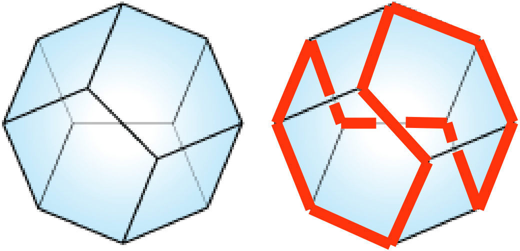
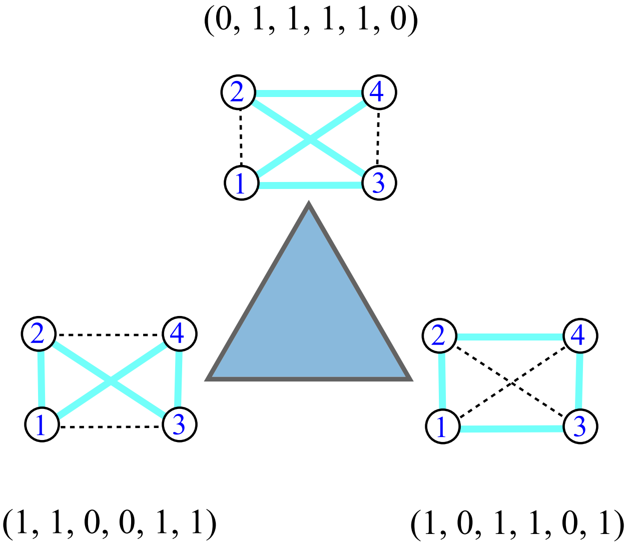
More Phylogenetic polytopes: filtering the STSP.

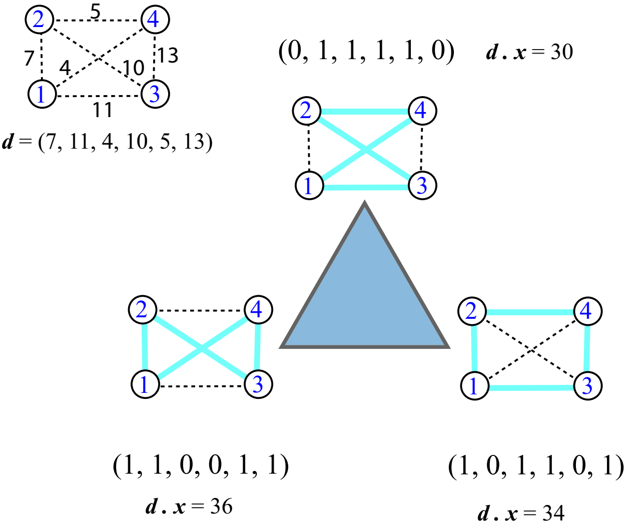
S. Forcey, L. Keefe, W. Sands. U. Akron.
S. Devadoss. U. San Diego



STSP



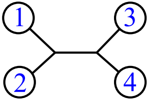
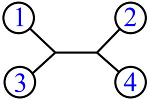
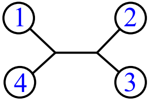
STSP


2 4 1 3
5 7 4 10 13 11
$\boldsymbol{d} = (7, 11, 4, 10, 5, 13)$
$(0, 1, 1, 1, 1, 0)$ $\boldsymbol{d} \cdot \boldsymbol{x} = 30$
$(1, 1, 0, 0, 1, 1)$
$\boldsymbol{d} \cdot \boldsymbol{x} = 36$
$(1, 0, 1, 1, 0, 1)$
$\boldsymbol{d} \cdot \boldsymbol{x} = 34$

The Balanced minimal evolution method: ex. tree metric.

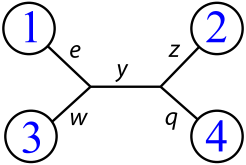
$$x(t)_{ij} = 2^{(n-1-p_{ij})}$$

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

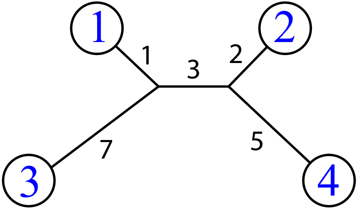
The Balanced minimal evolution method: ex. tree metric.

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

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

solve

$$e + y + z = 6$$
$$e + w = 8$$
$$e + y + q = 9$$
$$w + y + z = 12$$
$$z + q = 7$$
$$w + y + q = 15$$



Notes.

1) This is slow–better to use linear programming on the polytope: hence the search for facets.

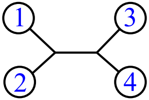
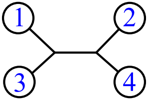
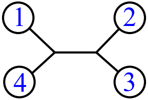
2) Notice that this method fixes the long branch problem.

3) The proof relies on the fact that our dot product calculates a multiple of the sum of the edge lengths.

4) Recall that the method returns an answer even if the distances are not a tree metric.

The Balanced minimal evolution method: ex. tree metric?

Given $\boldsymbol{d}$ = (7, 11, 4, 10, 5, 13), find the tree whose branches may be assigned lengths to achieve those distances.

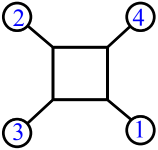
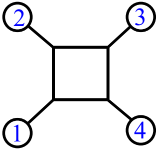
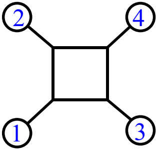
t	$x(t)$	$\boldsymbol{d}\cdot x(t)$
	(2, 1, 1, 1, 1, 2)	66
	(1, 2, 1, 1, 2, 1)	64
	(1, 1, 2, 2, 1, 1)	70

…but

solving for the edges gives no solution.

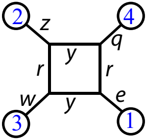
The Balanced minimal evolution method: ex. tree metric?

Given $\boldsymbol{d} = (7, 11, 4, 10, 5, 13)$, find the tree whose branches may be assigned lengths to achieve those distances.

S	$x(S)$	$\boldsymbol{d}\cdot x(S)$
	(0, 1, 1, 1, 1, 0)	30
	(1, 0, 1, 1, 0, 1)	34
	(1, 1, 0, 0, 1, 1)	36

The Balanced minimal evolution method: ex. tree metric?

Given $\boldsymbol{d} =$ (7, 11, 4, 10, 5, 13), find the tree whose branches may be assigned lengths to achieve those distances.

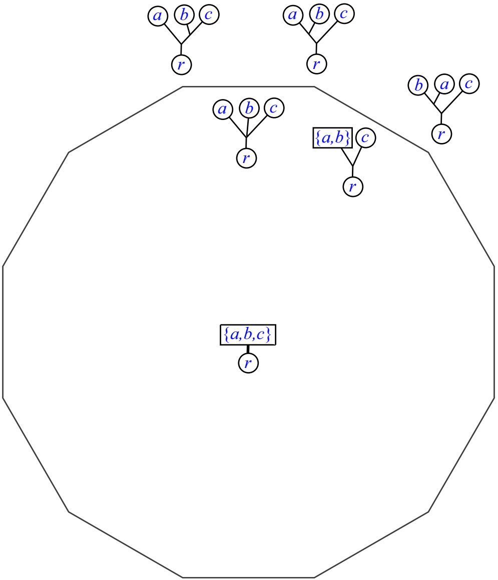
S	$x(S)$	$\boldsymbol{d} \cdot x(S)$
	(0, 1, 1, 1, 1, 0)	30

solve

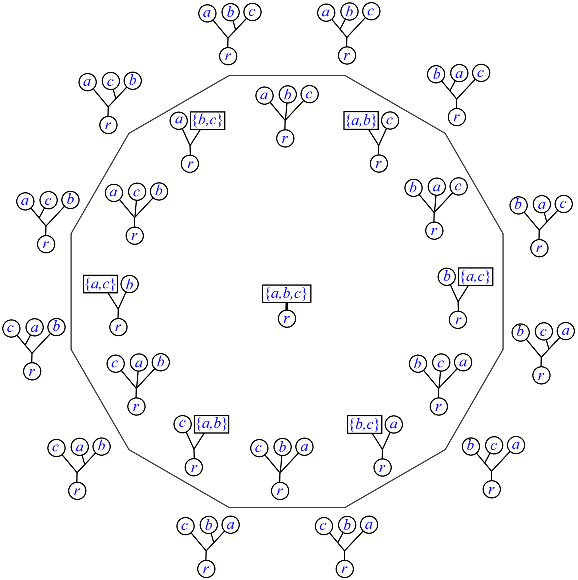
$e + y + z + r = 7$
$e + r + w = 11$
$e + y + q = 4$
$w + y + z = 10$
$z + r + q = 5$
$w + y + q + r = 13$

$r = 3$
$w = 7$
$x = 1$
$y = 2$
$z = 1$
$q = 1$

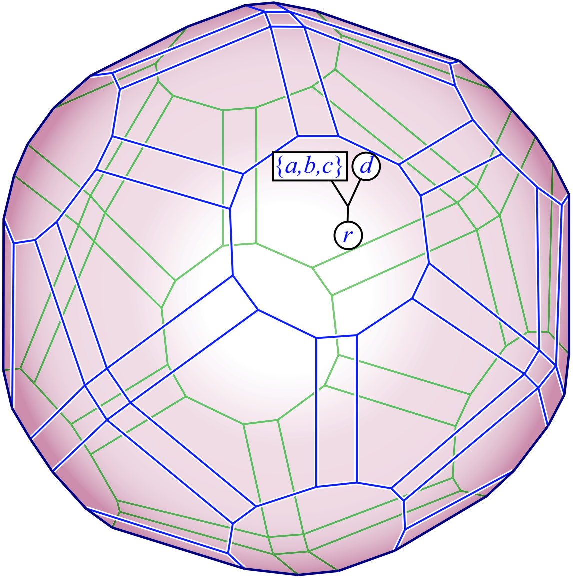
Permutoassociahedron $\mathcal{KP}_2$



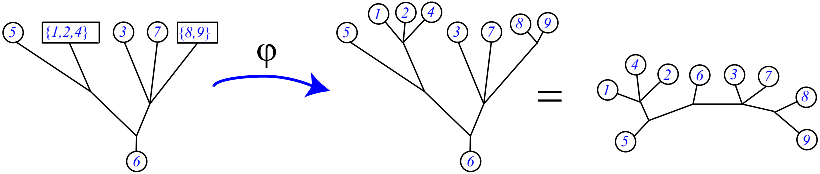
Permutoassociahedron $\mathcal{KP}_2$



Permutoassociahedron $\mathcal{KP}_3$



Projection to BME(n)



