Forestation in hypergraphs: linear k-trees

Ojas Parekh
Department of Mathematics and Computer Science
Emory University, Atlanta, USA
ojas@mathcs.emory.edu

Submitted: Sep 14, 2001; Accepted: Sep 1, 2003; Published: Sep 23, 2003
MR Subject Classifications: 05C65, 05E99

Abstract

We present a new proof of a result of Lovász on the maximum number of edges in a k-forest. We also apply a construction used in our proof to generalize the notions of a k-hypertree and k-forest to a class which extends some properties of trees, to which both specialize when k=2.

1 Introduction

Let $X = [n]$ and $\mathcal{F}$ be a k-uniform hypergraph on X. We say an edge $e \in \mathcal{F}$ *crosses* a k-partition, $X = X_1 \dot\cup \cdots \dot\cup X_k$, if $|e \cap X_i| = 1$ for $1 \leq i \leq k$. $\mathcal{F}$ is a *k-forest* if for each $e \in \mathcal{F}$ there is some k-partition $X = X_1^e \dot\cup \cdots \dot\cup X_k^e$ such that e is the unique edge crossing it. What is the maximum number of edges in $\mathcal{F}$?

This problem was initially posed to László Lovász by Ronald Graham [2]. Lovász's novel algebraic proof appeared in [3] in 1979, and our proof remains algebraic in nature; however, it relies on homogeneous multilinear polynomials over $\mathbb{F}_2$ rather than tensors. The reader is encouraged to consult [1] for an introduction to and extensive applications of linear algebra in combinatorics.

Theorem 1.1. *A k-forest $\mathcal{F}$ on X has at most $\binom{n-1}{k-1}$ edges.*

Proof. We open with a few definitions. By $\mathbb{P}_{k-1}^{n-1}$ we mean the space of multilinear homogeneous polynomials of degree $k-1$ in $\mathbb{F}_2[x_1, \ldots, x_{n-1}]$. We make use of the shorthand $p(\mathbf{x})$ to denote $p(x_1, \ldots, x_{n-1})$, where $\mathbf{x} = (x_1, \ldots, x_{n-1}) \in \mathbb{F}_2^{n-1}$ and $p \in \mathbb{P}_{k-1}^{n-1}$. Finally, for $e \in \mathcal{F}$, $\mathbb{1}_e$ denotes the incidence vector of e.

For each edge $e \in \mathcal{F}$ we pick a k-partition $\pi_e = (X_1^e, \ldots, X_k^e)$, such that e is the unique edge crossing it. For simplicity we assume X_1^e contains the element n. We then define a polynomial,

$$p_e(x_1, \ldots, x_{n-1}) = \prod_{i=2}^{k} \sum_{j \in X_i^e} x_j.$$

For each e in $\mathcal{F}$, p_e is in $\mathbb{P}^{n-1}_{k-1}$, hence it suffices to demonstrate the independence of these polynomials. To that end we seek to show that if $e, f \in \mathcal{F}$, then $p_e(\mathbb{1}_{f\setminus\{n\}}) = 1$ if and only if $f = e$. We have

$$p_e(\mathbb{1}_{f\setminus\{n\}}) = \prod_{i=2}^{k}(|f \cap X_i^e| \bmod 2).$$

Clearly $p_e(\mathbb{1}_{e\setminus\{n\}}) = 1$. If $f \neq e$ there must be some i for which $|f \cap X_i^e| = 0$, since f does not cross π_e. In this case there also exists a $j \neq i$ such that $|f \cap X_j^e| \bmod 2 = 0$. Thus $p_e(\mathbb{1}_{f\setminus\{n\}}) = 0$. $\square$

Our agenda for the remainder of the paper is to first consider a generalization of k-forests which preserves certain properties of forests and to then proceed to compare our generalization with existing ones.

