A Combinatorial Analysis of k- pseudo Trees

A.Ehrenfeucht, D. Haussier, E. Knill

CU-CS-509-91 January 1991

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

A Combinatorial Analysis of k-pseudo Trees

A.Ehrenfeucht, D.Haussler, E. Knill

CU-CS-509-91 January 1991

Department of Computer Science
University of Colorado at Boulder
Campus Box 430
Boulder, Colorado 80309-0430 USA

(303) 492-7514
(303) 492-2844 Fax

A Combinatorial Analysis of k-pseudo Trees

A. Ehrenfeucht, D. Haussler, E. Knill

Abstract

A k-pseudo tree is a family of sets such that each member U contains at least one element not contained in any member V incomparable to U, and the intersection of $k+1$ incomparable members is empty. We show that the maximum cardinality of a k-pseudo tree consisting of subsets of an n-element set is $(k+1)n - (k+1)k/2$.

Definitions: Let $\mathcal{F}$ be a family of subsets of an n-set X. We will consider $\mathcal{F}$ partially ordered by the subset relation. Thus, $\{U_1, \ldots, U_n\}$ is a *chain* if the U_i are linearly ordered by containment and it is an *antichain* if no U_i is contained in any U_j with $j \neq i$. U and V are *comparable* if either U is contained in V or V is contained in U, else they are *incomparable*.

The *center* of a member U of $\mathcal{F}$ is the set of all elements of U not contained in any member of $\mathcal{F}$ incomparable to U.

$\mathcal{F}$ is a *pseudo tree* if the center of everyone of its members is non-empty and it contains all the singletons and X.

$\mathcal{F}$ satisfies the *k-intersection property* if every antichain in $\mathcal{F}$ of size greater than k has empty intersection.

$\mathcal{F}$ is a *k-pseudo tree* if $\mathcal{F}$ is a pseudo tree and satisfies the k-intersection property.

Further information on families of sets and partial orders can be found in [1], [2] and [4]. [3] discusses commonly studied intersection properties of structures.

Observation 1 *For any familiy $\mathcal{F}$ on X containing X and the singletons, if U is X or a singleton, the center of U is itself.*

Observation 2 *$\mathcal{F}$ is a 1-pseudo tree if and only if it is a tree, that is, if and only if any two of its members are either disjoint or comparable and it contains the singletons (the leaves) and X (the root).*

Observation 3 *In a pseudo tree, the centers of the members of an antichain must be mutually disjoint and also disjoint from the intersection.*

Therefore, an antichain of size k with non-empty intersection contains at least $k+1$ elements in its union.

Observation 4 *A pseudo tree on an n-set is an $n-1$-pseudo tree.*

This implies that if $\mathcal{F}$ is a k-pseudo tree on an n-set we may assume that $k < n$.

Theorem 5 *The maximum cardinality of a k-pseudo tree $\mathcal{F}$ on an n-set X is $(k+1)n - (k+1)k/2$.*

Proof of the lower bound:

Example: Let $X = \{1, 2, \ldots, n\}$ with $n > k$.

$$\mathcal{F} = \{\{i\} \mid i \leq n\} \bigcup \{\{1, \ldots, j\} \cup \{j+l\} \mid 1 \leq j, 1 \leq l \leq k, j+l \leq n\}$$

Note that the second component of $\mathcal{F}$ consists of initial segments of X with one of the next k singletons adjoined.

The cardinality of $\mathcal{F}$ is $n+kn-(k+(k-1)+\ldots+1) = (k+1)n-(k+1)k/2$.

For any set U in $\mathcal{F}$ observe that if V is in $\mathcal{F}$ and V contains the greatest element of U, then either $V \subset U$ or $U \subset V$. Therefore the center of U contains U's greatest member and we can conclude that $\mathcal{F}$ satisfies the second condition on k-pseudo trees.

To show that $\mathcal{F}$ satisfies the k-intersection property suppose that $\{U_1, \ldots, U_l\}$ is an antichain. Let x_i be the greatest member of U_i. Since x_i is in the center of U_i, observation 3 implies that the x_i are distinct. Without loss of generality assume that $x_1 \leq x_2 \leq \ldots \leq x_l$. Then l must be less than or equal to k, for if u and v are the greatest members of U and V respectively, and $v - u > k$ then $U \subset V$ or $U \cap V$ is empty.

