Indian Journal of Natural Sciences

www.tnsroindia.org.in ©IJONS

Vol.14 / Issue 81 / Dec / 2023 International Bimonthly (Print) – Open Access ISSN: 0976 – 0997

RESEARCH ARTICLE

# Construction of tree from $n$ x $n$ distance matrix that preserves the distance given in the matrix, $n \geq 4$

Renjini Raveendran .P[1]* and Beena. S[2]

[1]Assistant Professor, Department of Mathematics, All Saints' College, (Affiliated to University of Kerala), Thiruvananthapuram, Kerala, India

[2]Principal, NSS College, (Affiliated to University of Kerala), Nilamel, Kerala, India

Received: 16 July 2023 Revised: 22 Sep 2023 Accepted: 06 Nov 2023

***Address for Correspondence**
**Renjini Raveendran .P**
Assistant Professor,
Department of Mathematics,
All Saints' College, (Affiliated to University of Kerala),
Thiruvananthapuram, Kerala, India.
E.mail-: renjiniraveendran7@gmail.com



## ABSTRACT

The combination of graph theory and biology is the focus of this work. In order to understand how different species or taxa have evolved over time and how life on Earth has changed, phylogenetic analysis seeks to reveal these relationships. Here we make use of graph theory concepts along with that of phylogenetic analysis. Phylogenetic trees are commonly utilized to accomplish this. Tree diagram plays an important role in Phylogenetic analysis. The interconnections and inter distance of various taxa can be visualise easily with the help of these diagrams. Trees are connected acyclic graphs. The Distance matrix gives pair wise distance between the set of taxa X. Here we derived a method for the construction of tree from $n$ x $n$ distance matrix that preserves the distance given in the matrix, $n \geq 4$.

**Keywords**: Distance Matrix, Tree, Phylogenetic tree.

## INTRODUCTION

Method for the Construction of tree from $n$ x $n$ distance matrix that preserves the distance given in the matrix, $n \geq 4$.
$D = (x_{ij}), i, j = 1,2,3 \ldots n$

Here $x_{ij}$ denote the distance between $x_i$ *and* $x_j$ . Also $x_{ij} = x_{ji}$

$$\begin{pmatrix} 0 & x_{12} & x_{13} & \ldots & x_{1n} \\ x_{21} & 0 & x_{23} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n2} & \ldots & \ldots & 0 \end{pmatrix}$$

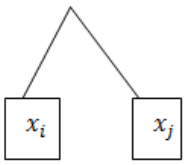
Consider the groups of 3 taxa. There are $nC_3$ groups in all. Form equations connecting each group and solve for $x_i$.
For the group $\{x_i, x_j, x_k\}$
The equations are
$x_i + x_j = x_{ij}$,
$x_j + x_k = x_{jk}$,
$x_k + x_i = x_{ki}$
Solve these equations to find the value of $\{x_i, x_j, x_k\}$. The corresponding values of taxa can be obtained from solving the groups of 3.Each $x_i$ can be obtained $nC_3 - (n-1)C_3$ times. From among the obtained values of $x_i$ , select the one which is least positive. Label it as $x_i^*$

| $x_i$ | The distinct values of $x_i$ | $x_i^*$ |
|---|---|---|
| $x_1$ | Distinct values of $x_1$ | $x_1$ * |
| $x_2$ | | |
| ⋮ | | |
| $x_n$ | | |

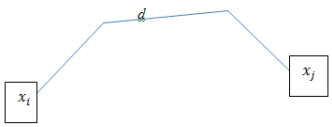
Now consider the pair of taxa. There are $nC_2$ pairs.

| $\{x_i, x_j\}$ | $x_i^* + x_j^*$ | $x_{ij}$ | $d = \lvert x_{ij} - (x_i^* + x_j^*)\rvert$ |
|---|---|---|---|
| $\{x_1, x_2\}$ | | | |
| $\{x_1, x_3\}$ | | | |
| $\{x_1, x_4\}$ | | | |
| $\{x_2, x_3\}$ | | | |
| $\{x_2, x_4\}$ | | | |
| ⋮ | ⋮ | ⋮ | ⋮ |

