Strad Research DOI: 10.37896/sr10.7/002 ISSN: 0039-2049

ALGORITHM TO CHECK THE NON-TRIVIAL SPLITS ARE CIRCULAR, TO FIND THE CIRCULAR ORDERING AND A TECHNIQUE FOR THE CONSTRUCTION OF OUTER LABELED PLANAR SPLIT NETWORK.

RENJINI RAVEENDRAN P[1] AND BEENA S

ABSTRACT. The concept of Splits plays an important role in phylogenetic analysis. It divides taxa on their distinctive features. Based on various conditions, splits may be compatible, circular or weakly compatible. If the Split is circular, it can be represented by split networks which are outer labeled and planar. Suppose we are having a collection of trivial and nontrivial splits. This work explains an algorithm to check whether the non-trivial splits from the collection are circular. It gives the circular ordering and provides a technique for the construction of outer labeled planar split network of non-trivial splits as well as Split network for the collection of trivial and non-trivial splits.

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 x_i represents some species, group or individual organism whose evolutionary history is of interest to us. A split is any bipartitioning of X into two non-empty subsets A and B of X, such that $A \cap B = \phi$ and $A \cup B = X$. Split Graphs and Split networks can be constructed using the collection of splits. In a split network, one or more edges are used to represent a split, S. The deletion of all these edges produces exactly two connected components, each component was provided with the split parts of S.

The concept of circular split reduces the complication of Split Networks. If the taxa in X can be placed around a circle in such a way that each split $S = A/B$ can be realized by a line through the circle that separates the plane into two

[1]*Renjini Raveendran P (corresponding author)*

2010 *Mathematics Subject Classification.* 05C90.

Key words and phrases. Split graph, Split network, Outer labeled planar graph.

half-planes, one containing all taxa in A and the other containing all taxa in B is called Circular Split.

Suppose we are having a collection of trivial and nontrivial splits. In this work, an algorithm is introduced to find out whether the non-trivial splits are circular, to find out their circular ordering and also the representation of these non-trivial splits to an outer labeled planar split network. That is these splits can be represented in the plane such that no two edges intersect and all labeled nodes lie on the outside of the graph. This work also gives an idea for the construction of split network of the collection of both trivial and non trivial splits.

I **Preliminary**

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 $X = \{x_1, \cdots x_n\}$ 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 split $S = A/B$ is a bipartition of a set of taxa X into two non-empty subsets A, B with $A \cap B = \phi$ and $A \cup B = X \cdot A$ and B are called the two split parts of S.

Definition 1.5 A split graph consists of a finite, simple, connected, bipartite graph $G = (V, E)$, together with an edge coloring $\sigma : E \to K$ that is surjective and isometric.

Definition 1.6 Let **S** be a set of splits on X. A split network $N = (V, E, \sigma, \lambda)$ that represents **S** is given by a split graph $(G = (V, E), \sigma : E \to \mathbf{S})$ and a node labelling $\lambda : X \to V$, with the property that for every split $S = A/B$ in **S** we have ,$A = \bigcup_{v \in {V_s}^0} \lambda^{-1}(v)$ and $B = \bigcup_{v \in {V_s}^1} \lambda^{-1}(v)$ or in other words, deletion of all edges of color S produces a graph consisting of precisely two connected components, one containing all nodes

Definition 1.7 A set of splits **S** on X is called circular, if there exists a linear ordering $(x_1, ..., x_n)$ of the elements of X for **S** such that each split $S \in \mathbf{S}$

has the form. $S = \{x_p, x_{p+1}, \cdots, x_q\}/\{X - \{x_p, x_{p+1}, \cdots, x_q\}\}$ for appropriately chosen $1 < p \leq q \leq n$.

Definition 1.8 Let G be a graph in which some of the nodes are labeled. G is outer-labeled planar, if there exists a drawing of G in the plane such that no two edges intersect and all labeled nodes lie on the outside of the graph.

Theorem 1.9 Circular implies outer-labeled planar

Theorem 1.10 (Deletion of split produces two components)
Let $(G = (V, E), \sigma : E \to K)$ be a split graph. For any color $c \in K$, the graph G^C obtained by deleting all edges of color c consists of precisely two connected components.

II Algorithm to check the non-trivial splits are circular, to find the circular ordering and a technique for the construction of outer labeled planar split network.

This section explains the step by step procedure to find the circular ordering and the construction of outer labeled planar split network.