Proof of the upper bound: By identifying some of the sets in $\mathcal{F}$ with their centers in a one-to-one way, and then identifying the remaining sets in $\mathcal{F}$ with elements of X in a k-to-one way, we first establish a coarse upper bound of $(k+2)n-1$. To reduce this to the bound given in the theorem, we will define a notion of deficiency for arbitrary centers of fixed size. The deficiency measures the minimum reduction of the size of $\mathcal{F}$ relative to the coarse bound due to such a center.

Observation 6 *Let A be any subset of X. The set of members of $\mathcal{F}$ with center A forms a chain.*

By definition, if U intersects the center of V, then U must be comparable to V.

Lemma 7 *If U is contained in V and U intersects the center of V, then the center of U is contained in the center of V.*

Proof of Lemma 7: If x is any element not in the center of V, there is a set W in $\mathcal{F}$ incomparable to V with $x \in W$. W does not intersect the center of V, so W must be incomparable to U. Therefore x is not in the center of U.□

Corollary 8 *$\mathcal{C}$, the set of centers of members of $\mathcal{F}$, is a tree. That is, any two centers are either comparable or disjoint.*

Proof of Corollary 8: If A and B are centers of U and V respectively, and A and B have non-empty intersection, U and V must be comparable. The result now follows by Lemma 7. □

Lemma 9 *If $\mathcal{G}$ is a tree on X, then the cardinality of $\mathcal{G}$ is at most $2n-1$.*

Proof of Lemma 9: Consider the members of $\mathcal{G}$ smallest sets first. That is, write $\mathcal{G} = \{U_1, \ldots, U_l\}$ where U_i is maximal in $\mathcal{G}_i = \{U_1, \ldots, U_i\}$ and the singletons are listed first. Let $\mathcal{M}_i$ be the set of maximal members of $\mathcal{G}_i$. Suppose $i \geq n$. Then $\mathcal{M}_i$ is a partition of X. Since the sets in $\mathcal{G}$ are distinct and U_{i+1} is maximal in $\mathcal{G}_{i+1}$, $\mathcal{M}_{i+1}$ is strictly coarser than $\mathcal{M}_i$. We obtain $l - n \leq n - 1$, thus $l \leq 2n-1$.□

If A is in $\mathcal{C}$, and U is the smallest set in $\mathcal{F}$ with center A (see Observation 6), identify U with A. This identification (which is one-to-one) accounts for $|\mathcal{C}| \leq 2n-1$ many sets in $\mathcal{F}$.

For each set not yet accounted for, we will select one of its element not in its center in such a way that if the same element is associated with l many sets in $\mathcal{F}$, then these sets form an antichain.

Lemma 10 *Let $V_1, \ldots, V_r$ be a decreasing chain of sets in $\mathcal{F}$ such that for any j, V_{j+1} is maximal below V_j with respect to the sets in $\mathcal{F}$ intersecting the center of V_j. If $W \in \mathcal{F}$ intersects the center of V_r and is not properly contained in V_r, then either the center of W contains the center of V_r or W properly contains V_1.*

Proof of Lemma 10: By Lemma 7, the centers of the V_j form a decreasing chain. Suppose that W does not contain V_1. Since W intersects the center of every V_j, W is comparable to every V_j. In particular, it must be contained in V_1. Let j be the largest index such that $W \subset V_j$. If $j \neq r$ we have $V_j \supset W \supset V_{j+1}$, so by the maximality assumption, $W = V_j$ or $W = V_{j+1}$, and the result follows.□

Lemma 11 *Let U be a set in $\mathcal{F}$ not identified with its center. Let V be the greatest set in $\mathcal{F}$ contained in U such that the center of V is the same as the center of U. Then there is a $x \in U \setminus V$ contained in the center of every member W of $\mathcal{F}$ properly contained in U with $x \in W$.*

