ISSN : 0950-0707

Construction of a tree with n taxa on an n−dimensional space where n = 2,3, that preserves distance as in a given distance matrix.

Renjini Raveendran P

ASSISTANT PROFESSOR, DEPARTMENT OF MATHEMATICS ALL SAINTS' COLLEGE, UNIVERSITY OF KERALA, THIRUVANANTHAPURAM

Beena S

PRINCIPAL, NSS COLLEGE NILAMEL

Abstract - Graph theory can be applied to Biology for better results. Tree diagram plays an important role in Phylogenetic analysis. The interconnections and inter distance of various taxa can be visualize easily with the help of these diagrams. Trees are connected acyclic graphs. Suppose we are given distance matrix D of order 2 × 2 (in the case of 2 taxa) or 3 × 3 (in the case of 3 taxa). This work explains the construction of a tree in R^2 or R^3 space from the given distance matrix D such that the taxa preserves the distance same as that of the given matrix D. The paper include theorems to find out the number of tree structures and isomorphic tree structures possible in each case and also the transformation from one tree structure to another on the basis of operations on graph.

Keywords – Phylogenetic analysis, Distance matrix, graph operations.

I. Introduction

Mathematics has its application in almost all subjects including Biology. In Phylogenetic Analysis, the role of Mathematics is inevitable. Phylogeny deals with the evolutionary history of a set of taxa X. $X = \{x_1, \cdots, x_n\}$ denote a set of taxa, in which each taxon xi represents some species, group or individual organism whose evolutionary history is of interest to us. Tree plays an important role in phylogenetic analysis. The Distance matrix gives pairwise distance between the set of taxa X. From this, we can form equations whose solution can be represented in 2-space or 3-space depending on the number of taxa. Then we construct the tree in 2-space or 3-space.The tree preserve the distance same as that of the Distance matrix. Here in this work, we are converting a biological data to a Tree in Cartesian space where various calculations can be carried out mathematically. The paper include theorems to find out the number of tree structures and isomorphic tree structures possible in each case and also the transformation from one tree structure to another on the basis of operations on graph like fusion of vertices , contraction of edges and so on. The mathematical formula for number of distinct tree structures and the number of distinct tree structures up to isomorphism for n taxa are derived for n = 2,3.

II. **Preliminaries**

Definition 1.1

A *graph* G is a pair G= (V, E) consisting of a finite set V and a set of 2-element subsets of V. The elements of V and E are called Vertices and Edges respectively.

Definition 1.2 X ={ $x_1,...,x_n$}to denote a set of *taxa*, in which each taxon x_i represents some species, group or individual organism whose evolutionary history is of interest to us

Definition 1.3 A *phylogeny* describes the evolutionary history of a set of taxa.

Definition 1.4 A pair of vertices (u,v) in a graph G is said to be fused if the two vertices are replaced by a single vertex denoted by (uv) such that all the edges incident on either u or v or on both are made incident on (uv). If from the new graph, the loop corresponding to the edge e is deleted, then we obtain the contraction of the edge e

A) Method of constructing a tree in R^2 space, from a given matrix D on a set of taxa X = {x_1, x_2}

Suppose we are given a distance matrix D on set of taxa **X** = {x_1, x_2}

D = $\begin{pmatrix} 0 & k \\ k & 0 \end{pmatrix}$, k > 0

From **D,** it is clear that x_1 and x_2 are k distance apart.

Consider the equation $x_1+x_2 =k$

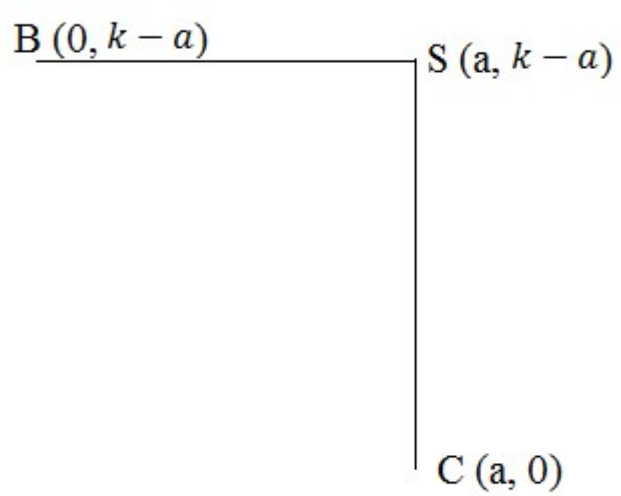
This system of equation has infinitely many solutions.

