Tropical median consensus trees

with an introduction to tropical convexity

Michael Joswig

TU Berlin & MPI MiS, Leipzig

Séminaires CMLS-CMAP, École Polytechnique, 06 Nov 2024

joint w/ Andrei Comăneci
Lars Kastner
Georg Loho
Benjamin Schröter
and others

1. **Tropical Convexity**
   - max-plus linear algebra
   - regular subdivisions of products of simplices
   - Maslov dequantization
   - applications

2. **A Tropical Fermat–Weber Problem**
   - an asymmetric distance function

3. **Optimal Transport**
   - a dual pair of linear programs
   - application: phylogenetic trees

4. **How Good Is This Method?**
   - theoretically
   - computations on actual & synthetic data

Tropical convexity

$(\mathbb{T}, \oplus, \odot)$ = tropical semiring (with respect to max)

- $\mathbb{T} := \mathbb{R} \cup \{-\infty\}$, $\oplus :=$ max and $\odot := +$

A set $S \subset \mathbb{R}^n$ is a tropical cone if

$$\lambda \odot x \oplus \mu \odot y \in S \quad \text{for all } x, y \in S \text{ and } \lambda, \mu \in \mathbb{R} \ .$$

- tropical projective torus $\mathbb{R}^n/\mathbb{R}\mathbf{1}$

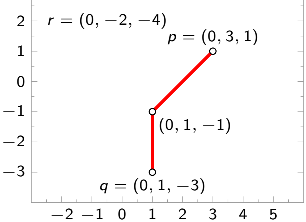
Definition (Develin & Sturmfels 2004)

A set $S' \subset \mathbb{R}^n/\mathbb{R}\mathbf{1}$ is tropically convex if it is the image of a tropical cone under the canonical projection $x \mapsto x + \mathbb{R}\mathbf{1}$.

- tropical polytope = finitely generated tropically convex set
- max-plus linear algebra:
  Cuninghame-Greene 1979, Gaubert 1992, Baccelli et al. 2002, ...

Tropical line segments

Pick $p, q \in \mathbb{R}^d$. Up to relabeling, assume:



$$q_1 - p_1 \;\geq\; q_2 - p_2 \;\geq\; \dots \;\geq q_d - p_d \;.$$

With $r_i := q_i - p_i$ we have

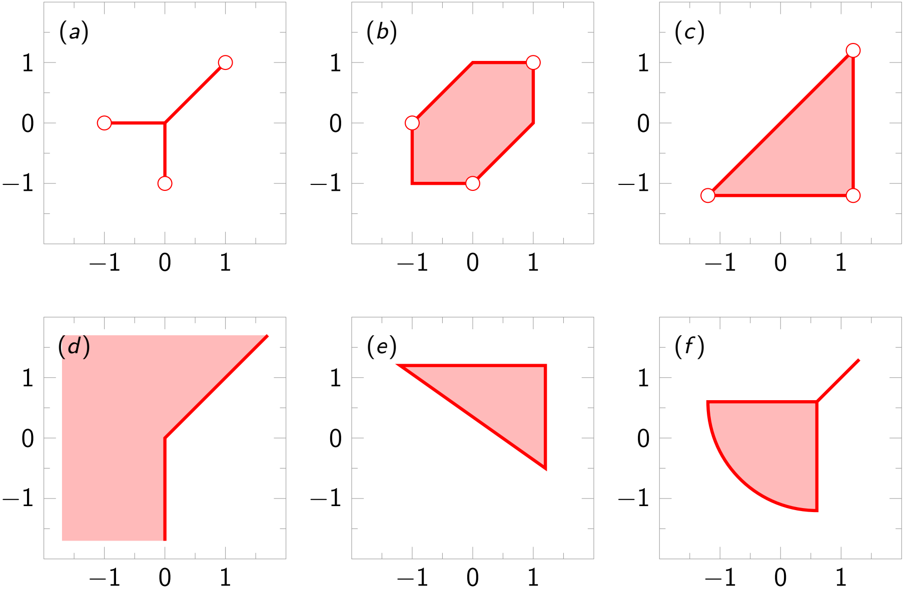
$$\begin{aligned}
(r_1 \odot p) \oplus (0 \odot q) &= (r_1 + p_1, r_1 + p_2, \dots, r_1 + p_d) \;, \\
(r_2 \odot p) \oplus (0 \odot q) &= (q_1, r_2 + p_2, r_2 + p_3, \dots, r_2 + p_d) \;, \\
\vdots \qquad & \quad \vdots \qquad \vdots \\
(r_{d-1} \odot p) \oplus (0 \odot q) &= (q_1, q_2, \dots, q_{d-1}, r_{d-1} + p_d) \;, \\
(r_d \odot p) \oplus (0 \odot q) &= (q_1, q_2, q_3, \dots, q_d) \;=\; q \;.
\end{aligned}$$