Proof of Lemma 11: To find the desired x, let W_1 be maximal below U among members of $\mathcal{F}$ below U intersecting $U \setminus V$. Since the singletons are in $\mathcal{F}$, such sets exist. Since W_1 is incomparable to V, the center of W_1 is contained in $U \setminus V$. Let $W_1, \ldots, W_r$ be a decreasing chain starting with W_1 satisfying the conditions given in Lemma 10 with W_r minimal in $\mathcal{F}$. Suppose Y is properly contained in U and intersects the center of W_r. By minimality of W_r, Y is not properly contained in W_r. By the maximality condition on W_1, Y does not properly contain W_1. Hence, by Lemma 10, the center of Y contains the center of W_r. It follows that any x in the center of W_r satisfies the desired conditions (note that W_r must be a singleton).□

For any U in $\mathcal{F}$ not identified with its center, we select any one of the x satisfying the conclusion of Lemma 11. By the lemma, if x is selected for l many distinct sets, these sets must form an antichain, so, by the k-intersection property, no x is selected more than k times. Since we have now accounted for all the sets in $\mathcal{F}$, we obtain the primary upper bound of $\mid \mathcal{F} \mid \leq b_p = 2n - 1 + kn = (k+2)n - 1$.

Definition: To reduce b_p we associate with each $x \in X$ a positive integer $\mathrm{d}(x)$, the *deficiency* due to x, in such a way that $\mid \mathcal{F} \mid$ is at least the sum of all these deficiencies less than b_p. Let $s(x)$ be the number of times x is selected for sets in $\mathcal{F}$. Clearly, we can let $\mathrm{d}(x)$ be any number less than or equal to $k - s(x)$. Let l be the maximum size of any antichain in $\mathcal{F}$ with x in its intersection. If x is selected exactly l times, we may set the deficiency of x to be as great as $k + 1 - s(x)$. This is due to a forced reduction in the size of the tree of centers (Lemma 12 below). Since in either case the given value is at least $k + 1 - l$ we define $\mathrm{d}(x) = k + 1 - l$.

Lemma 12 *Let X' be the set of all $x \in X$ selected exactly l times, where l is the maximum size of any antichain containing x in its intersection. Then* $\mid \mathcal{C} \mid \leq 2n - 1 - \mid X' \mid$.

Proof of Lemma 12: For each x in X', let C_x be the least set in $\mathcal{C}$ containing both x and the centers of the (antichain of) sets for which x was selected. Since the center of X is X, such a set exists. We will show that C_x has at least $r+2$ many children (i.e. members of $\mathcal{C}$ maximal below C_x), where r is the number of y's in X' for which $C_y = C_x$. An analysis of the proof of the bound on the size of trees then shows that $\mid \mathcal{C} \mid$ is reduced by at least r for this C_x, and the result follows.

Suppose that $x \in X'$, and $C_1, \ldots, C_l$ are the centers of the sets $U_1, \ldots, U_l$ in $\mathcal{F}$ for which x was selected. First we observe that $l \geq 2$: Otherwise, the sets in the pseudo tree containing x form a chain, which implies that x is in the center of every one of these sets, so that x is never selected. Since the maximum size of an antichain containing x is at least 1, it follows that x is not in X', contrary to assumption.

Lemma 12.1 *If $C \in \mathcal{C}$ contains x and intersects at least one of the C_i then $C \supset C_x$.*

Proof of Lemma 12.1: Without loss of generality, assume that C intersects C_1. Remember that the C_i are disjoint, and x is not contained in any of the C_i. Let W be any set in $\mathcal{F}$ with center C. Since x is in the center of W, all of the U_i are comparable to W. If W were contained in U_1, by Lemma 7, C would be contained in C_1, which is not the case. Hence W contains U_1. Since U_1 is incomparable to any of the other U_i, W also contains all the other U_i. Since all the U_i contain x which is in the center of W, again by Lemma 7, all the C_i are contained in C. The result now follows by definition of C_x. □

Lemma 12.2 *If $C \in \mathcal{C}$ contains all the C_i, then $C \supset C_x$.*

Proof of Lemma 12.2: Again, let W be a set in $\mathcal{F}$ with center C. W contains all the U_i, for else (by comparability), its center would be contained in one of the C_i, but there are at least two of these, both contained in C. Suppose that x is not in C. Let V be a set in $\mathcal{F}$ incomparable to W containing x. Since C is disjoint from V (otherwise it wouldn't be the center of W), V is incomparable to all the U_i, so V together with the U_i form an antichain with non-empty intersection containing x, contradicting the maximality assumption on l. Hence x must be in C. □

The last two lemmas show that C_x has at least three children: one containing x, and at least two partitioning $\{C_1, \ldots, C_l\}$, since $l \geq 2$. It remains to consider what happens when $C_x = C_y$ for $y \neq x$. Let $V_1, \ldots, V_{l'}$ be the sets for which y was selected.

