(n-1)/2]$. Iterate this procedure for all vertices x~ that are adjacent to x in G but that do not lie on C, then for all vertices x~~ that are adjacent to a vertex x~ from the previous step but that also do not lie on C such that no vertex that has already been visited is re-visited, etc. If the number of vertices drawn in blue is still $< [(n-1)/2]$, draw edge (x, u) in blue, where u is the only vertex of C besides v that is adjacent to x on C. If the number of blue edges drawn is now $[(n-1)/2]$, stop here for a moment. If the number of blue edges drawn is still $< [(n-1)/2]$, repeat the above "blue-edge-adding" procedure using u as the starting point this time. Keep adding blue edges until a total of exactly $[(n-1)/2]$ blue edges have been drawn. Note that this is possible by the claim (part (i)), since for each vertex q of G which does not lie on C, there are at least 2 internally disjoint paths in G which start at q and end at non-adjacent vertices of C (the reason for this non-adjacency is the maximality of C). Note that a new edge of C is drawn in blue only when all "neighbors" (that have not already been visited and that are not on C) of the last vertex that was visited on C have been exhausted (i.e., visited) and the number of edges drawn is $< [(n-1)/2]$. Moreover, note that the above procedure produces a branch B of G which is rooted at v such that $\text{size}(B) = [(n-1)/2]$.

Now, the number of vertices of G that do not lie on B is $n - ([(n-1)/2] + 1)$. If n is odd, $n - ([(n-1)/2] + 1) = n - (n-1)/2 - 1 = (2n - n + 1 - 2)/2 = (n-1)/2$ is the number of vertices of G that do not lie on B. If n is even, $n - ([(n-1)/2] + 1) = n - (n-2)/2 - 1 = (2n - n + 2 - 2)/2 = n/2$ is the number of vertices of G that do not lie on B.

Now, construct a branch B~ rooted at v which is disjoint from B except for vertex v by following the same procedure as that followed in constructing branch B above, but this time start by drawing edge (v, y) and continuing until no new edge can be drawn without creating a cycle. (Note that the restriction in the construction of B that size(B) = [(n-1)/2] has been lifted in the construction of B~; i.,e., there is no special requirement on the size of B~.) If the construction of B~ exhausts all the vertices of G, then branches B and B~ together form a spanning tree T of G. If there are vertices of G that lie on neither B nor B~, then there must be one or more branches B~~, B~~~, etc. that can be constructed, starting at v, which are pairwise disjoint from each other, B, and B~, except for vertex v, and such that every vertex of G lies on exactly one of these branches. The reason for this is that for each vertex $t \in V(G)$ which does not lie on branch B, there is at least one path P_t in G from t to either a vertex of C that does not lie on B or to v, such that $V(P_t) \cap V(B) = \emptyset$ or $\{v\}$ (if not, either the 2-connectedness of G would be violated or t would have to lie on branch B). If t also does not lie on B~, then t must lie on a branch emanating from v which is determined by neither x nor y, since t was not reached by B and could not be reached by B~. Note that in this case, each vertex on the path in T from t to v, other than vertex v, does not lie on branch B or branch B~, by the way we constructed these 2 branches. Hence, every vertex of G lies on a unique branch of T. So, T is a spanning tree of G rooted at vertex v. Recall that size(B) = [(n-1)/2]. If n is odd, then size(B) =

(n-1)/2. This implies that the sum of the sizes of all branches other than B emanating from v in T is (n-1) - (n-1)/2 = (n-1)/2. By Lemma 4.1.2, vertex v is the centroid of T (in case n is odd). If n is even, then size(B) = (n-2)/2. This implies that the sum of the sizes of all branches other than B emanating from v in T is (n-1) - (n-2)/2 = (2n - 2 - n + 2)/2 = n/2. Now, if the number of branches of T, other than branch B, which emanate from v is 1, then there are exactly two branches of T which emanate from v: one (namely, B) of size (n-2)/2, and one of size n/2. Since n/2 = (n-2)/2 + 1, v is one of a pair of adjacent vertices of T in the centroid of T in this case, by Lemma 4.1.1. If, instead, the number of branches of T, other than B, is > 1, then the size of any one of these branches is at most (n-2)/2, and the sum of the sizes of all of them is n/2. Hence, by Lemma 4.1.2, v is the centroid of T in this case. So, whether n is odd or even, vertex v is in the centroid of T. That is, the tree T that we have constructed is a spanning tree of G such that vertex v is in the centroid of T, and we are done. □

(⇐) Conversely, suppose that G is a graph on n ≥ 3 vertices which is not 2-connected. Then, G is not a cyclic block, so G is either a disconnected graph or a 1-connected graph.

If G is disconnected, then the number of components of G is > 1. This implies that G contains no spanning tree, so the theorem is clearly true in this case.

If, on the other hand, G is 1-connected, then either G contains one or more 1-valent vertices, or it doesn't. If x is a 1-valent vertex of G, then x must also be a 1-valent vertex of any spanning tree T of G. And, since the order of G is ≥ 3, the order of any spanning tree T of G is ≥ 3. This implies that vertex x is not in the centroid of any spanning tree T of G, since the centroid of any tree of order ≥ 3 does not contain any of its 1-valent vertices. That is, if G contains a 1-valent vertex, then that vertex is not in the centroid of any spanning tree of G. Otherwise, the valence of each vertex of G is ≥ 2. Since G is only 1-connected, G contains a cut-vertex v. So, G-v is disconnected (and non-trivial since the order of G is ≥ 3). Note that we can partition G into a set of 2 or more induced subgraphs of G as follows. For each component Q of G-v, consider the subgraph Q~ of G induced by V(Q) U {v}. Note that G can be recovered by simply amalgamating all such induced subgraphs Q~ of G at vertex v. Now, since $n \geq 3$, at least one of the induced subgraphs mentioned above must be such that its order, excluding vertex v, is $< n/2$. Let Q_1~ be such an induced subgraph of G of minimum order. Note that since G contains no 1-valent vertices, the order of Q_1~, including vertex v, is at least 3.

Claim: If w is any vertex of Q_1~ which is adjacent to v in Q_1~, then there is no spanning tree of G such that w is in the centroid of that spanning tree.

Proof of Claim: Note that any path from w to a vertex $y \in V(G)\backslash V(Q_1\sim)$ must pass through vertex v. If T is any spanning tree of G, then the size of a branch B of greatest size emanating from w in T is > n/2, since B contains every vertex of $V(G) \backslash V(Q_1\sim) \cup \{v\}$. That is, size(B) > n/2. So, the weight a(w) of vertex w in T is > n/2. Now, consider vertex v. Since each branch emanating from v in T is contained in one and only one of the induced subgraphs Q~ mentioned above (since v is a cut-vertex of G), the maximum possible size of a branch B_v of greatest size emanating from v in T is given by (the number of vertices of an induced subgraph [say, $Q_2\sim$, where $Q_2\sim \neq Q_1\sim$] of G of greatest order) - 1, since v is the vertex of T from which branch B_v emanates. Therefore, $\text{size}(B_v) \leq \text{order}(Q_2\sim) - 1 < \text{order}(Q_2\sim) \leq \text{size}(B)$, since branch B, which emanates from w in T, includes every vertex of each induced subgraph $Q\sim \neq Q_1\sim$ of G, and vertex v is one of these vertices. That is, $\text{size}(B_v) < \text{size}(B)$, and since B_v is a branch of greatest size emanating from v in T and B is a branch of greatest size emanating from w in T, we have, by definition, a(v) < a(w) in T, which proves that w is not in the centroid of T. Hence, we have shown that if T is any spanning tree of G, vertex w is not in the centroid of T. Note that we could have concluded that w is not in the centroid of T much earlier in the proof, when we proved that a(w) > n/2 in T, since we could have, at that point, invoked the fact that the weight of a vertex in the centroid of a tree T on $n \geq 2$ vertices must be $\leq n/2$. □

The following theorem provides a sufficiency condition for each vertex of a graph G to be the centroid of some spanning tree of G.

Theorem 4.1.2: If G is a 3-connected graph, then each vertex of G is the centroid of some spanning tree of G.

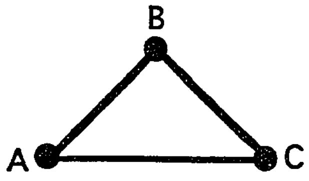
Proof: Let $v \in V(G)$. Since G is 3-connected, there are at least 3 internally disjoint paths in G from v to any other vertex of G. In particular, this means that $\text{val}(v) \geq 3$.

We'll show now how to construct a spanning tree T of G for which vertex v is the centroid of T. Let w be any vertex which is adjacent to v in G. Consider the graph G-w. Since G is 3-connected, G-w is 2-connected. So, by Theorem 4.1.1, there is a spanning tree T* of G-w such that vertex v is in the centroid of T*. Note that this means that the size of a branch B of greatest size emanating from v in T* is either exactly 1 greater than or ≤ the sum of the sizes of all the other branches emanating from v in T*.

Now, consider graph G once again. G is obtained from G-w by replacing vertex w and all the edges of G that were incident with vertex w in G. In particular, recall that v and w are adjacent in G. So, draw tree T* (described above) in G, and then add edge (v, w) to T*. This produces a new tree T which contains all the vertices of G and is, therefore, a spanning tree of G. In addition, edge (v, w) constitutes a branch of size 1 emanating from v in T, so the size of this branch is ≤ the size of each branch of T emanating from v. Thus, by the way we constructed T, a branch B of greatest size emanating from v in T* is also a branch of greatest size emanating from v in T. And, since size(B) in T* is either exactly 1 greater than or ≤ the sum of the sizes of all the other branches emanating from v in T*, we have size(B) in T is ≤ the sum of the sizes of all the other branches emanating from v in T. This implies that vertex v is the centroid of T, by Lemma 4.1.2, completing the proof. ☐

Remark 4.1.1: Observe that the converse of Theorem 4.1.2 is clearly false, since each vertex of any odd cycle C_{2n+1} (e.g., a triangle or a pentagon), which is only 2-connected, is certainly the centroid of some spanning tree of C_{2n+1}. In fact, the unique spanning tree of which it is the centroid is a path, symmetrically placed with respect to that vertex. See Figure 4.1.1.

Figure 4.1.1: Each vertex of triangle ABC is the centroid of the spanning tree obtained by deleting the side (edge) opposite that vertex from triangle ABC.



In Section 4.2, we will see whether theorems similar to those proved in the present section hold true when the word "centroid" is replaced by the word "center".

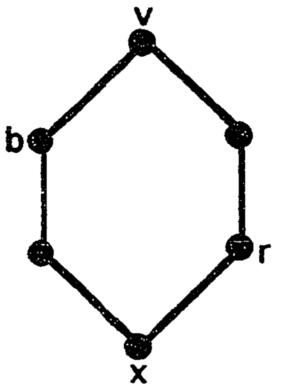
4.2 Spanning Trees of a Connected Graph and the Center

In this section, we will discover that 2-connectedness is a sufficient, but not necessary, condition for each vertex of a graph G to be in the center of some spanning tree of G. Here's the formal statement and proof of this theorem.

Theorem 4.2.1: If G is a 2-connected graph, then each vertex of G is in the center of some spanning tree of G.

Proof: Let G be 2-connected, and let $v \in V(G)$. Since G is 2-connected, each pair of vertices of G lie on a common cycle of G. So, given any vertex $w \in V(G)\backslash\{v\}$, there is a cycle in G that both v and w lie on. Let C_k be a cycle of greatest size (= k) within G that vertex v lies on. Since $k \geq 3$, there is a vertex $x \in V(C_k)$ such that $d(v,x) \geq d(v,y)$ in C_k for all $y \in V(C_k)$. If k is even, then $d(v,x) = k/2$ in C_k. If k is odd, then $d(v,x) = (k-1)/2$ in C_k. Note that C_k consists of 2 internally disjoint paths joining vertices v and x. Let P_1 be a geodesic in C_k joining vertices v and x. Let P_2 be the path in C_k which starts at v, is internally disjoint from P_1, and which terminates at the uniquely determined vertex r which is adjacent to x in C_k and which does not lie on P_1 (see Figure 4.2.1). Note that $\text{length}(P_1) = \text{length}(P_2) + 1$, if k is even, and $\text{length}(P_1) = \text{length}(P_2)$, if k is odd. In addition, $d_{C_k}(v,x) \geq e_G(v)$.