Theorem

If $x \leq y$ as faces in the face lattice of $\mathcal{KP}_n$, then $\varphi(x) \leq \varphi(y)$ as faces in the face lattice of $\mathcal{P}_n$, the BME polytope.

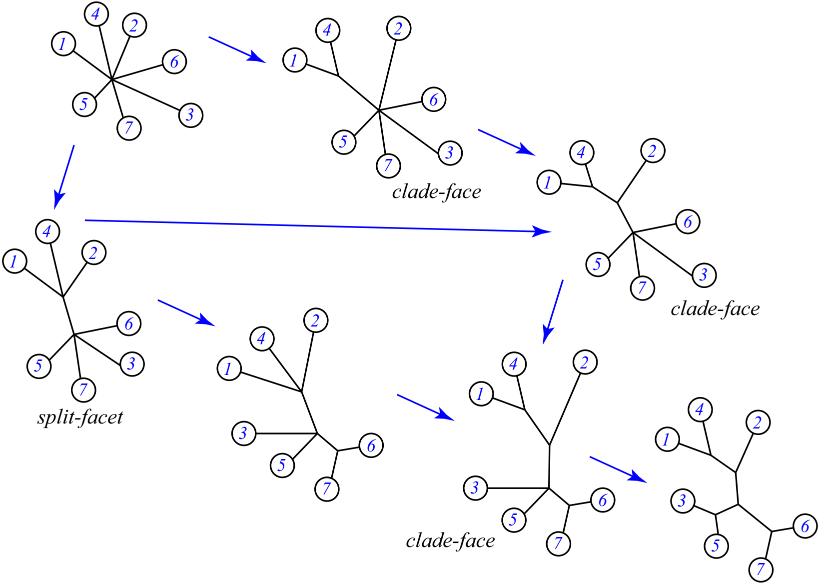
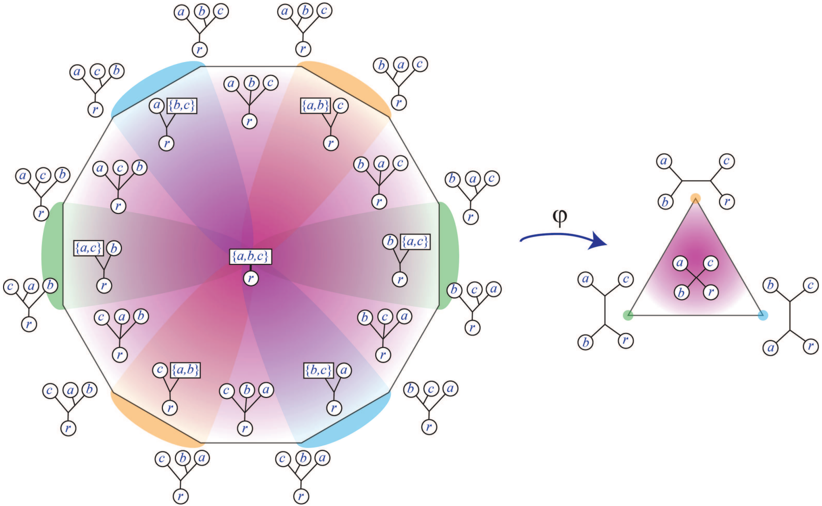


Figure: Examples of chains in the lattice of tree-faces of the BME polytope $\mathcal{P}_9$.

Projection to BME(3)

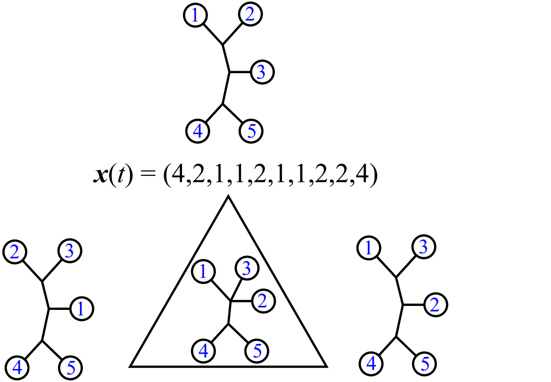


Now we show how the target of the map φ is actually the BME polytope.

Theorem

For each non-binary phylogenetic tree t with n leaves there is a corresponding face $F(t)$ of the BME polytope BME(n). The vertices of $F(t)$ are the binary phylogenetic trees which are refinements of t.

proof idea.



