Optimality of Consecutive and Nested Tree Partitions

G. J. Chang, F. K. Hwang

Department of Applied Mathematics, National Chiao Tung University, Hsinchu, 30050 Taiwan, Republic of China

Received 11 August 1995; accepted 2 April 1997

Abstract: We consider the problem of partitioning the vertex-set of a tree to *p* parts to minimize a cost function. Since the number of partitions is exponential in the number of vertices, it is helpful to identify small classes of partitions which also contain optimal partitions. Two such classes, called consecutive partitions and nested partitions, have been well studied for the set partition problem, which is a special case of the tree-partition problem when the tree is a path. We give conditions on the optimality of these classes on tree partitions and also extend our results to tree networks. © 1997 John Wiley & Sons, Inc. Networks **30:** 75–80, 1997

1. INTRODUCTION

Let *T* denote a tree with vertex-set *V* and edge-set *E*. Consider a *p*-partition π which partitions *V* into *p* disjoint nonempty *parts* V_1, \ldots, V_p . The problem is to find an optimal partition for a given *p* which minimizes the cost $C(\pi)$. A *p*-partition is called a *p*-partition if empty parts are allowed.

Suppose that |V| = n. Then, the number of *p*-partitions is the Bell number

$$\frac{1}{p!} \sum_{k=0}^{p-1} (-1)^k \binom{p}{k} (p-k)^n.$$

which grows extremely fast. The number of \overline{p} -partitions is p^n , which is, of course, an even bigger number. Thus,

it is impractical to search an optimal partition by brute force. One approach to significantly reduce the load of searching is to identify a much smaller class of partitions which still contains an optimal partition. Two classes which have been extensively studied in set partition (which can be viewed as a special case of the tree partition when T is a path) are consecutive partitions and nested partitions. Hwang and Chang [4] extended the notion of these classes to graphs. They showed that the number of consecutive partitions is polynomial in n and also gave a linear time algorithm to count the number of nested 2partitions.

In a graph partition problem, it is typical to partition the vertices into subsets to maximize certain internal relations within the subsets. For example, if each vertex represents a component in an integrated circuit and an edge between two components represents the fact that they need be connected, then one would like to partition the components into subsets (chips) of bounded sizes (a chip cannot carry too many components) such that the number of edges external to the chips are minimized. Unfortunately, this problem is well known [3] to be NP-complete even for a partition into two parts. In this paper, we study the

Correspondence to: F. K. Hwang

Contract grant sponsor: National Science Council of the Republic of China

Contract grant number: NSC 82-0208-M009-050

Contract grant sponsor: DIMACS, Rutgers University

graph partition problem for simple graphs, like trees. We give conditions for the optimality of consecutive and nested *p*-partitions. We then extend these results to tree networks, i.e., when points on edges (not necessarily vertices) can be considered. The motivation of this extension is that many set-partition problems involve points lying between set elements (e.g., the mean of a set of numbers). So, the network model is more suitable for them than is the graph model. In particular, we extend our results to the case that the underlying set is a multiset and disprove a conjecture of Boros and Hwang in a clustering problem.

2. SOME PRELIMINARY REMARKS

Let T(V, E) be a tree and let l(e) > 0 denote the length of an edge e. The distance d(u, v) between two vertices u and v is simply the length of the path u - v. We say that T has general lengths if for every vertex v there do not exist two vertices u and w such that d(u, v) = d(w, v). A convex hull H(S) of a subset $S \subseteq V$ is a minimal connected subgraph induced by any subset of V containing S. For a tree, H(S) is the unique minimum Steiner tree of S. A subset S is said to penetrate another disjoint subset S' if $S \cap H(S') \neq \emptyset$. A partition is consecutive if no part penetrates another part; it is *nested* if penetration defines an acyclic digraph with the parts as vertices and an edge from part V_i to part V_i if V_i penetrates V_i .