Suppose we are having a set of Splits. Consider non-trivial Splits. They are n in number on the set of taxa X whose cardinality is m.

(1) Form a binary matrix (a_{ij}) of order $n \times m$, where

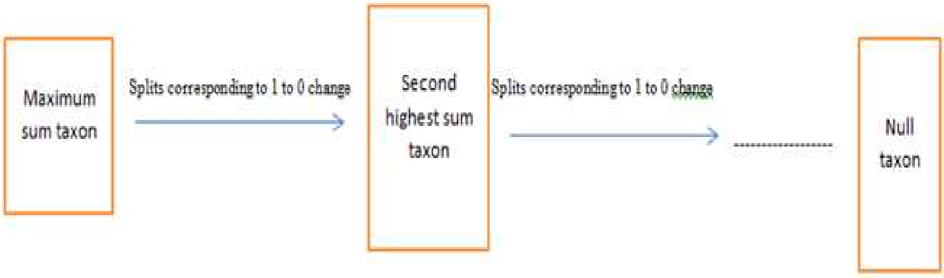
$$a_{ij} = \begin{cases} 1, \text{if the } j^{th} \text{ taxon belongs to } \ i^{th} \text{ split} \\ 0, \text{otherwise} \end{cases}$$

(2) Add the column wise entries and consider the maximum sum and the next highest.

(3) Point out the change of 1 to 0 and note the corresponding split/splits for this change.

(4) Compare the second highest and third highest, third highest and fourth highest and so on upto comparison with a null taxon whose column wise sum is zero (all entries in column are 0). Sometimes these types of taxon will be there; otherwise we have to consider a null taxon. While comparing, only 1 to 0 change is allowed. Does not consider any taxa which changes previous 0 entry to 1, even though its column sum is next highest. If any entry of column is 0, it will remain as such until null taxon is reached.

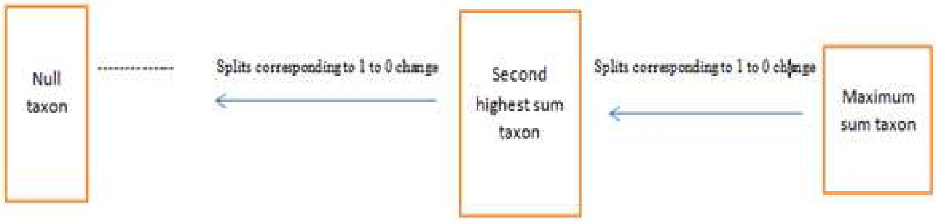
(5) From step (2) onwards the changes have to represent in a pictorial form.



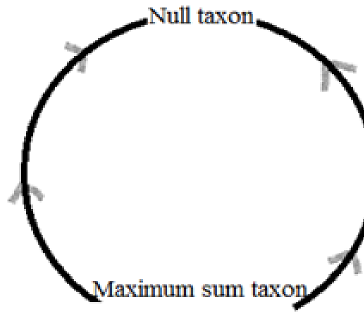
(6) Some taxon with a certain sum must be left out. Now compare maximum sum taxon to the next highest (which didn't choose earlier).

(7) A similar process will be repeated until we reach the null taxon.

(8) Pictorially represent these steps in opposite direction.



(9) Join two parts of the pictorial representation. This gives the circular ordering of given splits. After obtaining this circular ordering, remove the null taxon as it has no role in further construction. (If we provide it.)



(10) Draw a circle and divide the circle into ($2\times$ number of splits) parts equally by drawing radial lines.(each angle measures $[360/(2n)]^{\circ}$).

(11) Represent taxon on the endpoints of radial lines and split on arc parts.

(12) If the same splits occur on opposite sides of the part of the circle, join the end points of the adjacent radial lines. Remove the circular outline to obtain outer labeled planar split network.

(13) If we want to represent a trivial split also, then we have to extend the radial line where we represent the taxa and write the taxon on its end and corresponding trivial split on the extended edge.

(14) If Step (12) doesn't hold in the circular ordering, we have to construct a outer labeled planar split network.

(15) Draw a circle and draw radial lines to divide it into (number of splits-1) parts, that is $(n-1)$ parts.

(16) Construct a parallelogram on each radial line resulting in $(n-1)$ parallelograms.(This is useful for representing $(n-1)$ split as it is having $2n-2$ outer edges.

(17) Extend any radial line to a distance equal to radius of circle and construct two same sized parallelograms as earlier sharing the extended radii as its common edge. Then it represents a polygon having $2n$ outer edges.