$\boldsymbol{x}(t) = (4,2,1,1,2,1,1,2,2,4)$

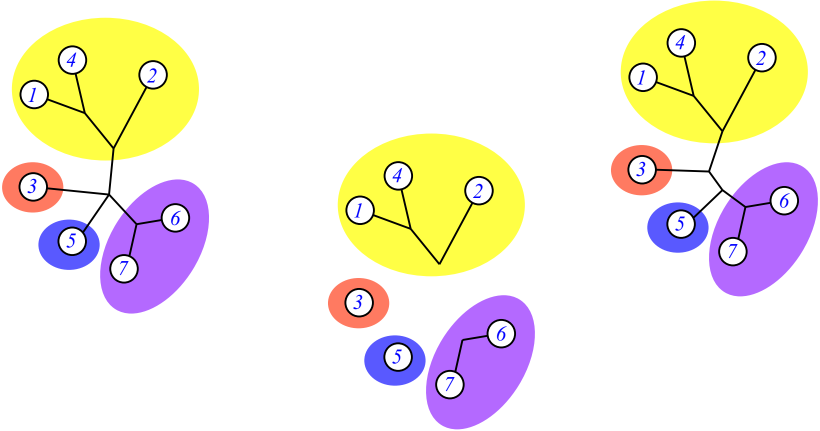
$\boldsymbol{x}(t) = (2,2,2,2,4,1,1,1,1,4)$

$\boldsymbol{x}(t) = (2,4,1,1,2,2,2,1,1,4)$

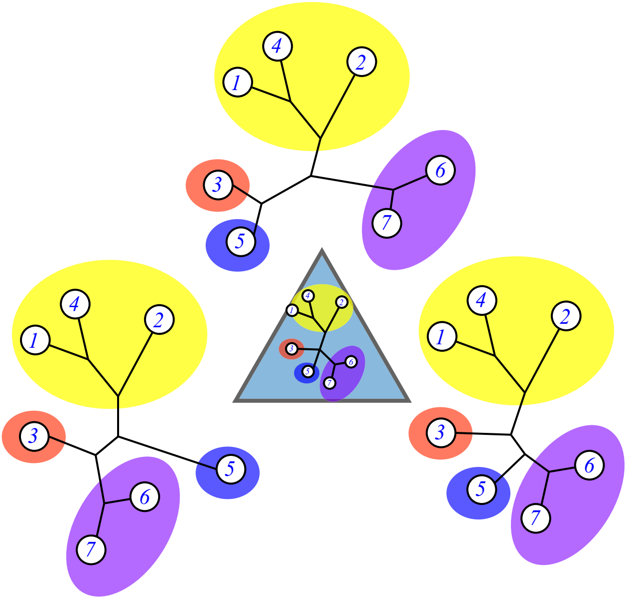
Theorem

For t an n-leaved phylogenetic tree with exactly one node ν of degree $m > 3$, the tree face $F(t)$ is precisely the clade-face $F_{C_1,\ldots,C_m}$, defined in [H,H,Y], corresponding to the collection of clades $C_1, \ldots, C_p$ which result from deletion of ν. Thus $F(t)$ is combinatorially equivalent to the smaller dimensional BME polytope BME(m).

Clade face



Clade face

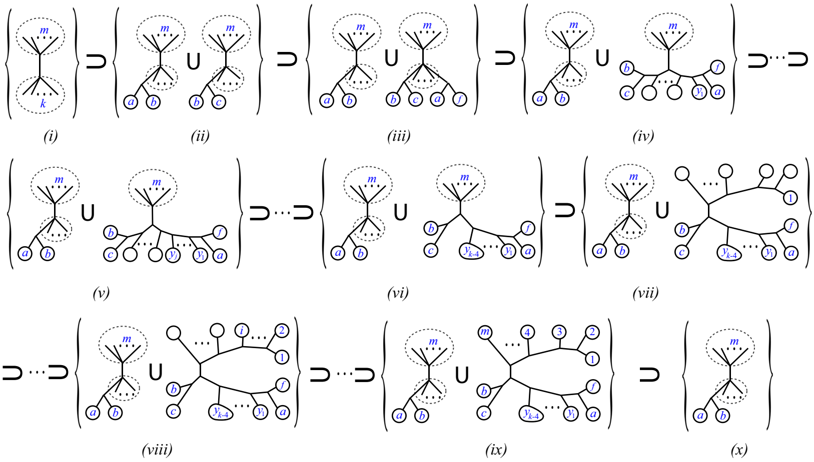


Split facets.

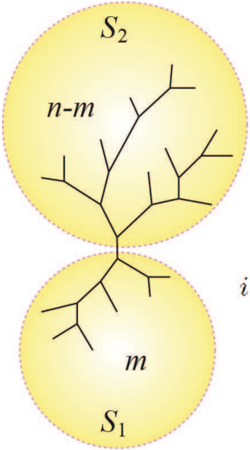
Theorem

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 facets.



Split facets.