If $d = 0$ *for the pair* $\{x_i, x_j\}$, the part of tree connecting these two taxa has the structure as follows:




If $d \neq 0$ *for the pair* $\{x_i, x_j\}$, the part of tree connecting these two taxa has the structure as follows:




The parts of the tree connecting each pair of taxa can be obtained in this manner. Combining all these together, we get the tree that preserves the given distance matrix $D$.Sometimes we need to split the edge length to obtain the tree. This can be seen in the Illustration of the Construction of a tree from a Distance Matrix of order 5x5
Some deductions from this method:

1) The closely related taxa can be identified from the value of d in this method.
2) If $d = 0$ *for the pair* $\{x_i, x_j\}$, then $x_i$ *and* $x_j$ are closely related taxa. That is they are siblings from the same parent node.

3) The tree obtained in this way is a weighted tree.
4) The total weight of the tree can be calculated using the equation

$$W = \sum_{i=1}^{n} x_i^{*} + \text{maximum value of d obtained for each pair of taxa.}$$

**Illustration: Construction of a tree from a Distance Matrix of order 4x4**

Given a distance matrix D showing the inter-distance between four taxa $\{x_1, x_2, x_3, x_4\}$. The entry $x_{ij}$ denote the distance between $x_i$ $and x_j$.

$$D = \begin{pmatrix} 0 & 2 & 4 & 6 \\ 2 & 0 & 4 & 6 \\ 4 & 4 & 0 & 6 \\ 6 & 6 & 6 & 0 \end{pmatrix}$$

From among this 4 taxa, form groups of 3 taxa. There are $4C_3 = 4$ such groups.

1) $\{x_1, x_2, x_3\}$
2) $\{x_1, x_2, x_4\}$
3) $\{x_2, x_3, x_4\}$
4) $\{x_1, x_3, x_4\}$

Solve the system of equations connecting each group as follows:

1) $\{x_1, x_2, x_3\}$

$x_1 + x_2 = x_{12} = 2$
$x_2 + x_3 = x_{23} = 4$
$x_1 + x_3 = x_{13} = 4$

The solution obtained is $x_1 = 1,\ x_2 = 1,\ x_3 = 3$

2) $\{x_1, x_2, x_4\}$

The solution obtained is $x_1 = 1,\ x_2 = 1,\ x_4 = 5$

3) $\{x_2, x_3, x_4\}$.

The solution obtained is $x_2 = 2,\ x_3 = 2,\ x_4 = 4$

4) $\{x_1, x_3, x_4\}$

The solution obtained is $x_1 = 2,\ x_3 = 2,\ x_4 = 4$

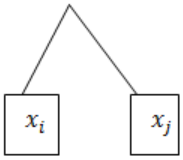
The value of each $x_i$ for $i = 1,2,3,4$ is obtained $4C_3 - (4-1)C_3 = 3$times
The distinct values obtained for each $x_i$for $i = 1,2,3,4$ is shown in table below:
Let $x_i^{*}$denote the minimum positive value of $x_i$for $i = 1,2,3,4$

| $x_i$ | The distinct values of $x_i$ | $x_i^{*}$ |
|---|---|---|
| $x_1$ | 1,2 | 1 |
| $x_2$ | 1,2 | 1 |
| $x_3$ | 2,3 | 2 |
| $x_4$ | 4,5 | 4 |

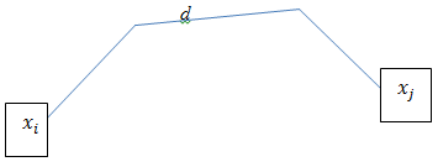
Now consider the pair of taxa. There are $4C_2 = 6$pairs can be obtained.

