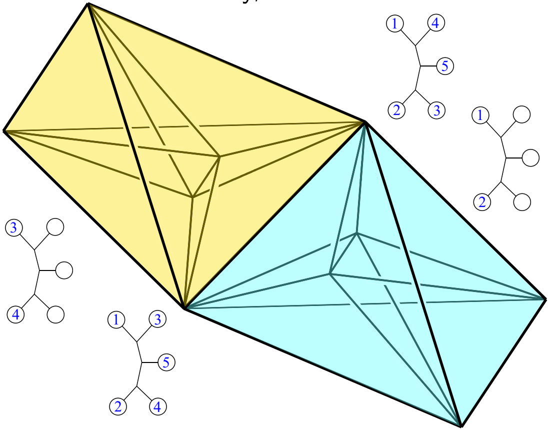
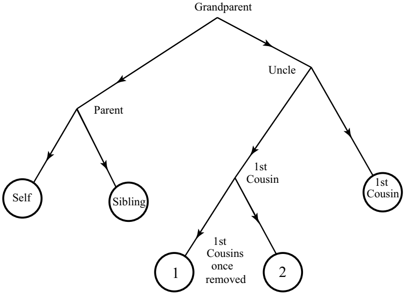
Using 5 dimensions to identify a first cousin once removed.

S. Forcey, U. Akron.

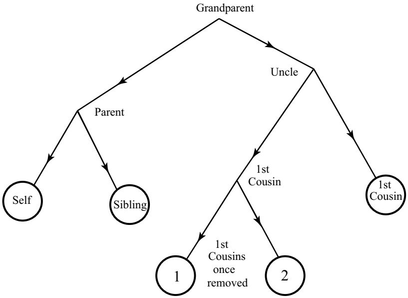


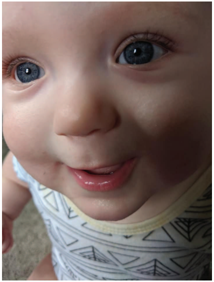
Trees



Family tree.

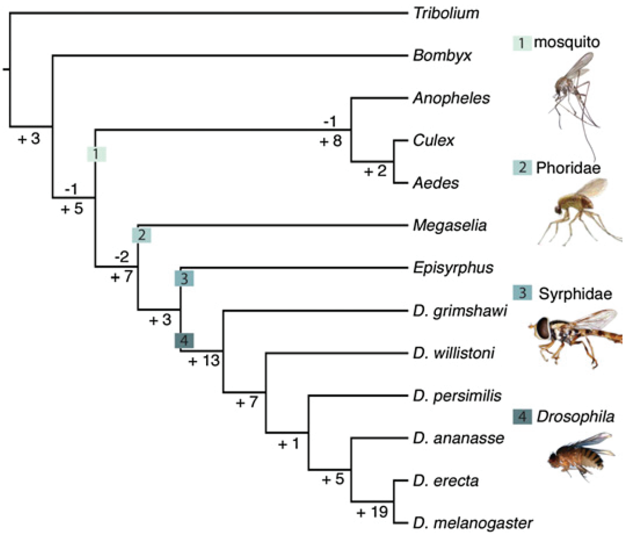
Trees





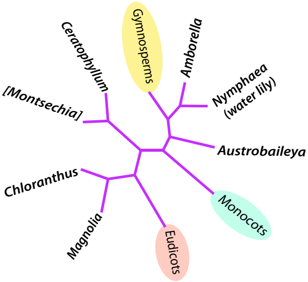
Family tree.

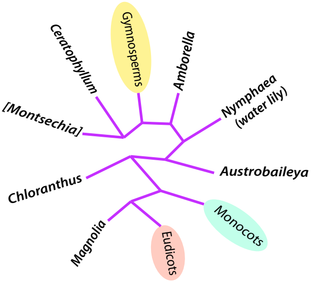
Trees

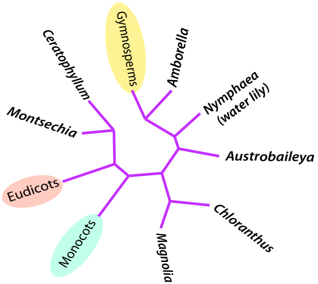


Episodic radiations in the fly tree of life, Wiegmann et.al. PNAS 2011

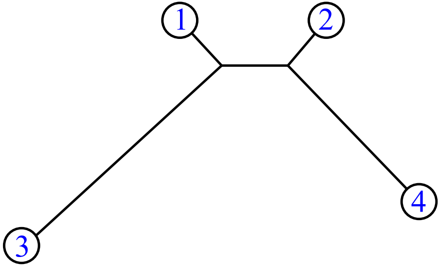
Trees


Ceratophyllum
Gymnosperms
Amborella
[Montsechia]
Nymphaea
(water lily)
Austrobaileya
Chloranthus
Monocots
Magnolia
Eudicots