2 Linear k-trees

In light of the result of the previous section a natural question arises. What can one say about the maximum k-forests, those with exactly $\binom{n-1}{k-1}$ edges? We could begin by considering small examples. It is not difficult to verify that a 2-forest is indeed a forest. In this case any maximal forest is a tree, which one may define in several ways. A basic result in graph theory is that a graph which exhibits any two of

(i) acyclicity

(ii) exactly $n-1$ edges

(iii) connectivity

necessarily exhibits the third.

We already have analogues of (i) and (ii) that we could use in defining a *k-tree* for $k > 2$, and one might conjecture that for a k-uniform hypergraph $\mathcal{H}$ on X, any two of

(i') $\mathcal{H}$ is a k-forest.

(ii') $\mathcal{H}$ has exactly $\binom{n-1}{k-1}$ edges.

(iii') For each k-partition of X, $\mathcal{H}$ contains an edge that crosses it.

implies the third. Unfortunately this is not true.

Counterexample 2.1. *The 3-uniform hypergraph*

$$\mathcal{H} = \{\{1,2,4\}, \{1,2,5\}, \{1,3,5\}, \{1,4,5\}, \{2,3,4\}, \{3,4,5\}\}$$

over $\{1,2,3,4,5\}$ *satisfies (ii') and (iii') but not (i').*

Why does this generalization fail? Conditions (i')-(iii') extend the notion of a cut in a graph, which is implicit in (i)-(iii), to that of a k-partition. We do have that a 2-partition is indeed a cut; however, this is not the whole story. The proof of Theorem 1.1 offers some insight into the matter. The multilinear polynomial space $\mathbb{P}_1^{n-1}$ consists entirely of polynomials which correspond to 2-partitions; however, the reader may verify that an analogous statement is not true for even $\mathbb{P}_2^{n-1}$. Guided by this discrepancy, we say an edge $e \in \mathcal{H}$ *crosses* a *polynomial* $p \in \mathbb{P}_{k-1}^{n-1}$ if $p(\mathbb{1}_{e\setminus\{n\}}) = 1$, and we relax (i'): The hypergraph $\mathcal{H}$ is a *linear k-forest* (vs. a k-forest) if for each edge $e \in \mathcal{H}$, there is a polynomial $p_e \in \mathbb{P}_{k-1}^{n-1}$ (vs. a k-partition) such that e is the unique edge in $\mathcal{H}$ crossing p_e. We accordingly strengthen (iii'): The hypergraph $\mathcal{H}$ is *linearly k-connected*, or simply k-connected, if for each polynomial $p \in \mathbb{P}_{k-1}^{n-1}$, there is an edge $e \in \mathcal{H}$ which crosses p. The scrutinizing reader might have sensed something amiss in the preceding definitions. The polynomial space $\mathbb{P}_{k-1}^{n-1}$ is defined with respect to a distinguished element $n \in X$.

Lemma 2.2. *A hypergraph is a linear k-forest or k-connected independently of the choice of distinguished element used in defining $\mathbb{P}_{k-1}^{n-1}$.*

Proof. Let $p(x_1, \ldots, x_{n-1}) \in \mathbb{P}_{k-1}^{n-1}$. We will demonstrate a $p'(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \mathbb{P}_{k-1}^{n}$ such that $\{e \in \binom{X}{k} \mid p(\mathbb{1}_{e\setminus\{n\}}) = 1\} = \{e \in \binom{X}{k} \mid p'(\mathbb{1}_{e\setminus\{i\}}) = 1\}$. We divide p by x_i to yield $p = x_i q + r$ where $q \in \mathbb{P}_{k-2}^{n-1}$, $r \in \mathbb{P}_{k-1}^{n-1}$, and neither contain the variable x_i. We can represent q as a sum of monomials, that is there exist sets $Y_s \in \binom{X\setminus\{i,n\}}{k-2}$ for s in some index set S such that $q = \sum_{s\in S}\prod_{j\in Y_s} x_j$. Notice that an edge crosses the polynomial $(\prod_{j\in Y_s} x_j)(\sum_{j\notin Y_s\cup\{i\}} x_j)$ if and only if it crosses the monomial $x_i(\prod_{j\in Y_s} x_j)$. This provides us the construction we seek, and we set

$$p'(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = r + \sum_{s\in S}\left(\prod_{j\in Y_s} x_j\right)\left(\sum_{j\notin Y_s\cup\{i\}} x_j\right).$$

□

We will henceforth use $\mathbb{P}_{k-1}^{n-1}$ to refer to a multilinear polynomial space in $n-1$ variables, the indices of which will be clear from context. We now have the following.