Let $\pi = (V_1, \ldots, V_p)$ be a partition of V and let K be a k-subset of $\{1, \ldots, p\}$. Define $V(K) = \bigcup_{i \in K} V_i$. Consider a partition property Q, such as consecutiveness or nestedness, which is *hereditary*, i.e., if π has property Q, then for any k and any k-subset K, $\{V_i : i \in K\}$ is a partition of V(K) with property Q. Q is said to be kconsistent if whenever Q holds for any subset K then Q holds for π . A class M of partitions of V is said to be k-Q-sortable locally if for any partition $\pi \in M$ and any k-subset K there exists a partition $\pi' \in M$ which is obtained from π by sorting K into property Q. Q is said to be *k*-sortable if for any class M which is *k*-Q-sortable locally there exists a partition in M with property Q. Hwang et al. [5] proved that k-sortable implies k-consistent. Note that if Q is a k-sortable property then the existence of an optimal *p*-partition with property Q can be established by verifying that the class of optimal p-partitions is k-Q-sortable locally, thus transferring a global condition to a local one.

It is easily seen that consecutiveness is 2-consistent for the tree. Chang et al. [2] proved that it is also 2sortable. Therefore, the optimality of consecutiveness can be proved by inspecting some local (pairwise) condition. However, the subgraph induced by the vertices in $V_k \cup$ V_j may be disconnected—hence, it has no consecutive partition. Therefore, we have to bring in $H(V_k \cup V_j)$ as the underlying connected graph, which may contain vertices not in $V_k \cup V_j$. So when we are dealing with partitions



Fig. 1. A nonnested partition with pairwise nestedness.

of a proper subset $S \subset V$, it is understood that the underlying graph has vertex set $S' \supseteq S$.

Surprisingly, the example in Figure 1 shows that nestedness is not 2-consistent for the tree, hence, not k-sortable, for any $k \ge 2$. To establish the optimality of nested partitions, we need to take a whole different approach by looking at things globally.

3. OPTIMAL PARTITIONS ON TREES

All results in this section deal with the tree T(V, E).

Theorem 1. Suppose that the cost of a partition $\pi = \{V_1, \dots, V_p\}$ is

$$C(\pi) = \sum_{i=1}^{p} \sum_{e \in H(V_i)} l(e).$$

An optimal p-partition can be obtained by deleting a set L of p - 1 longest edges of T. If L is unique, then it is the unique optimal p-partition.

Proof. Let $\pi = \{V_1, \ldots, V_p\}$ be an optimal partition. Since $H(V_i)$ is connected for each *i*, at most p - 1 edges can be missing from $\bigcup_{i=1}^{p} H(V_i)$. Hence,

$$C(\pi) \geq \sum_{e \in \bigcup_{i=1}^{p} H(V_i)} l(e) \geq \sum_{e \notin L} l(e).$$

On the other hand, deleting L yields p components C_1 , ..., C_p . The partition $\pi' = \{C_1, \ldots, C_p\}$ has cost

$$C(\pi') = \sum_{e \notin L} l(e). \qquad \blacksquare$$

Corollary 1. All optimal p-partitions (unique if L is unique) are consecutive.

The distance d(x, y) between two vertices x and y is simply the sum of l(e) over all e on the path from x to y. For a given function f and a subset $S \subseteq V$, a vertex $c \in V$ is called a *weighted f-centroid vertex* of S if

$$\sum_{x\in S} w_x f(d(x, c)) = \min_{y\in V} \sum_{x\in S} w_x f(d(x, y)),$$

where $w_x > 0$ denotes a weight associated with vertex x. In particular, c_i denotes the weighted f-centroid vertex of an nonempty part V_i . We define $\sum_{x \in V_i} w_x f(d(x, c_i))$ = 0 if V_i is an empty part. Note that $\sum_{x \in S} w_x f(d(x, c))$ is a general measure for weighted deviations of a set of points. For example, when T is a path, $f(d) = d^2$ and w_x = $(|S| - 1)^{-1}$, then c is the mean and $\sum_{x \in S} w_x f(d(x, c))$ the variance.

Lemma 1. Suppose that

$$C(\pi) = \sum_{i=1}^{p} W_i \sum_{x \in V_i} w_x f(d(x, c_i)),$$

where $W_i > 0$ is a weight associated with *i*th part of π . Let $\pi = (V_1, \ldots, V_p)$ denote an optimal \overline{p} -partition. Suppose that $v \in V_k$. Then, $W_k f(d(v, c_k)) \leq W_j f(d(v, c_j))$ for all $j \neq k$.