Ceratophyllum
Gymnosperms
Amborella
[Montsechia]
Nymphaea
(water lily)
Austrobaileya
Chloranthus
Monocots
Magnolia
Eudicots

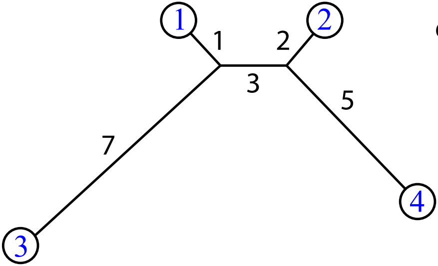

Ceratophyllum
Gymnosperms
Amborella
Montsechia
Nymphaea
(water lily)
Austrobaileya
Eudicots
Monocots
Magnolia
Chloranthus

The Balanced minimal evolution method: ex. tree metric.



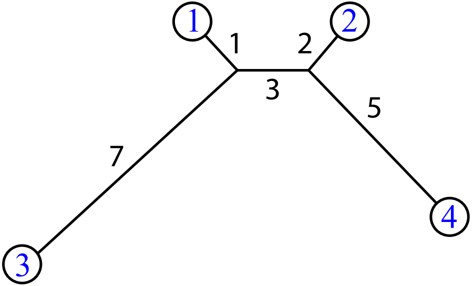
Definition: A *phylogenetic tree*, hereafter *tree*, is a tree with labeled leaves, unlabeled vertices of degree 3 or larger, and without degree 2 vertices. A *rooted tree* has a distinguished leaf.

The Balanced minimal evolution method: ex. tree metric.



$$
\begin{array}{c}
\text{distance 1-2} \\
\begin{bmatrix} 0 & 6 & 8 & 9 \\ 6 & 0 & 12 & 7 \\ 8 & 12 & 0 & 15 \\ 9 & 7 & 15 & 0 \end{bmatrix}
\end{array}
$$

The Balanced minimal evolution method: ex. tree metric.



$$\begin{bmatrix} 0 & 6 & 8 & 9 \\ 6 & 0 & 12 & 7 \\ 8 & 12 & 0 & 15 \\ 9 & 7 & 15 & 0 \end{bmatrix}$$

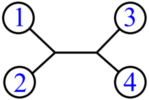
$$\boldsymbol{d} = \langle 6, 8, 9, 12, 7, 15 \rangle$$

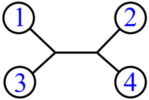
Now: if we are given d, (experiment, measurement), can we recover the original tree?

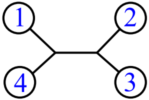
The Balanced minimal evolution method: ex. tree metric.

$$x(t)_{ij} = 2^{(n-1-p_{ij})}$$

t $\qquad$ $x(t)$

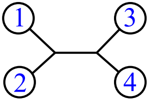
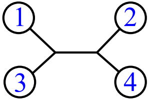
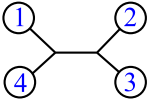






The Balanced minimal evolution method: ex. tree metric.

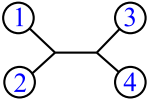
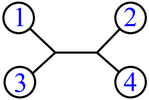
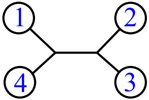
$$x(t)_{ij} = 2^{(n-1-p_{ij})}$$

t	$x(t)$	$\boldsymbol{d} \cdot x(t)$
	$\langle 2, 1, 1, 1, 1, 2 \rangle$	
	$\langle 1, 2, 1, 1, 2, 1 \rangle$	
	$\langle 1, 1, 2, 2, 1, 1 \rangle$	