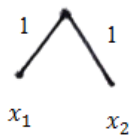
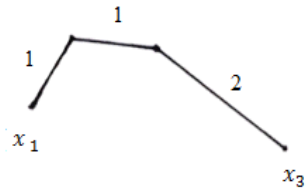
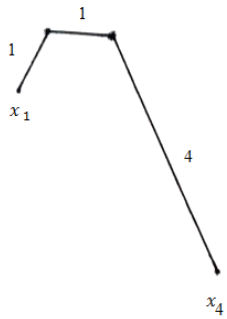
| $\{x_i, x_j\}$ | $x_i^{*} + x_j^{*}$ | $x_{ij}$ | $d = \lvert x_{ij} - (x_i^{*} + x_j^{*}) \rvert$ | |
|---|---|---|---|---|
| $\{x_1, x_2\}$ | 2 | 2 | 0 | Closely related pair of taxa |
| $\{x_1, x_3\}$ | 3 | 4 | 1 | |
| $\{x_1, x_4\}$ | 5 | 6 | 1 | |
| $\{x_2, x_3\}$ | 3 | 4 | 1 | |
| $\{x_2, x_4\}$ | 5 | 6 | 1 | |
| $\{x_3, x_4\}$ | 6 | 6 | 0 | Closely related pair of taxa |

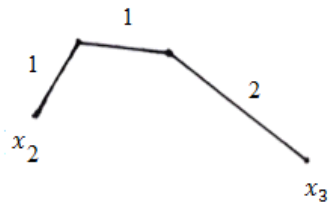
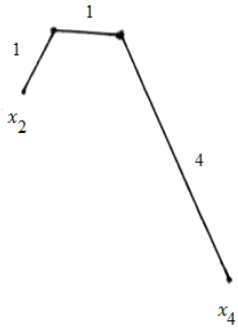
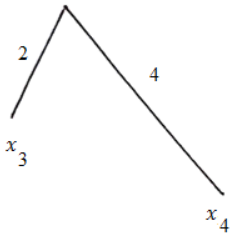
If $\boldsymbol{d = 0\ for\ the\ pair}\ \{x_i, x_j\}$, the part of tree connecting these two taxa has the structure as




If $d \neq 0\ for\ the\ pair\ \{x_i, x_j\}$, the part of tree connecting these two taxa has the structure as follows:




| $\{x_i, x_j\}$ | Structure of the Part of tree connecting $x_i$ and $x_j$ |
|---|---|
| $\{x_1, x_2\}$ |   |
| $\{x_1, x_3\}$ |   |
| $\{x_1, x_4\}$ |   |

| | |
|---|---|
| $\{x_2, x_3\}$ |   |
| $\{x_2, x_4\}$ |   |
| $\{x_3, x_4\}$ |   |

Combining all these together we get a weighted tree diagram that satisfies the given distance matrix

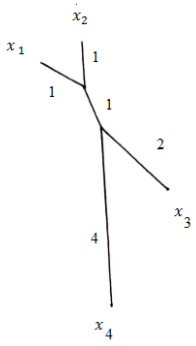



**Illustration: Construction of a tree from a Distance Matrix of order 5x5**

Given a distance matrix D showing the inter-distance between five taxa$\{a, b, c, d, e\}$. The entry $x_{ij}$ denote the distance between taxa $x_i$ *and* $x_j$.

$$D = \begin{pmatrix} 0 & 11 & 10 & 9 & 15 \\ 11 & 0 & 3 & 12 & 18 \\ 10 & 3 & 0 & 11 & 17 \\ 9 & 12 & 11 & 0 & 8 \\ 15 & 18 & 17 & 8 & 0 \end{pmatrix}$$

From among this 5 taxa, form groups of 3 taxa. There are $5C_3 = 10$such groups.
They are as follows:

1) $\{a, b, c\}$
2) $\{a, b, d\}$
3) $\{a, b, e\}$
4) $\{b, c, d\}$
5) $\{b, c, e\}$
6) $\{e, c, d\}$
7) $\{a, c, d\}$
8) $\{a, c, e\}$
9) $\{a, e, d\}$
10) $\{b, e, d\}$

Solve the system of equations connecting each group as follows:

1. $\{a, b, c\}$
$a + b = 11$
$b + c = 3$
$a + c = 10$
The solution obtained is $a = 9, b = 2, c = 1$
2. $\{a, b, d\}$
$a + b = 11$
$b + d = 12$
$a + d = 9$
The solution obtained is $a = 6, b = 5, d = 3$
1. $\{a, b, e\}$
The solution obtained is $a = -1, b = 2, e = 16$
2. $\{b, c, d\}$
The solution obtained is$c = 1, b = 2, d = 10$
3. $\{b, c, e\}$
The solution obtained is$c = 1, b = 2, e = 16$
4. $\{e, c, d\}$
The solution obtained is$c = 10, d = 1, e = 7$
5. $\{a, c, d\}$
The solution obtained is$a = 4, c = 6, d = 5$
6. $\{a, c, e\}$
The solution obtained is$a = 4, c = 6, e = 11$
7. $\{a, e, d\}$
The solution obtained is$a = 8, d = 1, e = 7$
8. $\{b, e, d\}$
The solution obtained is$b = 11, d = 1, e = 7$

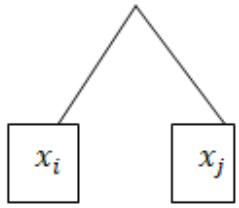
The value of each taxon is obtained $5C_3 - (5 - 1)C_3 = 6$ times
The distinct values obtained for each taxon is shown in table below:

| Taxon | The distinct values of each taxon |
|---|---|
| $a$ | 9,6,-1,4,8 |
| $b$ | 2,5,11 |
| $c$ | 1,10,6 |
| $d$ | 3,10,1,5 |
| $e$ | 16,7,11 |

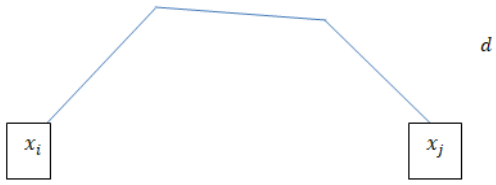
| Taxon | The distinct values of each taxon | Minimum positive value among the distinct values |
|---|---|---|
| $a$ | 9,6,-1,4,8 | $a* = 4$ |
| $b$ | 2,5,11 | $b* = 2$ |
| $c$ | 1,10,6 | $c* = 1$ |
| $d$ | 3,10,1,5 | $d* = 1$ |
| $e$ | 16,7,11 | $e* = 7$ |

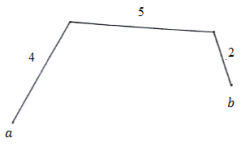
| $\{x_i, x_j\}$ | $x_i^* + x_j^*$ | $x_{ij}$ | $d = \lvert x_{ij} - (x_i^* + x_j^*)\rvert$ | |
|---|---|---|---|---|
| $\{a, b\}$ | 6 | 11 | 5 | |
| $\{a, c\}$ | 5 | 10 | 5 | |
| $\{a, d\}$ | 5 | 9 | 4 | |
| $\{a, e\}$ | 11 | 15 | 4 | |
| $\{b, c\}$ | 3 | 3 | 0 | Closely related taxa |
| $\{b, d\}$ | 3 | 12 | 9 | |
| $\{b, e\}$ | 9 | 18 | 9 | |
| $\{c, d\}$ | 2 | 11 | 9 | |
| $\{c, e\}$ | 8 | 17 | 9 | |
| $\{e, d\}$ | 8 | 8 | 0 | Closely related taxa |

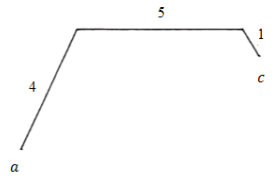
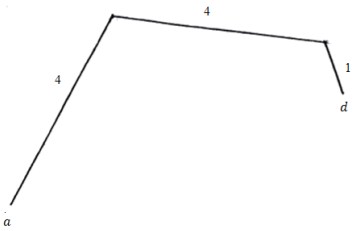
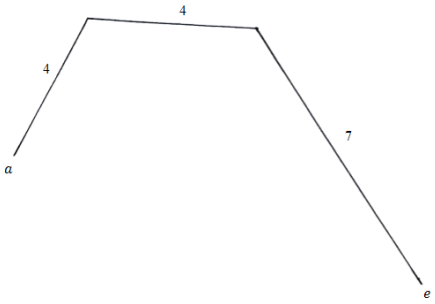
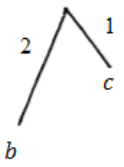
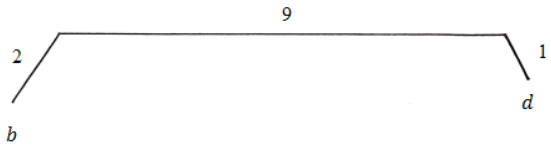
**If $d = 0$ *for te pair* $\{x_i, x_j\}$,** the part of tree connecting these two taxa has the structure as follows:

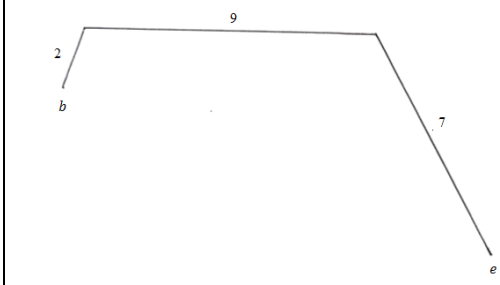
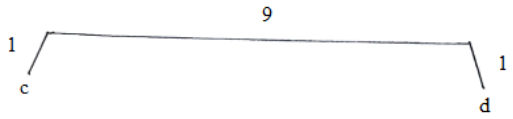
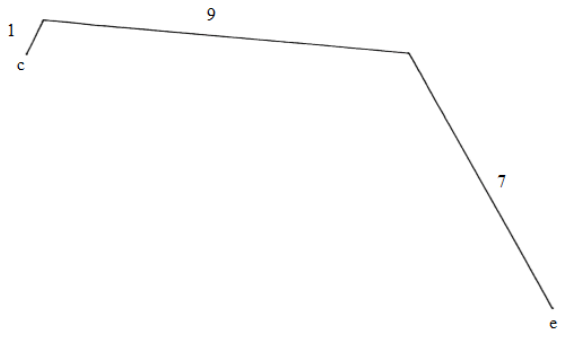
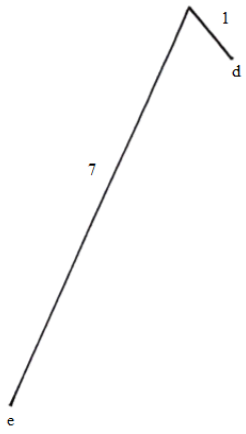



If $d \neq 0$ *for the pair* $\{x_i, x_j\}$, the part of tree connecting these two taxa has the structure as follows:

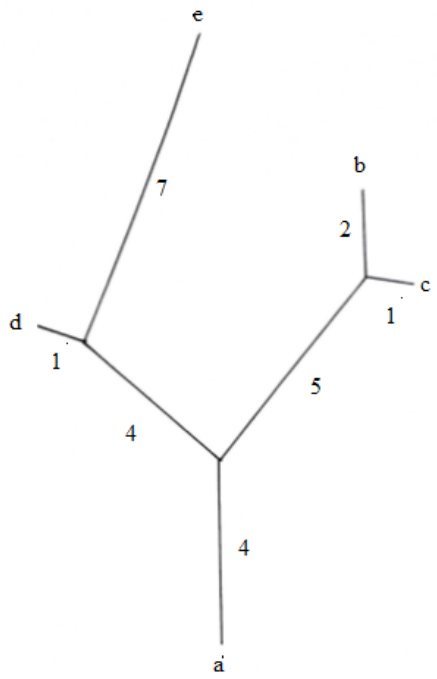



| $\{x_i, x_j\}$ | Structure of the Part of tree connecting $x_i$ *and* $x_j$ |
|---|---|
| $\{a, b\}$ |  |

| | |
|---|---|
| $\{a, c\}$ |   |
| $\{a, d\}$ |   |
| $\{a, e\}$ |   |
| $\{b, c\}$ |   |
| $\{b, d\}$ |   |

| | |
|---|---|
| $\{b, e\}$ |   |
| $\{c, d\}$ |   |
| $\{c, e\}$ |   |
| $\{e, d\}$ |   |

Combining all these together we get a weighted tree diagram that satisfies the given distance matrix.




## CONCLUSION

Construction of trees plays an important role in phylogenetic analysis. From Biological data we obtain the inter distance between organisms. From the biological data we cannot say any relation among the various taxa under consideration. But using this method, the value of d predicts the relationship. If d=0 for a pair of taxa means they are closely related or they evolve from a single parent. These data can be mathematically converted to trees that help us to view the position of organisms at the leaves of the tree.

## 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.