Proof. Suppose to the contrary that there exists a $v \in V_k$ such that $W_k f(d(v, c_k)) > W_j f(d(v, c_j))$. Let $\pi' = \{V'_1, \ldots, V'_p\}$ be a partition obtained from π by switching v to V_j .

$$C(\pi') = \sum_{i=1}^{p} W_i \sum_{x \in V'_i} w_x f(d(x, c'_i))$$

$$\leq \sum_{i=1}^{p} W_i \sum_{x \in V'_i} w_x f(d(x, c_i))$$

$$< \sum_{i=1}^{p} W_i \sum_{x \in V_i} w_x f(d(x, c_i)) = C(\pi),$$

contradicting the assumption that π is optimal.

Corollary 2. Suppose that f is increasing and either W_i = W for i = 1, ..., p, or f(0) = 0. Then, Lemma 1 holds for the *p*-partition.

Proof. Suppose that $|V_k| = 1$. Then, $c_k = v$ and

$$W_k f(d(v, v)) < W_i f(d(v, c_i))$$

by the assumptions on f and W_i . Suppose that $|V_k| \ge 2$. Then, $|V'_k| \ge 1$ and π' remains a *p*-partition. The proof of Lemma 1 remains valid.

Lemma 2. Suppose that $C(\pi) = \sum_{i=1}^{p} W_i \sum_{x \in V_i} w_x f(d(x, c_i))$, where *f* is nonnegative and ln *f* is strictly concave increasing. Then, V_k does not penetrate V_i in an optimal

 \bar{p} -partition if either $W_k < W_j$ or $W_k = W_j$ but T has general lengths.

Proof. Suppose to the contrary that π is an optimal *p*-partition such that V_k penetrates V_j . Let $z \in V_k \cap H(V_j)$ and let *y* be a vertex in V_j such that *z* lies on the path *y* $-c_j$. From Lemma 1,

$$W_k f(d(z, c_k)) \le W_j f(d(z, c_j)), \qquad (1)$$

$$W_j f(d(y, c_j)) \le W_k f(d(y, c_k)).$$
(2)

Since $W_k < W_j$ (or $W_k \le W_j$ and *T* has general lengths) and *f* is increasing, (2) implies that

$$f(d(y, c_j)) < f(d(y, c_k)), d(y, c_j) < d(y, c_k).$$
(3)

From (1) and (2), we also have that

$$\frac{f(d(z, c_k))}{f(d(z, c_j))} \le \frac{W_j}{W_k} \le \frac{f(d(y, c_k))}{f(d(y, c_j))} \,. \tag{4}$$

It follows that

$$\ln f(d(z, c_k)) + \ln f(d(y, c_j))$$

$$\leq \ln f(d(z, c_j)) + \ln f(d(y, c_k)).$$
(5)

Since z is between y and c_i ,

$$d(y, c_j) = d(y, z) + d(z, c_j).$$

Furthermore,

$$d(z, c_k) \ge d(y, c_k) - d(y, z)$$

> $d(y, c_j) - d(y, z)$ (6)
= $d(z, c_j)$.

Therefore,

$$d(z, c_j) < \min\{d(z, c_k), d(y, c_j)\}$$

< max{d(z, c_k), d(y, c_j)} < d(y, c_k).

Furthermore,

$$d(z, c_j) + d(y, c_k) = d(z, c_j) + d(y, z) + d(z, c_k)$$
$$= d(y, c_i) + d(z, c_k).$$

From the strict concavity of $\ln f$,

$$\ln f(d(z, c_k)) + \ln f(d(y, c_j)) > \ln f(d(z, c_j)) + \ln f(d(y, c_k)),$$
(7)

contradicting (5). Hence, V_k cannot penetrate V_j in π .

Corollary 3. If f(0) = 0 is added to the condition $W_k < W_j$, then Lemma 2 applies to *p*-partitions.

Theorem 2. Let

$$C(\pi) = \sum_{i=1}^{p} W_i \sum_{x \in V_i} w_x f(d(x, c_i)),$$

where *f* is nonnegative and ln *f* is strictly concave increasing. Then, every optimal \overline{p} -partition is nested if either the W_i are all distinct or *T* has general lengths.