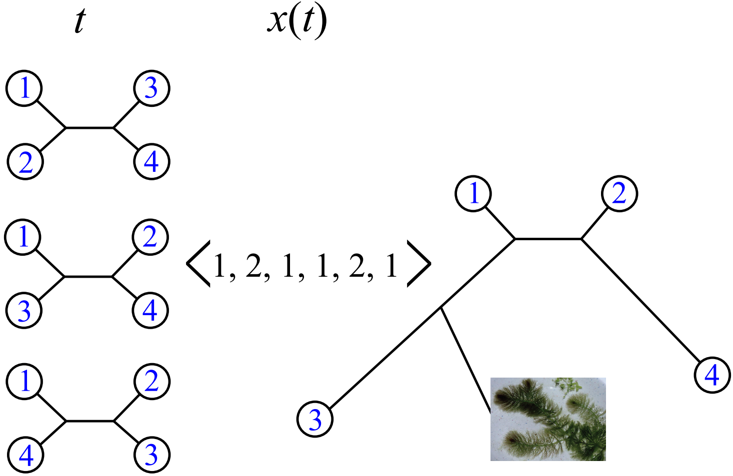
The Balanced minimal evolution method: ex. tree metric.

$$x(t)_{ij} = 2^{(n-1-p_{ij})}$$

Given $\boldsymbol{d} =$ (6, 8, 9, 12, 7, 15), find the tree whose branches may be assigned lengths to achieve those distances.

t	$x(t)$	$\boldsymbol{d}\cdot x(t)$
	(2, 1, 1, 1, 1, 2)	78
	(1, 2, 1, 1, 2, 1)	72
	(1, 1, 2, 2, 1, 1)	78

The Balanced minimal evolution method: ex. tree metric.



Theorems

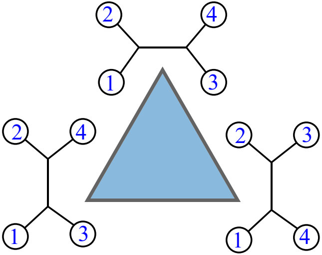
1) The BME method gives the unique tree if $\mathbf{d}$ is a *tree metric*. (L. Pauplin, 2000)

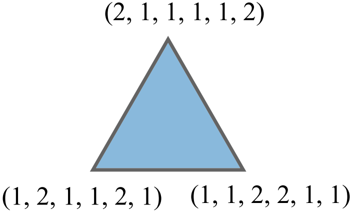
2) The BME method is *statistically consistent* (R. Desper, O. Gascuel, 2004.)

3) The BME vectors $\mathbf{x}(t)$ are the vertices of a polytope sequence which exhibits some recursion: subsequent terms have faces equivalent to prior terms. (D. Haws, T. Hodges, R. Yoshida, L. Pachter, P. Huggins, K. Eickmeyer, 2008.)

4) The BME problem is NP-hard, even when restricted to metric instances. (S. Fiorini, G. Joret, 2012.)

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





Statistics.

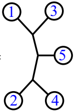
- Dimensions (start $n = 3$): $0, 2, 5, 9, 14 \ldots \binom{n}{2} - n$

$$\text{vertices } \mathbf{x}(t) \text{ obey } \sum_{\substack{i=1 \\ i \neq j}}^{n} x_{ij} = 2^{n-2} \text{ for } j = 1, \ldots, n$$

- Number of Vertices in n^{th} polytope: $1, 3, 15, 105, \ldots (2n-5)!!$
- Number of Facets: $0, 3, 52, 90262 \ldots$ OPEN
- f-vectors: $1, 3, 3, 1, 15, 105, 250, 210, 52, 1, 105, 5460 \ldots$

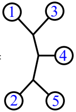
The Balanced minimal evolution polytope $\mathcal{P}_5$.