Figure 4.2.1: Path P_1 starts at vertex v, moves down the left side of cycle C_k, and terminates at vertex x. Path P_2 starts at vertex v, moves down the right side of C_k, and terminates at vertex r.



Cycle C_k

We will now construct a spanning tree T of G such that vertex v is in the center of T. Starting from vertex v, draw paths P_1 and P_2 in G. Note that if paths P_1 and P_2 together contain all the vertices of G, then we are done and v is either one of a pair of adjacent vertices in the center of T, if k is even, or the center of T, if k is odd. Otherwise, draw all edges of G that are incident with v and do not intersect C_k at any vertex other than v. If no such edge exists in G, skip the remainder of this paragraph and proceed to the next paragraph. If none of these paths (of length 1) just drawn is of length $\geq [k/2]$, where $[k/2]$ = the greatest integer $\leq k/2$, then at each endpoint reached by an edge drawn in the previous step, draw all edges that are incident with that endpoint (i.e., vertex) other than those that have an endpoint on C_k or those that re-visit a vertex that has already been visited by an edge. Iterate this process until either there is a path in $G \backslash E(C_k)$ of length [k/2] or no edge can be drawn without either intersecting C_k or visiting a previously visited vertex of G.

Now, starting at the unique vertex (say, b) which is adjacent to v on path P_1, do the same thing that we did in the previous paragraph (in which we started at vertex v), but this time stop when the length of any path drawn in this step which emanates from b is [k/2] - 1 (since the corresponding path which starts at vertex v would have length [k/2]). Note, of course, that at no time do we ever re-visit any vertex of G which has already been visited by an edge.

Next, we start at the unique vertex (say, d) which is adjacent to v on P_2 and do the same thing that we did starting at vertex b in the previous paragraph. We then move to the unique vertex b~ which is 2 units away from v on P_1 and iterate the above procedure, always stopping when the length of any path drawn, starting from v, has length [k/2] (or less, if this process produces only paths of length < [k/2]). Then move to the unique vertex d~ which is 2 units away from v on P_2 and iterate the above procedure. Continue this process, moving alternately from the unique vertex b~~ which is 3 units away from v on P_1, then to the unique vertex d~~ which is 3 units away from v on P_2, etc., until either all vertices of G have been visited by an edge or until we have performed this iterative procedure at every vertex of both P_1 and P_2.

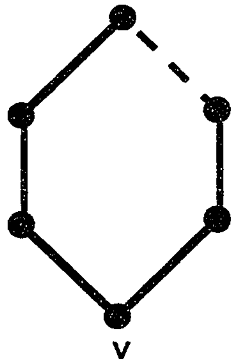
Claim: The above procedure produces a spanning tree T of G.

Proof of Claim: Let $t \in V(G)$. If t lies on either P_1 or P_2, then t clearly lies on the subgraph of G produced by the above procedure. If t lies on neither P_1 nor P_2, then there must be at least 2 internally disjoint paths in G from t to 2 distinct vertices of cycle C_k, by the 2-connectedness of G. This means that vertex t lies on a cycle C together with vertex v in G which includes one or more edges of cycle C_k. By the maximality of cycle C_k, size(C) $\leq$ k. This implies that d(v,t) on cycle C is $\leq [k/2]$. Thus, vertex t lies on the subgraph of G produced by the above procedure. That is, for each vertex $t \in V(G)$, t lies on the subgraph of G generated by the above procedure. So, the subgraph of G so generated is a spanning subgraph of G. And, because the algorithm used to construct this spanning subgraph of G calls for the addition of one or more new edges, at each stage, from a vertex of G already visited to one or more vertices of G that have not been visited previously, the spanning subgraph of G produced is a tree (say, T). So, the above procedure produces a spanning tree T of G, proving the claim.

Finally, note that vertex v is an endpoint of both paths P_1 and P_2. If size(C_k) = k is even, then length(P_1) = k/2 and length(P_2) = (k/2) - 1. If size(C_k) = k is odd, then length(P_1) = (k-1)/2 = length(P_2). And, by the way we constructed the spanning tree T of G, every path emanating from vertex v in T is of length $\leq [k/2]$ (where $[k/2] = k/2$ if k is even, and $[k/2] = (k-1)/2$ if k is odd). This means that if k is even, two longest internally disjoint paths emanating from v in T either differ in length by exactly 1, or have exactly the same length (namely, k/2). And, if k is odd, two longest internally disjoint paths emanating from v in T have exactly the same length (namely, (k-1)/2). Hence, by Theorems 3.1.1 and 3.1.2, v is in the center of T, regardless of the value of k. This completes the proof of the theorem. ☐

Remark 4.2.1: Note that if a graph G is 2-connected, it is not true in general that each vertex $v \in V(G)$ is the center of some spanning tree of G. For example, if $G = C_6$ (or if G is any even cycle), then for any vertex $v \in V(G)$, v is not the center of any spanning tree of G (although v is in the center of the spanning tree T of G shown below in Figure 4.2.2).

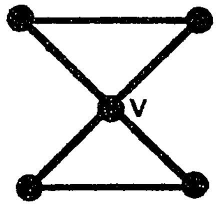
Figure 4.2.2: Vertex v is in the center of the spanning tree T (whose edges are indicated with solid line segments) of C_6. Observe that v is not the center of any spanning tree of C_6.



C_6 and one of its spanning trees T

Remark 4.2.2: The converse of Theorem 4.2.1 is false, as exemplified by the graph G of Figure 4.2.3.

Figure 4.2.3 : Graph G is only 1-connected (since v is a cut-vertex of G), but each vertex of G is in the center of some spanning tree of G.



Graph G

The next theorem indicates that 3-connectedness is a sufficient condition for each vertex of a graph G to be the center of some spanning tree of G.

Theorem 4.2.2: If G is a 3-connected graph, then each vertex of G is the center of some spanning tree of G.

Proof: If G is 3-connected, then G is certainly also 2-connected, so we can repeat the construction of paths P_1 and P_2 found in the first paragraph of the proof Theorem 4.2.1.

Now, if size(C_k) = k is odd, then size(P_1) = (k-1)/2 = size(P_2). In this case, simply repeat the construction of the spanning tree T of G found in the first four paragraphs of the proof of Theorem 4.2.1 (which, again, is possible because G is 3-connected). Note that since k is odd, two longest internally disjoint paths emanating from v in T have the same length (namely, (k-1)/2), so by Theorem 3.1.1, v is the center of T. Hence, in case k is odd, we are done.

If, on the other hand, k is even, we modify the construction of spanning tree T of G as follows. First, repeat the construction of paths P_1 and P_2 found in the first paragraph of the proof of Theorem 4.2.1. Since k is even, length(P_1) = k/2 and length(P_2) = (k/2) - 1. Next, repeat the procedure described in the second, third, and fourth paragraphs of the proof of Theorem 4.2.1, but do not alternate back and forth from a vertex of P_1, then to a vertex of P_2, then back to a vertex of P_1, etc. Instead, this time start at vertex v, and move along only the vertices of path P_2, one by one, at each step making sure that each path generated, starting from v, is of length k/2, or less than k/2 if the construction does not allow for any path from v of length k/2. The reason for this modification in the procedure is that we would like, if possible, to construct a spanning tree of G which contains a path P_3 that starts at v and is internally disjoint from P_1 such that length(P_3) = k/2. After carrying out this procedure at each vertex of path P_2, perform the same procedure at the vertices of path P_1, moving along P_1 from vertex v to vertex x. Note that this modified procedure produces a spanning tree (say, T~) of G for the same reasons that the original procedure produces a spanning tree T of G. Furthermore, if any path starting from v in T~ is internally disjoint from path P_1 (which also starts from v) and has length k/2, then v is the center of T~, by Theorem 3.1.1, and we are done. If there is a path $\neq P_1$ of length k/2 in T~ which is internally disjoint from path P_2, then remove the unique edge of P_1 which is incident with vertex x and add the unique edge of cycle C_k which is incident with vertex x and the endpoint $r \neq v$ of path P_2. This produces a spanning tree of G in which 2 longest internally disjoint paths which start at v both have length k/2, so v is the center of this tree. If there is no path of length k/2 in T~ which starts at v and is internally disjoint from P_1, then each path starting at v in T~ which is internally disjoint from path P_1 is of length $\leq (k/2) - 1$.

If each path besides path P_2 which starts at v in T~ and is internally disjoint from path P_1 is of length $< (k/2) - 1$, then we can modify the construction of T~ as follows. First, remove the unique edge (say, e) of path P_1 which is incident with endpoint x of path P_1. Since G is 3-connected, $\text{val}(x) \geq 3$ (in G). Thus, there must be an edge (say, e~) incident with vertex x such that e~ is not an edge of cycle C_k (although it is possible that both endpoints of e~ lie on C_k). After removing edge e from T~, add edge e~. Note that this produces a new spanning tree T~~ of G. And, since 2 longest internally disjoint paths which start at vertex v in T~ (namely, $P_1 - x$ and P_2) have the same length (namely, $((k/2) - 1)$), vertex v is the center of tree T~~, and we are done in this case.

If, however, the length of each longest path which starts at v in T~ and is internally disjoint from path P_1 is exactly $((k/2) - 1)$, then remove the unique edge e of path P_1 which is incident with endpoint x of path P_1. Since $\text{val}(x) \geq 3$ in G, there is an edge e~ incident with vertex x such that e~ is not an edge of cycle C_k. So, x is one endpoint of e~, and let y be the other endpoint of e~. That is, e~ = (x, y). After removing edge e from T~, add edge e~ to produce a new tree T~~ which also spans G. If the unique path in T~~ from vertex v to vertex x is of length $\leq$ (k/2) - 1, then v is the center of T~~, and we are done. If, however, the unique path in T~~ from vertex v to vertex x is of length exactly k/2, then remove edge e~ and join x to any vertex of G to which it is adjacent such that the unique path from v to x in the resulting spanning tree of G is of length $\leq$ (k/2) - 1, if possible. If this is possible, then v is the center of the spanning tree T~~~ of G so constructed, since the lengths of two longest paths in T~~~ which start at v (namely, paths P_1 - x and P_2) are the same (namely, (k/2) - 1). Otherwise, return to spanning tree T~~ in which d(v, y) is ((k/2) - 1). Since G is 3-connected, val(y) $\geq$ 3. This implies that vertex y is incident with at least one edge e~~ of G which is not an edge of T~~, since y is a 2-valent vertex in T~~. Let t be the only vertex besides x which is adjacent to y in T~~. Also, let e~~ = (y, z). Note that although e~~ is not an edge of T~~, both y and z are vertices of T~~, since T~~ is a spanning tree of G.

Now, if d(v, z) in T~~ is < (k/2) - 2, remove edge (y, t) from T~~ and add edge e~~ instead. Also, remove the unique edge of T~~ that is incident with vertex x and add edge e~ instead. This produces a new spanning tree of G such that 2 longest internally disjoint paths which start at v in this tree are both of length ((k/2) - 1). Hence, v is the center of this spanning tree.

If d(v, z) in T~~ is ((k/2) - 1), remove edge (y, t) from T~~ and add edge e~~ instead. This creates a new spanning tree T# of G.

If vertices y and x lie on different branches of T# which start at v, then v is the center of T# because 2 longest internally disjoint paths in T# which start at v have the same length (namely, k/2). If, however, vertices y and x lie on the same branch of T#, remove the unique edge of path P_1 which is incident with vertex x and add an edge from the endpoint r of path P_2 which is 1-valent in T# (i.e., the endpoint of P_2 which $\neq$ v) to vertex x. This produces another spanning tree T## of G in which 2 longest internally disjoint paths which start at v have the same length (namely, k/2). Hence, v is the center of T##, and we are done in this case.

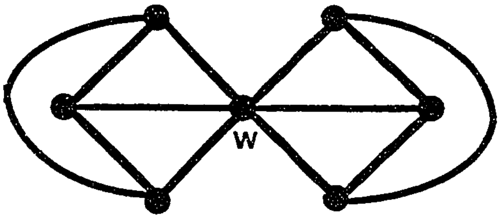
Finally, if d(v, z) in T~~ is ((k/2) - 2), there are 2 cases to be considered.

Case (i): If vertex z does not lie on cycle C_k, then add edge e~~ and remove the unique edge of T~~ which is incident with z and a vertex u of T~~ such that d(v, u) = (k/2) - 3 in T~~. This creates a spanning tree T^+ of G in which a longest path P_4 which starts at v has length either k/2 or ((k/2) + 1). If length(P_4) = k/2 and paths P_4 and P_1 are internally disjoint, then v is the center of T^+ because length(P_4) = length(P_1) = k/2. If paths P_4 and P_1 are not internally disjoint, then remove the unique edge of P_1 which is incident with vertex x and add an edge from the endpoint r $\neq$ v of path P_2 to vertex x. This produces a spanning tree T^{++} of G in which 2 longest internally disjoint paths in T^{++} which start at v have the same length (namely, k/2). Thus, v is the center of T^{++}. If, on the other hand, length(P_4) = ((k/2) + 1), then either P_4 and P_1 are internally disjoint, or they are not. If P_4 and P_1 are internally disjoint, remove the unique edge of P_2 which is incident with endpoint r of P_2 and add edge (x, r) instead. The spanning tree of G thus constructed is such that 2 longest internally disjoint paths which start at v in this tree have the same length (namely, (k/2) + 1)), so v is the center of this tree. If P_4 and P_1 are not internally disjoint, remove the unique edge of path P_1 which joins the vertices that are at distances ((k/2) - 1) and ((k/2) - 2), respectively, from v on P_1 and add edge (r, x). This produces a spanning tree of G in which 2 longest internally disjoint paths which start at v have length ((k/2) + 1), so v is the center of this tree.

Case (ii): If vertex z does lie on cycle C_k, then either vertex t also lies on C_k or it does not. If t does not lie on C_k, we can remove the unique edge of T~~ which is incident with t and a vertex of T~~ which is at distance ((k/2) - 3) from v in T~~. We can then add edge (y, z), producing a spanning tree of G such that a longest path in this tree starting at v has length either k/2 or ((k/2) + 1). So, just as we did in Case (i) above, in either case we can construct a spanning tree of G whose center is vertex v. If, on the other hand, t also lies on C_k, let q be the unique vertex of P_1 which is adjacent to vertex x on P_1. Then, by the 3-connectedness of G, there is an edge e_1 incident with q in G but not in T~~. Similarly, there is an edge e_2 incident with r in G but not in T~~. If e_1 is also incident with any vertex of G other than the vertex g at distance ((k/2) - 2) from v on P_2 or if e_2 is also incident with any vertex of G other than the vertex h at distance ((k/2) - 2) from v on P_1, a spanning tree of G whose center is v can be constructed by strategically adding certain edges to and removing certain edges from T~~. If, however, $e_1 = (q, g)$ and $e_2 = (r, h)$ and there are no other edges of G incident with either q or r in G but not in T~~, observe that either vertex y is adjacent in G to a vertex other than g, h, and x, or vertex x is adjacent in G to a vertex other than q, r, or y, or both, by the 3-connectedness of G. In any case, a spanning tree of G can be constructed in which 2 longest internally disjoint paths which start at v have the same length (namely, either ((k/2) - 1), k/2, or ((k/2) + 1)). Hence, v is the center of this spanning tree. This completes the proof, since every possibility has now been addressed. □

The graph H of Figure 4.2.4 demonstrates that 3-connectedness is definitely not a necessary condition for each vertex of a graph to be the center of some spanning tree of that graph, since H is only 1-connected!

Figure 4.2.4: Graph H is only 1-connected (since w is a cut-vertex of H), but each vertex of H is the center of some spanning tree of H.

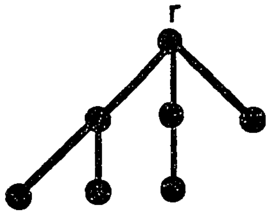


Graph H

Before moving on to Theorem 4.2.3, we state the following definition.

Definition 4.2.1: A **rooted tree** T is a directed tree (i.e., a tree, each of whose edges is directed with an arrow) with some distinguished vertex r (called the **root**) of T such that T contains an r-v (directed) path for each vertex v of T. Note that a rooted tree T with root r contains no v-r (directed) path for each vertex $v \neq r$ of T. Figure 4.2.5 illustrates the concept of a rooted tree.

Figure 4.2.5: T is a rooted tree with root r. Note that when r is placed at the top of T, it is customary to omit the arrows (which all point downward) on the directed edges of T.



Tree T

Definition 4.2.2: The **height** h(T) of a rooted tree T with root r is given by: $h(T) = \max_{v \in V(T)} (\text{length}(P_{rv}))$.

Therefore, the height of the tree T of Figure 4.2.3 is 2 (i.e., h(T) = 2).

We are now prepared to state and prove Theorem 4.2.3.

Theorem 4.2.3: If G is a connected graph, then a minimum height rooted spanning tree of G must be rooted at a central vertex of G.

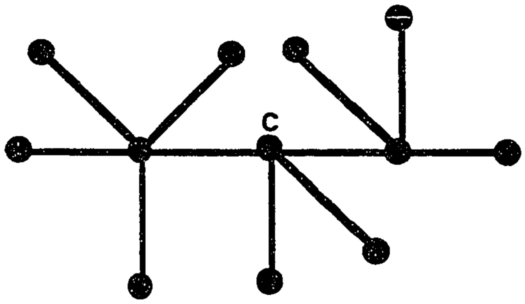
Proof: Suppose that v is a vertex of G which is not in the center of G such that there is a minimum height spanning tree of G rooted at v. Let e(v) be the eccentricity of v in G. Also, let c be a vertex in the center of G, and let e(c) be the eccentricity of c in G. Then, since v is not in the center of G, e(v) > e(c). Hence, any spanning tree of G rooted at v must have height ≥ e(v). On the other hand, there exists a spanning tree of G rooted at c of height e(c). Since e(v) > e(c), we see that a minimum height spanning tree of G could not possibly be rooted at v, and the theorem is proved. ☐

In the following section, we shall present a variety of examples of trees in order to illustrate several facts about the centroid and center of a tree.

4.3 Examples of Trees, Their Centroids and Centers

The purpose of this section is to present several specific examples of trees and connected graphs that further illuminate the concepts of centroid, centroid sequence, center, and eccentricity sequence.

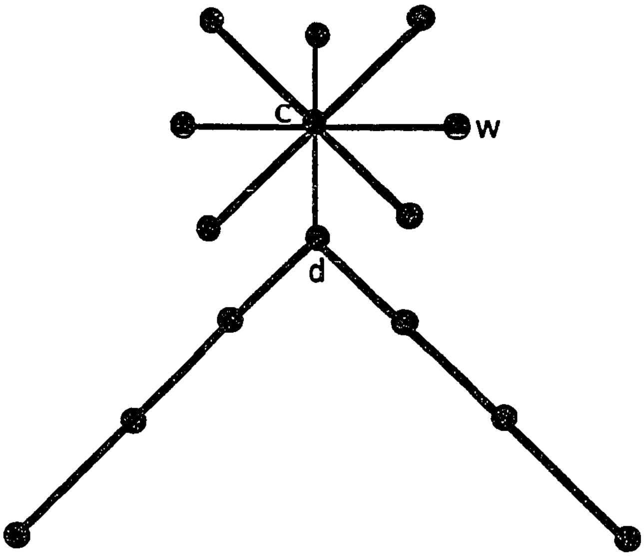
Figure 4.3.1:



Caterpillar C

Example 4.3.1: Figure 4.3.1 shows a caterpillar C, i.e., a tree of order n ≥ 3 whose pruned tree is a (possibly degenerate) path. Observe that the centroid (vertex c) of C lies on the backbone of C. In fact, the centroid of any caterpillar must lie on its backbone.

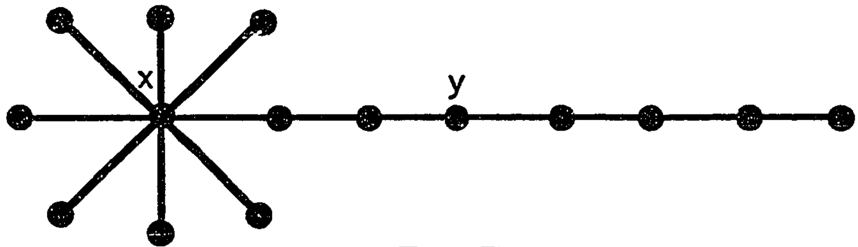
Figure 4.3.2:



Tree T

Example 4.3.2: The tree T of Figure 4.3.2 illustrates the point that the centroid of a tree need not lie on any diameter of T. Note that vertex c is the centroid of T, but c lies on no diameter of T. Recall that the center of a tree lies on every diameter of that tree (see Theorems 3.1.3 and 3.1.4).

Figure 4.3.3:



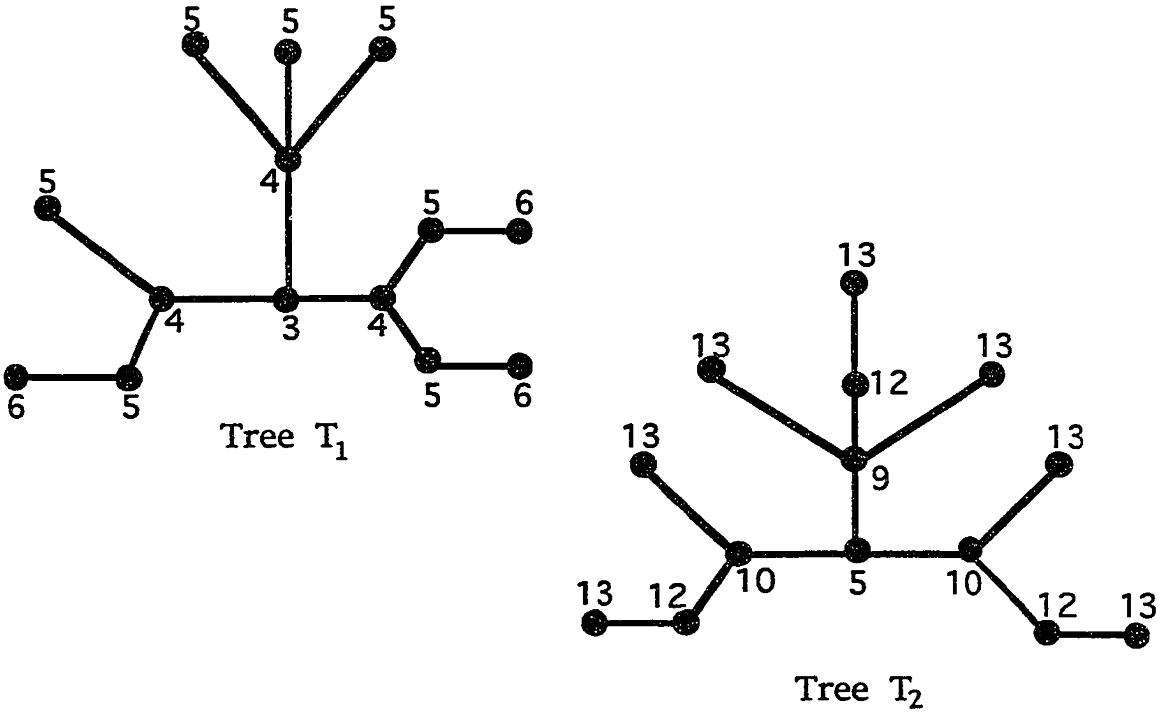
Tree T

Example 4.3.3: Figure 4.3.3 contains a tree T whose centroid (vertex x) and whose center (vertex y) are not adjacent in T. This example provides the motivation for the following theorem, which shows that for each non-negative integer n, there is a tree whose centroid and center are at distance n from each other.

Theorem 4.3.1: For each integer $n \geq 0$, there exists a tree T_n such that the centroid of T_n consists of a single vertex x, the center of T_n consists of a single vertex y, and x and y are at distance n from each other in T_n (i.,e., $d(x,y) = n$).

