Facets of Balanced Minimal Evolution polytopes.

Stefan Forcey, Logan Keefe, William Sands. U. Akron.



Stefan Forcey, Logan Keefe, William Sands. U. Akron.

Facets of Balanced Minimal Evolution polytopes.

The goal of *phylogenetics* is to take a finite set of data structures and to construct a branching diagram that explains how its elements are related. Biological sets are usually referred to collectively as taxa-populations, species, individuals or genes-and elements are assumed to be related genetically and chronologically. The diagram we will be concerned with is a binary tree with labeled leaves, known as a phylogenetic tree. Precisely, we consider a cycle-free simple graph with nodes (vertices) that are either of degree one (touching a single edge) or degree three, and with a set of distinct items assigned to the nodes of degree 1-the leaves. The nodes of degree 3 are unlabeled, and can be thought of as representing speciation events.

We study a method called *balanced minimal evolution*. This method begins with a given set of *n* items and a symmetric (or upper triangular) square $n \times n$ dissimilarity matrix whose entries are numerical dissimilarities, or distances, between pairs of items. From the dissimilarity matrix the balanced minimal evolution (BME) method constructs a binary tree with the *n* items labeling the *n* leaves. This BME tree has the property that the distances between its leaves most closely match the given distances between corresponding pairs of taxa. More precisely: Let the set of *n* distinct species, or taxa, be called *S*. For convenience we will often let $S = [n] = \{1, 2, ..., n\}$. Let a vector **d** be given, having $\binom{n}{2}$ real valued components the distances d_{ij} between unordered pairs of distinct taxa $i, j \in S$. There is a vector $\mathbf{x}(t)$ for each binary tree *t* on leaves *S*, also having $\binom{n}{2}$ components $x_{ij}(t)$, one for each pair $\{i, j\} \subset S$. These components are ordered in the same way for both vectors, and we will use the lexicographic ordering: $\mathbf{d} = \langle d_{12}, d_{13}, \ldots, d_{1n}, d_{23}, d_{24}, \ldots, d_{n-1,n} \rangle$. We define: $x_{ij} = 2^{n-2-l(i,j)}$. where l(i, j) is the number of internal nodes (degree 3 vertices) in the path from leaf *i* to leaf *j*.

Introduction

The BME tree for the vector **d** is the binary tree *t* that minimizes $\mathbf{d} \cdot \mathbf{x}(t)$ for all binary trees on leaves S. The value of setting up the question in this way is that it becomes a linear programming problem. The convex hull of all the vectors $\mathbf{x}(t)$ for all binary trees t on S is a polytope BME(S), hereafter also denoted BME(n) or \mathcal{P}_n . The vertices of \mathcal{P}_n are precisely the (2n-5)!! vectors $\mathbf{x}(t)$. Minimizing our inner product over this polytope is equivalent to minimizing over its vertices, which correspond to the phylogenetic trees. This method is consistent, and statistically consistent. In other words, if a sequence of distance matrices approaches a distance matrix whose entries are exactly the summed edge lengths of paths between leaves of a given binary tree T, then the BME trees on that sequence are guaranteed (in the limit) to match the given tree topology of T.

Phylogenetic tree.



Rooting with *Ceratophyllum*, photo: Christian Fischer (CC).









d = $\langle 6, 8, 9, 12, 7, 15 \rangle$









The Balanced minimal evolution polytope \mathcal{P}_4 .





The Balanced minimal evolution polytope \mathcal{P}_5 .



Figure: Two sample vertex trees of \mathcal{P}_5 with their respective coordinates shown beneath, followed by all 15 vertex points calculated for n=5, and the *f*-vector for \mathcal{P}_5 as found by polymake.

Intersecting cherries facet: $x_{ab} + x_{bc} - x_{ac} \le 8$.



Intersecting cherry flag: $x_{ab} + x_{bc} - x_{ac} \le 8$.



Intersecting cherries facet flag: $x_{ab} + x_{bc} - x_{ac} \le 2^{n-3}$



Caterpillar facet: $x_{ab} \ge 1$.



Figure: On the left is a facet of \mathcal{P}_5 with each vertex labeled by the caterpillar tree. On the right is the Birkhoff polytope B(3) with vertices labeled by the corresponding permutation matrices.



Intersection.



Necklace, or cyclic ordering facets.