$t =$



$\vec{c}(t) = (1, 4, 1, 2, 1, 4, 2, 1, 2, 2)$

$t =$



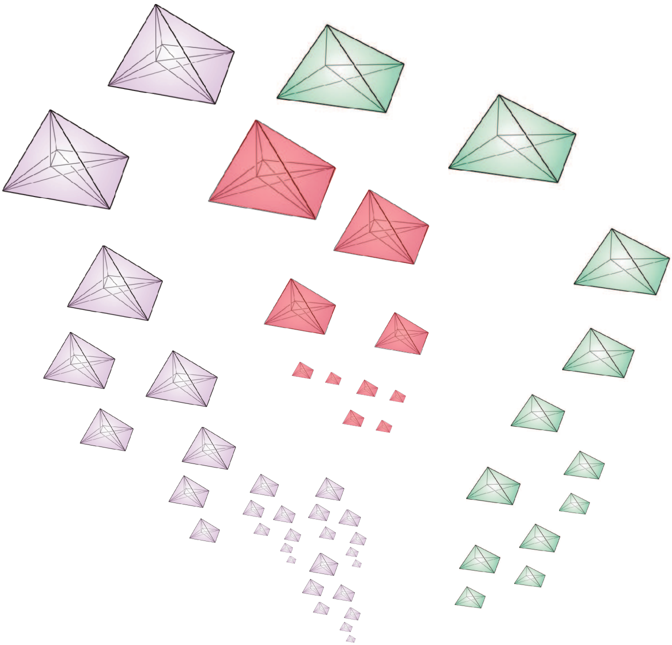
$\vec{c}(t) = (1, 4, 2, 1, 1, 2, 4, 2, 1, 2)$

$$
\begin{array}{l}
\{(4,1,1,2,1,1,2,4,2,2) \\
(4,2,1,1,2,1,1,2,2,4) \\
(4,1,2,1,1,2,1,2,4,2) \\
(2,1,4,1,2,2,2,1,4,1) \\
(2,2,2,2,1,4,1,1,4,1) \\
(1,4,1,2,1,4,2,1,2,2) \\
(1,2,1,4,2,4,1,2,2,1) \\
(2,1,1,4,2,2,2,4,1,1) \\
(1,1,2,4,4,2,1,2,1,2) \\
(1,1,4,2,4,1,2,1,2,2) \\
(2,2,2,2,4,1,1,1,1,4) \\
(2,4,1,1,2,2,2,1,1,4) \\
(1,4,2,1,1,2,4,2,1,2) \\
(1,2,4,1,2,1,4,2,2,1) \\
(2,2,2,2,1,1,4,4,1,1)\}
\end{array}
$$

$\vec{f} = \left\langle 15, 105, 250, 210, 52, 1 \right\rangle$

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.

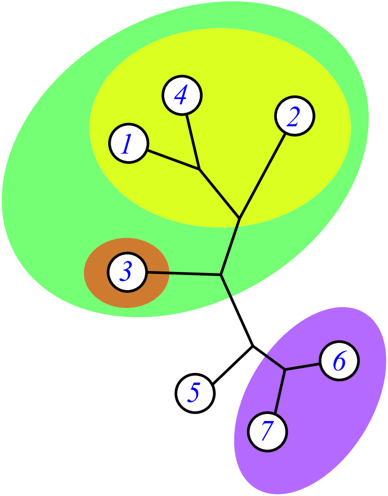
The Balanced minimal evolution polytope $\mathcal{P}_5$.



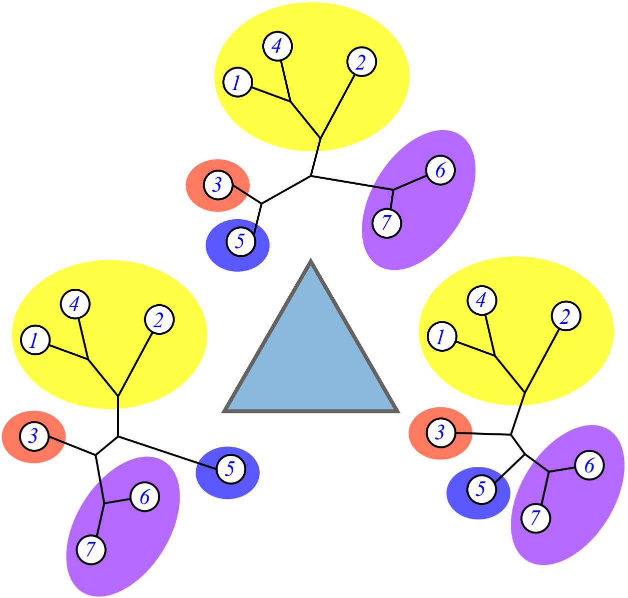
Definitions.