Case I a):

Choose positive values for x_1 and x_2.

Let the solution be S (a, $k-a$) where a > 0

Fix two points B (0, $k-a$) and C (a, 0).The three points can be represented graphically as:

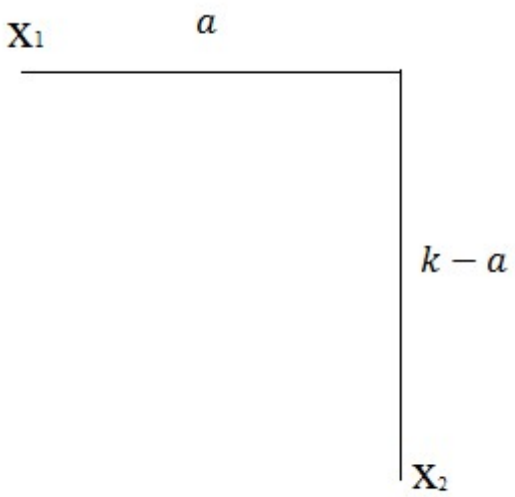


The taxa were placed in such a way that i^{th} taxon is in the coordinate where i^{th} entry is 0.

Solution is (a, $k-a$)

Taxa x_1 is placed in B (0, $k-a$)

Taxa x_2 is placed in C (a,0)

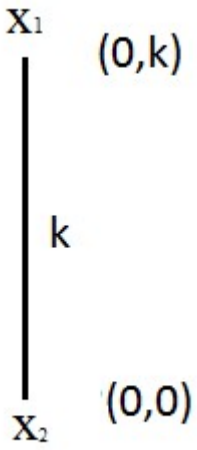


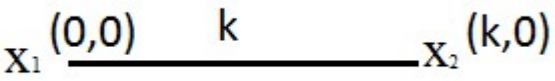
Tree

By distance formula BS = a and SC = k-a. Taxa x_1 and x_2 are k units apart (same as that in D)

Case: IIa)

Suppose we choose solution as S (0,k) or S(k,0). Then the tree has one of the following forms





Note:

If we choose the solution as positive values, then we obtain a tree with 3 nodes and 2 edges

If we choose the solution in such a way that one of the value is 0, the the tree obtained has 2 nodes and 1 edge.

A) Method of constructing a tree in R^3 space, from a given matrix D on a set of taxa X = {x_1, x_2,x_3}

Suppose we are given a distance matrix **D** on X = {x_1, x_2, x_3}

$$\mathbf{D} = \begin{pmatrix} 0 & k_1 & k_2 \\ k_1 & 0 & k_3 \\ k_2 & k_3 & 0 \end{pmatrix}$$, $x_i x_j = k_p \neq 0$ for $i \neq j$ and $k_p > 0$ for p=1, 2, 3

Consider the system of equations:

$x_1+x_2= k_1$

$x_1+x_3 = k_2$

$x_2+x_3=k_3$

The above equations can be solved and obtain the solution as

$x_1 = \frac{k_1+k_2-k_3}{2}$

$x_2 = \frac{k_1+k_3-k_2}{2}$

$x_3 = \frac{k_2+k_3-k_1}{2}$

Case Ia)

Suppose the solution we obtained is

S $\left(\frac{k_1+k_2-k_3}{2}, \frac{k_1+k_2-k_3}{2}, \frac{k_2+k_3-k_1}{2}\right)$

Fix three points A= $\left(0, \frac{k_1+k_2-k_3}{2}, \frac{k_2+k_3-k_1}{2}\right)$

B = $\left(\frac{k_1+k_2-k_3}{2}, 0, \frac{k_2+k_3-k_1}{2}\right)$ and

C = $\left(\frac{k_1+k_2-k_3}{2}, \frac{k_1+k_2-k_3}{2}, 0\right)$

A is obtained by substituting 0 in first coordinate of the solution so A represent x_1

B is obtained by substituting 0 in second coordinate of the solution so B represent x_2

C is obtained by substituting 0 in third coordinate of the solution so C represent x_3

Thus we obtain a tree with 4 nodes and 3 edges, in which the three taxa preserves the distance as in **D**

Case IIB

Consider the case when $k_1+k_2 = k_3$ or $k_1+k_3 = k_2$ or $k_2+k_3 = k_1$;

Resulting in $x_1 = 0$ or $x_2=0$ or $x_3=0$

Suppose the solution obtained has any of the following forms

I
i) (0,b,c) or
ii)(a,0,c) or
iii)(a, b,0).
Then we can construct tree as follows:

i)Check the position where 0 occurs. In this case 1st coordinate is 0.
So x_1 represents (0, b, c)
The point corresponds to x_2 is obtained by substituting second coordinate the obtained solution to 0
x_2 takes the position of (0, 0, c)
The point corresponds to x_3 is obtained by substituting third coordinate of obtained solution to 0
x_3 takes the position of (0, b, 0)
Similar case follows for ii) and iii)
The tree thus obtained has 3 nodes and 2 edges. The tree thus obtained has taxa at the nodes that preserve the distance as in **D**

Note:
If the solution consists of three positive values, then the tree representation has 4 nodes and 3 edges.
If the solution has one 0 value, then the tree representation has 3 nodes and 2 edges

Illustration: 1

$$\mathbf{D} = \begin{bmatrix} 0 & 4 & 3 \\ 4 & 0 & 2 \\ 3 & 2 & 0 \end{bmatrix}$$

$x_1+x_2= 4$
$x_1+x_3 = 3$
$x_2+x_3=2$
The solution is I (2.5,1.5, 0.5)
Fix points in such a way that
x_1(0,1.5,0.5)
x_2(2.5,0,0.5)
x_3(2.5,1.5,0)

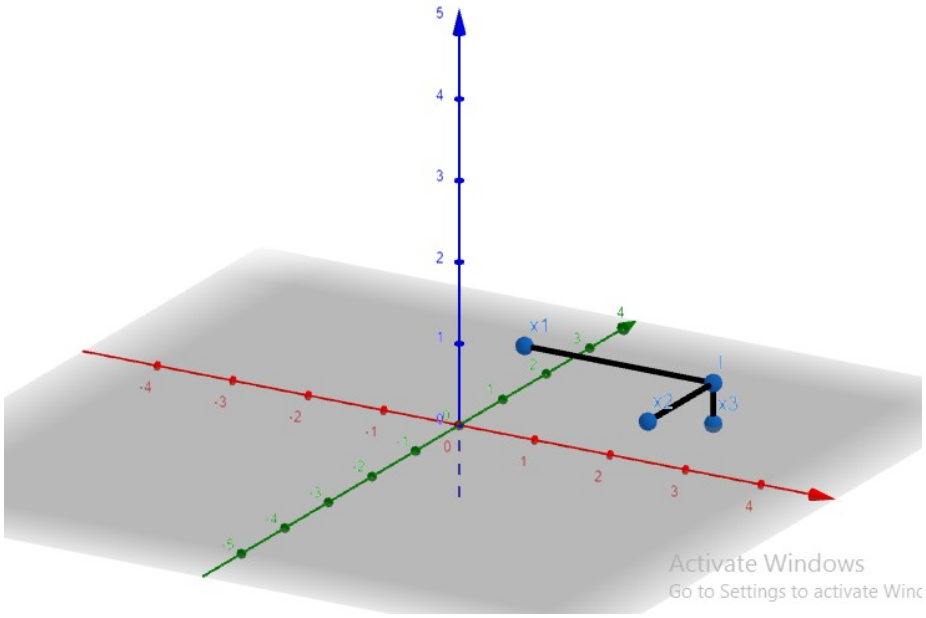


Fig: represents is a tree with 4 nodes and 3 edges

Distance between x_1 and x_2 = (Distance between x1 and I) + (Distance between I and x_2)
= 2.5 +1.5 (by distance formula)
= 4 = a_{12} in **D** = a_{21} in **D**

Similarly,
Distance between x_1 and x_3 = 3
Distance between x_2 and x_3 = 2

Illustration: 2

Let **D** = $\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \end{bmatrix}$

$x_1+x_2= 1$
$x_1+x_3 = 2$
$x_2+x_3=3$
The solution is I(0,1,2)
In the solution as the first coordinate is 0, this point corresponds to the taxon x_1(0,1,2)
As described in Case IIB, x_2(0,0,2) and x_3(0,1,0)

