European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 08, 2020

On Metric and Ultrametric Trees

Dr. R. Sivaraman
Associate Professor, Department of Mathematics, D. G. Vaishnav College, Chennai, India
National Awardee for Popularizing Mathematics among masses
Email: rsivaraman1729@yahoo.co.in

ABSTRACT-
In this paper, I will introduce the concept of phylogenetic trees and study their properties. In particular, the main focus of the paper is to study the concept of metric and ultra – metric trees that exist from phylogenetic trees. Finally I will prove a theorem which will provide a way for determining the root of a ultrametric tree. These concepts have interesting applications in study of genetics and various other related branches of science.

Keywords: Taxa, Phylogenetic Tree, Rooted Binary Tree, Metric Tree, Metric Space, Ultrametric Tree

1. Introduction

Phylogenetics is the research area focusing mainly on determining the genetic lineages and relationships that exist between various kinds of species. Classic phylogenetics deals mainly with physical or morphological features like size, color, number of legs etc. Modern phylogeny uses information extracted from genetic material - mainly DNA and protein sequences. The characters used are usually the DNA or protein sites (a site means a single position in the sequence) after aligning several such sequences, and using only blocks which were conserved in all the examined species. The most convenient way of presenting phylogenetic information is by using a phylogenetic tree. In a phylogenetic tree, we often think of each vertex as representing a species, with the edges indicating lines of direct evolutionary relationships existing between them. In this paper, I will introduce the concept of Phylogenetic trees, Metric and Ultrametric trees and provide some theorems regarding them.

2. Definitions

2.1 Graph and Tree

A Graph G consists of pair of sets (V, E) where V is the set of all vertices also called nodes and E represent the set of all edges that can exist between particular pair of vertices. We can denote a graph by $G = (V, E)$. A Graph G is said to be connected if there exists a path between every pair of distinct vertices in G. Otherwise G is said to be disconnected.

A cycle is defined as a closed path. A Graph G is said to be acyclic if contains no cycle. A Graph G is said to be a Tree if it is connected and acyclic.

2.2 Binary Tree

A tree T is said to be binary tree if it has a unique vertex (node) of degree 2 and all remaining vertices have either degree 1 or 3.

The unique vertex of degree 2 is called the root of the binary tree and the vertices of degree 3 are called internal vertices and those of degree 1 are called pendant vertices or leaves of the binary tree. With

respect to characterizing binary tree as phylogenetic tree, the root of the tree (which has degree two) is considered as the common ancestor to all the remaining vertices of the tree.

2.3 Phylogenetic Trees

Let *X* denote a finite set of taxa or labels. Then a phylogenetic *X* – Tree is defined to be a tree of the form *T* = (*V*, E) together with a one – one correspondence $\phi : X \to L$ where $L \subseteq V$ denotes the set of leaves of the tree. The mapping ϕ is called the labeling map. Such a tree may also be called as leaf – labeled tree. We notice that the labeling map simply assigns each of the taxa to a different leaf of the tree, so that every leaf is assigned with a label.

2.4 Metric Tree

A Metric tree is a rooted or unrooted tree *T* = (*V*, E) together with a function $w : E \to [0, \infty)$. From definition of w, we see that w assigns non-negative numbers to each edge of *T*. If $e \in E$ is an edge of *T* then $w(e)$ is defined as the length or weight of the edge *e*.

When we do not specify the length or weight for the edges of the tree that we consider, then we can call such tree as a Topological Tree.

2.5 Distance between vertices

Let *T* = (*V*, E) be a metric tree. Let $v_1, v_2 \in V$ be any two vertices of *T*. We define a distance between any two vertices of *T* as $d(v_1, v_2) = \sum_e w(e)$ (2.1) where the weights $w(e)$ is taken for all edges on the path e between v_1 and v_2.

That is, the distance between two vertices in a tree is the sum of lengths of the edges that exist between v_1 and v_2.

3.1 Theorem 1

If *T* = (*V*, E) is a metric tree, then (V, d) is a metric space.

Proof: To show that (V, d) is a metric space we have to prove that the distance function *d* as defined in (2.1) forms a metric meaning that it satisfies all the axioms for a metric in a metric space.

(i) From (2.1), we note that $d(v_1, v_2) = \sum_e w(e)$. Since by definition the length or weight of any edge e in a tree is non-negative $w(e) \geq 0$ for all $e \in E$. Since sum of non-negative numbers is also non-negative, it follows that $d(v_1, v_2) = \sum_e w(e) \geq 0$ for all $v_1, v_2 \in V$. Thus *d* is non-negative. Further we note that if all edges of *T* have positive length then $d(v_1, v_2) = 0$ if and only if $\sum_e w(e) = 0$ if and only if $v_1 = v_2$.

(ii) First, we note that in a tree, there exists a unique path between any pair of vertices. Thus, if $v_1 v_i v_j \cdots v_r v_2$ is a unique path between v_1 and v_2 then retracing along the same path we get $v_2 v_r \cdots v_j v_i v_1$

as a unique path between v_2 and v_1. Since the path between v_1 and v_2 is same in either direction, the sum of lengths along traversing the edges between them will also be equal. Hence, $d(v_1, v_2) = \sum_e w(e) = d(v_2, v_1)$ proving that *d* is symmetric.

(iii) Let $v_1, v_2, v_3 \in V$ be any three vertices in the tree *T* = (*V*, E). Let $P_1 : v_1 v_i v_{i+1} v_{i+2} \cdots v_{i+k} v_2$ be a unique path between v_1 and v_2 and $P_2 : v_2 v_j v_{j+1} v_{j+2} \cdots v_{j+r} v_3$ be a unique path between v_2 and v_3. Then we observe that $P_1 \bigcup P_2$ is the unique path between v_1 and v_3. If any of the edges in $P_1 \bigcup P_2$ have zero length then by property (i) those two vertices can be merged together to a single vertex in $P_1 \bigcup P_2$. Hence $|P_1 \bigcup P_2| \le |P_1| + |P_2|$. Now by definition (2.1), we have

$$d(v_1, v_3) = \sum_{e \in P_1 \bigcup P_2} w(e) \le \sum_{e \in P_1} w(e) + \sum_{e \in P_2} w(e) = d(v_1, v_2) + d(v_2, v_3)$$

Thus, $d(v_1, v_3) \le d(v_1, v_2) + d(v_2, v_3)$. This establishes the triangle inequality property of *d*.

We also note that if all the edges have positive lengths then none of the vertices in either P_1 or P_2 will be merged and by uniqueness of the path in a tree, we will then have $|P_1 \bigcup P_2| = |P_1| + |P_2|$ giving $d(v_1, v_3) = d(v_1, v_2) + d(v_2, v_3)$.

Thus from (i), (ii) and (iii) we see that *d* is a metric. Hence (V, d) is a metric space.

This completes the proof.

3.2 Definition

A rooted metric tree is said to be ultametric tree if all its leaves are equidistant from its root, where the distance is measured using the metric of the metric tree as defined in (2.1)

3.3 Molecular Clock Trees

In phylogenetics, we often ultrametric trees as molecular clock trees. This is because if mutation occurs at a constant rate over all time and lineages, i.e. is clocklike, then many methods of inference from sequences will produce such trees in idealized circumstances. Edge lengths should be interpreted as measures of how much mutation has occurred, so that with clock-like mutation we simply scale elapsed time by a constant mutation rate to get lengths. But one should be careful – even if a tree is ultrametric, it need not have been produced under a molecular clock. For instance mutation rates could increase in time, uniformly over all lineages, and each would show the same total mutation from root to leaf.