- A *clade* is a sub-tree of a phylogenetic tree which is a connected component after deleting a single interior edge. (It contains all the leaves of a single ancestor, for rooted trees).
- A *cherry* is a clade with only two leaves.
- A pair of *intersecting cherries* $\{a, b\}$ and $\{b, c\}$ have intersection in one leaf b, and thus cannot exist both on the same tree.
- A *caterpillar* is a tree with only two cherries.
- A *split* of the set of n leaves for our phylogenetic trees is a partition of the leaves into two parts, one part called S with m leaves and another with the remaining $n - m$ leaves. A tree *displays* a split if each part makes up the leaves of a *clade*.
- A *tube* is a connected subgraph. A clade is a specialized tube. A *tubing* is a set of nested or disconnected tubes. Any set of clades on a rooted tree form a tubing.

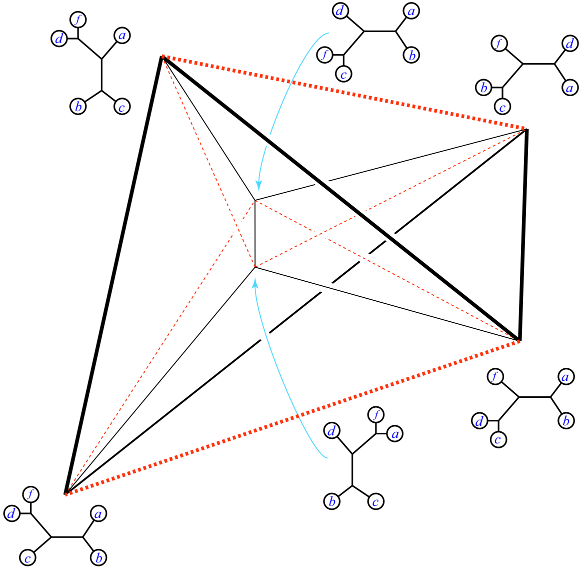
Definitions.



Clade face: K. Eickmeyer et al.



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



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

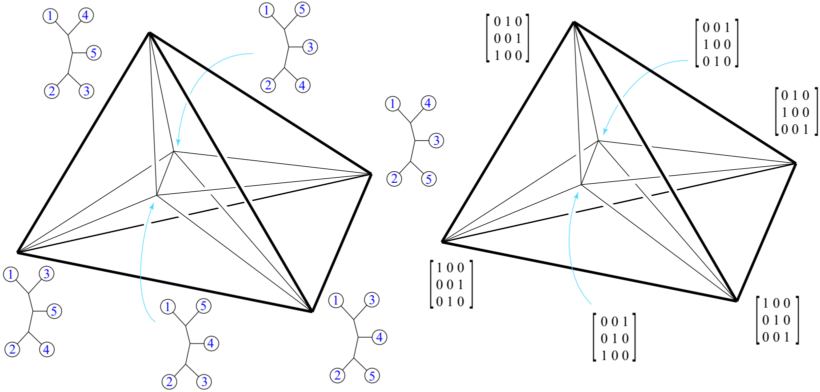
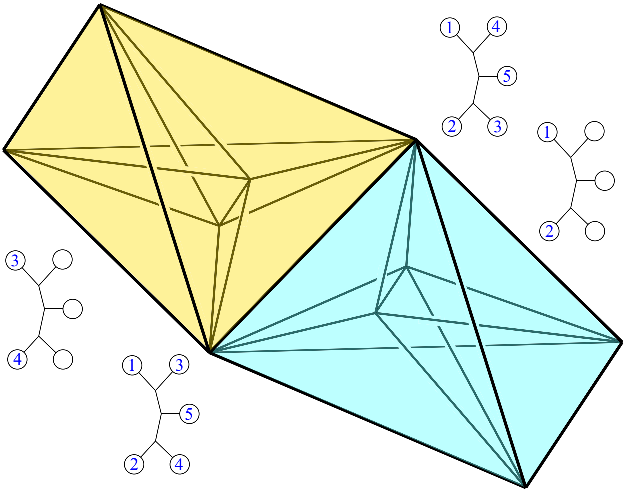


Figure: On the left is a facet of $\mathcal{P}_5$ with each vertex labeled by the caterpillar tree. On the right is the Birkhoff polytope $B(3)$ with vertices labeled by the corresponding permutation matrices.

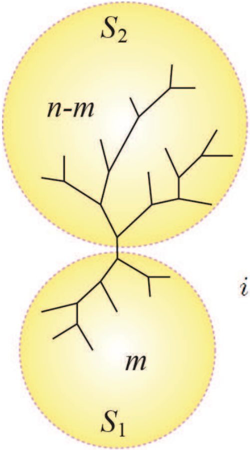
Intersection.