Proof. Consider the partial order *P* defined on W_i by the "greater than" relation, i.e., W_i and W_j are comparable in *P* if and only if one is greater than the other. Since V_k can penetrate V_j in an optimal *p*-partition only if $W_k > W_j$ in *P*, the penetration relation on the parts also defines a partial order (which is a suborder of *P*). Thus, an optimal partition is nested.

Corollary 4. Suppose that f(0) = 0 is added to the condition that the W_i are all distinct. Then, Theorem 2 applies to *p*-partitions.

Corollary 5. Suppose that $W_i = W$ for i = 1, ..., p, T is of general lengths and ln *f* is strictly concave increasing. Then, every optimal *p*-partition is consecutive.

For consecutiveness, we can use its 2-sortable property to get rid of the nuisance condition that T is of general lengths and also relax the condition on f.

Theorem 3. Suppose that

$$C(\pi) = \sum_{i=1}^{p} \sum_{x \in V_i} w_x f(d(x, c_i)),$$

where f is increasing. Then, there exists a consecutive optimal p-partition.

Proof. Let $\pi = (V_1, \ldots, V_p)$ denote a nonconsecutive optimal *p*-partition. Since consecutiveness is 2-consistent, there exist two parts V_k and V_j such that V_k penetrates V_j . By the 2-sortable property, it suffices to show that we can always sort $V_k \cup V_j$ into a consecutive 2-partition without increasing the cost (hence, preserving optimality). Define $Z = V_k \cap H(V_j)$. We prove Theorem 3 by induction on |Z|.

Let $z \in Z$. Then, there exists a $y \in V_j$ such that z lies on the path $y - c_j$. By Lemma 1, $d(z, c_k) \le d(z, c_j)$ and $d(y, c_j) \le d(y, c_k)$. On the other hand,

$$d(y, c_i) = d(y, z) + d(z, c_i)$$

$$\geq d(y, z) + d(z, c_k) = d(y, c_k)$$

Therefore, $d(y, c_j) = d(y, c_k)$, which, in turn, forces $d(z, c_k) = d(z, c_j)$.

Let $Y = \{y \in V_j: z \text{ lies on the path } y - c_j\}$. Then, *Y* is a proper subset of V_j . Switch *Y* to V_k . Then, *z* is no longer in *Z*. Furthermore, since all vertices on the opposite side of c_j with respect to *z* are in V_k , this switch does not introduce any new vertex into *Z*. Hence, |Z| decreases by at least one. Repeating this operation, eventually *Z* becomes our empty set, or V_j and V_k become consecutive to each other.

4. AN EXTENSION TO NETWORKS

Note that when T is a path then the tree-partition problem is reduced to the well-studied set-partition problem where the vertices can be represented by a set of real numbers. In such a problem, a centroid, e.g., a mean, can be a real number not in the given set. To cover these cases, we need to allow a centroid to be a point in the graph, not necessarily a vertex. Therefore, we need to deal with distances between a point and a vertex.

We now extend the definition of length so that the distance of any two points, not necessarily vertices, on the tree can be quantified. An edge (u, v) is interpreted as a straight line between u and v and we assume that no two edges cross in this representation. A point w on the edge (u, v) can be represented by $w = \lambda u + (1 - \lambda)v$ for some $0 \le \lambda \le 1$, and $l(u, w) = \lambda l(u, v)$. d(x, y) is defined as before except if x(y) is not a vertex then the edge involving x(y) is a partial edge.

For a given function f and a subset $S \subseteq V$, a point c is called a *weighted f-centroid* of S if

$$\sum_{x\in S} w_x f(d(x, c)) = \min_{y\in T} \sum_{x\in S} w_x f(d(x, y)).$$

For a network, the notion of general lengths must be redefined as "for each vertex x and points y, z, d(x, y) = d(x, z) implies y = z." Clearly, the assumption of general lengths is much less tenable in networks than in graphs. But, otherwise, the proof of Theorems 2 and 3 do not depend on c_i being a vertex. Therefore, we have

Theorem 4. Let

$$C(\pi) = \sum_{i=1}^{p} W_{i} \sum_{x \in V_{i}} w_{x} f(d(x, c_{i})).$$

Suppose that the W_i are all distinct and $\ln f$ is strictly concave increasing. Then, every optimal \bar{p} -partition is nested.