A molecular clock assumption can be biologically reasonable in some circumstances, for instance, if all taxa are reasonably closely related and one suspects little variation in evolutionary processes during the evolutionary period under study. Other times it is less probable, if more distantly related taxa are in the tree, and the circumstances under which they evolved may have changed considerably throughout the tree.

4.1 Theorem 2

Let *T* be a rooted ultrametric tree *T* with positive edge lengths. Let v_1, v_2 be any two leaves of such that $d(v_1, v_2)$ is maximum among distances between leaves in *T*. Then there exists a unique vertex *r* in *T* such that $d(v_1, r) = d(r, v_2) = \frac{d(v_1, v_2)}{2}$ (4.1) .

Proof: We first note that if *T* has more than one leaf, then *r* will be an internal vertex. Since in a tree every edge is cut-edge, deleting *r* and its incident edges from *T* will produce at least two connected components of *T*.

Suppose if v_1, v_2 are vertices in two different components of $T - \{r\}$, then the path from v_1 to v_2 must pass through *r*. Since the vertices v_1, r, v_2 are distinct, from triangle inequality property of the metric *d* established in theorem 1, we have $d(v_1, v_2) = d(v_1, r) + d(r, v_2)$ (4.2).

Since *T* is a ultrametric tree, all the leaves are equidistant from its root. Hence $d(v_1, r) = d(r, v_2)$. Also from (4.2) we have $d(v_1, v_2) = 2d(v_1, r)$ (4.3). Now equations (4.2) and (4.3) prove (4.1). Since the edge lengths are positive, all the vertices are distinct and the root *r* satisfying (4.1) must be unique.

If we assume v_1, v_2 are vertices in same connected component of $T - \{r\}$, then the path between v_1 and v_2 does not pass through *r*. Thus by triangle inequality property of the metric *d* between vertices we have $d(v_1, v_2) < d(v_1, r) + d(r, v_2)$ (4.4) . Since the root is equidistant between the leaves, we get $d(v_1, r) = d(r, v_2)$. Hence (4.4) becomes, $d(v_1, v_2) < 2d(v_1, r)$ implying that $d(v_1, v_2)$ is not maximum among distances between leaves in *T*. Thus the vertices v_1, v_2 in two different components of $T - \{r\}$, proving (4.1). This completes the proof.

5. Conclusion

Through the introduction of trees in usual Graph Theory, I had defined phylogenetic trees and through that I had introduced Metric tree. In theorem 1 of section 3.1, I proved that the distance between two vertices in a metric tree is in fact a metric making (V, d) a metric space. This idea enabled to define ultrametric tree and view the idea of molecular clock trees. The concept of ultrametric tree helped us in locating the unique root through theorem 2 of section 4.1. This theorem has wide implications and applications in genetic studies and in investigation of origin of viral diseases. Thus, the ideas discussed in this paper forms a vital part of understanding phylogenetics research and provide new insight in to the investigations for further research in this discipline.

REFERENCES

[1] L.J. Billera, S.P. Holmes, K. Vogtmann, Geometry of the space of phylogenetic trees, Adv. Appl. Math., 27 (4) (2001), pp. 733-767.

[2] Alex Gavryushkin, AlexeiJ.Drummond, The space of ultrametric phylogenetic trees, Journal of Theoretical Biology, 403(2016), 197–208

[3] G. Cardona, M. Llabrés, F. Rosselló, G. Valiente, Nodal distances for rooted phylogenetic trees, J. Math. Biol., 61 (2) (2010), pp. 253-276.

[4] D.M. Hillis, T.A. Heath, K.S. JohnAnalysis and visualization of tree space, Syst. Biol., 54 (3) (2005), pp. 471-482.

[5] M. Owen, Computing geodesic distances in tree space, SIAM J. Discrete Math., 25 (4) (2011), pp. 1506-1529.

[6] M.R. Bridson, A. Haeiger, Metric Spaces of Non-Positive Curvature, 319, Springer (1999)