Theorem

Let t be a phylogenetic tree with $n > 5$ leaves which has exactly two nodes ν and μ, with degrees both larger than 3. Then the trees which refine t are the vertices of a facet of the BME polytope $\mathcal{P}_n$.

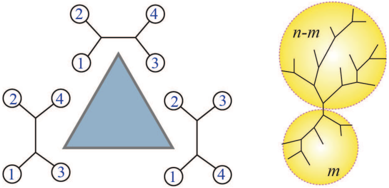
Split faces; split facets.



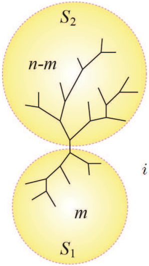
$$\sum_{i<j,\ \text{leaves } i,j \in S_1} x_{ij} \leq (m-1)2^{n-3}$$

Split faces; split facets.

Question. If we use branch and bound to optimize on the region bounded by split faces of the BME polytope, are we guaranteed to get a valid tree?



Splitohedron.



$$\sum_{i<j,\ \text{leaves } i,j \in S_1} x_{ij} \leq (m-1)2^{n-3}$$

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 $[1, 2^{n-3}]^{\binom{n}{2}}$.

Features of the BME polytope $\mathcal{P}_n$

number of species	dim. of $\mathcal{P}_n$	vertices of $\mathcal{P}_n$	facets of $\mathcal{P}_n$	facet inequalities (classification)	number of facets	number of vertices in facet
3	0	1	0	-	-	-
4	2	3	3	$x_{ab} \geq 1$	3	2
				$x_{ab} + x_{bc} - x_{ac} \leq 2$	3	2
5	5	15	52	$x_{ab} \geq 1$ (caterpillar)	10	6
				$x_{ab} + x_{bc} - x_{ac} \leq 4$ (intersecting-cherry)	30	6
				$x_{ab} + x_{bc} + x_{cd} + x_{df} + x_{fa} \leq 13$ (cyclic ordering)	12	5
6	9	105	90262	$x_{ab} \geq 1$ (caterpillar)	15	24
				$x_{ab} + x_{bc} - x_{ac} \leq 8$ (intersecting-cherry)	60	30
				$x_{ab} + x_{bc} + x_{ac} \leq 16$ $(3,3)$-split	10	9
n	$\binom{n}{2} - n$	$(2n-5)!!$	?	$x_{ab} \geq 1$ (caterpillar)	$\binom{n}{2}$	$(n-2)!$
				$x_{ab} + x_{bc} - x_{ac} \leq 2^{n-3}$ (intersecting-cherry)	$\binom{n}{2}(n-2)$	$2(2n-7)!!$
				$x_{ab} + x_{bc} + x_{ac} \leq 2^{n-2}$ $(m,3)$-split, $m \geq 3$	$\binom{n}{3}$	$3(2n-9)!!$
				$\sum_S x_{ij} \leq (m-1)2^{n-3}$ $(m, n-m)$-split S, $m > 2, n > 5$	$2^{n-1} - \binom{n}{2} - n - 1$	$(2(n-m)-3)!! \times (2m-3)!!$

Splitohedron.

```
polytope > print $p->VERTICES;

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

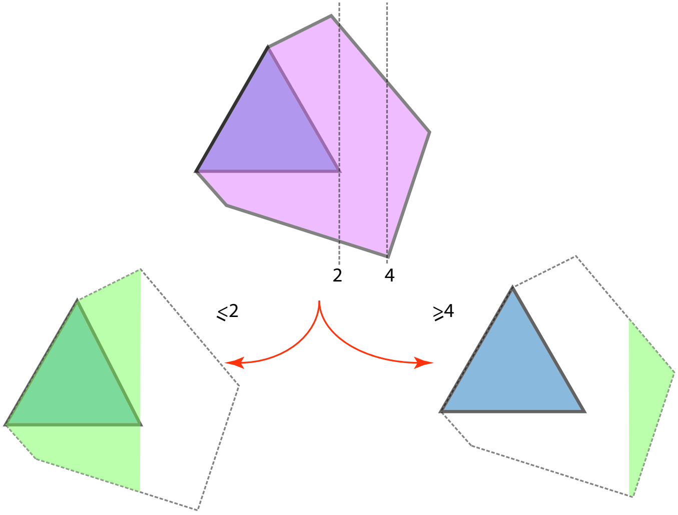
Splitohedron.