Lemma 12.3 *If C is the center of W in $\mathcal{F}$ and C contains both x and y, then C contains $C_x = C_y$.*

Proof of Lemma 12.3: First suppose that for some i, y is not contained in U_i. Then U_i is contained in W, and by Lemma 7, C_i is contained in C. Lemma 12.1 now implies that C contains C_x. Suppose then that y is contained in U_i. By the k-intersection property, U_i must be comparable to at least one of the V_j. Suppose $U_i \subset V_j$ (the other case is symmetric). By Lemma 11, y is in C_i, so by Lemma 12.1, C, which also contains x, contains C_x. □

Let $x_1, \ldots, x_r$ be all the x such that $C_x = C$. By the above lemmas, C must have at least one child for each x_i, none of which may intersect any of the centers of sets for which one of the x_i was selected. In addition C must have at least two children for, say, the centers of the sets for which x_1 was selected. This adds up to at least $r + 2$ many children, as desired. □

Let $\mathrm{d}(m, l)$ be the minimum possible value of the sum of the deficiencies of elements in the center of a member U of a k-pseudo tree where m is the cardinality of the center, and l is the maximum size of any antichain in the pseudo tree with U contained in its intersection. We may assume that $1 \leq l \leq k$.

Lemma 13

$$\begin{aligned} \mathrm{d}(m,l) &\geq m-1 \\ &+ (k+1-l)+(k+1-l-1)+\ldots+(k+1-l-\min(m-1,k-l)) \end{aligned}$$

In particular, this shows that the total deficiency of X, the center of which is X, is at least $\mathrm{d}(n,1) = n-1+(k(k+1))/2$, so $|\mathcal{F}| \leq b_p - \mathrm{d}(n,1) = (k+1)n - k(k+1)/2$ thus completing the proof of the theorem.

Proof of Lemma 13: If $l = k$ the lemma asserts $\mathrm{d}(m,k) \geq (m-1) + (k-k+1) = m$. Since for any x, $\mathrm{d}(x) \geq 1$, this part of the lemma is true, so we may assume $l < k$.

The remainder of the proof proceeds by induction on the first argument. Consider first $m = 1$. Let U be a set in a k-pseudo tree with center $C = \{x\}$. Suppose the maximum size of an antichain containing U in its intersection is l. By Lemma 7, every member of the pseudo tree contained in U which contains x, has x as its center. Therefore, if $U_1, \ldots, U_r$ is an antichain with x in its intersection, each U_i contains U so that $r \leq l$. It follows that $\mathrm{d}(x) = k+1-l$. By generality, $\mathrm{d}(1,l) \geq k+1-l$, which is as desired.

Let U be a set in a k-pseudo tree with center C of cardinality $m > 1$. We may assume that U is the least set in the pseudo tree with this center, for if C is contained in the intersection of an antichain of size greater than one, so is every set in the pseudo tree with center C (if it contains one of the members of the antichain, then by Lemma 7 its center must also contain the center of that member).

Every set in the pseudo tree contained in U intersecting C has its center strictly contained in C. Since the singletons are in the pseudo tree and $m > 1$, C has at least two children in the tree of centers. For each child of C, the greatest set in the pseudo tree with this child as its center is contained in U. Let $\mathcal{U}$ consist of these sets, one for each child of C.

Observe that $\mathcal{U}$ has at least two maximal members U_1 and U_2, respectively: If V is maximal in $\mathcal{U}$, and C' is the center of V, let x be any element in $C \setminus C'$. Let V' be a set maximal with respect to the familiy of sets in the pseudo tree incomparable to V containing x. Since x is not in the center of V, such sets exist. Since $C' \subset C$ is not contained in V' and x is in C, V' is contained in U, so its center is contained in C, and by maximality, V' is in $\mathcal{U}$. V' is also maximal in $\mathcal{U}$.