Proof: If T_n is constructed by first drawing a path P of odd length $(2n+1)$ and then attaching $(2n+1)$ leaves to one of the endpoints of P, then the centroid of T_n will consist of a single vertex x, the center of T_n will consist of a single vertex y, and $d(x,y) = n$. Why? The reason is that the centroid of T_n will be the unique vertex x that has (2n+1) leaves attached to it (since its weight is (2n+1), which is strictly less than the weight of any other vertex in T_n), and the center of T_n will be the unique vertex y which divides each of the (2n+1) diameters D of T_n, consisting of path P together with one of the (2n+1) leaves that were attached to one of the endpoints of P to form T_n, into 2 paths, P_1 and P_2, of equal length. Note that such a vertex y exists and is unique in T_n because each of the (2n+1) diameters D of T_n has length $(2n+1) + 1 = 2n + 2$, which is even, so there exists a unique vertex y which lies on D and divides D into 2 paths P_1 and P_2, each of length (n+1). And, y is the center of T_n because $e(y) < e(v)$ for all $v \in V(T_n)\backslash\{y\}$. Finally, since the distance from y to any of the (2n+1) leaves attached to x is (n+1), we have $d(x,y) = n$. See Figure 4.3.3. □

Figure 4.3.4: Each vertex of Tree T_1 below (left) is labeled with its eccentricity in T_1. Each vertex of Tree T_2 below (right) is labeled with its weight in T_2.

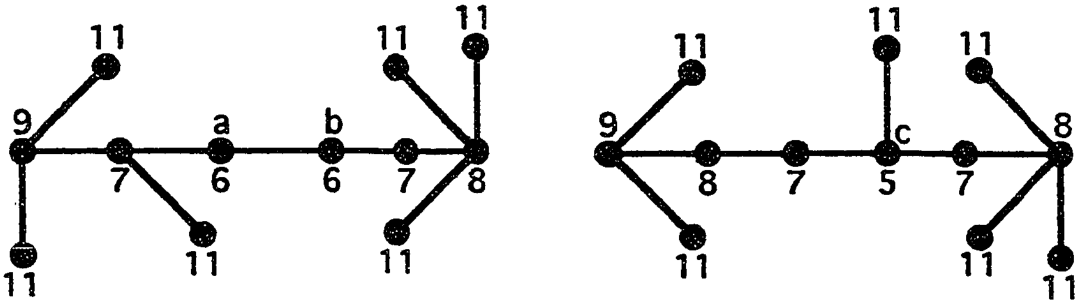


Tree T_1

Tree T_2

Example 4.3.4: The two trees of Figure 4.3.4 have the same centroid sequence {13, 13, 13, 13, 13, 13, 13, 12, 12, 12, 10, 10, 9, 5}, eccentricity sequence {3, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6}, and valence sequence {4, 3, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1}, but these two trees are not isomorphic. This example demonstrates that a centroid sequence, an eccentricity sequence, and a valence sequence (all 3 together!) are not sufficient to uniquely determine a tree.

Figure 4.3.5: In Tree T_1 and Tree T_2 below, the weight of each vertex is indicated next to that vertex.



Tree T_1 Tree T_2

Example 4.3.5: The two trees of Figure 4.3.5 have the same valence sequence $S = \{4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1\}$, but these trees are of different **centroid type** (i.e., the centroid of T_1 consists of a pair of adjacent vertices a and b, while the centroid of T_2 consists of a single vertex c). Thus, a valence sequence alone is not sufficient to uniquely determine the centroid type of a tree.

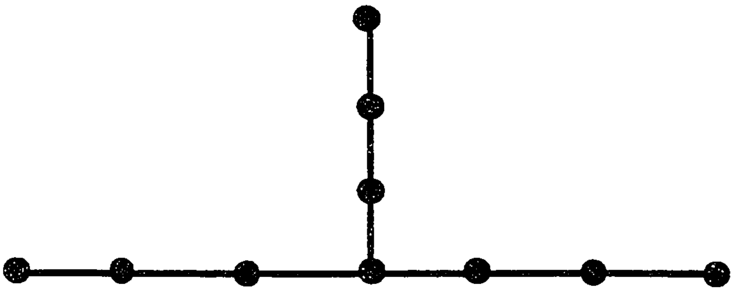
Before moving on to the next example, we make the following **definition**.

Definition 4.3.1 : An **r-star** S_r ($r \geq 0$) is a tree which is obtained as follows:

(i) draw the star on 4 vertices;

(ii) now, place r new vertices on each of the 3 edges of the star drawn in (i).

Figure 4.3.6:



Tree S_2 is called a 2-star.

Example 4.3.6: The tree S_2 of Figure 4.3.6 is a 2-star. Note that S_2 must be pruned two times in order to reduce it to a caterpillar. More generally, an r-star must be pruned r times in order to reduce it to a caterpillar. The following two theorems make connections between an r-star and its centroid sequence.

Theorem 4.3.2: If $A = \{a_1, a_2, ..., a_n\}$ is the centroid sequence of an r-star S_r ($r \geq 0$), then S_r is the only tree, up to isomorphism, whose centroid sequence is A.

Proof: The r-star S_r ($r \geq 0$) is a tree on $n = 3r + 4$ vertices. Its centroid sequence is

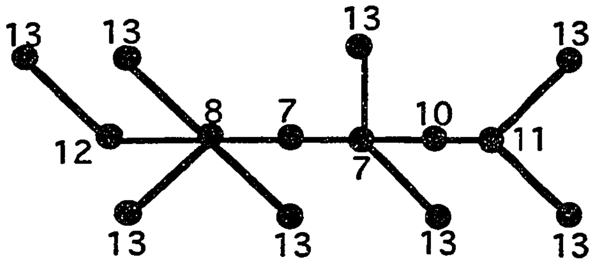
$A = \{3r+3, 3r+3, 3r+3, 3r+2, 3r+2, 3r+2, ..., 2r+3, 2r+3, 2r+3, r+1\}$, where each positive integer j such that $2r+3 \leq j \leq 3r+3$ appears exactly 3 times in A, and the positive integer (r+1) appears exactly once in A. So, the complementary sequence of S_r is

$B = \{1, 1, 1, 2, 2, 2, ..., r+1, r+1, r+1, 2r+3\}$, where each positive integer k such that $1 \leq k \leq r+1$ appears exactly 3 times in B, and the positive integer (2r+3) appears exactly once in B. Note that B is clearly a satisfiable sequence, since each positive integer $k > 1$ which appears in B three times can be satisfied by the positive integer $(k-1)$ which also appears in B three times. And, the positive integer $(2r+3)$, which appears exactly once in B, can be satisfied by two of the three occurrences of the positive integer $(r+1)$ in B, leaving one occurrence of $(r+1)$ and the $(2r+3)$ as the only non-crossed-out elements of B. And, this is the only way that B can be a satisfiable sequence, because the three 2's (present if $r \geq 1$) require the three 1's for satisfaction, then the three 3's (present if $r \geq 2$) require the three 2's for satisfaction (since the three 1's have already been crossed out to satisfy the three 2's), etc., then the three $(r+1)$'s require the three r's for satisfaction (since all elements of B that are $< r$ have already been crossed out), and finally the single $(2r+3)$ requires two of the three $(r+1)$'s for satisfaction (since all elements of B that are $< (r+1)$ have already been crossed out), leaving one occurrence of $(r+1)$ and the $(2r+3)$ as the only non-crossed-out elements of B. Since there is only one way for B to be satisfied, there is only one tree, up to isomorphism, whose complementary sequence is B, and, hence, whose centroid sequence is A. And, since A is the centroid sequence of the r-star S_r $(r \geq 0)$, we see that S_r is the only tree, up to isomorphism, whose centroid sequence is A. □

Theorem 4.3.3: For each positive integer r, there exists a finite sequence of positive integers $A = \{a_1, a_2, \ldots, a_n\}$ which is realizable as the centroid sequence of a tree such that every tree T whose centroid sequence is A must be pruned exactly r times to reduce T to a caterpillar.

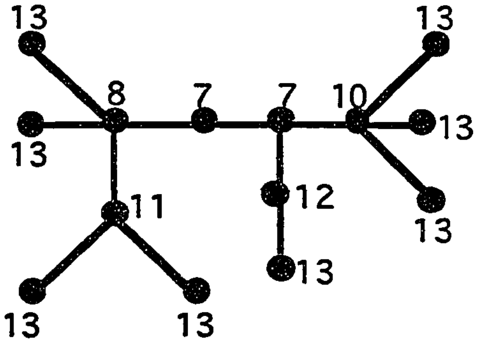
Proof: Let r be any positive integer, and let $A = \{a_1, a_2, ..., a_n\}$ be the centroid sequence of an r-star S_r. Then, by Theorem 4.3.2, S_r is the only tree on n vertices, up to isomorphism, whose centroid sequence is A. And, S_r must be pruned exactly r times to reduce it to a caterpillar (more specifically, to reduce it to the star on 4 vertices). □

Figure 4.3.7:



Caterpillar T_1

(a)

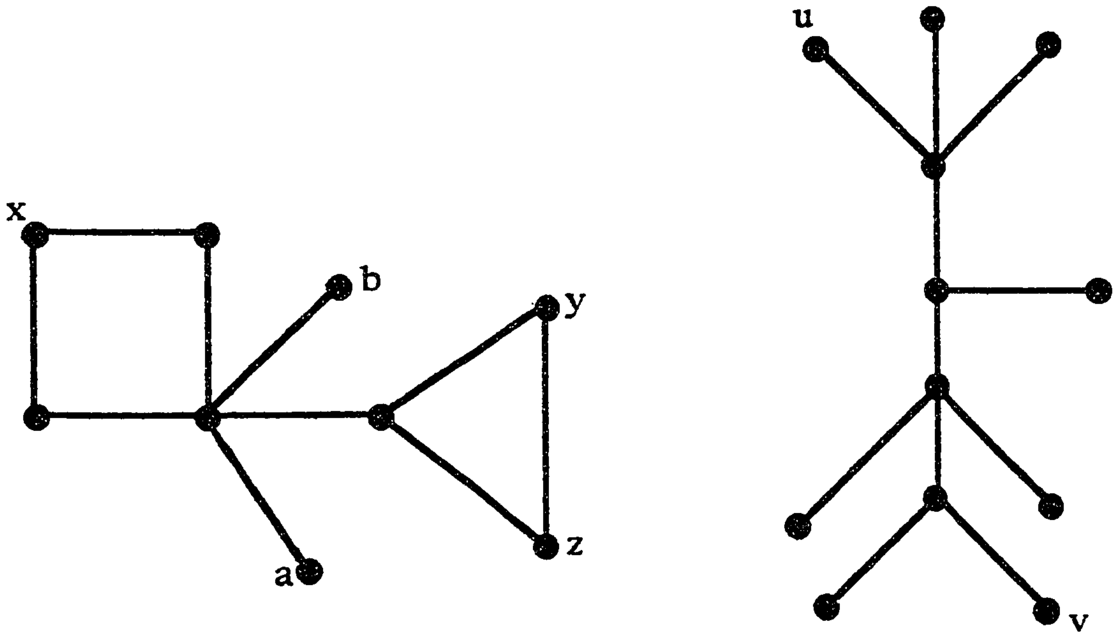


Tree T_2

(b)

Example 4.3.7: The two trees of Figure 4.3.7 have the same centroid sequence A = {13, 13, 13, 13, 13, 13, 13, 13, 12, 11, 10, 8, 7, 7}, but T_1 is a caterpillar and T_2 is not a caterpillar. For a more detailed discussion of this phenomenon, see Theorem 2.3.4.

Figure 4.3.8:



Graph G
(a)

Tree T
(b)

Example 4.3.8: In the connected graph G of Figure 4.3.8(a), the endpoints (e.g., vertices x and y) of any diameter of G are not 1-valent vertices, even though G contains 1-valent vertices a and b. Thus, the endpoints of a diameter of a connected graph which is not a tree need not be 1-valent vertices. By contrast, the endpoints of a diameter of a non-trivial tree T must both be

1-valent vertices. See Figure 4.3.8(b), in which 1-valent vertices u and v are endpoints of a diameter of T.

Figure 4.3.9: In Square Q and Graph G below, the Slater number of each vertex is indicated next to that vertex.

Square Q
(a)

Graph G
(b)

Example 4.3.9: The centroid of a connected graph which is not a tree may consist of more than 2 vertices. In fact, the centroid of the square Q of Figure 4.3.9(a) consists of all of its 4 vertices. Additionally, the centroid of a connected graph which is not a tree can consist of a pair of non-adjacent vertices, as in the graph G of Figure 4.3.9(b), in which the centroid of G consists of non-adjacent vertices u and v.