Theorem 2.3. *For $\mathcal{H}$, a k-uniform hypergraph on X, any two of*

(i) *$\mathcal{H}$ is a linear k-forest.*

(ii) *$\mathcal{H}$ has exactly $\binom{n-1}{k-1}$ edges.*

(iii) *$\mathcal{H}$ is k-connected.*

implies the third.

Proof.

(i),(ii) implies (iii): For each edge e let p_e be a polynomial for which e is the unique edge in $\mathcal{H}$ crossing p_e. For a polynomial $p \in \mathbb{P}_{k-1}^{n-1}$, let $\mathcal{H}(p)$ denote $\{e \in \mathcal{H} \mid e \text{ crosses } p\}$. As in the proof of Theorem 1.1 we have the independence of the polynomials p_e for $e \in \mathcal{H}$, hence $|\{p_e \mid e \in \mathcal{H}\}| = \binom{n-1}{k-1}$ by (ii). The set $\{p_e \mid e \in \mathcal{H}\}$ is a basis for $\mathbb{P}_{k-1}^{n-1}$, so for any $q \in \mathbb{P}_{k-1}^{n-1}$ we must have $\mathcal{H}(q) \neq \emptyset$.

(ii),(iii) implies (i): First we establish $p \neq q$ implies $\mathcal{H}(p) \neq \mathcal{H}(q)$, for polynomials $p, q \in \mathbb{P}_{k-1}^{n-1}$. Proceeding by contrapositive, if $\mathcal{H}(p) = \mathcal{H}(q)$ then $\mathcal{H}(p+q) = \emptyset$, hence $p = q$. There are exactly $2^{\binom{n-1}{k-1}} - 1$ polynomials in $\mathbb{P}_{k-1}^{n-1}$ and $|\mathcal{H}| = \binom{n-1}{k-1}$, so by (iii) $\{\mathcal{H}(p) \mid p \in \mathbb{P}_{k-1}^{n-1}\} = 2^{\mathcal{H}} \setminus \{\emptyset\}$, where $2^{\mathcal{H}}$ is the powerset of $\mathcal{H}$. Thus for each edge $e \in \mathcal{H}$, $\{e\} \in \{\mathcal{H}(p) \mid p \in \mathbb{P}_{k-1}^{n-1}\}$.

(iii),(i) implies (ii): From the proof of the first part we have (i) implies $|\mathcal{H}| \leq \binom{n-1}{k-1}$; from that of the second we have (iii) implies $|\mathcal{H}| \geq \binom{n-1}{k-1}$.

$\square$

We are finally in position to call a hypergraph $\mathcal{T}$ that satisfies any two conditions above a linear k-tree. The third part of the proof of the theorem hints at two other characterizations of linear k-trees.

Theorem 2.4.

(i) *Every k-connected hypergraph contains a linear k-tree.*

(ii) *Every linear k-forest is contained in a linear k-tree.*

Proof.

(i): For the sake of contradiction, let $\mathcal{H}$ be a minimal k-uniform hypergraph over X that is k-connected but does not contain a linear k-tree. We let $\mathbb{P}(H)$ represent $\{p \in \mathbb{P}_{k-1}^{n-1} \mid \text{some } e \in H \text{ crosses } p\}$, where $H \subseteq \mathcal{H}$; we omit braces for singleton arguments. Since $\mathcal{H}$ is not a linear k-forest, there is some $e \in \mathcal{H}$ such that $\mathbb{P}(e) \subseteq \mathbb{P}(\mathcal{H} \setminus \{e\})$, hence $\mathcal{H} \setminus \{e\}$ is also a counterexample.