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

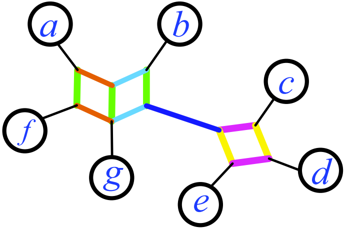
Features of the BME polytope BME(n)

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, $m > 2, n > 5$	$2^{n-1} - \binom{n}{2} - n - 1$	$(2(n-m)-3)!! \times(2m-3)!!$

Definitions

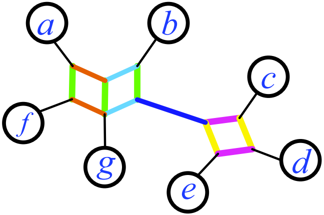
A *split network* is a collection of splits of a set of leaves.
A *split network diagram* represents each split with a set of parallel edges.
A *circular split network*, also known as a planar split network, is a network whose diagram can be drawn on the plane without crossing edges.
A network of *compatible* splits is one whose diagram is a tree.
A *binary* split network is one whose diagram has vertices of degree three (or one, for the leaves) only.

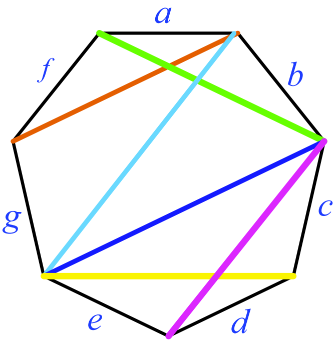
Definitions.



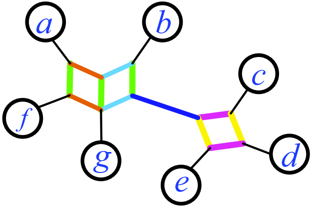
$\{a, f\} | \{b, c, d, e, g\}$
$\{a, b\} | \{c, d, e, f, g\}$
$\{a, f, g\} | \{b, c, d, e\}$
$\{a, b, f, g\} | \{c, d, e\}$
$\{a, b, e, f, g\} | \{c, d\}$
$\{a, b, c, f, g\} | \{d, e\}$

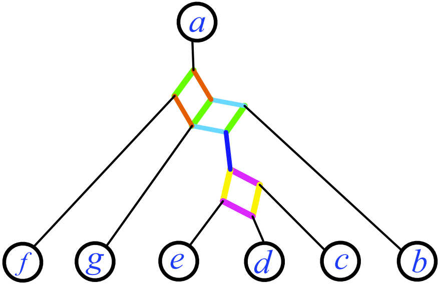
Definitions.





Definitions.



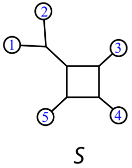


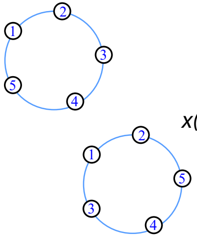
More polytopes.

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

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

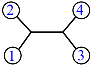
where k is the number of (non-leaf edge) *bridges* in the diagram. (These are non-crossing diagonals in the multitriangulation).



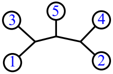


x(S) = (2, 1, 0, 1, 1, 0, 1, 2, 0, 2)

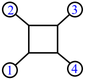
Split network vectors.



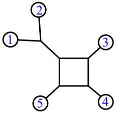
(2, 1, 1, 1, 1, 2)



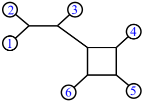
(1, 4, 1, 2, 1, 4, 2, 1, 2, 2)

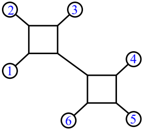


(1, 0, 1, 1, 0, 1)



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



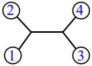


(4, 2, 1, 0, 1, 2, 1, 0, 1, 2, 0, 2, 4, 0, 4)

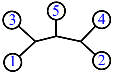
(2, 0, 1, 0, 1, 2, 0, 0, 0, 1, 0, 1, 2, 0, 2)

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

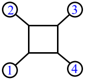
Split network vectors.



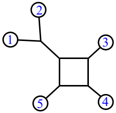
(2, 1, 1, 1, 1, 2)



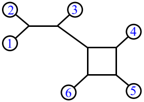
(1, 4, 1, 2, 1, 4, 2, 1, 2, 2)



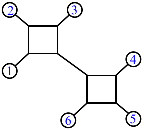
(1, 0, 1, 1, 0, 1)



(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 STSP 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.

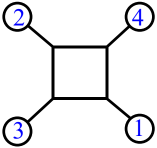
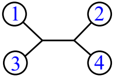
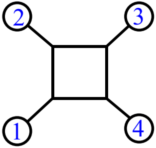
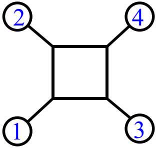
Idea: a split network distance vector d (seen as a linear functional) from a split network S (with edge lengths) and $j \geq k$ bridges will be simultaneously minimized at the vertices of $\mathrm{BME}(n, k)$ which correspond to the split networks that S *resolves*.

S resolves S' means that some splits of S' are collapsed (the parallel edges are assigned length zero) to achieve S.

A filtration of split networks.

Specifically: A tree metric d (as linear functional) is minimized simultaneously at the vertices of the $\mathrm{STSP}(n) = \mathrm{BME}(n, 0)$ which correspond to the cycles with which d is compatible

A filtration of split networks.