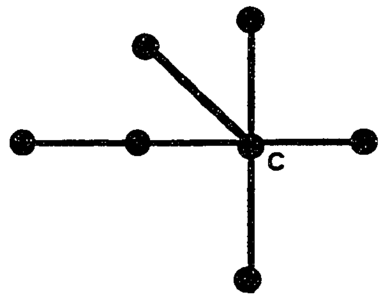
Figure 4.3.10:

Tree T
(a)

Tree T~
(b)

Example 4.3.10: In Figure 4.3.10(a), the centroid of tree T is vertex v, and $a_T(v) = 2$. In Figure 4.3.10(b), the centroid of tree T~ is vertex v~, and $a_{T\sim}(v\sim) = 2$. That is, $a_T(v) = 2 = a_{T\sim}(v\sim)$. Also, by the theorem of Zelinka (see Section 1.6), vertex v is the median of T and vertex v~ is the median of T~. Note, however, that while $a_T(v) = a_{T\sim}(v\sim)$, $s_T(v) \neq s_{T\sim}(v\sim)$, since $s_T(v) = 5$ and $s_{T\sim}(v\sim) = 6$. This example illustrates the fact that given two trees, each with the same number of vertices, a vertex v in the centroid (or, equivalently, in the median) of the first tree and a vertex v~ in the centroid of the second tree may have the same centroid number, and yet have different statuses (i.e., the status of v in T may differ from that of v~ in T~).

Figure 4.3.11: In tree T below, vertex c is the centroid of T. Observe that there are 5 branches that start at c, and these branches have sizes 2,1,1,1, and 1, respectively.



Tree T

Example 4.3.11: If T is a tree on $n \geq 3$ vertices whose centroid consists of a single vertex $c \in V(T)$, then two branches B_1 and B_2 of T which start at c and whose sizes are greatest amongst the sizes of all branches of T which start at c need not be of the same size [i.e., size(B_1) need not = size(B_2)], as in the tree T of Figure 4.3.11. Compare Theorem 3.1.1 and Theorem 4.3.4.

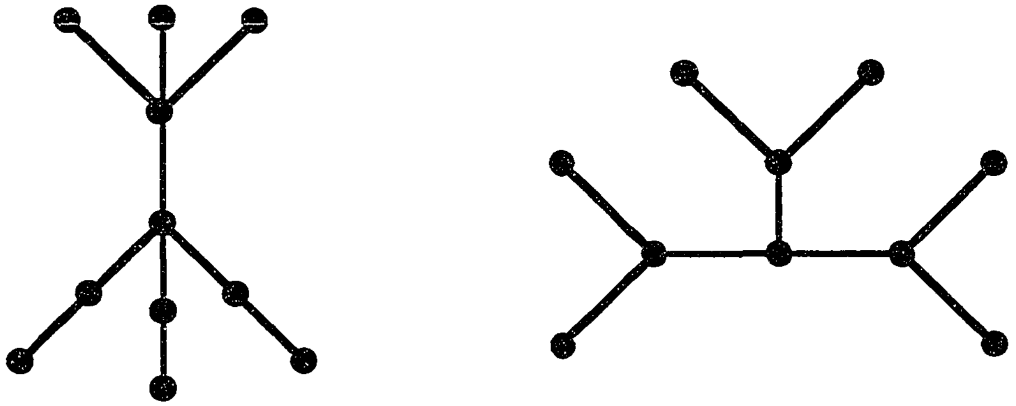
Theorem 4.3.4: Let T be a tree on $n \geq 3$ vertices, and let $v \in V(T)$. If there are two distinct branches B_1 and B_2 of T which start at v such that size(B_1) = size(B_2) and size(B_1) $\geq$ size(B) for all branches B of T that start at v, then v is the centroid of T.

Proof: Suppose, on the contrary, that v is not the centroid of T. Let $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ be the centroid sequence of T. Then, $a(v) \geq a_{n-1} \geq a_n > 0$. And since, by Theorem 2.2.1, $a_{n-1} + a_n = n$, we have $a(v) \geq n/2$. This implies that a branch B~ of T of greatest size emanating from v must have size at least n/2. So, the sum of the sizes of all branches other than B~ emanating from v in T is $\leq$ (n - 1) - n/2 = (2n - 2 - n)/2 = (n - 2)/2. In particular, the size of any single branch B# $\neq$ B~ of T emanating from v is $\leq$ (n - 2)/2, contradicting the hypothesis that there are two distinct branches B_1 and B_2 of T which start at v such that size(B_1) = size(B_2) and size(B_1) $\geq$ size(B) for all branches B of T that start at v. Hence, v must be <u>the</u> centroid of T. $\square$

We now define a new sequence which can be associated with the 1-valent vertices, or **leaves**, of a tree on n $\geq$ 2 vertices. Recall that a tree of order n $\geq$ 2 has at least two 1-valent vertices.

<u>Definition 4.3.2</u>: Let T be a tree on n $\geq$ 2 vertices, and let $X = \{x_1, x_2, ..., x_k\}$ ($k \leq n$) be the set of leaves of T. Observe that there are ${}_kC_2$ (i.e., "k choose 2") ways to choose a pair of distinct vertices x_i and x_j from X such that $1 \leq i, j \leq k$. Define the **leaf distance sequence** $D = \{d_1, d_2, ..., d_{{}_kC_2}\}$ of T to be the sequence of distances in T between each pair of leaves of T, where D is arranged in non-decreasing order (i,.e., $d_1 \leq d_2 \leq ... \leq d_{{}_kC_2}$).

Figure 4.3.12:



Tree T_1

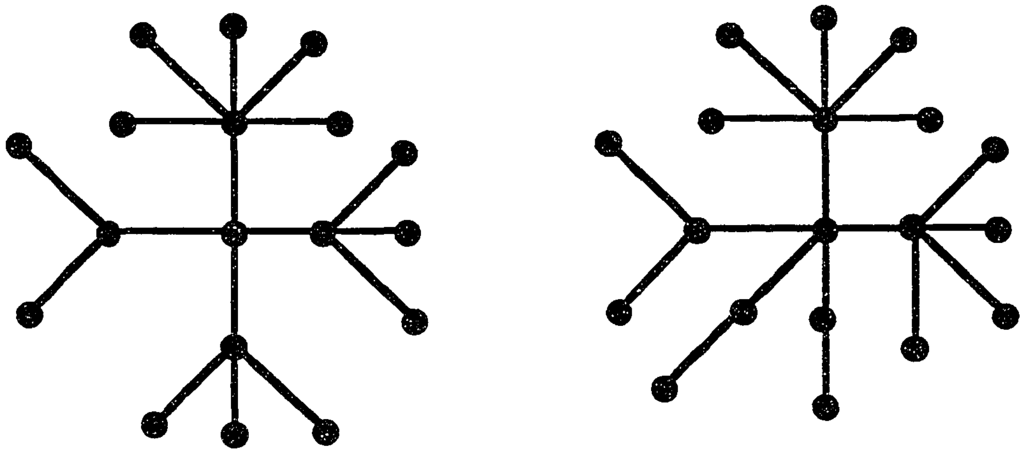
(a)

Tree T_2

(b)

Example 4.3.12: The tree T_1 of Figure 4.3.12(a) contains six leaves, so the leaf distance sequence D of T_1 contains ${}_6C_2 = 15$ elements. The leaf distance sequence of T_1 is D = {2, 2, 2, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4}. Observe that D is also the leaf distance sequence of the tree T_2 of Figure 4.3.12(b), but T_2 is not isomorphic to T_1. This example illustrates the fact that the leaf distance sequence of a tree is insufficient to uniquely determine that tree.

Figure 4.3.13:



Tree T

(a)

Tree T~

(b)

Example 4.3.13: Tree T of Figure 4.3.13(a) and tree T~ of Figure 4.3.13(b) have the same leaf distance sequence D = {2, 2, ..., 2, 4, 4, ..., 4}, where "2" occurs 17 times in D and "4" occurs 61 times in D, and the maximum valence of any vertex of T is the same as the maximum valence of any vertex of T~ (namely, 6), but T and T~ are non-isomorphic trees.

We conclude this section with one more example. Before doing so, we make the following two definitions.

Definition 4.3.3: Let T be a tree on n vertices, and let v be a vertex of T. We define the **mean branch weight at v**, denoted **MBW(v)**, as follows: MBW(v) = |E(T)| / val(v). That is, MBW(v) = (n-1)/ val(v).

Definition 4.3.4: Let T be a tree on n vertices, and let v be a vertex of T. Define the **total branch weight deviation from the mean at v, denoted TD(v),** as follows:

$$\mathrm{TD}(v) = \sum_{i=1}^{\mathrm{val}(v)} |\mathrm{MBW}(v) - \mathrm{size}(B_{v_i})|,$$

where $\mathrm{size}(B_{v_i})$ is the size of the i^{th} branch of T which starts at v. Thus, TD(v) is the sum of the deviations from the mean branch weight at v of the size of each branch of T which starts at v.

Example 4.3.14: Refer, once again, to the tree T of Figure 4.3.2. Note that vertex c is the centroid of T.

$\mathrm{TD}(c) = (7)(14/8 - 1) + |14/8 - 7| = (7)(3/4) + 21/4 = 42/4$
$= 10.5.$

Vertex d is a non-1-valent vertex of T which is not in the centroid of T, and $\mathrm{TD}(d) = (2)(|14/3 - 3|) + |14/3 - 8| = (2)(5/3) + 10/3 = 20/3 \approx 6.67$. So, $\mathrm{TD}(d) < \mathrm{TD}(c)$. This example demonstrates that if a vertex c is in the centroid of a tree T, then it need not be true that $\mathrm{TD}(c) \le \mathrm{TD}(v)$ for each v V(T). In fact, given any 1-valent vertex $w \in V(T)$, $\mathrm{TD}(w) = |(n-1) - (n-1)| = 0 \le \mathrm{TD}(v)$ for each $v \in V(T)$. So, the total branch-weight deviation from the mean at a vertex v of a tree T does not necessarily take its minimum value at the centroid of T.

4.4 More About the Centroid, Centroid Sequence, Center, and Eccentricity Sequence of a Tree

We now present a collection of theorems that provide more information about the centroid, centroid sequence, center, and eccentricity sequence of a tree.

Theorem 4.4.1: Let $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ $(n \geq 2)$ be the centroid sequence of a tree T. Then, the sum of any two elements of A exceeds any single element of A.

Proof: Since A is the centroid sequence of tree T, $|V(T)| = n$ and the elements of A are arranged in non-increasing order (i.e., $a_1 \geq a_2 \geq \cdots \geq a_{n-1} \geq a_n$). Also, $a_{n-1} + a_n = n$. Hence, for any $i, j \in \{1, 2, ..., n\}$ such that $i \neq j$, $a_i + a_j \geq a_{n-1} + a_n = n > a_k$ for all $k \in \{1, 2, ..., n\}$, since the weight of each vertex of T is bounded above by $(n - 1)$ = the weight of each 1-valent vertex of T. □

We will need to recall the following two definitions from Section 1.5.

(i) A **star** S is a tree on $n \geq 3$ vertices whose pruned tree S^* is the trivial tree.

(ii) A non-trivial **path** P is a tree on $n \geq 2$ vertices that contains exactly two 1-valent vertices.

We now introduce the following notation to facilitate the statement and proof of Theorems 4.4.2 through 4.4.5.

Definition 4.4.1: Let T be a tree on n vertices, and let $A_T = \{a_1, a_2, ..., a_n\}$ be the centroid sequence of T. Define $\sum(A_T)$ to be the sum of the elements of A_T. That is, $\sum(A_T) = \sum_{a_i \in A_T} a_i$.

Theorem 4.4.2: If S is the star on $n \geq 2$ vertices, then $\sum(A_S) \geq \sum(A_T)$ for any tree T on the same number n of vertices.

Proof: If n = 2 or 3, the result is clear because the star is the only tree, up to isomorphism, on n = 2 or 3 vertices.