Split faces; split facets.



n	dim.	vertices	facets	facet types	number of	number of
					facets	vertices
						in facet
3	0	1	0	-	-	-
4	2	3	3	caterpillar	3	2
				intersecting cherries	3	2
5	5	15	52	caterpillar	10	6
				intersecting cherries	30	6
				necklace	12	5
6	9	105	90262	caterpillar	15	24
				intersecting cherries	60	30
				(3,3)-split	10	9
<i>n</i> > 6	$\binom{n}{2} - n$	(2n-5)!!	?	caterpillar	(ⁿ ₂)	(n-2)!
				intersecting cherries	$\binom{n}{2}(n-2)$	2(2n-7)!!
				(m,3)-split	$\binom{n}{3}$	3(2 <i>n</i> - 9)!!
				(m, n - m)-split,	$2^{n-1} - \binom{n}{2}$	(2m - 3)!!
				m > 3	-n-1	$\times (2(n-m)-3)!!$



Together with the equalities $\sum_{j \neq i} x_{ij} = 2^{n-2}$, we take the caterpillar, intersecting-cherry, and split inequalities.

Splitohedron.



Theorem: the Splitohedron is a bounded polytope that is a relaxation of the BME polytope.

Proof: The split-faces include the cherries where the inequality is $x_{ij} \leq 2^{n-3}$, and the caterpillar facets have the inequality $x_{ij} \geq 1$, thus the resulting intersection of halfspaces is a bounded polytope since it is inside the hypercube $[1, 2^{n-3}]^{\binom{n}{2}}$.

polytope > print \$p->VERTICES;

- 12222114411
- 12222141141
- 1 4/3 8/3 8/3 4/3 8/3 4/3 8/3 4/3 4/3 8/3
- 1 4/3 8/3 8/3 4/3 4/3 8/3 8/3 4/3 8/3 4/3
- 14112112422
- 1 8/3 4/3 4/3 8/3 8/3 4/3 4/3 8/3 4/3 8/3 1 8/3 4/3 8/3 4/3 8/3 4/3 4/3 4/3 8/3 8/3 1 2 2 2 2 4 1 1 1 1 4
- 1 8/3 8/3 4/3 4/3 4/3 8/3 4/3 4/3 8/3 8/3 1 8/3 8/3 4/3 4/3 4/3 4/3 8/3 8/3 4/3 8/3
- 12411222114
- 1 4/3 4/3 8/3 8/3 8/3 4/3 8/3 8/3 4/3 4/3
- 1 4/3 8/3 4/3 8/3 4/3 8/3 8/3 8/3 4/3 4/3

polytope > print \$p->VERTICES;

```
11214241221
```

- 11241214221
- 11421124212
- 11124421212
- 1112772121
- 1 1 1 4 2 4 1 2 1 2 2
- 11412142122
- 12141222141

1 8/3 4/3 8/3 4/3 4/3 4/3 8/3 8/3 8/3 4/3

12114222411

1 4/3 4/3 8/3 8/3 8/3 8/3 4/3 4/3 4/3 8/3 4/3 1 4/3 8/3 4/3 8/3 8/3 8/3 8/3 4/3 4/3 4/3 8/3 1 4 1 2 1 1 2 1 2 4 2

14211211224

1 8/3 4/3 4/3 8/3 4/3 8/3 4/3 8/3 8/3 4/3

12222114411

12222141141

- 1 4/3 8/3 8/3 4/3 8/3 4/3 8/3 4/3 8/3
- 1 4/3 8/3 8/3 4/3 4/3 8/3 8/3 4/3 8/3 4/3 1 4 1 1 2 1 1 2 4 2 2

1 8/3 4/3 4/3 8/3 8/3 4/3 4/3 8/3 4/3 8/3 1 8/3 4/3 8/3 4/3 8/3 4/3 4/3 4/3 8/3 8/3 1 2 2 2 2 4 1 1 1 1 4

1 8/3 8/3 4/3 4/3 4/3 8/3 4/3 4/3 8/3 8/3 1 8/3 8/3 4/3 4/3 4/3 4/3 8/3 8/3 4/3 8/3 1 2 4 1 1 2 2 2 1 1 4

12411222114

1 4/3 4/3 8/3 8/3 8/3 4/3 8/3 8/3 4/3 4/3 1 4/3 8/3 4/3 8/3 4/3 8/3 8/3 8/3 4/3 4/3 Advertisement:

http://www.math.uakron.edu/~sf34/hedra.htm
Questions and comments?