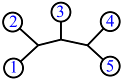
S	$x(S)$	$\boldsymbol{d}\cdot x(S)$
	$\langle 0, 1, 1, 1, 1, 0 \rangle$ $\langle 6, 8, 9, 12, 7, 15 \rangle$	36 
	$\langle 1, 0, 1, 1, 0, 1 \rangle$	42
	$\langle 1, 1, 0, 0, 1, 1 \rangle$	36

Corollary

Every circular split network with k bridges corresponds to a face of each $\mathrm{BME}(n, j)$ polytope for $j \leq k$.

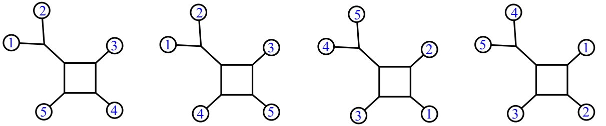
A filtration of split networks.

Every circular split network with k bridges corresponds to a face of each BME(n, j) polytope for $j \leq k$.

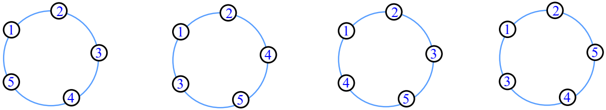


is a vertex in BME(5,2): (4, 2, 1, 1, 2, 1, 1, 2, 2, 4)

and a face with 4 vertices in BME(5,1):

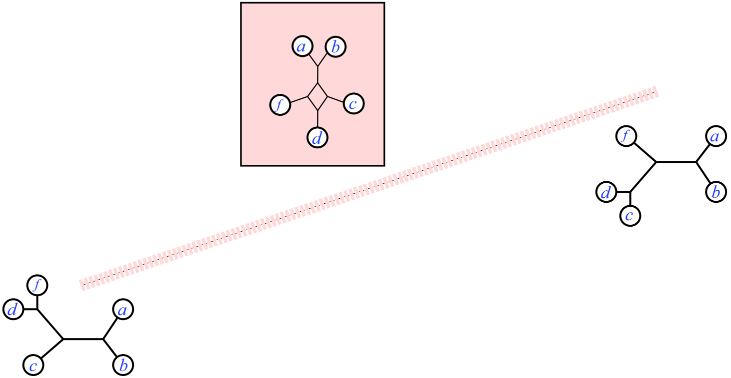


and a face with 4 vertices in BME(5,0):

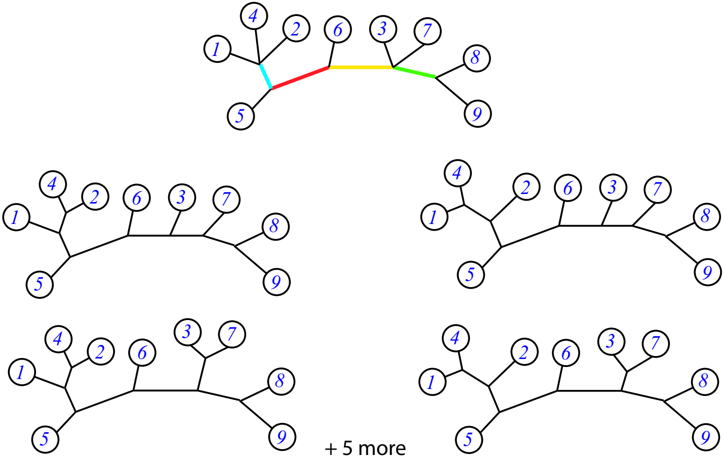


Thanks for day 2! New question tomorrow...

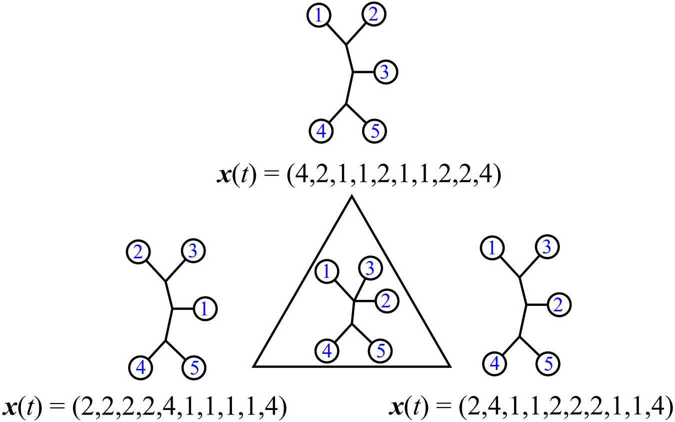
Question 1. Which split networks correspond to faces
(and especially facets)
of the Balanced Minimal Evolution polytope?



