Total cophenetic index and p-adic valuations

Harry Richman

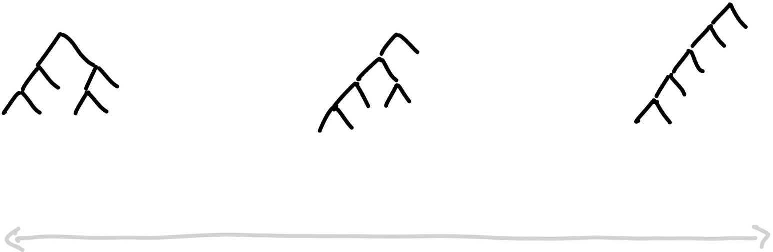
National Center for Theoretical Sciences, Taipei

Fred Hutch Cancer Center, Seattle

International Workshop on Prob., Comb., and Algo. Phylogenetics
NCCU, Taipei
2 April 2026

Tree balance

Question: How “balanced” is a tree?

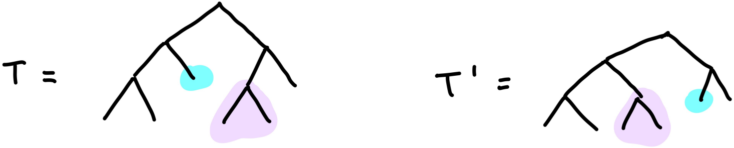


Balance indices:

- Colless
- Sackin
- Total cophenetic (Mir–Rosselló–Rotger)

Tree balance

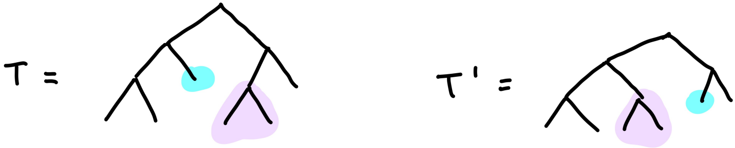
- Note (MRR): total cophenetic index has better resolution



	T	T'
Colless	2	2
Sackin	16	16
Tot. cophenetic	8	9

Tree balance

- Note (MRR): total cophenetic index has better resolution



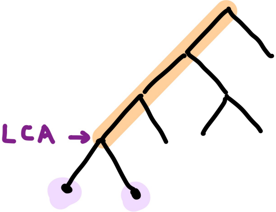
	T	T'	
Colless	2	2	☹
Sackin	16	16	☹
Tot. cophenetic	8	9	😀

Tree balance: total cophenetic index

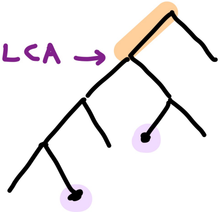
The cophenetic value of leaves x, y is

$$\phi(x,y) = \mathrm{dist}_T(\mathrm{root}, \mathrm{LCA}(x,y))$$

Ex. $\phi(x,y) = 3$



Ex. $\phi(x,y) = 1$



The total cophenetic index is the sum of coph. values over leaf pairs

$$\Phi(T) = \sum_{\substack{x,y \in L(T) \\ x \neq y}} \phi(x,y)$$

Tree balance: total cophenetic index

Question: What are min. (and max.) values of $\Phi(T)$?

For binary trees:

n	2	3	4	5	6	7	8	9	10	11	12
$\min \Phi(T)$	0	1	2	5	8	12	16	23	30	38	46

Tree balance: total cophenetic index

Question: What are min. (and max.) values of $\Phi(T)$?

For binary trees:

n	2	3	4	5	6	7	8	9	10	11	12
$\min \Phi(T)$	0	1	2	5	8	12	16	23	30	38	46

Theorem (Mir–Rosselló–Rotger, 2013)

For binary phylogenetic trees $\mathcal{BT}_n$,

$$\min\{\Phi(T) : T \in \mathcal{BT}_n\} = \sum_{m=0}^{n-1} \sum_{i \geq 1} \left\lfloor \frac{m}{2^i} \right\rfloor$$

Tree balance: total cophenetic index

For binary trees:

n	2	3	4	5	6	7	8	9	10	11	12
$\min \Phi(T)$	0	1	2	5	8	12	16	23	30	38	46

Theorem (Mir–Rosselló–Rotger, 2013)

For binary phylogenetic trees $\mathcal{BT}_n$,

$$\min\{\Phi(T) : T \in \mathcal{BT}_n\} = \sum_{m=0}^{n-1} \sum_{i \geq 1} \left\lfloor \frac{m}{2^i} \right\rfloor$$

Proof sketch

- Show that minimum is achieved on max-balanced tree
- Use recursion $f(n+1) - f(n) = \lfloor n/2 \rfloor + f(\lfloor n/2 \rfloor + 1) - f(\lfloor n/2 \rfloor)$

Tree balance: total cophenetic index

Theorem (Mir–Rosselló–Rotger, 2013)

For binary phylogenetic trees,

$$\min\{\Phi(T) : T \in \mathcal{BT}_n\} = \sum_{m=0}^{n-1} \sum_{i \geq 1} \left\lfloor \frac{m}{2^i} \right\rfloor$$

Question: Can we generalize to (some) multifurcating trees?

Tree balance: total cophenetic index

Theorem (Mir–Rosselló–Rotger, 2013)

For binary phylogenetic trees,

$$\min\{\Phi(T) : T \in \mathcal{BT}_n\} = \sum_{m=0}^{n-1} \sum_{i \geq 1} \left\lfloor \frac{m}{2^i} \right\rfloor \overset{!}{=} \sum_{m=0}^{n-1} \mathrm{val}_2(m!)$$

Here, $\mathrm{val}_2(m!)$ is the largest number of 2's that divide $m!$, a.k.a. the "2-adic valuation"

Ex.

- $\mathrm{val}_2(1024) = \mathrm{val}_2(2^{10}) = 10$
- $\mathrm{val}_2(1000) = \mathrm{val}_2(2^3 \cdot 125) = 3$

Idea: p-adic valuation val_p is a "p-exclusive" logarithm

Tree balance: total cophenetic index

Theorem (Mir–Rosselló–Rotger, 2013)

For binary phylogenetic trees,

$$\min\{\Phi(T) : T \in \mathcal{BT}_n\} = \sum_{m=0}^{n-1} \sum_{i \geq 1} \left\lfloor \frac{m}{2^i} \right\rfloor \overset{!}{=} \sum_{m=0}^{n-1} \mathrm{val}_2(m!)$$

Question: Can we generalize to (some) multifurcating trees?

Let $\mathcal{MT}_{n,b}$ be phylogenetic trees with n leaves and $(\leq b)$-branching

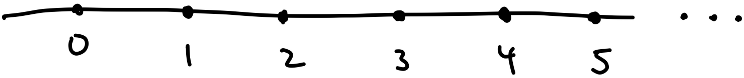
Theorem (R, 2026+)

If b is prime, then

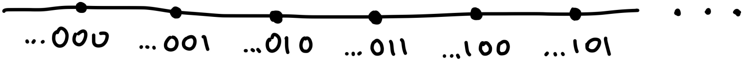
$$\min\{\Phi(T) : T \in \mathcal{MT}_{n,b}\} = \sum_{m=0}^{n-1} \mathrm{val}_b(m!)$$

p-adic valuation

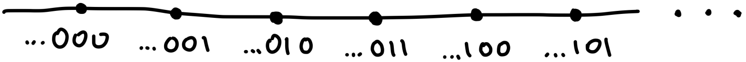
Question: How to order numbers?



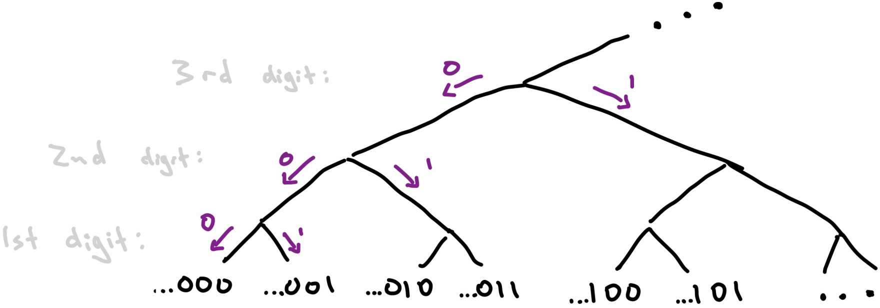
- $\rightsquigarrow$ in binary



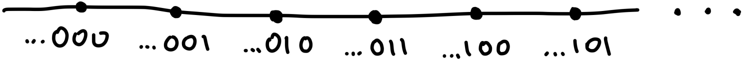
p-adic valuation: $p = 2$



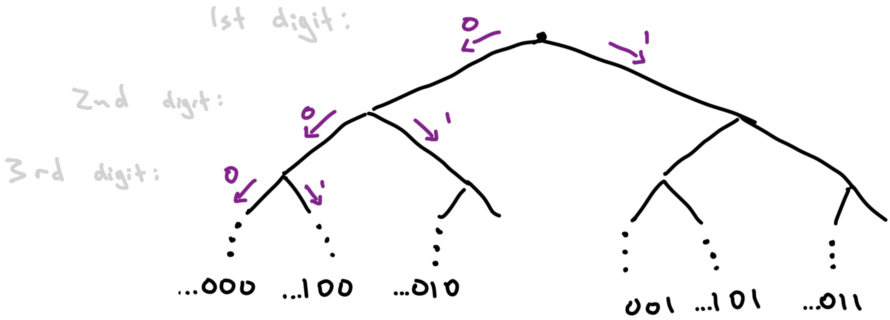
Question: How to order numbers, using binary expansion digits? 😐



p-adic valuation: $p = 2$

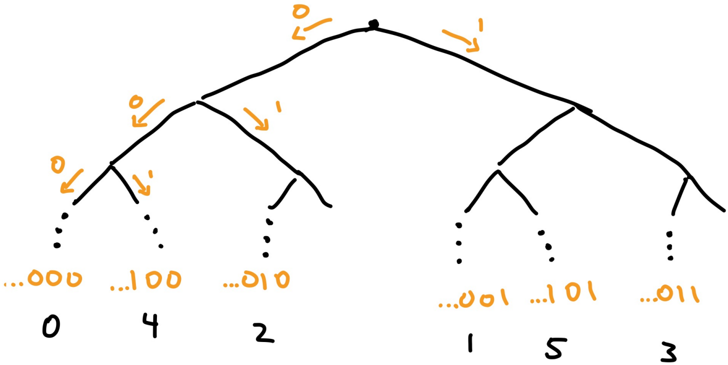


Question: How to order numbers, using binary expansion digits? 😀



p-adic valuation: $p = 2$

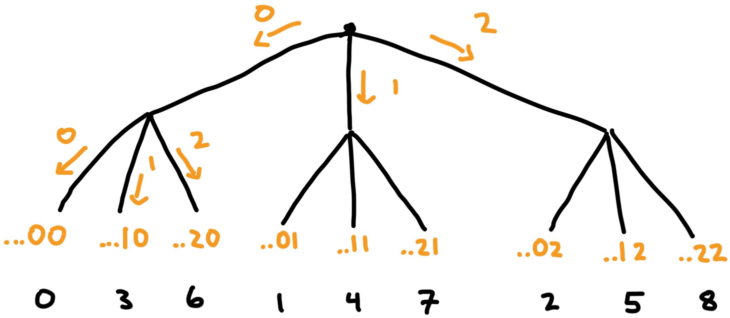
Answer: 2-adic tree on natural numbers $\{0, 1, 2, \ldots\}$



Observation: In 2-adic tree, cophenetic value $\phi(a, b) = \mathrm{val}_2(a - b)$

p-adic valuation: $p = 3$

Answer: 3-adic tree on natural numbers $\{0, 1, 2, \ldots\}$



Observation: In p-adic tree, cophenetic value $\phi(a, b) = \mathrm{val}_p(a - b)$

p-adic valuations: reproving MRR on cophenetic index

Recall: want to show

$$\min\{\Phi(T) : T \in \mathcal{BT}_n\} = \sum_{m=0}^{n-1} \mathrm{val}_2(m!) = \mathrm{val}_2\left(\prod_{m=0}^{n-1} m!\right)$$

Proof sketch.

- Find $a_1, \dots a_n \in \mathbb{Z}$ such that $T = Tree_{(\text{2-adic})}(a_1, \dots, a_n)$
- $\Phi(T) = \sum_{i<j} \mathrm{val}_2(a_j - a_i) = \mathrm{val}_2\left(\prod_{i<j}(a_j - a_i)\right)$
- Suffices to show

$$\prod_{m=0}^{n-1} m! \quad \text{always divides} \quad \prod_{i<j}(a_j - a_i)$$

p-adic valuation: reproving MRR on cophenetic index

Claim

$$\prod_{m=0}^{n-1} m! \quad \text{always divides} \quad \prod_{i<j}(a_j - a_i)$$

- Vandermonde determinant

$$\prod_{i<j}(a_j - a_i) = \det \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ 1 & a_3 & a_3^3 & & a_3^{n-1} \\ \vdots & & & \ddots & \vdots \end{pmatrix}$$

- Row-operations to binomial-coefficient matrix

$$\prod_{i<j}(a_j - a_i) = \left(\prod_{m=0}^{n-1} m!\right) \cdot \det \begin{pmatrix} 1 & \binom{a_1}{1} & \binom{a_1}{2} & \cdots & \binom{a_1}{n-1} \\ 1 & \binom{a_2}{1} & \binom{a_2}{2} & & \binom{a_2}{n-1} \\ 1 & \binom{a_3}{1} & \binom{a_3}{2} & & \binom{a_3}{n-1} \\ \vdots & & & \ddots & \vdots \end{pmatrix}$$

p-adic valuation: multifurcating case of cophenetic index

Let $\mathcal{MT}_{n,b}$ be phylogenetic trees with n leaves and $(\leq b)$-branching

Theorem (R, 2026+)

If b is prime, then

$$\min\{\Phi(T) : T \in \mathcal{MT}_{n,b}\} = \sum_{m=0}^{n-1} \mathrm{val}_b(m!)$$

Proof sketch. Replace val_2 with val_b in previous proof

p-adic valuation: multifurcating case of cophenetic index

Let $\mathcal{MT}_{n,b}$ be phylogenetic trees with n leaves and $(\leq b)$-branching

Theorem (R, 2026+)

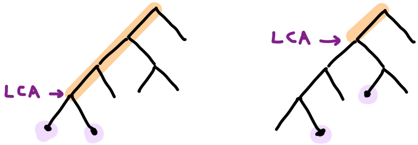
If b is prime, then

$$\min\{\Phi(T) : T \in \mathcal{MT}_{n,b}\} = \sum_{m=0}^{n-1} \mathrm{val}_b(m!) \quad = \sum_{m=0}^{n-1} \sum_{i \geq 1} \left\lfloor \frac{m}{b^i} \right\rfloor$$

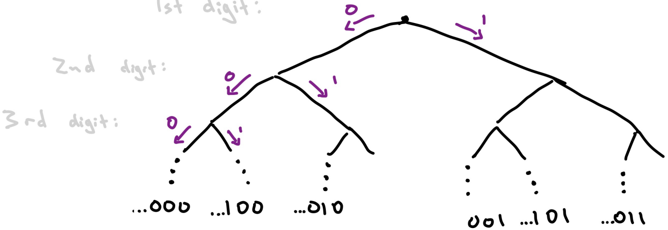
Proof sketch. Replace val_2 with val_b in previous proof

Problem: Can we drop the assumption that b is prime?

- ⚠ For non-prime b, generally $\sum_{m=0}^{n-1} \mathrm{val}_b(m!) \neq \mathrm{val}_b\left(\prod_{m=0}^{n-1} m!\right)$

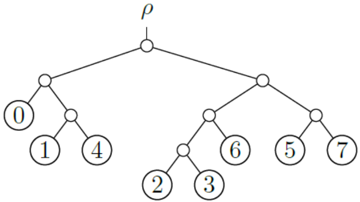

LCA
LCA

Thank you!


1st digit:
2nd digit:
3rd digit:
0
1
0
1
0
1
...000
...100
...010
001
...101
...011

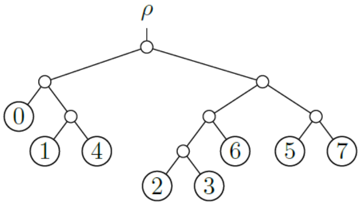
How to tell your computer about phylogenetic trees?

- Newick format: ID's and parentheses



$\leftrightarrow$ `((0,(1,4)),(((2,3),6),(5,7)));`

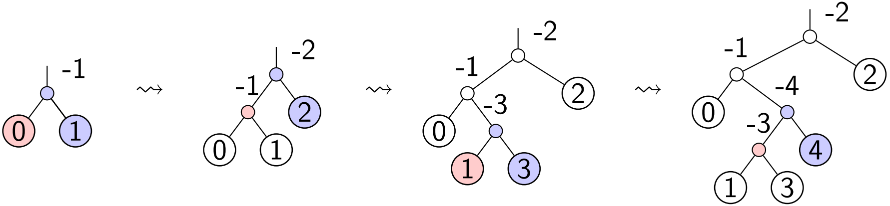
- 💡 OLA encoding: list of integers



$\leftrightarrow$ `(0,-1,2,1,-3,-3,5)`

OLA = "Ordered Leaf Attachment"

Example: tree with encoding (0,-1,1,-3)



Pros:

- linear-time encoding and decoding
- convex region in encoding space
- somewhat-low distortion vs. SPR distance
- stable under adding leaves