(ii): For the sake of contradiction, let $\mathcal{H}$ be a maximal k-uniform hypergraph over X that is a linear k-forest but is not contained in a linear k-tree. Since $\mathcal{H}$ is not k-connected, there is some $p \in \mathbb{P}_{k-1}^{n-1}$ such that $\mathcal{H}(p) = \emptyset$. Let $f \in \binom{X}{k}$ be some set such that $p(\mathbb{1}_{f\setminus\{n\}}) = 1$, and for $e \in \mathcal{H}$ let p_e be a polynomial such that $\mathcal{H}(p_e) = \{e\}$. We set $p'_f = p$, and for each edge $e \in \mathcal{H}$, we set

$$p'_e = \begin{cases} p_e + p & \text{if } p_e(\mathbb{1}_{f\setminus\{n\}}) = 1 \\ p_e & \text{otherwise} \end{cases},$$

which renders e the unique edge in $\mathcal{H} \cup \{f\}$ crossing p'_e and $\mathcal{H} \cup \{f\}$ a counterexample.

$\square$

Thus we may also think of linear k-trees as maximal linear k-forests or minimally k-connected hypergraphs.

3 All trees are not created equal

A linear k-tree is only one of a multitude of possible generalizations of trees to hypergraphs; in this section we explore the connection between linear k-trees and a generalization which exists in the literature.

The combinatorial structure known as a k-hypertree was introduced in [4] as a tool for developing Bonferroni type inequalities. A k-hypertree is a k-uniform hypergraph $\mathcal{T}$ on X such that for $k = 2$, $\mathcal{T}$ is a tree with vertex set X and for $k \geq 3$, $\mathcal{T}$ is defined recursively as follows:

(i) If $X = \{1, \ldots, k\}$ then $\mathcal{T}$ has a unique edge $\{1, \ldots, k\}$.

(ii) If $|X| \geq k+1$ then there exists an element $i \in X$ such that if $e_1, \ldots, e_q$ denote all edges containing i then $e_1 \setminus \{i\}, \ldots, e_q \setminus \{i\}$ induce an $(k-1)$-hypertree with vertex set $X \setminus \{i\}$ and the remaining edges of $\mathcal{T}$ induce a k-hypertree with vertex set $X \setminus \{i\}$. A k-hypertree has exactly $\binom{n-1}{k-1}$ edges.

The notion was augmented [5] by imposing a total ordering μ on X, yielding several nice characterizations of k-hypertrees which generalize properties of trees. We show that linear k-trees generalize k-hypertrees. We denote the classes of linear k-trees and k-hypertrees on X by $\mathcal{LKT}(k,n)$ and $\mathcal{HT}(k,n)$ respectively.

Theorem 3.1. *$\mathcal{HT}(k,n) \subset \mathcal{LKT}(k,n)$.*

Proof. We show inclusion by induction. We have that $\mathcal{HT}(k,k) = \mathcal{LKT}(k,k)$, so let us consider some $\mathcal{T} \in \mathcal{HT}(k,n)$ for $k < n$. Since $|\mathcal{T}| = \binom{n-1}{k-1}$, by Theorem 2.3 it suffices to show $\mathcal{T}$ is k-connected. Let $l \in X$ be an element such that $\mathcal{T}_l = \{e \setminus \{l\} \mid l \in e \in \mathcal{T}\}$ and $\mathcal{T}_{\bar{l}} = \{e \in \mathcal{T} \mid l \notin e\}$ are respectively $(k-1)$- and k-hypertrees over $X \setminus \{l\}$.