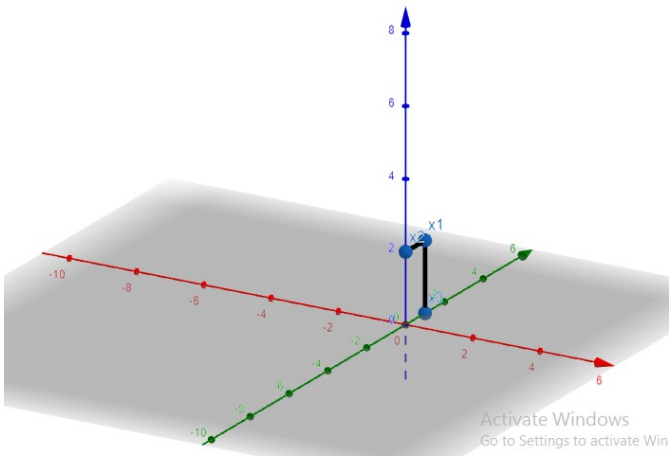


Fig is a tree with 3 nodes and 2 edges.
Distance between x_1 and x_2 = 1
Distance between x_1 and x_3 = 2
Distance between x_2 and x_3 = 3

Theorem:I

Given a distance matrix D showing the distance between taxa from the set $X = \{x_1,x_2,\ldots x_n\}$, $x_ix_i = 0$, $x_ix_j = k_p \neq 0$ for $i \neq j$ and $k_p > 0$ for p=1, 2, 3….,n. For n=2,3, the relation between taxa can be shown by trees in $\mathbf{R}^2$ or $\mathbf{R}^3$ space from the given distance matrix **D** such that the taxa preserves the distance same as that of the given matrix **D**. The number of distinct tree structures obtained for n taxa is given by the expression $nC_n + nC_{(n-1)}$, n= 2,3.

Proof:

From D, we can form equations. Solution of these equations may be in $\mathbf{R}^2$ or $\mathbf{R}^3$ depending on the number of taxa n. Sometimes the solution contains n nonzero positive entries or sometimes n-1. If the solution has 2 or more zero entries, then it violate the condition that $x_ix_i = 0$, $x_ix_j = k_p \neq 0$ for $i \neq j$ and $k_p > 0$.

For n nonzero positive entries, there are nC_n possibilities and for (n-1) nonzero positive entries, there are $nC_{(n-1)}$ possibilities.

Thus the number of distinct tree structures obtained for n taxa is given by the expression $nC_n + nC_{(n-1)}$, n= 2, 3.

Theorem: II

Given a distance matrix D showing the distance between taxa from the set $X = \{x_1,x_2,\ldots x_n\}$, $x_ix_i = 0$, $x_ix_j = k_p \neq 0$ for $i \neq j$ and $k_p > 0$ for p=1, 2, 3….,n. For n= 2,3, the relation between taxa can be shown by trees in $\mathbf{R}^2$ or $\mathbf{R}^3$ space from the given distance matrix **D** such that the taxa preserves the distance same as that of the given matrix **D**. The number of distinct tree structures upto isomorphism obtained for n taxa is given by the expression $nC_n + 1/n\ [nC_{(n-1)}]$, n= 2,3.

Proof:

From D, we can form equations. Solution of these equations may be in $\mathbf{R}^2$ or $\mathbf{R}^3$ depending on the number of taxa n. Sometimes the solution contains n nonzero positive entries or sometimes n-1. If the solution has 2 or more zero entries, then it violate the condition that $x_ix_i = 0$, $x_ix_j = k_p \neq 0$ for $i \neq j$ and $k_p > 0$.

For n nonzero positive entries, there are nC_n possibilities and for (n-1) nonzero positive entries, there are $nC_{(n-1)}$ possibilities. These $nC_{(n-1)}$ gives n isomorphic tree structures.(Only difference is for the labelling of taxa. That is if 0 occurs at the i^{th} coordinate then that point will be labelled by i^{th} taxa). Thus upto isomorphism there will be $1/n\ [nC_{(n-1)}]$ trees.

Thus the number of distinct tree structures upto isomorphism obtained for n taxa is given by the expression $nC_n + 1/n[nC_{(n-1)}]$, n= 2, 3.

Theorem:III

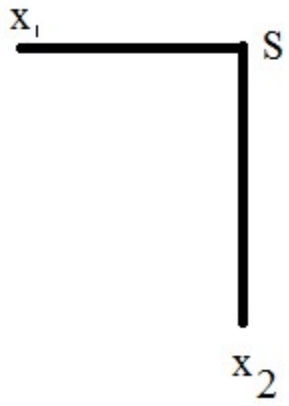
Given a distance matrix D showing the distance between taxa from the set $X = \{x_1,x_2,\ldots x_n\}$, $x_ix_i = 0$, $x_ix_j = k_p \neq 0$ for $i \neq j$ and $k_p > 0$ for p=1, 2, 3….,n. For n=2,3 the relation between taxa can be shown by trees in $\mathbf{R}^2$ or $\mathbf{R}^3$ space from the given distance matrix **D** such that the taxa preserves the distance same as that of the given matrix **D**.