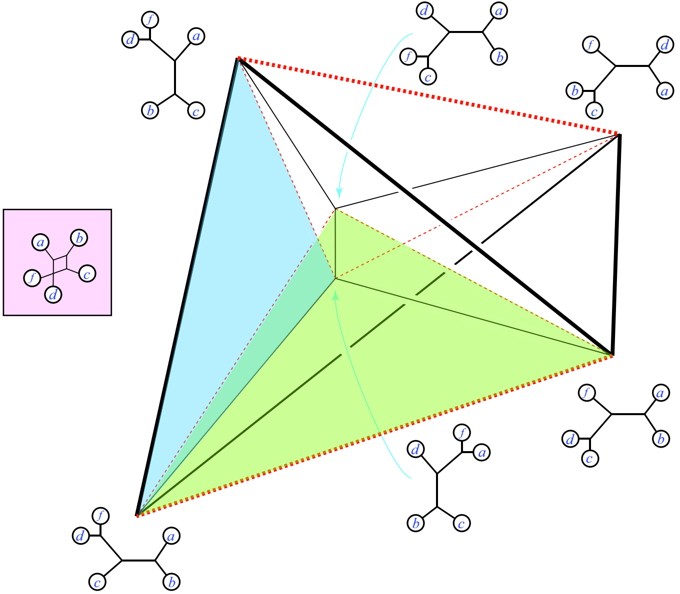
A1. any set of compatible splits.



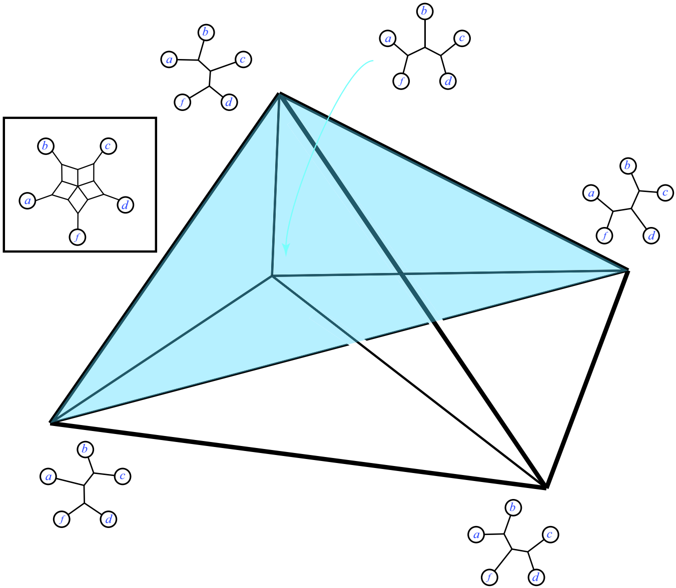
A1. any set of compatible splits.



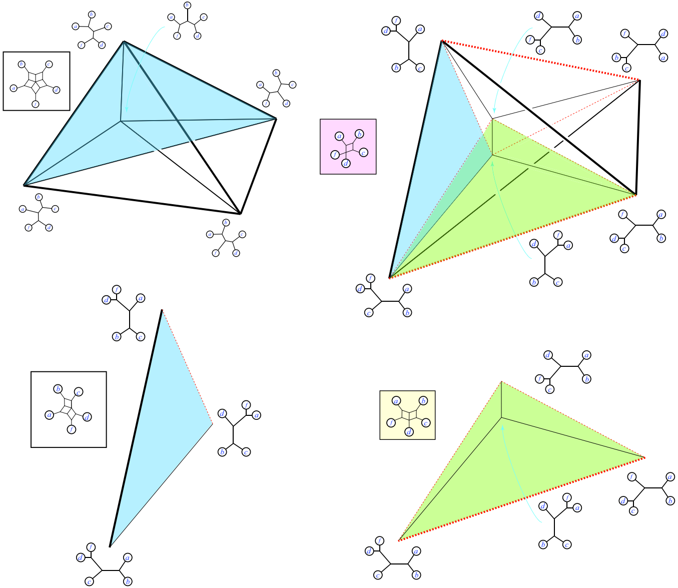
A1. Intersecting cherry splits



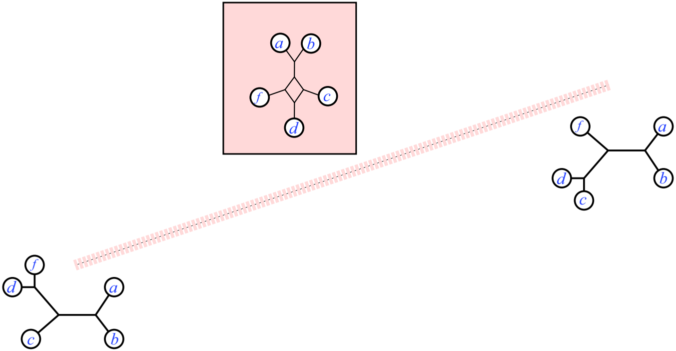
A1: Cyclic splits for $n = 5$



A1: Four split networks.

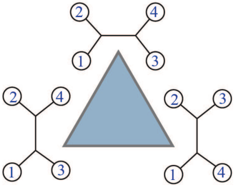


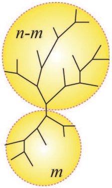
A1: Nearest Neighbor Interchange.



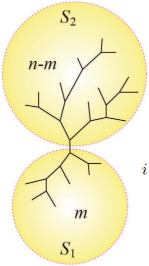
Q2: Split faces; split facets.

Question 2. 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.



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

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 $x_{ij} \le 2^{n-3}$, and the caterpillar facets have the inequality $x_{ij} \ge 1$, 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
```

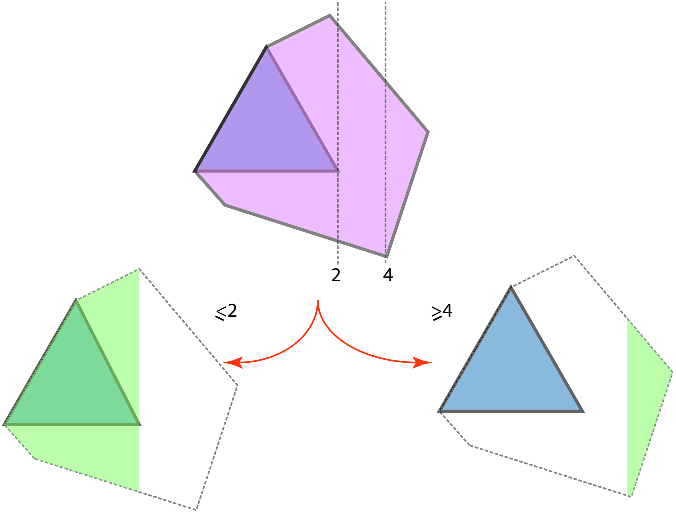
Splitohedron.

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.



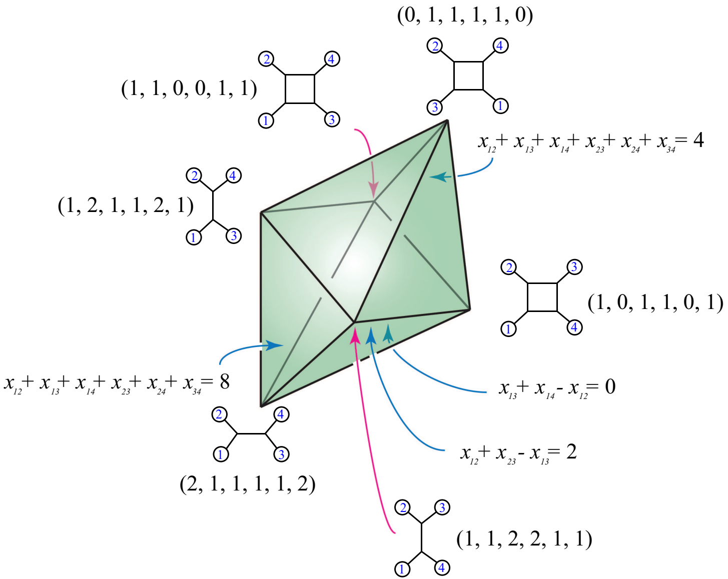
A2: So far so good!

- 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`

Or...

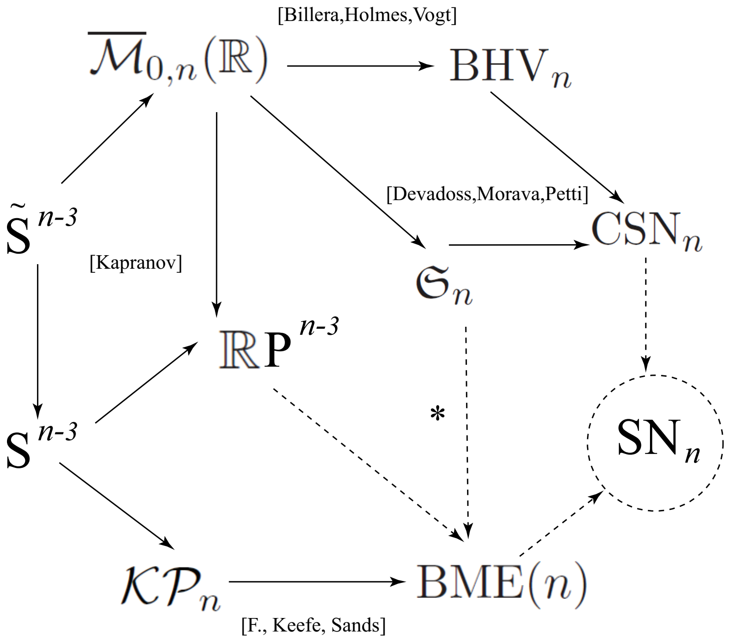
We might propose an extension of the BME polytope which is the the convex hull of all vectors $\eta(S)$ for binary split systems S on a set of size n.
This new polytope has vertices corresponding to all the binary split systems.
These binary split systems come in two varieties: the binary phylogenetic 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)
$x_{12} + x_{13} + x_{14} + x_{23} + x_{24} + x_{34} = 8$
$x_{13} + x_{14} - x_{12} = 0$
$x_{12} + x_{23} - x_{13} = 2$
(2, 1, 1, 1, 1, 2)
(1, 1, 2, 2, 1, 1)

Next.


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

Thanks so much!