Note that $(r_1 + p_1, r_1 + p_2, \dots, r_1 + p_d) = r_1 \odot p$ equals p in $\mathbb{R}^d/\mathbb{R}\mathbf{1}$.

Proposition

The tropical line segment $\mathrm{tconv}(p, q) \subset \mathbb{R}^d/\mathbb{R}\mathbf{1}$ *is the union of at most* $d-1$ *ordinary line segments.*

Max-tropically convex sets in the plane $\mathbb{R}^3/\mathbb{R}\mathbf{1}$



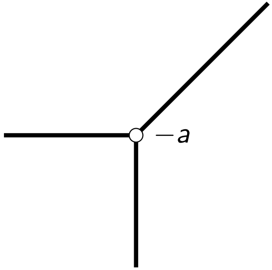
Tropical hyperplanes with respect to max and min

tropical linear form a on $\mathbb{R}^d$ with

$$a(x) \;=\; a_1 \odot x_1 \oplus \cdots \oplus a_d \odot x_d$$

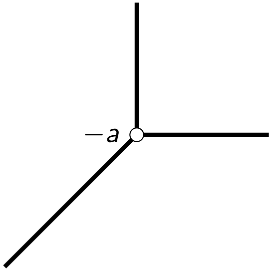
vanishes where the maximum/minimum is attained at least twice

$\mathbb{T}_{\max} = (\mathbb{R} \cup \{-\infty\}, \max, +)$



$\mathcal{T}^{\max}(a)$

$\mathbb{T}_{\min} = (\mathbb{R} \cup \{\infty\}, \min, +)$



$\mathcal{T}^{\min}(a)$

- $-\min(x, y) = \max(-x, -y) \implies \mathcal{T}^{\min}(a) = -\mathcal{T}^{\max}(-a)$

The structure theorem of tropical convexity

Let $V \in \mathbb{R}^{d \times n}$.

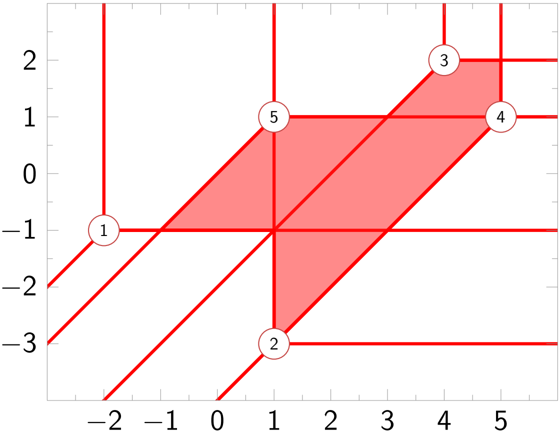
Theorem (Develin & Sturmfels 2004)

1. *The polyhedral decomposition $\mathcal{S}_V$ of $\mathbb{R}^d/\mathbb{R}\mathbf{1}$, which is formed by the regions of the* min*-tropical hyperplane arrangement A_V, is dual to the (lower) regular subdivision $\Sigma(V)$, where V is considered as a height function on the vertices of the ordinary polytope $\Delta_{d-1} \times \Delta_{n-1}$.*
2. *The* max*-tropical polytope* $\operatorname{tconv}(V)$ *agrees with the union of the bounded cells of the polyhedral complex $\mathcal{S}_V$.*

- Ardila & Develin 2007, Horn 2012: nonregular subdivisions
- Fink & Rincón 2015, J. & Loho 2016: $V \in \mathbb{T}^{d \times n}$

Example: a max-tropical pentagon ($d = 3$, $n = 5$)

$$V = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -2 & 1 & 4 & 5 & 1 \\ -1 & -3 & 2 & 1 & 1 \end{pmatrix}$$