Corollary 6. If f(0) = 0 is added to the condition the W_i are distinct, then Theorem 4 applies to *p*-partitions.

Theorem 5. Suppose that

$$C(\pi) = \sum_{i=1}^p \sum_{x \in V_i} w_x f(d(x, c_i)),$$

where f is increasing. Then, there exists a consecutive optimal \bar{p} -partition.

When the tree is a path, we can improve Theorems 4 and 5. We first state a lemma:

Lemma 3. Suppose that T is a path and

$$C(\pi) = \sum_{i=1}^{p} W_i \sum_{x \in V_i} w_x f(d(x, c_i)),$$

where *f* is increasing. Then, $W_k = W_j$ implies that V_k does not penetrate V_j .

Proof. As we proved in Theorem 3, $W_k = W_j$ implies that

$$d(y, c_k) = d(y, c_i)$$

and

$$d(z, c_k) = d(z, c_j).$$

But for two given points, the point equidistant from both of them is unique. This contradicts the fact that $y \neq z$.

Corollary 7. Suppose that *T* is a path and

$$C(\pi) = \sum_{i=1}^{p} \sum_{x \in V_i} w_x f(d(x, c_i)),$$

where f is increasing. Then, every optimal p-partition is consecutive.

We are also able to eliminate the nuisance condition that the W_i are distinct in Theorem 4.

Theorem 6. Suppose that *T* is a path and

$$C(\pi) = \sum_{i=1}^{p} W_i \sum_{x \in V_i} w_x f(d(x, c_i)),$$

where f is nonnegative and $\ln f$ is strictly concave increasing. Then, every optimal \bar{p} -partition is nested.

Proof. If
$$W_k < W_j$$
, then V_k does not penetrate V_j by

Lemma 2, and if $W_k = W_j$, then by Lemma 3. The proof now is analogous to that of Theorem 2.

Corollary 8. If, furthermore, f(0) = 0, then Theorem 6 applies to *p*-partitions.

5. A COUNTEREXAMPLE TO A CONJECTURE IN A CLUSTERING PROBLEM

Can Theorem 4 be strengthened to cover "consecutiveness?" The answer is no even when T is a path. In particular, the following examples disprove a conjecture of Boros and Hwang [1] in a clustering problem that there always exists a consecutive optimal p-partition if

$$C(\pi) = \sum_{i=1}^{p} w_i \sum_{x \in V_i} (x - c_i)^2,$$

where $w_i > 0$.

In a set-partition problem, usually, numbers in the set can repeat themselves, i.e., the given set is a multiset. Furthermore, if two intervals intersect at a unique point (which is necessarily a boundary point for both intervals), then they are still considered disjoint. In terms of the treepartition problem, the partition is on a set $N = \{n_1, \ldots, n_m\}$, where *N* maps into *V*, and a partition $\pi = \{N_1, \ldots, n_m\}$, where *N* maps into *V*, and a partition $\pi = \{N_1, \ldots, n_m\}$ with $\bigcup_{i=1}^{p} N_i = N$. If $N_i \cap H(W_j)$ at a single boundary point of $H(N_j)$, then N_i is not considered to penetrate N_j . For easier presentation, our first example is on a multiset. Note that two identical numbers are considered two different elements and can be partitioned into different parts.

We need to inspect carefully how the multiset extension affects our results. Note that arguments using the general lengths are of suspicion since for two points y and z in N

$$d(x, y) \le d(x, z)$$

no longer forces a strict inequality as we could have y = z. We also need to interpret that "x lies on the path y - z" to mean that x is an internal point of the interval [y, z]. Finally, there can exist more than two points in N with equal distance to two given points, but these points in N must correspond to the same vertex in V. By noting these and changing π_i to N_i , then the results reported in this paper (with the deletion of references to general strengths) except Theorem 1 and Corollary 1 remain valid.