```
polytope > print $p->VERTICES;

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

1 2 2 2 2 1 1 4 4 1 1
1 2 2 2 2 1 4 1 1 4 1
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
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
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
```

BnB.



A2: So far so good!

- We tested up to $n = 10$, with and without noise.
- Results are completely accurate...
- We need to find a way to break it! MatLab code available: `http://www.math.uakron.edu/~sf34/class_home/research.htm`

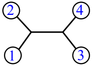
More polytopes.

For any circular split system S, $\mathbf{x}(S)$ is a vector whose *ij*-component is the number of circular orderings consistent with that system for which i and j are adjacent.

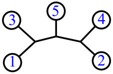
$$\text{These vertices } \mathbf{x}(t) \text{ obey } \sum_{\substack{i=1 \\ i \neq j}}^{n} x_{ij} = 2^{k+1} \text{ for } j = 1, \ldots, n$$

where k is the number of *bridges* in the diagram.

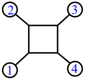
Split network vectors.



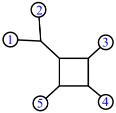
(2, 1, 1, 1, 1, 2)



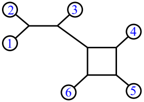
(1, 4, 1, 2, 1, 4, 2, 1, 2, 2)



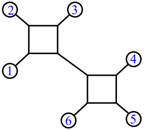
(1, 0, 1, 1, 0, 1)



(2, 1, 0, 1, 1, 0, 1, 2, 0, 2)



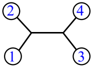
(4, 2, 1, 0, 1, 2, 1, 0, 1, 2, 0, 2, 4, 0, 4)



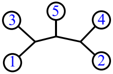
(2, 0, 1, 0, 1, 2, 0, 0, 0, 1, 0, 1, 2, 0, 2)

Notes: Agrees with previous $x(t)$. Gives TSP when there are no bridges.

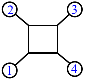
Split network vectors.



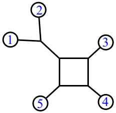
(2, 1, 1, 1, 1, 2)



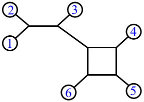
(1, 4, 1, 2, 1, 4, 2, 1, 2, 2)

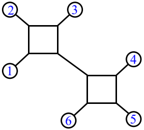


(1, 0, 1, 1, 0, 1)



(2, 1, 0, 1, 1, 0, 1, 2, 0, 2)





(4, 2, 1, 0, 1, 2, 1, 0, 1, 2, 0, 2, 4, 0, 4)

(2, 0, 1, 0, 1, 2, 0, 0, 0, 1, 0, 1, 2, 0, 2)

Notes: Agrees with previous $x(t)$. Gives TSP when there are no bridges.