If $n \geq 4$, there are n elements in the centroid sequence A of any tree T on n vertices, so there are $n - 2 \geq 4 - 2 = 2$ elements in $A\backslash\{a_{n-1}, a_n\}$. For the star S on $n \geq 4$ vertices, $a_i = n - 1$ for all $a_i \in A\backslash\{a_{n-1}, a_n\}$. And, since the greatest possible weight that any vertex of any tree on n vertices can have is (n-1), we see that for $n \geq 4$, the $i^{\underline{th}}$ element a_i of the centroid sequence A_S of the star S on n vertices is $\geq$ the $i^{\underline{th}}$ element $a_i\sim$ of the centroid sequence A_T of any tree on n vertices for each i = 1, 2, ..., n-2.
That is,

(1) $$(n - 1) = a_i \geq a_i\sim \quad \text{for each } i = 1, 2, ..., n-2.$$

Also, the sum of the last two elements of the centroid sequence of any tree (including the star) on $n \geq 4$ vertices is n, so

(2) $$a_{n-1} + a_n = n = a_{n-1}\sim + a_n\sim.$$

Inequality (1) and equation (2) together imply that $\sum(A_S) \geq \sum(A_T)$ in case $n \geq 4$, and the theorem is proved. ☐

Remark 4.4.1: Observe that equality can hold in Theorem 4.4.2 even if T is not a star. For example, if S is the star on 4 vertices and T is the path on 4 vertices, then $\sum(A_S) = 10 = \sum(A_T)$.

Theorem 4.4.3: If S is the star on $n \geq 2$ vertices, then

$$\sum(A_S) = (n - 1)^2 + 1 = n^2 - 2n + 2.$$

Proof: If S is the star on $n = 2$ vertices, then $\sum(A_S) = 1 + 1 = 2 = (2)^2 - 2(2) + 2$. If S is the star on $n \geq 3$ vertices, then the centroid sequence of S is $A_S = \{n-1, n-1, \ldots, n-1, 1\}$, where the positive integer $(n-1)$ occurs $(n-1)$ times in A_S because S contains $(n-1)$ 1-valent vertices. So, $\sum(A_S) = (n - 1)(n - 1) + 1 = (n - 1)^2 + 1 = n^2 - 2n + 2$. ☐

Corollary 4.4.1: If T is any tree on $n \geq 2$ vertices, then

$$\sum(A_T) \leq (n - 1)^2 + 1.$$

Proof: Combine the results of Theorems 4.4.2 and 4.4.3. ☐

Theorem 4.4.4: If P is the path on n vertices, then $\sum(A_P) \leq \sum(A_T)$ for any tree T on the same number n of vertices.

Proof: We'll prove this theorem by induction on the number n of vertices of T. The theorem is clearly true for n = 1, 2, and 3, since the only trees, up to isomorphism, on n = 1, 2, and 3 vertices are paths. So, assume that $\sum(A_P) \leq \sum(A_T)$ is true for all trees on n = k vertices, where $k \geq 3$. To complete the proof, we must show that $\sum(A_P) \leq \sum(A_T)$ continues to hold true for all trees on n = k + 1 vertices. So, let P be the path on (k + 1) vertices, and let T be any tree on (k + 1) vertices. Observe that since $k \geq 3$, path P and tree T each have at least two 1-valent vertices (actually, P has exactly two 1-valent vertices). Let u be a 1-valent vertex of P, and let v be a 1-valent vertex of T. Then, if we prune u from P, we obtain the path P~ on k vertices, and if we prune v from T, we obtain a tree T~ on k vertices. And, by our induction hypothesis, $\sum(A_{P\sim}) \leq \sum(A_{T\sim})$. Now, we can re-attach vertex u to path P~ to recover path P. If the order k of path P~ is even, then the weights of k/2 of the k vertices of P~ increase by 1 in P, and there is a new vertex (namely, u) of weight k in P which wasn't in P~, so $\sum(A_P) = \sum(A_{P\sim}) + 3k/2$. If the order k of path P~ is odd, then the weights of (k + 1)/2 of the k vertices of P~ increase by 1 in P, and there is a new vertex (namely, u) of weight k in P which wasn't in P~, so $\sum(A_P) = \sum(A_{P\sim}) + (3k + 1)/2$. Similarly, we can re-attach vertex v to tree T~ to recover tree T. Note that the smallest number of vertices of T~ have their weights in T~ raised by 1 in T if v is a 1-valent vertex of T which lies on a branch B_c which starts from a vertex c in the centroid of T such that $\text{size}(B_c) \geq \text{size}(B)$ for all branches B of T which start at c. And, the greater the size of B_c, the smaller the number of vertices of T~ whose weights in T~ will be raised by 1 in T. Therefore, since the size of branch B_c in T~ is $\leq k/2$ if k is even and $\leq (k-1)/2$ if k is odd, the minimum number of vertices of T~ whose weights in T~ are raised by 1 in T is k/2, if k is even, and (k + 1)/2, if k is odd. And, since T has a vertex (namely, v) of weight k which wasn't in T~, we see that $\sum(A_T) \geq \sum(A_{T\sim}) + 3k/2$, if k is even, and $\sum(A_T) \geq \sum(A_{T\sim}) + (3k+1)/2$, if k is odd. Hence, if k is even, we have $\sum(A_P) = \sum(A_{P\sim}) + 3k/2 \leq \sum(A_{T\sim}) + 3k/2 \leq \sum(A_T)$, which implies that $\sum(A_P) \leq \sum(A_T)$. And, if k is odd, we have $\sum(A_P) = \sum(A_{P\sim}) + (3k+1)/2 \leq \sum(A_{T\sim}) + (3k+1)/2 \leq \sum(A_T)$, which implies that $\sum(A_P) \leq \sum(A_T)$. Hence, whether k is even or odd, $\sum(A_P) \leq \sum(A_T)$, which is exactly what we needed to show. □

Theorem 4.4.5: If P is the path on $n \geq 2$ vertices, then:

(i) $\sum(A_P) = (3n^2 - 2n)/4$, if n is even;

(ii) $\sum(A_P) = (3n^2 - 2n - 1)/4$, if n is odd.

Proof:

(i) If P is a path on an even number $n \geq 2$ of vertices, then its centroid sequence is {n-1, n-1, n-2, n-2, ..., n/2, n/2}. So,

$$\sum(A_P) = \sum_{i=n/2}^{n-1} 2i = (3n^2 - 2n)/4, \text{ if n is even.}$$

(ii) If P is a path on an odd number $n \geq 3$ vertices, then its centroid sequence is
$\{n-1, n-1, n-2, n-2, \ldots, (n+1)/2, (n+1)/2, (n-1)/2\}$. Thus,

$$\sum(A_P) = (n-1)/2 + \sum_{i=(n+1)/2}^{n-1} 2i = (3n^2 - 2n - 1)/4, \text{ if } n$$

is odd. ☐

Theorems 4.4.2 and 4.4.4 together show that the sum of the weights of the vertices of the star S on n vertices provides an upper bound, and the sum of the weights of the vertices of the path P on n vertices provides a lower bound for the sum of the weights of the vertices of any tree T on n vertices (i.e., $\sum(A_P) \leq \sum(A_T) \leq \sum(A_S)$). What can be said about the sum of the eccentricities of the vertices of a tree on n vertices? The answer to this question can be found in the three theorems that follow our next definition.

Definition 4.4.2: Let T be a tree on n vertices, and let $E_T = \{e_1, e_2, \ldots, e_n\}$ be the eccentricity sequence of T. Define $\sum(E_T)$ to be the sum of the elements of E_T. That is, $\sum(E_T) = \sum_{e_i \in E_T} e_i$.

Theorem 4.4.6: The sum of the eccentricities of the vertices of a tree on n vertices is bounded below by the sum of the eccentricities of the vertices of the star on n vertices and bounded above by the sum of the eccentricities of the vertices of the path on n vertices (i.e., $\sum(E_S) \leq \sum(E_T) \leq \sum(E_P)$).

<u>Proof</u>: If n = 1, 2, or 3, the result is clear because there is only one tree on n vertices, up to isomorphism, for each n = 1, 2, or 3. For $n \geq 4$, the eccentricity sequence of the star S on n vertices is $E_S = \{1, 2, ..., 2\}$, where the number of times that "2" occurs in E is (n - 1). Now, let T be any tree on $n \geq 4$ vertices. If $E_T = \{e_1, e_2, ..., e_n\}$ is the eccentricity sequence of T, then $e_1 \geq 1$ because each element of E_T is a positive integer. Furthermore, $e_2 > 1$ because the only tree whose eccentricity sequence contains more than one "1" is the star on 2 vertices. So, $e_2 \geq 2$. And, since the elements of E_T are arranged in non-decreasing order, we see that $e_i \geq 2$ for each $e_i \in E$ with i = 2, 3, ..., n. Therefore, each element e_i of E_T is $\geq$ the corresponding (i.e., the i<u>th</u>) element of E_S, which implies that $\sum(E_S) \leq \sum(E_T)$.

To complete the proof, we must show that for each tree on $n \geq 4$ vertices, $\sum(E_T) \leq \sum(E_P)$. We'll do this by induction on the number n of vertices of T. We have already seen that $\sum(E_T) \leq \sum(E_P)$ holds for n = 1, 2, or 3. Suppose that $\sum(E_T) \leq \sum(E_P)$ also holds for n = k, where $k \geq 3$. We must then show that $\sum(E_T) \leq \sum(E_P)$ also holds for n = k + 1. So, let P be the path on (k+1) vertices, and let T be any tree on (k+1) vertices. Since $n \geq 3$, P and T each contain at least two 1-valent vertices (actually, P contains exactly two 1-valent vertices). If we prune a 1-valent vertex u from P, we obtain the path P~ on k vertices, and if we prune a 1-valent vertex v from T, we obtain a tree $T\sim$ on k vertices. And, by our induction hypothesis, $\sum(E_{T\sim}) \leq \sum(E_{P\sim})$. Now, if k is even, then the eccentricity sequence $E_{P\sim}$ of P is given by $E_{P\sim} = \{k/2, k/2, (k+2)/2, (k+2)/2, \ldots, k-1, k-1\}$. And, if k is odd, then the eccentricity sequence $E_{P\sim}$ of $P\sim$ is given by $E_{P\sim} = \{(k-1)/2, (k+1)/2, (k+1)/2, (k+3)/2, (k+3)/2, \ldots, k-1, k-1\}$. This implies that the eccentricity sequence E_P of P is given by $E_P = \{k/2, (k+2)/2, (k+2)/2, (k+4)/2, (k+4)/2, \ldots, k, k\}$, if k is even, and by $E_P = \{(k+1)/2, (k+1)/2, (k+3)/2, (k+3)/2, \ldots, k, k\}$, if k is odd. This implies that $\sum(E_P) = \sum(E_{P\sim}) + k + k - k/2 = \sum(E_{P\sim}) + 3k/2$, if k is even, and $\sum(E_P) = \sum(E_{P\sim}) + k + k - (k-1)/2 = \sum(E_{P\sim}) + (3k+1)/2$, if k is odd.

Now, what is the greatest possible difference between $\sum(E_T)$ and $\sum(E_{T\sim})$, where T is any tree on (k+1) vertices and $T\sim$ is the tree obtained by pruning a 1-valent vertex from T? Observe that T may be recovered from $T\sim$ by replacing the 1-valent vertex, say, v, that was pruned from T. Note that the eccentricities of the greatest number of vertices of $T\sim$ will be raised if vertex v is re-attached to an endpoint of a longest path which emanates from a vertex in the center of $T\sim$. There are 2 cases to be considered. Case (i): If the center of $T\sim$ consists of a single vertex c, then there are at least 2 paths in $T\sim$ which emanate from c, are internally disjoint (i.e., share no edge and share only vertex c), and each of which has length e(c). So, if we attach 1-valent vertex v to the 1-valent endpoint of any of these 2 or more internally disjoint paths in $T\sim$ which emanate from c and have length e(c), where $e(c)$ is the eccentricity of c in $T\sim$, we can at best increase the eccentricity of every vertex of $T\sim$, with the exception of all the vertices that lie on path P_{cv} (which joins vertices c and v in T) other than the endpoints c (whose eccentricity is raised by 1) and v (which is a vertex which was not present in $T\sim$ and whose eccentricity in T is $1 + e(c) + e(c) = 1 + 2\cdot e(c)$), by 1. The reason for this is that in a tree, a longest path which emanates from a vertex in that tree must include the center of that tree. So, in this case, the greatest possible value for $\sum(E_T)$ is given by:

$\sum(E_T) = \sum(E_{T\sim}) + (k - e(c)) + (1 + 2\cdot e(c))$

$= \sum(E_{T\sim}) + k + e(c) + 1$ [where $(k - e(c))$ is the number of vertices of $T\sim$ whose eccentricities are raised by 1, and $(1 + 2\cdot e(c))$ is the eccentricity of v in T]

$\leq \sum(E_{T\sim}) + k + (k-1)/2 + 1$ [because the center of $T\sim$ consists of a single vertex]

$= \sum(E_{T\sim}) + (3k+1)/2.$

And, since $\sum(E_T)$ is an integer, $\sum(E_T) \leq \sum(E_{T\sim}) + [(3k+1)/2]$, where $[(3k+1)/2]$ = the greatest integer in $(3k+1)/2$.

Case (ii): If the center of $T\sim$ consists of 2 adjacent vertices c_1 and c_2, then there is a longest path P_1 which emanates from c_1 in $T\sim$ and a longest path P_2 which emanates from c_2 in $T\sim$ such that $\text{length}(P_1) = e(c_1) = e(c_2) = \text{length}(P_2)$. So, if we attach 1-valent vertex v to the 1-valent endpoint of P_1 (or P_2), we can at best increase the eccentricity of every vertex of T~, with the exception of all the vertices that lie on P_1 (or P_2) other than vertex c_1 (or c_2), whose eccentricity is raised by 1, by 1. Also, $e(v) = (1 + e(c_1)) + (e(c_2) - 1) = e(c_1) + e(c_2) = 2 \cdot e(c_1)$. Again, the reason underlying the above is that in a tree, a longest path which emanates from a vertex of that tree must contain the center of that tree. Therefore, in this case, the greatest possible value for $\sum(E_T)$ is given by:

$\sum(E_T) = \sum(E_{T\sim}) + (k - e(c_1)) + 2 \cdot e(c_1) = \sum(E_{T\sim}) + k + e(c_1)$

$\leq \sum(E_{T\sim}) + k + k/2$ [because the center of T~ consists of 2 adjacent vertices]

$= \sum(E_{T\sim}) + 3k/2.$

And since $\sum(E_T)$ is an integer, $\sum(E_T) \leq \sum(E_{T\sim}) + [3k/2]$, where $[3k/2]$ = the greatest integer in $3k/2$.

Hence, in either case,

$\sum(E_P) = \sum(E_{P\sim}) + [(3k+1)/2] \geq \sum(E_{T\sim}) + [(3k+1)/2] \geq \sum(E_T)$.

That is, $\sum(E_P) \geq \sum(E_T)$, and we are done. ☐

Theorem 4.4.7: If S is the star on $n \geq 3$ vertices, then

$$\sum(E_S) = 2(n - 1) + 1 = 2n - 1.$$

Proof: If $n \geq 3$, then the eccentricity sequence E of the star S on n vertices is $E = \{1, 2, ..., 2, 2\}$, where the number of times that the positive integer 2 occurs in E is $(n-1)$, since S contains $(n-1)$ 1-valent vertices, each of which is adjacent to the only non-1-valent vertex v of S. Hence, $\sum(E_S) = 2(n - 1) + 1$

$= 2n - 2 + 1 = 2n - 1.$ □

Theorem 4.4.8: If P is the path on $n \geq 2$ vertices, then:

(i) $\sum(E_P) = (3n^2 - 2n)/4$, if n is even;

(ii) $\sum(E_P) = (3n^2 - 2n - 1)/4$, if n is odd.

Proof: Let v be any vertex of P. Since P is a path on $n \geq 2$ vertices, a longest path P_v which starts at v in P is the same as a branch B_v of greatest size which starts at v in P. Therefore, the eccentricity e(v) and the weight a(v) of vertex v in P are the same (i.e., e(v) = a(v)). Since v is an arbitrary vertex of P, we have e(v) = a(v) for each $v \in V(P)$. Thus, the eccentricity sequence $E = \{e_1, e_2, ..., e_n\}$ and the centroid sequence $A = \{a_1, a_2, ..., a_n\}$ of P are equal multi-sets (i.e., E = A as multi-sets). Hence, $\sum(E_P) = \sum(A_P)$, and the theorem follows, by Theorem 4.4.5. □

The following theorem relates the sum of the eccentricities of the vertices of a tree T of order $n \geq 3$ to the sum of the valences of the vertices of that tree.

Theorem 4.4.9: The sum of the eccentricities of the vertices of a tree T on $n \geq 3$ vertices exceeds the sum of the valences of the vertices of that tree.

Proof: According to Theorem 4.4.6, if $\sum(E_S)$ = the sum of the eccentricities of the vertices of the star S on $n \geq 3$ vertices and $\sum(E_T)$ = the sum of the eccentricities of the vertices of any tree T on the same number n of vertices, then $\sum(E_S) \leq \sum(E_T)$. And, according to Theorem 4.4.7, for $n \geq 3$, $\sum(E_S) = 2(n - 1) + 1 = 2n - 1$.

Now, the sum of the valences of the vertices of any tree T on n vertices is given by $2 \cdot |E(T)| = 2(n - 1) = 2n - 2$. And, since for $n \geq 3$ we have $2n - 2 < 2n - 1 = \sum(E_S) \leq \sum(E_T)$, the theorem is proved. □

Our next theorem provides a new way to characterize a path.

Theorem 4.4.10: A tree T on $n \geq 2$ vertices with centroid sequence $A = \{a_1, a_2, ..., a_n\}$ and eccentricity sequence $E = \{e_1, e_2, ..., e_n\}$ is a path if and only if A = E as multi-sets.

Proof: (⇒) Suppose T is a path on $n \geq 2$ vertices. Then, for each vertex v of T, each subtree of T which starts at v is actually a maximal subpath of T which has v as an endpoint. Therefore, a subtree T~ of T which starts at v such that size(T~) ≥ size(T#) for all subtrees T# of T which start at v (of which there are at most 2) and a path P~ in T which starts at v such that length(P~) ≥ length(P#) for all paths P# in T which start at v are actually one and the same. Hence, $a(v) = e(v)$ for each vertex v of T, which implies that A = E as multi-sets (note that by the way we defined A and E, the elements of A and the elements of E do not appear in the same order unless T is the path on $n = 2$ vertices), and we are done in this direction. □

($\Leftarrow$) Conversely, suppose T is a tree on $n \geq 2$ vertices which is <u>not</u> a path. Then, T contains at least three 1-valent vertices, say, x, y, and z. And, the weight in T of each of these three 1-valent vertices of T is $(n-1)$. Therefore, the positive integer $(n-1)$ appears at least 3 times in the centroid sequence A of T.

Now, let P_x be a path in T which starts at vertex x such that $\text{length}(P_x) \geq \text{length}(P)$ for all paths P in T which start at x. Since T is not a path, path P_x does not include every edge of T. Therefore, $e(x) < n - 1$. Similarly, $e(y) < n - 1$, $e(z) < n - 1$, and the eccentricity in T of any other 1-valent vertex of T is $< n - 1$. And, if v is any non-1-valent vertex of T, $e(v)$ is clearly $< n - 1$ because each path in T which starts at v contains only 1 of the 2 or more edges of T that are incident with v. Thus, the positive integer $(n-1)$ does not appear in the eccentricity sequence E of T.

Since the positive integer $(n-1)$ appears at least 3 times in A but does not appear at all in E, $A \neq E$ as multi-sets. □

<u>Theorem 4.4.11</u>: Let T be a tree on $n \geq 2$ vertices. Let $A = \{a_1, a_2, \ldots, a_n\}$ be the centroid sequence of T and let $E = \{e_1, e_2, \ldots, e_n\}$ be the eccentricity sequence of T. Then,

$$\sum_{i=1}^{n} a_i \geq \sum_{i=1}^{n} e_i,$$

with equality holding if and only if T is a path.

Proof: According to Theorem 4.4.10, $A = E$ as multi-sets if and only if T is a path. So, if T is a path, then $A = E$. This implies that

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} e_i .$$

If T is not a path, then $A \neq E$ as multi-sets. Let w be a vertex of T. We'll compare e(w) with a(w). Now, e(w) = the length of a longest path P in T which emanates from vertex w. On the other hand, a(w) = the size of a subtree T~ of T which starts at vertex w such that size(T~) ≥ size(T#) for all subtrees T# of T that start at w. If P is a longest path in T emanating from vertex w, then P must terminate in some vertex $x \neq w$ of T. Note that path P is contained in one and only one subtree, say, T_w, of T which emanates from vertex w. Now, if size(T_w) = size(T~) and if subtree T_w and path P coincide, then e(w) = a(w). Otherwise, if size(T_w) < size(T~) or if subtree T_w contains one or more edges which do not lie on path P (or both), then a(w) > e(w). Putting these results together, we have a(w) ≥ e(w). Since w is an arbitrary vertex of T, we see that for each vertex $w \in V(T)$, a(w) ≥ e(w). Hence, if $V(T) = \{v_1, v_2, ..., v_n\}$, where $a(v_i) = a_i$ and $e(v_i) = e_i$ for each $v_i \in V(T)$, then

$$\sum_{i=1}^{n} a_i \geq \sum_{i=1}^{n} e_i .$$

Now, recall from the proof of Theorem 4.4.10 that if T is not a path, then for each 1-valent vertex y of T, a(y) > e(y). And, if T is not a path, V(T) contains at least three 1-valent vertices. So, for at least 3 distinct vertices $y_1, y_2, y_3 \in V(T)$, we have $a(y_k) > e(y_k)$ $(k = 1, 2, 3)$. Hence,

$$\begin{aligned}\sum_{i=1}^{n} e_i = \sum_{w \in V(T)} e(w) &= \sum_{w \in V(T)/\{y_1,y_2,y_3\}} e(w) + (e(y_1) + e(y_2) + e(y_3)) \\ &\leq \sum_{w \in V(T)/\{y_1,y_2,y_3\}} a(w) + (e(y_1) + e(y_2) + e(y_3)) \\ &< \sum_{w \in V(T)/\{y_1,y_2,y_3\}} a(w) + (a(y_1) + a(y_2) + a(y_3)) \\ &= \sum_{w \in V(T)} a(w) = \sum_{i=1}^{n} a(v_i).\end{aligned}$$

That is, if T is not a path, then

$$\sum_{i=1}^{n} e(v_i) < \sum_{i=1}^{n} a(v_i).$$

This gives us the desired conclusion that

$$\sum_{i=1}^{n} a(v_i) \geq \sum_{i=1}^{n} e(v_i),$$

with equality holding if and only if T is a path. ☐

The following theorem indicates a relationship between the number of 1-valent vertices of a non-trivial tree and the number of vertices of any weight k of that tree.

Theorem 4.4.12: Let T be a tree of order $n \geq 2$, and let $A = \{a_1, a_2, \ldots, a_n\}$ be the centroid sequence of T. For each positive integer k, let x_k = the number of elements of A that are equal to k. Then, $x_{n-1} \geq x_k$ for each positive integer k.