Let c_1 and c_2 be the cardinalities of the centers of U_1 and U_2 respectively. Let $\mathcal{U}_3$ be $\mathcal{U}$ excluding U_1 and U_2. Let U_3 be any maximal set in $\mathcal{U}_3$ and c_3 the cardinality of its center. Let $\mathcal{U}_4$ be $\mathcal{U}_3$ excluding U_3 and let U_4 be any maximal set in $\mathcal{U}_4$ and c_4 the cardinality of its center. Continue in this fashion obtaining U_i until $max(s_i, l) \geq k$ where $s_i = c_1 + \ldots + c_i$. Let r be the index of the last set thus obtained. For $r < i \leq t$, where $t = \mid \mathcal{U} \mid$, let $c_{r+1}, \ldots, c_t$ be the cardinalities of the centers of the remaining members of $\mathcal{U}$.

Lemma 13.1 *For $2 < i \leq r$, the maximal size of any antichain with U_i in its intersection is at most $max(s_{i-1}, l)$. For $i = 1$ or $i = 2$, this quantity is at most l.*

Proof of Lemma 13.1: Let $V_1, \ldots, V_j$ be an antichain in the pseudo tree with U_i in its intersection. All the members of the antichain are comparable to U, since they intersect U's center. Suppose V_h (say) contains U. Then, all the other members of the antichain contain U (else it would not be an antichain). In this case, $j \leq l$. So, suppose that all the V_h are contained in U. Since they intersect C, the centers of the V_h are (strictly) contained in C. In particular, each one is contained in or equal to at least on set in $\mathcal{U}$. By the maximality condition on U_1 and U_2, this is impossible for $i = 1$ or $i = 2$ unless $j = 1$. Suppose then that $i > 2$. Since U_i is maximal in $\mathcal{U}_i$, unless $j = 1$, the V_h are not contained in any member of $\mathcal{U}_i$. Let C_h be the center of V_h. C_h must be contained in C. Let U' be the member of $\mathcal{U}$ the center of which contains C_h. Then, since U' contains $V_h \supset U_i$, U' must be $U_{i'}$ for $i' < i$. Since h was arbitrary, the union of the centers of the V_h must be contained in the union of the centers of the $U_{i'}$ with $i' < i$, which has cardinality s_{i-1}. Since the centers of the V_h are disjoint, $j \leq s_{i-1}$. The result now follows. □

For $2 < i \leq r$, let $l_i = max(s_{i-1}, l)$. Let $l_1 = l_2 = l$. By choice of r, for all $i \leq r$, $l_i < k$. We can now bound the total deficiency d of the center of U from below:

$$d \geq \mathrm{d}(c_1, l) + \mathrm{d}(c_2, l) + \mathrm{d}(c_3, l_3) + \ldots + \mathrm{d}(c_r, l_r) + \mathrm{d}(c_{r+1}, k) + \ldots + \mathrm{d}(c_t, k)$$

The summands of the form $\mathrm{d}(c_i, k)$ are equal to c_i. Using the induction hypothesis:

$$\begin{array}{llll} d \geq & m - r & & \\ & +(k+1-l) & + & (k+1-l-1) + \ldots + (k+1-l-r_1) \\ & +1 & + & (k-l) + (k-l) + \ldots + (k+1-l-r_2) \\ & \vdots & & \\ & +1 & + & (k-l_r) + (k-l_r) + \ldots + (k+1-l_r-r_r) \end{array}$$

where $r_i = min(c_i - 1, k - l_i)$.