Consider the multiset $\{-1^{2b+100}, 1^{2b}, 9^b, 11^b, 100\}$, where x^y denotes y copies of x and b denotes a large number. Let p = 2 and

$$C(\pi) = \sum_{x \in V_1} (x - c_1)^2 + 2 \sum_{x \in V_2} (x - c_2)^2,$$

where c_i is the centroid of V_i . Thus, $w_1 = 1$, $w_2 = 2$, d(x, y) = |x - y| and $f(z) = z^2$. It is easily verified that *f* is nonnegative, f(0) = 0, and ln *f* is strictly concave increasing. By Corollary 8, an optimal 2-partition must be nested. We will show that $\pi^* = (V_1^*, V_2^*)$, where $V_1^* = \{-1^{2b+100}, 1^{2b}, 100\}$ and $V_2^* = \{9^b, 11^b\}$, a nested but not consecutive partition, is the unique optimal 2-partition. It is easily verified that $c_1^* = 0$, $c_2^* = 10$ and

$$F(\pi^*) = 2b + 100 + 2b + 100^2 + 2(b + b)$$

= 8b + 10100.

Let $\pi = (V_1, V_2)$ be an optimal 2-partition. Then,

Claim 1. $-1 \le \min\{c_1c_2\} \le 2$.

Proof. The lower bound is trivial. To prove the upper bound, suppose to the contrary that $\min\{c_1, c_2\} > 2$. Then, for b large enough,

 $C(\pi) > (2b + 100)(-1 - 2)^2 = 18b + 900$ > $C(\pi^*)$, contradicting the optimality of π .

Claim 2. $8 \le \max\{c_1, c_2\} \le 12.$

Proof. Suppose to the contrary that $\max\{c_1, c_2\}$ > 12 or <8. Then, for *b* large enough,

$$C(\pi) > b(1^2 + 3^2) = 10b > C(\pi^*),$$

contradicting the optimality of π .

We consider the two possible cases induced by the two claims:

CASE (i). $-1 \le c_1 \le 2, 8 \le c_2 \le 12$. Since,

$$(x - c_1)^2 < 2(x - c_2)^2$$

for $x \in \{-1, 1, 100\}$ and

$$(x - c_1)^2 > 2(x - c_2)^2$$

for $x \in \{9, 21\}, \pi = \pi^*$ by Lemma 1.

CASE (ii). $8 \le c_1 \le 12, -1 \le c_2 \le 2$. By Corollary 3,

it is easily verified that $V_1 = \{9^b, 11^b, 100\}$ and $V_2 = \{-1^{2b+100}, 1^{2b}\}$. Then,

$$F(\pi) > 10b > F(\pi^*),$$

contradicting the optimality of π . Therefore, we conclude that π^* is the only optimal 2-partition.

A continuity argument can obviously extend the above "multiset" example to "set." To be specific, we give the following "set" example. The proof is similar and omitted.

Consider the set $\{-100a, -2a, -2a + 1, \dots, -1, 1, 2, \dots, 2a, 19a, 19a + 1, \dots, 20a - 1, 20a + 1, 20a + 2, \dots, 21a, 100a \}$, where *a* is a large number. Then, $V_1^* = \{-100a, -2a, -2a + 1, \dots, -1, 1, 2, \dots, 2a, 100a\}$ and $V_2^* = \{19a, 19a + 1, \dots, 20a - 1, 20a + 1, \dots, 21a\}$ is the unique optimal 2-partition.

6. CONCLUSIONS

We gave sufficient conditions such that the searching of an optimal \overline{p} or *p*-partition on a tree can be restricted to a much smaller class, like the consecutive class and the nested class. When the tree is a path, our results cover the well-studied set partition. In particular, we used our result to construct a counterexample against a conjecture on consecutive optimal *p*-partitions.

If a partition can have any number of parts, it is called an *open partition*. Since an open partition must be a ppartition for some p, our results generalize to the class of open partitions.

REFERENCES

- [1] E. Boros and F. K. Hwang, On the optimality of nested partition. *SIAM J. Optim.*, **6** (1996) 1153–1162.
- [2] G. J. Chang, F. K. Hwang, and Y. C. Yao, Localizing combinatorial properties for partitions on block graphs. *J. Combin. Optim.*, to appear.
- [3] M. R. Garey and D. S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*. W. H. Freeman, San Francisco (1979).
- [4] F. K. Hwang and G. J. Chang, Enumerating consecutive and nested partitions for graphs. *Eur. J. Combin.*, to appear.
- [5] F. K. Hwang, U. G. Rothblum, and Y. C. Yao, Localizing combinatorial properties of partitions. *Discr. Math.*, 160 (1996) 1–23.