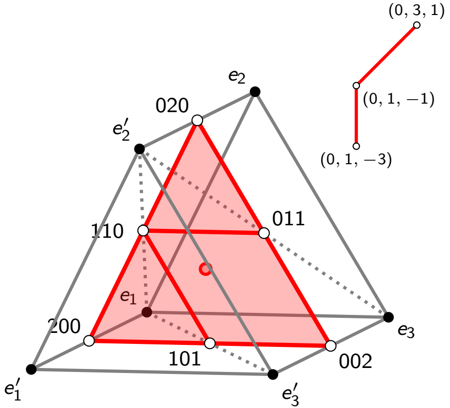


Regular subdivision of $\Delta_{d-1} \times \Delta_{n-1}$

dual to $\mathcal{S}_V$

Consider $V \in \mathbb{R}^{d \times n}$.

- append v_{ij} as additional coordinate to $e_i \times e'_j$
- take ordinary convex hull in $(d-1)+(n-1)+1$-space
- project down lower faces
  - polyhedral complex which subdivides $\Delta_{d-1} \times \Delta_{n-1}$



$(d = 3, n = 2)$

barycenter marks central cell

Computational classification

The known numbers of combinatorial types of regular triangulations of $\Delta_{d-1} \times \Delta_{n-1}$, up to Sym($d$)× Sym($n$)-symmetry:

$d \setminus n$	2	3	4	5	6	7
2	1	1	1	1	1	1
3		5	35	530	13 621	531 862
4			7 869	7 051 957		

- De Loera 1995: `PUNTOS`
- Rambau 2002–2020: `TOPCOM`
- Jordan, J. & Kastner 2018–2024: `mptopcom`

The fundamental theorem of tropical geometry

special case: hyperplane arrangement

The field of complex Puiseux series

$$\mathbb{K} = \mathbb{C}\{\{t\}\} = \left\{ \sum_{k=m}^{\infty} a_k \cdot t^{k/N} \mid m \in \mathbb{Z},\, N \in \mathbb{N}^{\times},\, a_k \in \mathbb{C} \right\}$$

is equipped with a valuation val, mapping to the lowest degree.

Theorem (Kapranov, Einsiedler & Lind 2005)

For a Laurent polynomial $f \in \mathbb{K}[x_1^{\pm}, \ldots, x_d^{\pm}]$ the tropical hypersurface $\mathcal{T}(\mathrm{trop}(f))$ equals the topological closure of the set $\mathrm{val}(V(f))$ in $\mathbb{R}^d$.

$$\mathrm{trop}(f)(X_1, \ldots, X_d) = \bigoplus_{u \in \mathrm{supp}(f)} \mathrm{val}(\gamma_u(t)) \odot X_1^{\odot u_1} X_2^{\odot u_2} \ldots X_d^{\odot u_d}$$

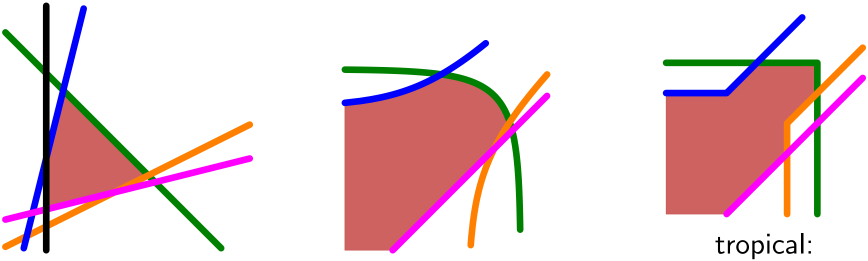
- for f a product of linear forms, the ordinary hypersurface $V(f)$ in $(\mathbb{C}^{\times})^d$ is a hyperplane arrangement

Maslov dequantization (applied to polyhedra)

Litvinov & Maslov 1996, Develin & Yu 2007

system of linear inequalities over $\mathbb{R}$ or $\mathbb{R}\{\{t\}\}$:

$x + y \leq 2$	$\max(X, Y) \leq 0$
$tx \leq 1 + t^2 y$	$1 + X \leq \max(0, 2 + Y)$
$2ty \leq 3 + 2t^3 x$	$1 + Y \leq \max(0, 3 + X)$
$x \leq t^2 y$	$X \leq 2 + Y$
$x, y \geq 0$	



ordinary: $t = 2$ — logarithmic: $\log_2(\cdot)$ — tropical: $\lim_{t\to\infty} \log_t(\cdot)$

Tropical convexity applications I

complexity theory

- Akian, Gaubert, Guterman 2012: feasibility of a tropical linear program equivalent to MEAN-PAYOFF
- Allamigeon, Benchimol, Gaubert & J. 2018: log-barrier interior point methods are not strongly polynomial
- Allamigeon, Gaubert & Vandame 2022: arbitrary self-concordant barrier functions

optimization

- J. & Schröter 2022: parametric shortest path algorithms
  - Cleveland et al. 2022: delay tolerant networks (NASA project)
- Gaubert & Vlassopoulos 2024: large language models

economics

- Shiozawa 2015: Ricardian theory of trade

Tropical convexity applications II

Theorem (Yuster & Yu 2007)

Tropical linear spaces are tropical polytopes in the tropical projective space $\mathbb{TP}^d \supseteq \mathbb{R}^d/\mathbb{R}\mathbf{1}$.

Theorem (Speyer 2008)

Uniform tropical linear spaces are equivalent to matroid decompositions of hypersimplices.

- Feichter & Sturmfels 2005; Ardila & Klivans 2006: Bergman fans
- Kapranov 1992; Keel & Tevelev 2006: Chow quotients of Grassmannians
- J., Sturmfels & Yu 2007: Bruhat–Tits buildings of type $\widetilde{A}$
- Adiprasito, Huh & Katz 2018: Hodge theory for combinatorial geometries

Fermat–Weber sets

The asymmetric tropical distance in $\mathbb{R}^n\mathcal{H}$ is given by

$$\mathrm{dist}_\triangle(x,y) \;=\; \sum_{i\in[n]}(y_i - x_i) - n\min_{i\in[n]}(y_i - x_i) \;=\; \sum_{i\in[n]}(y_i - x_i) + n\max_{i\in[n]}(x_i - y_i)\;,$$

where $x, y \in \mathbb{R}^n\mathcal{H}$.

- restrict to $\mathcal{H} = \{x \in \mathbb{R}^n \mid \sum x_i = 0\} \cong \mathbb{R}^n/\mathbb{R}\mathbf{1}$
- Amini & Manjunath 2010: Riemann–Roch for lattices

Now pick finite subset $V \subset \mathcal{H}$.

Definition (asymmetric tropical Fermat–Weber set)

$$\mathrm{FW}(V) \;=\; \arg\min_{x\in\mathbb{R}^n} \sum_{v\in V} \mathrm{dist}_\triangle(v,x)$$

- Lin & Yoshida 2018: symmetric tropical distance
- Sabol, Barnhill, Yoshida & Miura 2024

Asymmetric tropical Fermat–Weber sets are tropical polytopes

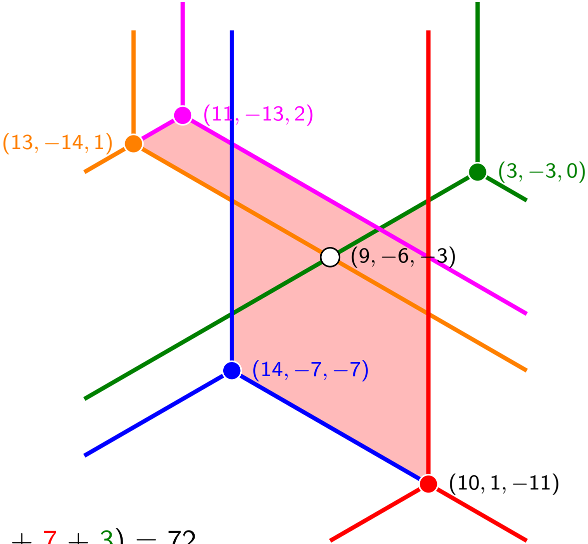
Theorem (Comăneci & J. 2024)

The Fermat–Weber set $\mathrm{FW}(V)$ *is a max-tropical polytope in* $\mathcal{H}$*, and it is contained in the max-tropical polytope* $\mathrm{tconv}(V)$.