(18) Represent the taxa and splits by looking at the circular order. First denote the taxa and split in outer nodes and edges. While moving from one taxa to another , if more splits are engaged, they can be represent in planar graph in any order. That is if splits $S3$ and $S1$ are responsible for a change from taxon a to b, we can represent these splits in the outer

edges of planar graph as $a, S1, S3, b$ or $a, S3, S1, b$. Remove the circular outline. This is an outer labeled planar split network.

(19) After considering Steps (1) to (8) , if any taxon is left behind again, which is not a part of both the pictorial representations in step (5) and Step (8), then the collection of splits is non- circular.

(20) This method can also be used for the construction of split network if the collection includes trivial splits, extend the nodes labeled with taxa and write the taxon on its end and corresponding trivial split on the extended edge to obtain the split network.

III. Illustration:

Case : I

Consider the set of Splits $\{S1, S2, S3\}$ where

$S1 = \{a, b, c\}/\{d, e, f\}$

$S2 = \{b, c, d\}/\{a, e, f\}$

$S3 = \{c, d, e\}/\{a, b, f\}$

	a	b	c	d	e	f
$S1$	1	1	1	0	0	0
$S2$	0	1	1	1	0	0
$S3$	0	0	1	1	1	0
SUM	1	2	3	2	1	0

c is the maximum sum taxon.

f is the null taxon.

Comparing c and b:

c	b
1	1
1	1
1	0

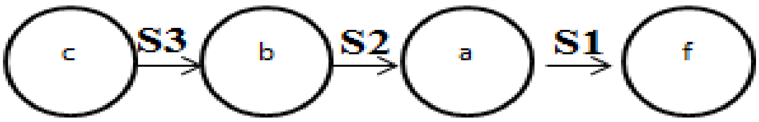
Considering c and b, 1 to 0 changes occurs corresponding to split $S3$.

b	a
1	1
1	0
0	0

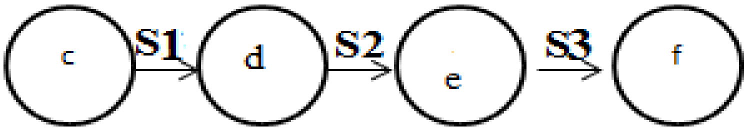
Considering b and a, 1 to 0 change occurs corresponding to split $S2$.

a	f
1	0
0	0
0	0

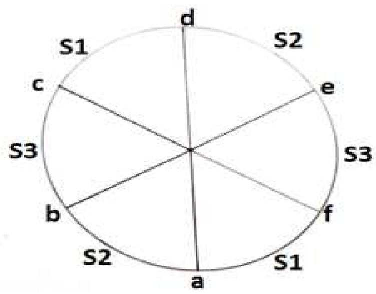
Considering a and f, 1 to 0 change occurs corresponding to split $S1$.
This can be represented as



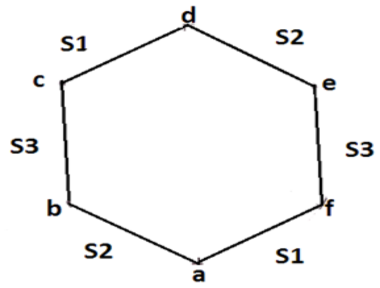
Similarly when comparing $(c, d), (d, e)$ and (e, f), we get



These two figures can be combined to get the Circular Ordering of the Splits



Here same splits appear oppositely, so join the ends of the radial lines and remove the circle to obtain the outer labeled planar split graph.



Case : II

Consider the set of splits $S = \{S1, S2, S3, S4, S5, S6\}$ where

$S1 = \{a, b, d, e, h\}/\{c, f, g\}$

$S2 = \{a, c, d, e, g, h\}/\{b, f\}$

$S3 = \{a, c, e, g\}/\{b, d, f, h\}$

$S4 = \{a, c, g\}/\{b, d, e, f, h\}$

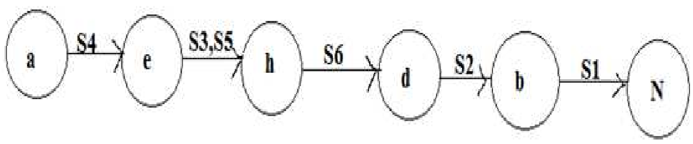
$S5 = \{a, c, e, f, g\}/\{b, d, h\}$

$S6 = \{a, e, h\}/\{b, c, d, f, g\}$

Convert to binary matrix and finding column wise sum. Last column denotes null taxa. Columns of taxa c and g are same.