A filtration of split networks.

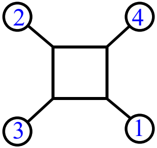
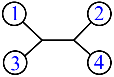
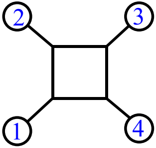
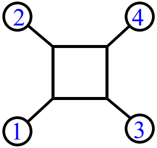
Definition. Let $\mathrm{BME}(n,k)$ be the convex hull of the split network vectors for the split networks having n leaves and k bridges.

Idea: a split network distance vector d (seen as a linear functional) from a split network (with edge lengths) and $j \geq k$ bridges will be simultaneously minimized at the vertices of $\mathrm{BME}(n,k)$ which correspond to the cycles which d resolves.

A filtration of split networks.

Specifically: A tree metric d (as linear functional) is minimized simultaneously at the vertices of the TSP which correspond to the cycles with which d is compatible

A filtration of split networks.

S	$x(S)$	$\boldsymbol{d}\cdot x(S)$
	$\langle 0, 1, 1, 1, 1, 0 \rangle$ $\langle 6, 8, 9, 12, 7, 15 \rangle$	36 
	$\langle 1, 0, 1, 1, 0, 1 \rangle$	42
	$\langle 1, 1, 0, 0, 1, 1 \rangle$	36

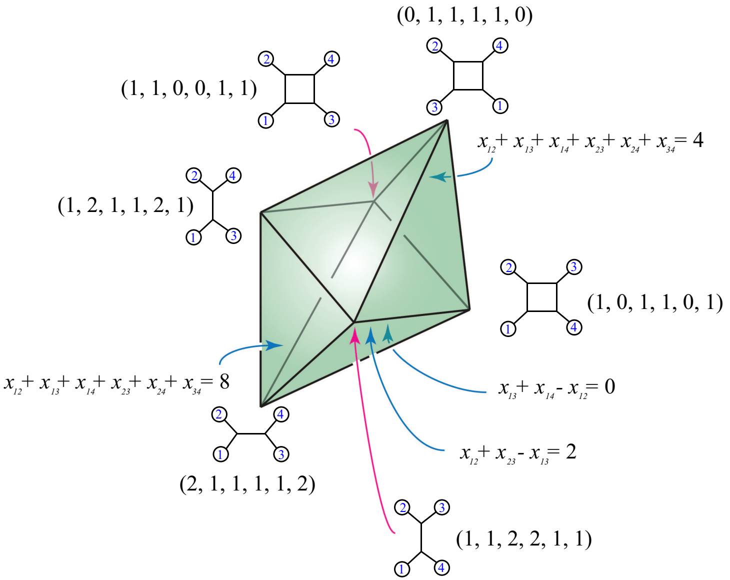
Or...

We might propose an extension of the BME polytope which is the the convex hull of all vectors $\eta(S)$ for binary split systems S on a set of size n.

This new polytope has vertices corresponding to all the binary split systems.

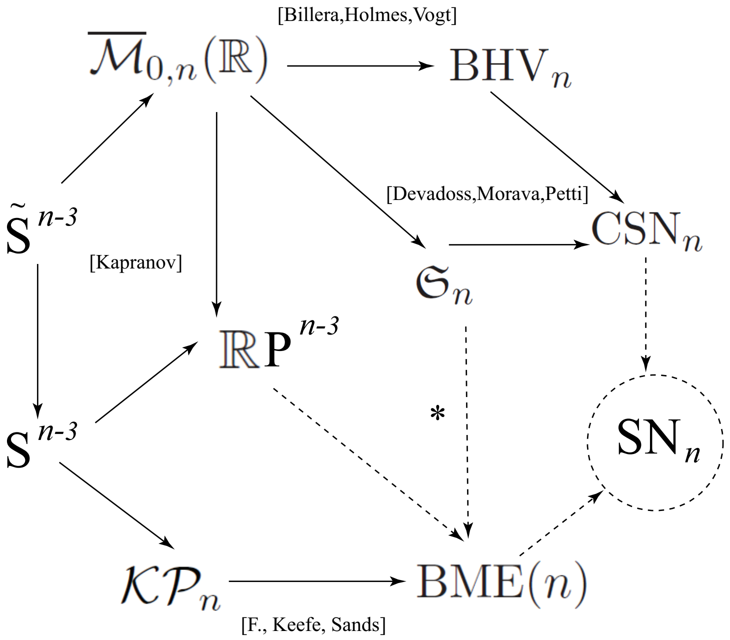
These binary split systems come in two varieties: the binary phylogenetic trees and the split systems for which any split is incompatible with at most one other split.

Next.



(0, 1, 1, 1, 1, 0)
(1, 1, 0, 0, 1, 1)
$x_{12} + x_{13} + x_{14} + x_{23} + x_{24} + x_{34} = 4$
(1, 2, 1, 1, 2, 1)
(1, 0, 1, 1, 0, 1)
$x_{12} + x_{13} + x_{14} + x_{23} + x_{24} + x_{34} = 8$
$x_{13} + x_{14} - x_{12} = 0$
(2, 1, 1, 1, 1, 2)
$x_{12} + x_{23} - x_{13} = 2$
(1, 1, 2, 2, 1, 1)

Next.


[Billera,Holmes,Vogt]
$\overline{\mathcal{M}}_{0,n}(\mathbb{R})$
BHV_n
[Devadoss,Morava,Petti]
$\tilde{S}^{n-3}$
CSN_n
[Kapranov]
$\mathfrak{S}_n$
$\mathbb{R}P^{n-3}$
*
S^{n-3}
SN_n
$\mathcal{KP}_n$
$\mathrm{BME}(n)$
[F., Keefe, Sands]

Thanks so much!