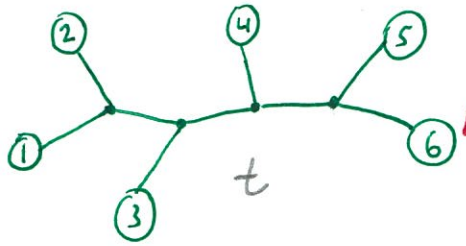
# Introduction to research projects in BME, Balanced Minimum Evolution method.




$n$ taxa (individuals, genomes) as leaves

$$\vec{x}(t) = \langle x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{23}, x_{24}, \dots, x_{56} \rangle$$

where $\boxed{x_{ij} = 2^{n - l_{ij} - 1}}$

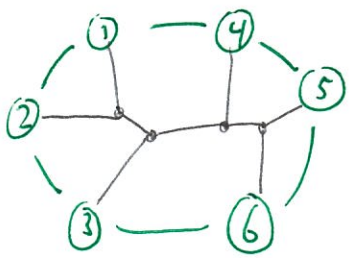
where $l_{ij}$ = length (# of edges) in path from leaf $i$ to leaf $j$.

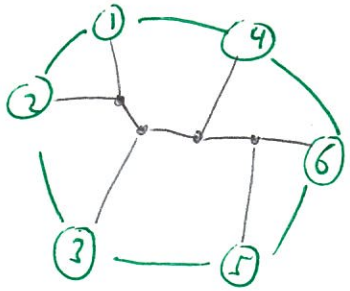
here $\vec{x}(t) = \langle 8, 4, 2, 1, 1, 4, 2, 1, 1, 4, 2, 2, 4, 4, 8 \rangle$.

---

Alternate formula:

$x_{ij}$ = # cycles using $1, \dots, n$ such that $t$ can be drawn with no crossings inside that cycle, with $i$ adjacent to $j$.




Both have 1, 4 adjacent, and these are the only such 2 cycles with $t$.

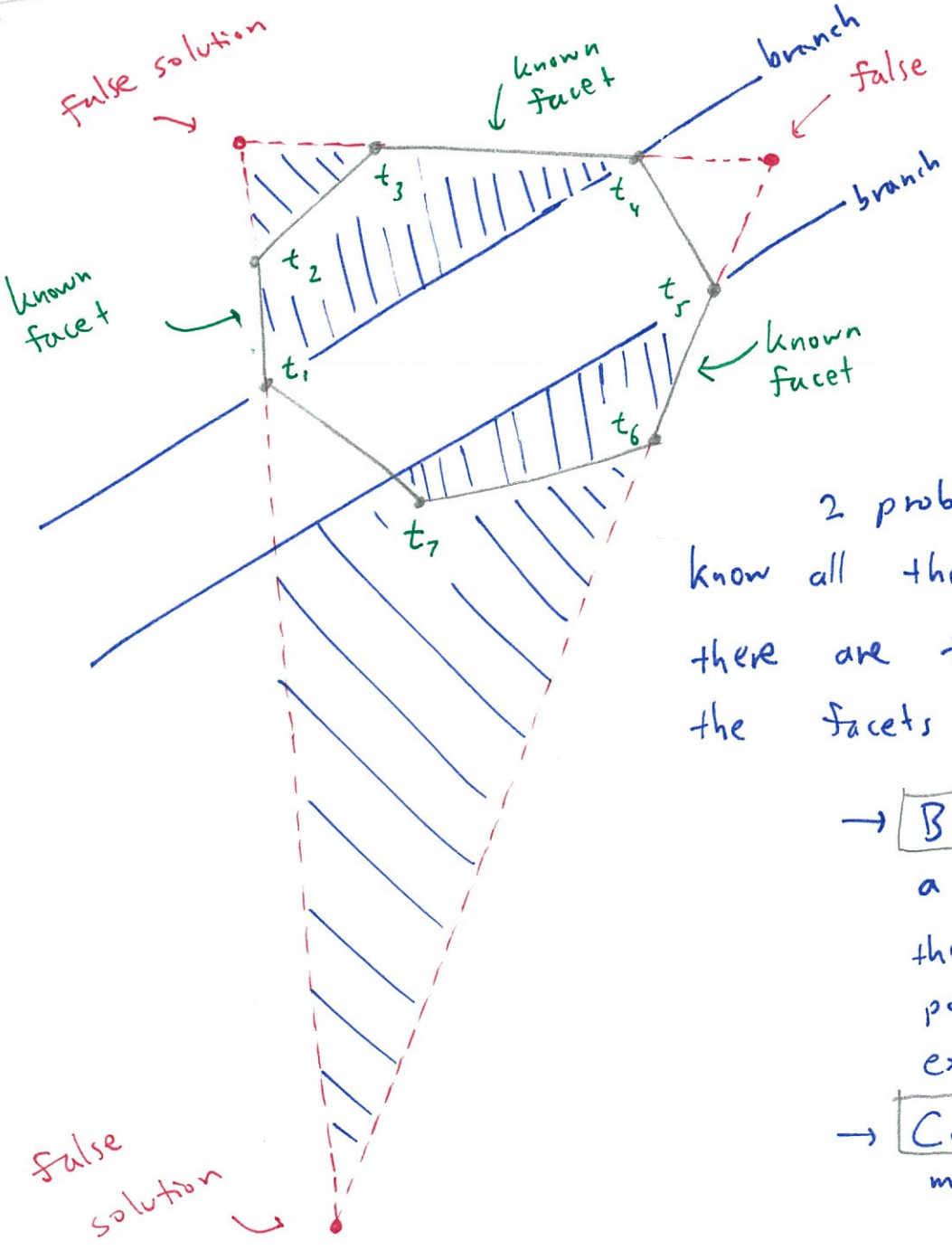
$\Rightarrow x_{14} = 2$.

Goal: given a list (vector, matrix) of pairwise distances $d_{ij}$, find $t$ such that the dot product $\vec{x}(t) \cdot \vec{d}$ is minimized.

---

Method: the vectors $\vec{x}(t)$ are the vertices of the BME(n) polytope. If we knew all the facets (bounding inequalities) then linprog (Matlab) would find the answer.

---




2 problems: we don't know all the facets, and there are too many of the facets we do know!

→ **Branch**: if we get a false solution then make two new polytopes both excluding it.

→ **Cut**: add in more known facets.

① Project:

Vary the parameters and decision making in Matlab, compare results.

- which initial facets to include?
- when and where, and how many, cuts to add during the branching process?
- when and where to fix equalities (reduce dimension)?
- test on real DNA data.

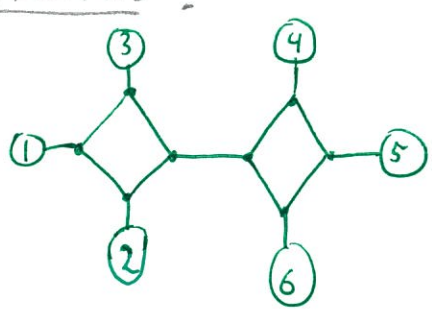
Existing code: www.math.uakron.edu/~sf34/hedra.htm#splito

---

② Project:

Adapt the Matlab code to find circular split networks.

$t =$




$\vec{x}(t) = \langle 2,2,0,0,0,0,1,0,1,\ 1,0,1,2,0,2 \rangle$

... using Alternate formula:

→ the above picture shows the only way in which 3, 4 are adjacent in the cycle.

→ here's another way that 1, 3 are adjacent: