10th International Congress on Industrial and Applied Mathematics, Poster Presentation at Waseda University, Tokyo, Japan. August 24, 2023

Constructing a Phylogenetic X-cactus from a Distance Matrix

Keita Watanabe[1], Momoko Hayamizu[2]

1 Department of Pure and Applied Mathematics, Graduate School of Fundamental Science and Engineering, Waseda University

2 Department of Pure and Applied Mathematics, Faculty of Science and Engineering, Waseda University

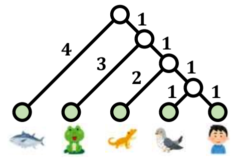
Background

Trees have been studied due to their usefulness in describing the evolutionary processes of organisms. However, when considering factors such as hybridization between different species, phylogenetic networks become necessary.

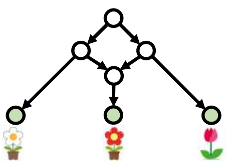
A distance matrix D on X for representing dissimilarities

	0	2	4	6	8
	2	0	4	6	8
	4	4	0	6	8
	6	6	6	0	8
	8	8	8	8	0

A realization of (X, D) $(d_G(x,y) = d(x,y))$



A phylogenetic network



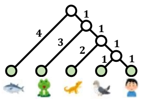
1965	1974	1987	2004
Uniqueness of phylogenetic tree [Hakimi & Yau]	Characterization of tree metric [Buneman]	Algorithm of constructing tree [Saitou & Nei]	Algorithm of constructing Network [Bryant & Moulton]

Tree metric

	0	2	4	6	8
	2	0	4	6	8
	4	4	0	6	8
	6	6	6	0	8
	8	8	8	8	0

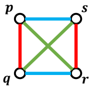
1 to 1

Tree

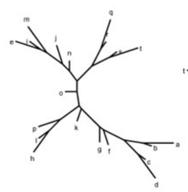


A distance d on X is tree metric $\Leftrightarrow$ d satisfies **"four-point condition"** such that $\forall p, q, r, s \in X$,

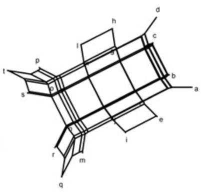
$$d(p,q) + d(r,s) \le \max\begin{cases} d(p,r) + d(q,s) \\ d(p,s) + d(q,r) \end{cases}$$



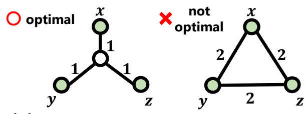
Neighbor-Joining algorithm



Neighbor-Net algorithm



We consider "optimal" realizations which are realizations of least total weight.



Saitou & Nei solved the following problem.

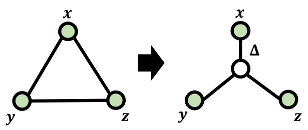
Problem 1

Propose an algorithm that satisfies the following conditions.

Input: A tree metric d on X

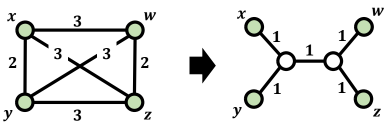
Output: The optimal realization of (X, d)

Define the operation "tightening".

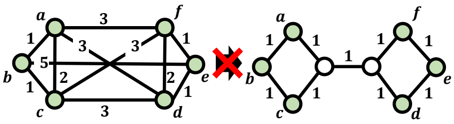


Determine how much each vertex "extends" from the set of other vertices using tightening.

$$\Delta := \min_{y,z \in X \setminus \{x\}} \left\{ \frac{1}{2}\left(d(x,y) + d(x,z) - d(y,z)\right) \right\}$$



If an input is a tree metric, the optimal realization can be obtained through tightening.



However, in cases where it is not a tree metric, the optimal realization might not be obtained.

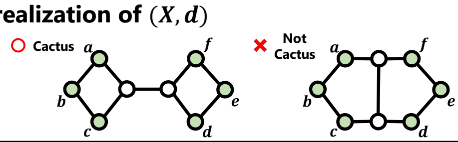
Problem 2

Propose an algorithm that satisfies the following conditions.

Input: A cactus metric d on X

Output: The optimal realization of (X, d)

Cactus:
Each edge belongs to at most one cycle

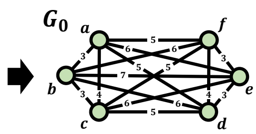


The Idea of proposed method

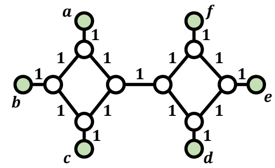
We demonstrate the proposal method with a small example.