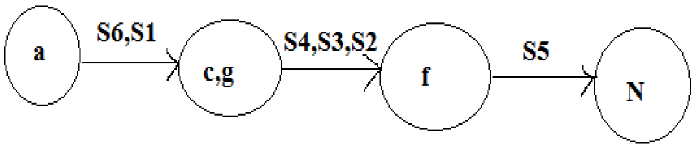
	a	b	c	d	e	f	g	h	N
$S1$	1	1	0	1	1	0	0	1	0
$S2$	1	0	1	1	1	0	1	1	0
$S3$	1	0	1	0	1	0	1	0	0
$S4$	1	0	1	0	0	0	1	0	0
$S5$	1	0	1	0	1	1	1	0	0
$S6$	1	0	0	0	1	0	0	1	0
Sum	6	1	4	2	5	1	4	3	0

Comparing taxa $(a, e), (e, h), (h, d), (d, b), (b, N)$

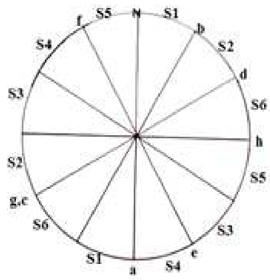


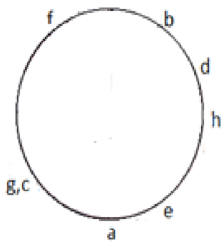
DOI: 10.37896/sr10.7/002
ISSN: 0039-2049

Again comparing taxa $(a,(g,c)),((g,c),f),(f,N)$

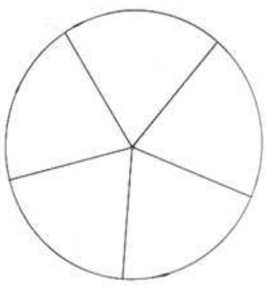


Combining the above two and represent it in a circle to get the circular ordering.

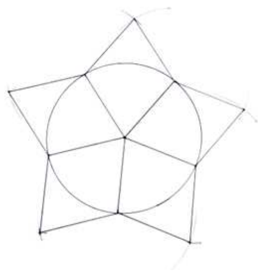




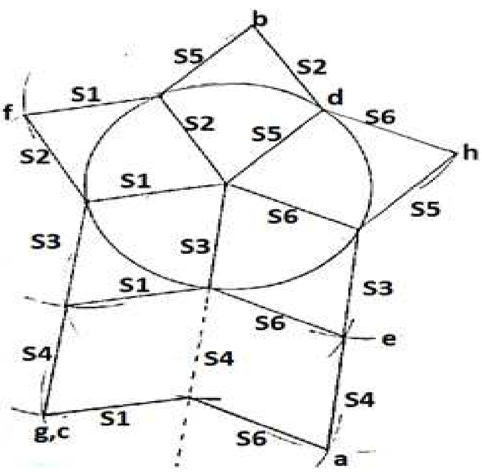
For outer labeled planar split network construction, divide the circle into 5 equal parts by drawing the radii having interior angles 72° each.



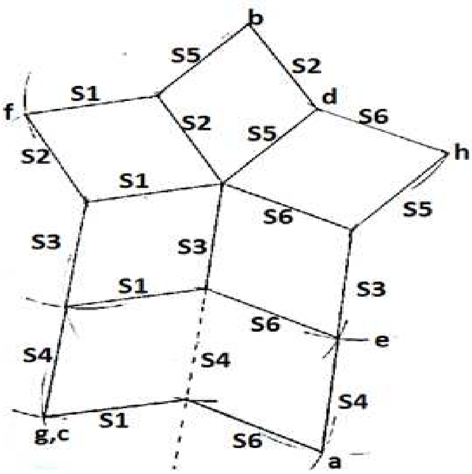
Then construct 5 parallelograms on these 5 radial lines whose sides are equal to the radius of the circle using compass.



Here the above structure have only 10 outer edges.To represent the 12 splits between the taxa, we have to construct two more parallolograms. So extend any radial line and construct two more parallelograms of same kind as earlier.Then it will have 12 outer edges.Represent the taxa in outer nodes and splits in outer edges using the circular ordering. Now label the inner edges with splits same as that of the splits represented parallel.