The tree obtained with one entry in the solution is 0, is the fusion of a pair of vertices followed by the contraction of an edge, from the tree obtained with no entry in the solution is 0

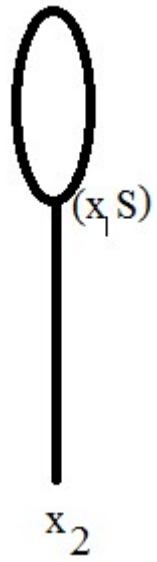
Proof:

For n = 2

Step I: The tree structure obtained when the solution S consist of non-zero positive values



Step II: The pair of vertices (x_1, S) in the above graph is fused and they are replaced by a single vertex (x_1S)



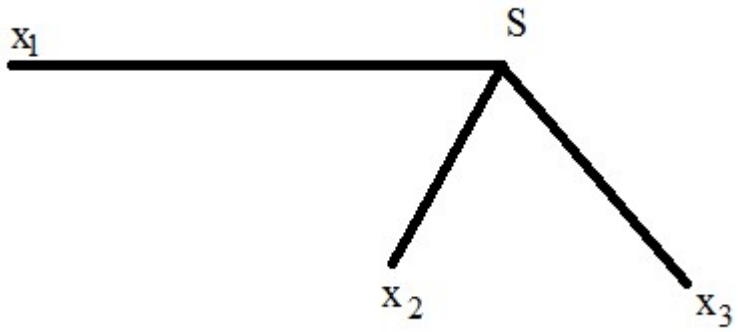
Step III: Contraction of loop

Step IV: Rename the vertex by x_1.

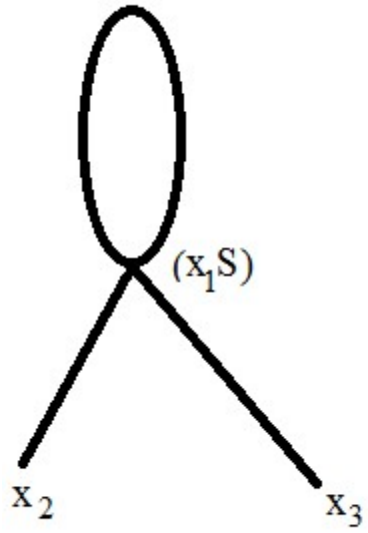
The graph below is the tree structure obtained when the solution consist of (n-1) non-zero positive values.

For n= 3

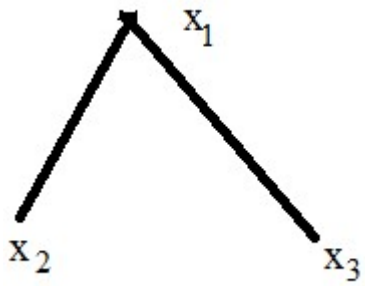
Step I: The tree structure obtained when the solution S consist of non-zero positive values



Step II: The pair of vertices (x_1, S) in the above graph is fused and they are replaced by a single vertex (x_1S) (x_1,x_2,x_3 were chosen on the basis of obtained solution)



Step III: Contraction of loop
Step IV: Rename the vertex by x_1.
The graph below is the tree structure obtained when the solution consist of (n-1) non-zero positive values.



IV.CONCLUSION

This work relates Biology and Mathematics. Mathematics always finds easier solutions to biological problems. Here we explained a way to construct tree on 2-space or 3-space based on given distance matrix. The interconnections or inter distance between various taxa can be shown graphically using the concepts of Graph theory

REFERENCES

[1] DR G SURESH SINGH: Graph Theory, PHI Private Limited, New Delhi (2010)

[2] NARSING DEO: Graph Theory with Applications to Engineering and Computer Science, PHI Learning.

[3] DANIEL H HUSON, REGULA RUPP, CELINE SCORNAVACCA: Phylogenetic Networks: Concepts, Algorithms and Applications, Cambridge University Press.

[4] T.A.BROWN : Genomes, 2nd Edition, BIOS Scientific Publishers Limited.

[5] GRAPH CONSTRUCTION USING : https://www.geogebra.org/3d?lang=en.