Proof: Observe that each element of A is a positive integer. Also, the number of elements of A that are equal to (n-1) is at least 2 (i.e., $x_{n-1} \geq 2$) since T has at least two 1-valent vertices. Therefore, if k is a positive integer such that either no element, exactly 1 element, or exactly 2 elements of A are equal to k, then clearly $x_{n-1} \geq x_k$. So, suppose k is a positive integer such that there are 3 or more elements of A that are equal to k. If $k = n - 1$, then obviously $x_{n-1} \geq x_k$. So, suppose $k \neq n-1$. Then, $k < n - 1$, so each vertex v of T whose weight $a(v) = k$ in T is a non-1-valent vertex of T. Since there are 3 or more elements of A that are equal to k, there are 3 or more vertices of T each of weight k in T. So, each vertex of weight k in T is not in the centroid of T, by Theorem 2.1.1. Therefore, each vertex of weight k in T has a unique branch of greatest size k which starts at that vertex in T, by Lemma 2.3.1. Let v be any vertex of T such that $a(v) = k$. Then, there is a unique branch B_v of greatest size k that starts at v in T, and there is (are) one or more branch(es) other than B_v which start at v in T, each of which contains at least one 1-valent vertex of T. Also, let w be any vertex other than v of T such that $a(w) = k$. Then, there is a unique branch B_w of greatest size k which starts at w in T, and there is (are) one or more branch(es) other than

B_w which start at w in T, each of which contains at least one 1-valent vertex of T. Moreover, the set of vertices that lie on all branches other than B_v that start at v in T is disjoint from the set of vertices that lie on all branches other than B_w that start at w in T, because if these two sets of vertices of T were not disjoint, then T would contain a cycle (i.e., there would be two internally disjoint paths joining vertices v and w in T). Hence, for each vertex of weight k in T, there is at least one 1-valent vertex (which lies on one of the one or more branches that are not of greatest size which start at that vertex of weight k in T) of T. And, for each pair of distinct vertices v and w of weight k in T, the set of 1-valent vertices that lie on all branches other than B_v which start at v in T is disjoint from the set of 1-valent vertices that lie on all branches other than B_w which start at w in T. In addition, there may be one or more 1-valent vertices of T that lie(s) on no branch of non-greatest size which starts at a vertex of weight k in T. Hence, the number of vertices of weight (n-1) in T (i.e., the number of 1-valent vertices of T) $\geq$ the number of vertices of weight k in T. This implies that $x_{n-1} \geq x_k$, and we are done. ☐

<u>Theorem 4.4.13</u>: Let T be a tree of order $n \geq 2$, and let $A = \{a_1, a_2, \ldots, a_n\}$ be the centroid sequence of T. For each positive integer k, let x_k = the number of elements of A that are equal to k. If $x_{n-1} = x_k$ for some positive integer $k \neq n - 1$, then $x_{n-1} = x_j$ for each j such that $k < j \leq n - 1$.

Proof: Suppose $x_{n-1} = x_k$ for some positive integer $k \neq n - 1$. Then, $k < n - 1$. Let v be any vertex of weight k in T, and let B_v be a branch of greatest size k which starts at v in T. Then, each vertex which lies on a branch other than B_v which starts at v in T is of weight > k in T, by Lemma 2.3.1. The total number of vertices that lie on all branches other than B_v which start at v in T and are of weight > k in T is $(n-1) - k$. Moreover, if v and w are distinct vertices of weight k in T, then the set of $(n-1) - k$ vertices of weight > k in T that lie on all branches other than B_v which start at v in T is disjoint from the set of $(n-1) - k$ vertices of weight > k in T that lie on all branches other than B_w (where B_w is a branch of greatest size which starts at w in T) which start at w in T (because if these sets of vertices were not disjoint, then T would contain a cycle). Thus, since there are x_k elements of A that are equal to k, there are at least $x_k((n-1) - k)$ distinct elements of A that are > k. (We say "at least" here because there may be one or more vertices of weight > k in T which lie(s) on no branch of non-greatest size which starts at a vertex of weight k in T.) Now, by Theorem 4.4.12, we know that $x_{n-1} \geq x_j$ for each positive integer j, and, in particular, for each positive integer j such that $k < j \leq n - 1$. Hence, for each positive integer j such that $k < j \leq n - 1$, there are at most x_{n-1} elements of A that are equal to j. Since the number of positive integers j such that $k < j \leq n - 1$ is $(n-1) - k$, we see that there are at most $x_{n-1}((n-1) - k)$ elements of A that are > k. Since $x_{n-1} = x_k$, by hypothesis, there are at least

$x_{n-1}((n-1) - k)$ elements of A that are $> k$ and at most $x_{n-1}((n-1) - k)$ elements of A that are $> k$. This implies that there are exactly $x_{n-1}((n-1) - k)$ elements of A that are $> k$. Hence, for each positive integer j such that $k < j \leq n - 1$, there are exactly x_{n-1} elements of A that are equal to j. This implies that for each positive integer j such that $k < j \leq n - 1$, $x_{n-1} = x_j$, proving the theorem. □

We now turn our attention to some results that relate the order of a tree to its centroid type.

Theorem 4.4.14: If T is a tree on an odd number of vertices, then the centroid of T consists of a single vertex.

Proof: If T is the trivial tree, the result is obvious. So, assume that T has $n \geq 3$ vertices, and let $A = \{a_1, a_2, ..., a_n\}$ be the centroid sequence of T. Then, $a_n \leq a_{n-1}$ and $a_{n-1} + a_n = n$. Since each element of A is a positive integer and since n is odd, a_{n-1} and a_n must be distinct (i.e., $a_{n-1} \neq a_n$). Thus, $a_n < a_{n-1}$, which implies that $a_n < a_i$ for each $a_i \in A \backslash \{a_n\}$, since the elements of A are arranged in non-increasing order. So, there is a vertex (say, v_n) in T whose weight a_n in T is strictly lower than the weight of any other vertex in T. Hence, the centroid of T consists of a single vertex (namely, v_n), and we are done. □

Remark 4.4.2: The converse of Theorem 4.4.12 is false. The star on 6 vertices provides a simple counter-example.

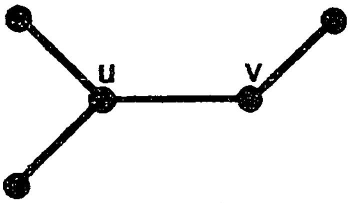
Theorem 4.4.15: If the centroid of T consists of a pair of adjacent vertices, then the order of T is even.

Proof: Let $A = \{a_1, a_2, ..., a_{n-1}, a_n\}$ be the centroid sequence of T. Note that since the centroid of T consists of a pair of adjacent vertices, $|V(T)| = n \geq 2$ and $a_{n-1} = a_n$. And, because A is the centroid sequence of T, $a_{n-1} + a_n = n$. Thus, $a_{n-1} = a_n = n/2$. Since each element of A is a positive integer, in particular, a_{n-1} and a_n are both positive integers. Therefore, $n/2$ is a positive integer, which implies that n is an even positive integer. Hence, the order of T is even. □

Remark 4.4.3: The converse of Theorem 4.4.13 is false. The star on 6 vertices again provides a simple counter-example.

Example 4.4.1: The tree T of Figure 4.4.1 illustrates that no relationship such as those described in Theorems 4.4.11 and 4.4.12 exists between the order of a tree and its center type (i.e., whether its center consists of a single vertex or a pair of adjacent vertices).

Figure 4.4.1: The center of tree T below consists of adjacent vertices u and v. Note that $|V(T)| = 5$ (i.e., the order of T is odd).



Tree T

Our last theorem of this section shows that there is a relationship between the last (i.e., greatest) element e_n in the eccentricity sequence $E = \{e_1, e_2, ..., e_n\}$ of a tree T on n vertices and the center type of T.

Theorem 4.4.16: Let $E = \{e_1, e_2, ..., e_n\}$ be the eccentricity sequence of a tree T on $n \geq 3$ vertices. Then:

(i) the center of T consists of a single vertex if and only if e_n is even;

(ii) the center of T consists of a pair of adjacent vertices if and only if e_n is odd.

Proof:

(i) ($\Rightarrow$) Suppose the center of T consists of a single vertex. Then, $e_1 < e_2$. By Theorem 3.2.1, we must have $e_n = 2 \cdot e_1$, and since e_1 is a positive integer, e_n is even.

($\Leftarrow$) Conversely, suppose that e_n is even. By Theorem 3.2.1, $e_n = 2 \cdot e_1$ and $e_1 < e_2$. Therefore, the center of T consists of a single vertex.

(ii) (⇒) Suppose, now, that the center of T consists of a pair of adjacent vertices. Then, $e_1 = e_2 < e_3$. By Theorem 3.2.1, we must have $e_n = 2 \cdot e_1 - 1$, and since e_1 is a positive integer, e_n is odd.

(⇐) Suppose, conversely, that e_n is odd. By Theorem 3.2.1, $e_n = 2 \cdot e_1 - 1$ and $e_1 = e_2 < e_3$. Hence, the center of T consists of a pair of adjacent vertices. □

4.5 Open Problems Related to the Subject Matter of the Present Dissertation

As is generally the case in the world of mathematics research, consideration of certain questions and, quite often, their answers has led to the formulation of new questions, not all of which have been addressed in this thesis. We conclude this section with an (incomplete) list of open problems related to the subject matter of the present doctoral dissertation.

1) What is the computational complexity of Algorithm 2.1.1 for locating the centroid of a tree?

2) What is the computational complexity of the problem of determining whether a given finite sequence of positive integers is realizable as the centroid sequence of a tree?

3) Is there a set of necessary and sufficient conditions for each vertex of a connected graph G to be:

  a) the centroid of some spanning tree of G?

  b) in the center of some spanning tree of G?

  c) the center of some spanning tree of G?

4) What can be said about the Slater sequence of an outerplanar graph? a complete graph? an r-regular graph (i.e., a graph whose vertices all have the same valence r)?

5) Under what conditions is the centroid sequence of a non-caterpillar tree T also the centroid sequence of another non-caterpillar tree T~ which is not isomorphic to T?

6) What other sequences can be associated with the vertices of a tree? a connected graph? any graph?

7) Can we construct an algebra of centroid sequences? of eccentricity sequences? For example, for each $n \geq 1$, is there a well-defined way to define a binary operation of "addition" of two centroid sequences A_1 and A_2 that each contain the same number n of elements?

8) Is there a pair of non-isomorphic trees, each of order $n \geq 2$, that have the same leaf distance sequence?

9) Can the leaf distance sequence of a tree be characterized?

10) For each positive integer $n \geq 2$, is there a way to generate all possible centroid sequences of a tree of order n ?

11) For which trees T does the operation of pruning preserve the centroid? That is, which trees T have the property that the pruned tree T' of T has the same centroid as T?

12) Is there an “interpolation theorem” of the following type?

Let $n \geq 4$ and let $\sum(A_S)$ = the sum of the weights of the vertices of the star of order n, $\sum(A_P)$ = the sum of the weights of the vertices of the path of order n, $\sum(E_S)$ = the sum of the eccentricities of the vertices of the star of order n, and $\sum(E_P)$ = the sum of the eccentricities of the vertices of the path of order n. Then,

(i) for each positive integer j such that $\sum(A_P) < j < \sum(A_S)$, there exists a tree T of order n such that $\sum(A_T) = j$;

(ii) for each positive integer k such that $\sum(E_S) < k < \sum(E_P)$, there exists a tree T of order n such that $\sum(E_T) = k$.

13) Which centroid sequences are realizable by a unique tree, up to isomorphism?

14) Is there an analogue of the Havel-Hakimi Theorem for centroid sequences?

15) Can two non-isomorphic (k,1) trees have the same leaf distance sequence? (Note: A (k,1) tree is a tree all of whose vertices have valence 1 or k, where k is a positive integer ≥ 2.)

BIBLIOGRAPHY

[1] B. Bollobás, *Graph Theory: An Introductory Course*, Springer-Verlag, New York, 1979.

[2] J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications*, Elsevier Science Publishing Co., Inc., New York, 1976.

[3] F. Buckley and F. Harary, *Distance in Graphs*, Addison-Wesley, New York, 1990.

[4] G. Chartrand and L. Lesniak, *Graphs and Digraphs*, Wadsworth and Brooks, California, 1986.

[5] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide To The Theory of NP-Completeness*, Freeman, San Francisco, 1979.

[6] A.J. Goldman, "Minimax Location of a Facility in a Network", *Transportation Science*, vol. 6, (1972), 407-418.

[7] A.J. Goldman, "Optimal Center Location in Simple Networks", *Transportation Science*, vol. 5, (1971), 212-221.

[8] C. Jordan, "Sur les assemblages de lignes", *J. Reine Angew. Math.*, vol. 70, (1869), 185-190.

[9] D. Konig, *Theorie der endlichen und undendlichen Graphen*, reprinted by Chelsea, New York, 1950.

[10] D. Konig, *Theory of Finite and Infinite Graphs*, translated by R. McCoart, Birkhauser, Boston, 1990.

[11] L. Lesniak, "Eccentric sequences in graphs", *Period. Math. Hungar.*, vol. 6, (1975), 287-293.

[12] P.J. Slater, "Maximin facility location", *J. Res. Natl. Bur. Stand.*, vol. 79B, (1975), 107-115.

[13] B. Zelinka, "Medians and peripherians of trees", *Arch. Math. (Brno)*, vol. 4, (1968), 87-95.