Now remove the circle to obtain the outer labeled planar split network representation.



In addition to the above splits, if we are given the trivial splits also, then that can also be included in the figure,

Let $S7 = \{a\}/\{b, c, d, e, f, g, h\}$

$S8 = \{b\}/\{a, c, d, e, f, g, h\}$

$S9 = \{c\}/\{a, b, d, e, f, g, h\}$

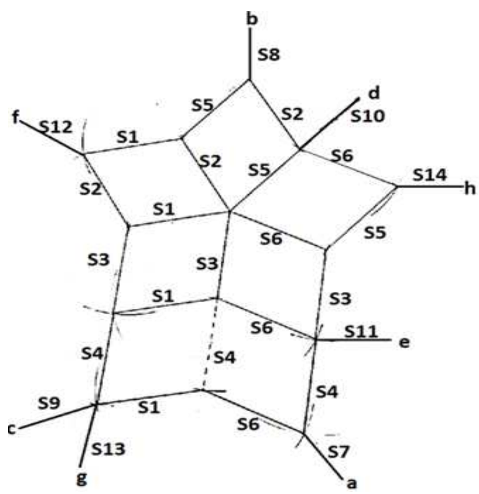
$S10 = \{d\}/\{a, b, c, e, f, g, h\}$

$S11 = \{e\}/\{a, b, c, d, f, g, h\}$

$S12 = \{f\}/\{a, b, c, d, e, g, h\}$

$S13 = \{g\}/\{a, b, c, d, e, f, h\}$

$S14 = \{h\}/\{a, b, c, d, e, f, g\}$



Theorem: Given a collection of m trivial and $(n+1)$ non-trivial splits on a set X of k taxa, where $m \leq k$.

a Suppose $n+1$ non-trivial split are proved to be circular and obtained the ordering by above algorithm. Then associated with this we can find outer labeled planar split network of $2(n+1)$ edges and $2(n+1)$ nodes or an outer labeled planar split network of $3n+5$ edges and $2(n+2)$ nodes.

b For representing all $n+1$ non trivial and m trivial splits, there exist either an outer labeled planar split network of $\{2(n+1)\}+m$ edges and $\{2(n+1)\}+m$ nodes or an outer labeled planar split network of $\{3n+5\}+m$ edges and $\{2(n+2)\}+m$ nodes.

Proof:

a . Consider the given collection of $n+1$ non trivial split on k taxa. When representing the circular ordering in a circle by dividing the circle into $\{360/(2(n+1))\}^\circ$ by drawing radii. If the ordering appears in such a way that the same splits occurs in exactly opposite arcs of the circle, then we can join the end points of the adjacent radii and remove the circular outline to obtain a regular polygon having $2(n+1)$ edges and $2(n+1)$ nodes. This is an outer planar split network.

Suppose the case is not as above. That is we cannot represent the circular ordering in such an arrangement around the circle.

Draw a circle and divide it into n parts by drawing radii at $(360/n)^\circ$ each. Construct parallelograms at each pair of radii, sides equal to the radius of the circle. From the end of each radial line, we can construct two outer edges of the parallelogram-resulting in a total of $2n$ outer edges (constructed sides of parallelogram) and n inner edges (radii of the circle). Now the geometric structure consist of n parallelograms, $2n$ outer edges, n inner edges, $2n+1$ nodes (1 node at the centre of the circle, n nodes as end points of radii and n nodes due to newly constructed sides of parallelogram). There are $2(n+1)$ splits from circular ordering which cannot be represent completely in outer edges of above geometric structure having only $2n$ outer edges. Extend any radial line and construct 2 more parallelograms having common edge (extended radii having measure same as that of radii) of same size as that of other parallelograms. Remove the circular outline. The constructed geometric figure has $2(n+1)$ outer edges, $2(n+1)$ outer nodes, $(n+3)$ inner edges and 2 inner nodes. That is it has $3n+5$ edges and $2(n+2)$ nodes. Label the outer nodes and edges with taxa and splits with the help of circular order obtained. This is an outer planar split network.

b . If we want to represent m trivial splits also in the outer labeled planar split network constructed above, extend the nodes labeled with each taxon and represent the corresponding trivial split in extended edge and taxon in the end of the same edge. By this construction we obtain a geometric structure having edges and nodes increased by m.

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.

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

BEENA S
PRINCIPAL
HHMSPB NSS COLLEGE FOR WOMEN
THIRUVANANTHAPURAM