- in fact, $\mathrm{FW}(V)$ dual to central cell in $\mathcal{S}(V)$
  - in particular, also a convex polytope in the ordinary sense!
- more information: e.g., sharp upper bound for $\dim \mathrm{FW}(V)$
- Cox & Curiel 2023: weighted Fermat–Weber points

Example

five points in the plane $\mathcal{H} \cong \mathbb{R}^3/\mathbb{R}\mathbf{1}$ with a unique Fermat–Webert point



$3 \cdot (4 + 5 + 5 + 7 + 3) = 72$

A linear program . . .

Consider $V = \{v_1, v_2, \ldots, v_m\} \subset \mathcal{H} = \mathbb{R}^n/\mathbb{R}\mathbf{1}$ finite. Then $x^* \in \mathcal{H}$ lies in FW(V) if and only if x^* minimizes

$$\sum_{i\in[m]} \mathrm{dist}_\triangle(v_i, x^*) \;=\; n \cdot \sum_{i\in[m]} \max_{j\in[n]}(v_{ij} - x_j^*)$$

Equivalently, (t^*, x^*) is an optimal solution of the LP

$$\begin{array}{lll} \text{minimize} & n\cdot(t_1+\cdots+t_m) & \\ \text{subject to} & v_{ij} - x_j \;\leq\; t_i\,, & \text{for } i\in[m] \text{ and } j\in[n] \\ & x_1+\cdots+x_n \;=\; 0 & \end{array} \tag{1}$$

. . . and its dual

Again we fix $V = (v_{ij}) \in \mathbb{R}^{m \times n}$. Then the following LP is dual to (1), with dual variables λ and y_{ij} for $i \in [m]$ and $j \in [n]$:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i\in[m]} \sum_{j\in[n]} v_{ij} \cdot y_{ij} \\
\text{subject to} \quad & \sum_{j\in[n]} y_{ij} \ = \ n \ , && \text{for } i \in [m] \\
& \lambda + \sum_{i\in[m]} y_{ij} \ = \ 0 \ , && \text{for } j \in [n] \\
& y_{ij} \ \geq \ 0 \ , && \text{for } i \in [m] \text{ and } j \in [n] \ .
\end{aligned}
\tag{2}
$$

- transportation problem
  - e.g., Tokuyama & Nakano (1995): $O(n^2 m \log^2 m)$ algorithm, for $m \geq n$

Dissimilarity maps and tree-like metrics

Definition

A symmetric $t \times t$-matrix $D = (\delta_{ij})$ is a dissimilarity map if $\delta_{ij} \geq 0$ and $\delta_{ii} = 0$ for all $i, j \in [t]$.

- D pseudometric $\iff \delta_{ik} \leq \delta_{ij} + \delta_{jk}$ for all $i, j, k \in [t]$
- D ultrametric $\iff \delta_{ik} \leq \max(\delta_{ij}, \delta_{jk})$ for all $i, j, k \in [t]$

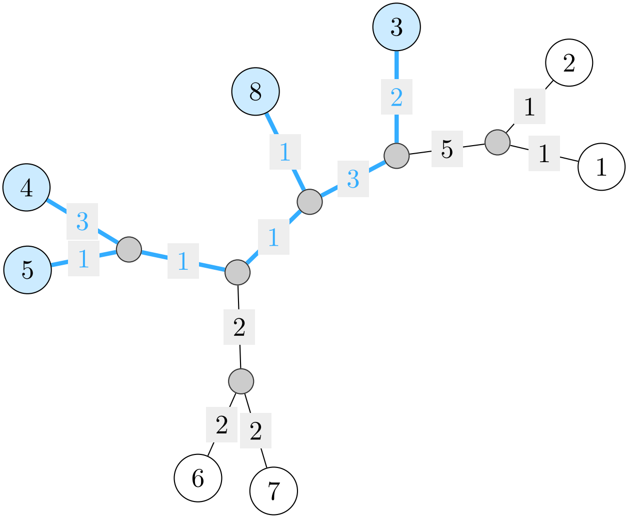
Theorem (Four-Point-Condition; see, e.g., Dress 1984)

A pseudometric D on the set $[t]$ is tree-like if and only if the maximum of the three numbers

$$\delta_{ij} + \delta_{k\ell}\,, \quad \delta_{ik} + \delta_{j\ell}\,, \quad \delta_{i\ell} + \delta_{jk}$$

is attained at least twice for all $i, j, k, \ell \in [t]$.

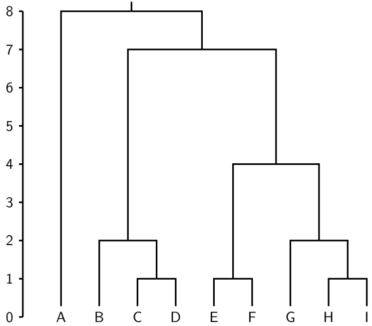
Example: Tree-like metric for $t = 8$



Ultrametric trees in tropical geometry

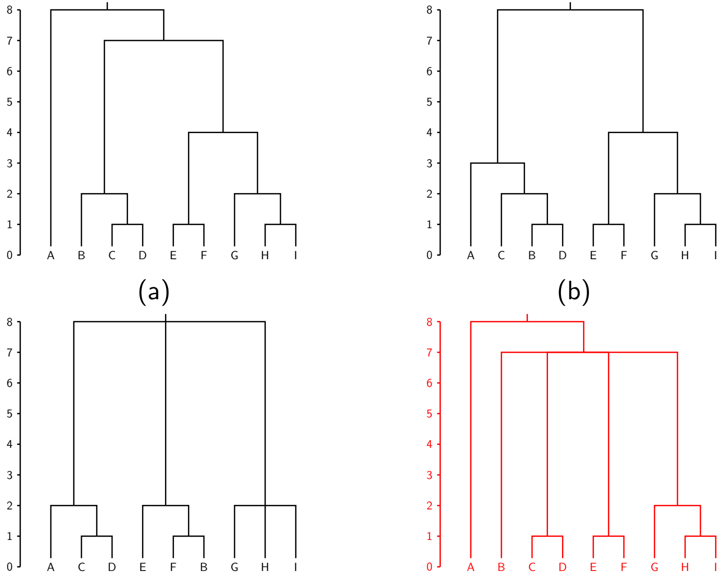
Proposition

D ultrametric $\iff$ corresponding tree is equidistant



- Billera, Holmes & Vogtmann 2001: space of equidistant trees $\mathcal{T}_t$
  - Lin, Sturmfels, Tang & Yoshida 2017: employ tropical convexity
- Ardila & Klivans 2006: D ultrametric $\iff$ D corresponds to a point in the Bergman fan of the complete graph K_t
- Speyer 2008: tropical linear spaces
  - Bergman fans arise as special cases

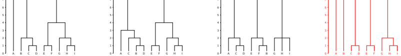
An (equidistant) consensus tree problem on $t = 9$ taxa



(a) (b)

(c) (d)

Tropical median consensus trees



Corollary (Comăneci & J. 2024)

Let $V \subset \mathcal{T}_t$ *be finite.*
Then the max-tropical polytope $\mathrm{FW}(V)$ *is contained in* $\mathcal{T}_t$*.*
Moreover, any two trees in $\mathrm{FW}(V)$ *share the same tree topology.*

Proof.

- Ardila & Klivans 2006: $\mathcal{T}_t$ tropically convex
- analyze covector decomposition $\mathcal{S}_V$ [Develin & Sturmfels 2004]

□

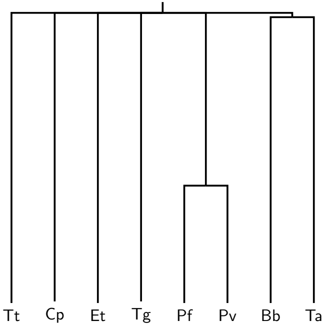
Idea: For a finite set of ultrametrics $V = \{D_1, D_2, \ldots, D_m\} \subset \mathbb{R}^{t \times t}$ pick a suitable point in $\mathrm{FW}(V)$ as a consensus tree; e.g., the ordinary average of the tropical vertices.