Preliminary Step: Input distance matrix

	a	b	c	d	e	f
a	0	3	4	5	6	5
b	3	0	3	6	7	6
c	4	3	0	5	6	5
d	5	6	5	0	3	4
e	6	7	6	3	0	3
f	5	6	5	4	3	0



Optimal realization



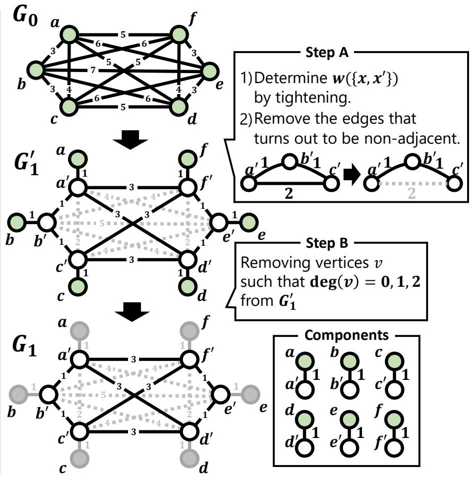
Main Step: Construct G^*

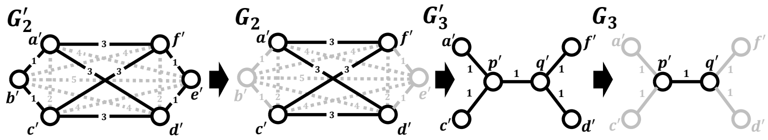
Repeat this "Main Step" until the number of vertices is 2 or fewer.

Step A: Construct G_1' from G_0
Determine how much each vertex "extends" from the set of other vertices using tightening.
Investigating adjacency relationships, including newly added vertices, for each vertex where $d(x,y) = d(x,t) + d(t,y)$ and x and y are not adjacent.

Step B: Remove vertices with $\deg(v) = 0, 1, 2$ and hold the removed elements as determined components of G^*.

Repeat this operation on G_1, and create G_2', G_2, G_3', G_3.
G_3 has 2 vertices, so the algorithm ends.

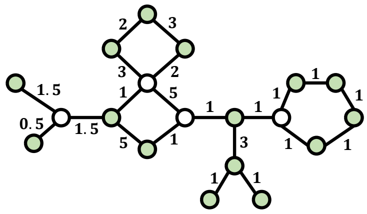




Examples

Illustrating successful examples using a Python program. This program runs on a laptop.

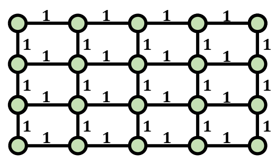
Example 1: A Cactus Metric



We were able to obtain the optimal realization fast using a cactus metric as the input.

Time: 0.2776 s

Example 2: A Grid (not cactus)



We were able to obtain the optimal realization fast not only a cactus metric.
The Neighbor-Net algorithm failed this example.

Time: 0.3961 s

Execution Environment

CPU: 11th Gen Intel(R) Core(TM) i7-1195G7 @ 2.90GHz 2.92 GHz
Python version: Python 3.11.4, Memory: 16.0GB, Operating System: Windows 11 22H2

Future works

Determining the metric spaces in which this method operates correctly. Furthermore, expanding to other metric spaces.

keita1126@akane.waseda.jp

GitHub (Programs)

References

[1] M. Hayamizu, K. T. Huber, V. Moulton, and Y. Murakami, "Recognizing and realizing cactus metrics," Information Processing Letters, vol. 157, p. 105916, 2020.
[2] S. L. Hakimi and S. S. Yau, "DISTANCE MATRIX OF A GRAPH AND ITS REALIZABILITY," Quarterly of applied mathematics, vol. 22, no. 4, pp. 305– 317, 1965.
[3] P. Buneman, "A Note on the Metric Properties of Trees," Journal of Combinatorial Theory, Series B, vol. 17, no. 1, pp. 48–50, 1974.
[4] N. Saitou and M. Nei, "The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees," Molecular biology and evolution, vol. 4, no. 4, pp. 406–425, 1987.
[5] D. Bryant and V. Moulton, "Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks," Molecular biology and evolution, vol. 21, no. 2, pp. 255–265, 2004.