We seek to show that for a polynomial $p(x_1, \ldots, x_{n-1}) \in \mathbb{P}^{n-1}_{k-1}$ there is some edge in $\mathcal{T} = \mathcal{T}_l \dot{\cup} \mathcal{T}_{\bar{l}}$ that crosses it. Note that we may assume $l \neq n$ by Lemma 2.2. Dividing by x_l gives us $p = x_l q(x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_{n-1}) + r(x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_{n-1})$. If $r \equiv 0$ then $e \cup \{l\}$ crosses $p = x_l q$ for some $e \in \mathcal{T}_l$, since $\mathcal{T}_l \in \mathcal{KT}(k-1, n-1)$ by the induction hypothesis. Otherwise some $e \in \mathcal{T}_{\bar{l}}$ crosses r, since $\mathcal{T}_{\bar{l}} \in \mathcal{KT}(k, n-1)$ by the induction hypothesis. In this case $l \notin e$, hence e crosses $p = x_l q + r$.

As for strict inclusion, we leave it to the reader to verify that

$$\begin{aligned} \mathcal{T} = \quad & \big\{\{1,2,3\}, \{1,2,4\}, \{1,2,6\}, \{1,4,5\}, \{1,5,6\}, \\ & \{2,3,5\}, \{2,3,6\}, \{3,4,5\}, \{3,4,6\}, \{4,5,6\}\big\} \end{aligned}$$

is a linear 3-tree but not a 3-hypertree. □

The class $\mathcal{LKT}(k,n)$ may be a practically significant generalization of $\mathcal{HT}(k,n)$. Given a cost function $c : \binom{X}{k} \to \mathbb{R}_+$, it is *NP*-complete to decide whether there is a k-hypertree of cost at most l for $n > k \geq 3$ [6]. This is known as the minimum spanning k-hypertree problem and for $k = 2$ reduces to the polynomial time solvable minimum spanning tree problem. Replacing 'k-hypertree' with 'linear k-tree' in the above definition drastically reduces the complexity of the problem. By Theorems 2.4 and 2.3 the linear k-forests on X comprise a matroid, hence we can apply a greedy algorithm to solve the minimum spanning linear k-tree problem in polynomial time for constant k.

We close by offering a conjecture. A k-tree is a k-forest of size $\binom{n-1}{k-1}$. We let $\mathcal{KT}(k,n)$ denote the class of k-trees on X. From Theorem 2.3, Counterexample 2.1, and the fact that $\{e \in \binom{X}{k} \mid 1 \in e\} \in \mathcal{HT}(k,n) \cap \mathcal{KT}(k,n)$, we derive the following properties.

$$\begin{aligned} \mathcal{KT}(k,n) &\subset \mathcal{LKT}(k,n) && (1) \\ \mathcal{HT}(k,n) &\setminus \mathcal{KT}(k,n) \neq \emptyset && (2) \\ \mathcal{HT}(k,n) &\cap \mathcal{KT}(k,n) \neq \emptyset && (3) \end{aligned}$$

Unfortunately these leave the precise interaction of $\mathcal{HT}(k,n)$ and $\mathcal{KT}(k,n)$ uncertain. Yet if one could show that for every $\mathcal{T} \in \mathcal{KT}(k,n)$ there is some $i \in X$ that is contained in exactly $\binom{n-2}{k-2}$ edges, then induction would yield the following.

Conjecture 3.2. $\mathcal{KT}(k,n) \subset \mathcal{HT}(k,n)$.

Acknowledgements

The author is grateful to Tom Bohman for exposing him to the tools and the trade.

References

[1] L. Babai and P. Frankl. *Linear Algebra Methods in Combinatorics*. Dept. of Computer Science, The Univ. of Chicago, Chicago, 1992.

[2] R. Graham. Personal Communication. 2000.

[3] L. Lovász. Topological and Algebraic Methods in Graph Theory. In *Graph theory and related topics (Proc. Conf., Univ. Waterloo, Waterloo, Ont., 1977)*, pp. 1–14, Academic Press, New York-London, 1979.

[4] I. Tomescu. Hypertrees and Bonferroni inequalities. *J. Combin. Theory Ser. B*, 41:209–217, 1986.

[5] I. Tomescu. Ordered h-Hypertrees. *Discrete Mathematics*, 195:241–248, 1992.

[6] I. Tomescu and M. Zimand. Minimum spanning hypertrees. *Discrete Appl. Math.*, 54.1:67–76, 1994.