The leading 1's together with the term $m-r$ sum up to $m-1$. It remains to show that the sum of the first $\min(m-1, k-l) + 1$ many remaining summands dominate the series $(k+1-l) + (k+1-l-1) + \ldots + (k+1-l-\min(m-1, k-l))$ consisting of the descending sequence of consecutive integers starting at $k+1-l$ with $\min(m, k-l+1)$ many terms. Here are the remaining summands again:

$$\begin{array}{lll} (k+1-l) & + & (k+1-l-1) + \ldots + (k+1-l-r_1) \\ & + & (k-l) + (k-l) + \ldots + (k+1-l-r_2) \\ \vdots & & \\ & + & (k-l_r) + (k-l_r) + \ldots + (k+1-l_r-r_r) \end{array}$$

This is greater than or equal to:

$$\begin{array}{llll} (1) & (k+1-l) & + & (k-l-1) + \ldots + (k-l-r_1) \\ (2) & & + & (k-l) + (k-l-1) + \ldots + (k-l-r_2) \\ & \vdots & & \\ (r) & & + & (k-l_r) + (k-l_r-1) + \ldots + (k-l_r-r_r) \end{array}$$

Note that all the terms are at least 0.

Let $r \geq i \geq 1$. By definition, $l_i \leq l + s_{i-1} + 1$. Thus, if $s_i \leq k - l$

$$c_i - 1 = s_i - s_{i-1} - 1 \leq k - l - s_{i-1} - 1 \leq k - l_i.$$

This inequality implies that $r_i = c_i - 1$. Since $s_i - l_i \leq c_i$, this in turn implies:

$$\begin{array}{rcr} k - l_{i+1} & = & k - \max(l, s_i) \\ & \geq & k - \max(l_i, s_i) \\ & = & k - l_i - max(s_i - l_i, 0) \\ & \geq & k - l_i - c_i \\ & = & k - l_i - r_i - 1 \end{array}$$

That is, the first term on the $i+1$'th line is at least as great as the last term on line i less one.

This shows that if $s_i \leq k-l$, the terms of this sum up to and including the last term of the $(i+1)$'th line, dominate consecutive non-negative decreasing integers starting with $k+1-l$ ('excess' terms dominate 0's).

The number of terms on the (i)'th line is $\min(c_i, k-l_i+1)$. Suppose that $c_i - 1 > k - l_i$. If $s_{i-1} < l$, $l_i = l$, so $c_i > k-l+1$, and line (i) has $k-l+1$ many terms. If $i=1$, since the first term on line (1) is $k+1-l$, the induction is complete. Else, the first term is $k-l$ (since $l_i = l$), so we simply add in the first term of line (1), and again we are done. If $s_{i-1} \geq l$, $l_i = s_{i-1}$, so $s_i = c_i + s_{i-1} > k - l_i + 1 + s_{i-1} = k+1$ hence $i = r$. We can subsume this case in the next one.

Let h be the least i such that $s_i > k-l$ or r if no such i exists. In either case, the terms in the first h lines dominate term by term the descending series of consecutive integers starting with $(k+1-l)$ and for $i < h$, the number of terms on line (i) is $r_i + 1 = c_i$. We have already dealt with the case where $c_i > k - l_i + 1$ with $i < r$, so we may assume this is not the case except possibly for $i = h = r$. The number of terms in lines (1) through (h) is therefore $c_1 + \ldots + c_{h-1} + r_h + 1 = s_{h-1} + min(c_h, k - l_h + 1)$. If $h \neq r$, the last term is c_h, so we get at least $s_h \geq k-l+1$ many terms the sum of which dominates $(k+1-l) + \ldots + 1$. If $h = r$ and $c_h \leq k - l_h + 1$ we get s_r many such terms. By construction, either $s_r \geq k \geq k-l+1$, or $s_r = m$ (which is the case where $m < k-l+1$), and we are done. Otherwise, by the previous paragraph, we may assume that $l_h = s_{h-1}$, so we get $s_{h-1} + k - s_{h-1} + 1 = k+1 \geq k-l+1$ many terms. □

The proof of the theorem is now complete.

References

[1] I. Anderson, *Combinatorics of Finite Sets*, Clarendon Press, Oxford, 1987.

[2] B. Bollobas, *Combinatorics*, Cambridge University Press, Cambridge, 1986.

[3] M. Simonovits, V. T. Sós, "Intersection theorems on structures", *Annals of Discrete Mathematics* **6** (1980) 301–313.

[4] R. P. Stanley, *Enumerative Combinatorics, Vol I*, Wadsworth&Brooks/Cole, Monterey Cal., 1986.