Example: Apicomplexa gene trees

$m = 268$ trees with $n = 8$ taxa

Kuo, Wares & Kissinger 2008: trees from 268 orthologous sequences with 8 species of protozoa:

- *Babesia bovis* (Bb), *Cryptosporidium parvum* (Cp), *Eimeria tenella* (Et), *Plasmodium falciparum* (Pf), *Plasmodium vivax* (Pv), *Theileria annulata* (Ta) and *Toxoplasma gondii* (Tg)
- outgroup: *Tetrahymena thermophila* (Tt)



Page, Yoshida & Zhang 2020:
tropical principal component analysis

- based on symmetric tropical distance

Computing tropical median consensus trees (from random)

Löbel, 2004: `https://www.zib.de/opt-long_projects/Software/Mcf/`

Table: Timings (in seconds @ quad core Intel Core i5-4590)

Leaves\Trees	50	100	150	200	250	300
5	0.04	0.06	0.07	0.08	0.09	0.11
10	0.11	0.16	0.23	0.26	0.31	0.36
15	0.33	0.45	0.57	0.69	0.79	0.91
20	0.87	1.08	1.29	1.50	1.70	1.92
25	4.13	16.55	50.81	11.15	3.89	382.89

Corollary (Comăneci & J. 2024)

Let $V \subset \mathcal{T}_n$ be a set of m equidistant trees on n leaves. Then

$$\dim \mathrm{FW}(V) \;\leq\; \min\left(\, n-1\,,\ \gcd\left(m, \tbinom{n}{2}\right) \right) - 1 \;.$$

Conclusion

- tropical median consensus trees are nice:
  - fast algorithm via transportation
  - regular (in the sense of Bryant, Francis & Steel 2017)
  - robust, Pareto and co-Pareto on triplets

Andrei Comăneci and Michael Joswig, Asymmetric tropical distances and power diagrams, Algebr. Comb. **6** (2023).

———, Tropical medians by transportation, Math. Program. **205** (2024).

Michael Joswig, Essentials of tropical combinatorics, Graduate Studies in Mathematics, vol. 219, American Mathematical Society, Providence, RI, 2021.

Regular consensus methods

Definition (Bryant, Francis & Steel 2017)

A consensus method $c : (T_1, \ldots, T_m) \mapsto T$ is called regular if the following conditions hold:

(U) $c(T, T, \ldots, T) = T$;

(A) $c(\ldots, T_i, \ldots, T_j, \ldots) = c(\ldots, T_j, \ldots, T_i, \ldots)$;

(N) permuting the taxa in the input trees results in the same permutation of the taxa in the consensus.

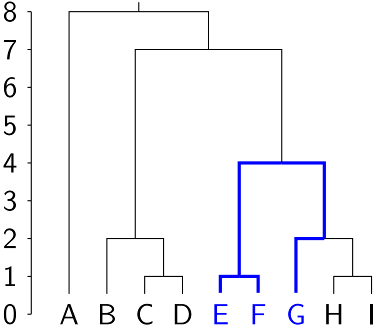
U = unanimity, A = anonymity, N = neutrality

Proposition

The tropical median consensus method is regular.

Rooted triplets

Let $i, j, k \in [n]$ be pairwise distinct taxa in some equidistant tree such that the lowest common ancestor of i and j is a proper descendant of the lowest common ancestor of i, j, and k. Then $ij|k$ form a rooted triplet.



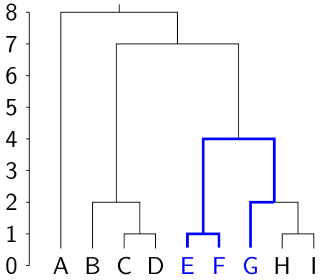
EF|G

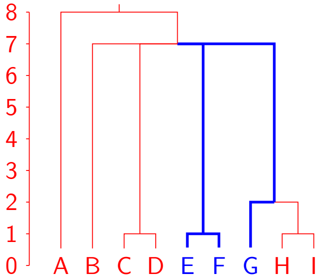
Pareto properties

- $r(D)$ = set of rooted triplets of tree associated with ultrametric D

A consensus method $(D_1, \ldots, D_m) \mapsto D$ is

- Pareto on rooted triplets if $\bigcap_{i\in[m]} r(D_i) \subseteq r(D)$;
- co-Pareto on rooted triplets if $r(D) \subseteq \bigcup_{i\in[m]} r(D_i)$.





Proposition

Any tropically convex consensus method is Pareto and co-Pareto on rooted triplets.